UNIVERSITATEA
DE MEDICINĂ ȘI FARMACIE
„VICTOR BABEȘ” DIN TIMIȘOARA

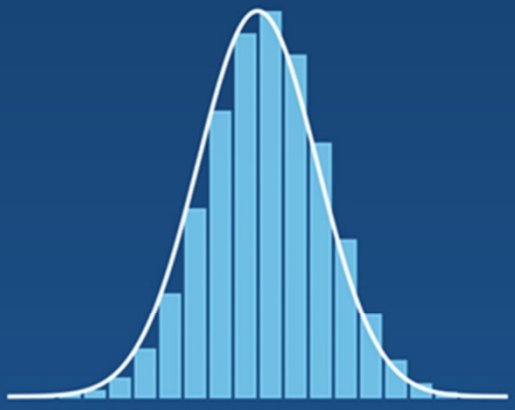
CĂLIN MUNTEAN

BIOSTATISTICĂ APLICATĂ

ÎN MEDICINA GENERALĂ

Date, metode, analiză și decizie. Orizont 2026

Curs și lucrări practice pentru studenții Facultății de Medicină





IC 95%	128,96-133,30
RR	2,62
AUC	0,85

MANUALE

Editura „Victor Babeș”
Timișoara, 2026

Editura „Victor Babeş”

Piața Eftimie Murgu nr. 2, cam. 316, 300041 Timișoara

Tel./Fax 0256 495 210

e-mail: evb@umft.ro

https://www.umft.ro/ro/organizare-evb/

Director: Prof. univ. dr. Sorin Ursoniu

Colecția: MANUALE

Coordonatori colecție: **Prof. univ. dr. Codruța Șoica**

Prof. univ. dr. Daniel Lighezan

Referent științific: Prof. univ. dr. Daliborca Vlad

Indicativ CNCSIS: 324

© 2026

Toate drepturile asupra acestei ediții sunt rezervate. Reproducerea parțială sau integrală a textului, pe orice suport, fără acordul scris al autorilor este interzisă și se va sancționa conform legilor în vigoare.

ISBN 978-606-786-652-0

Pacienții din exemple sunt fictivi. Setul de date didactic "mg2026" a fost generat prin simulare și nu provine din dosare de pacienți. Cifrele preluate din articole publicate sunt reproduse așa cum apar în rezumatele indexate în PubMed. Informațiile despre programe, versiuni și reglementări reflectă situația la 9 octombrie 2026. Volumul nu înlocuiește ghidurile clinice și reglementările în vigoare, care trebuie consultate în forma lor actualizată.

Cuprins

Prefață ... 11
Mulțumiri ... 12
Cum se citește acest volum ... 12
Lista abrevierilor ... 14
Partea I. Curs ... 16
Capitolul 1. Introducere în biostatistică: concepte fundamentale, rolul în cercetarea medicală și elemente de probabilități ... 17
1.1 De ce are nevoie un medic de biostatistică ... 17
1.2 Concepte de bază: unitate statistică, populație, eșantion ... 19
1.3 Setul didactic mg2026 ... 20
1.4 Eșantionarea: reprezentativitate și metode ... 21
1.5 Erorile sistematice (bias) ... 23
1.6 Statistica descriptivă și statistica inferențială ... 25
1.7 Etapele unui studiu statistic și întrebarea de cercetare ... 27
1.8 Datele în epoca dosarului electronic: legătura cu informatica medicală ... 28
1.9 Elemente de teoria probabilităților ... 31
1.10 Programele folosite în acest curs ... 36
1.11 Probleme rezolvate ... 37
1.12 Exerciții propuse ... 41
Capitolul 2. Descrierea și organizarea datelor: tipuri de variabile, indicatori descriptivi și reprezentări grafice ... 45
2.1 De la pacient la tabelul de date ... 45
2.2 Tabelele de frecvențe ... 49
2.3 Indicatorii tendinței centrale ... 51
2.4 Indicatorii de dispersie ... 54
2.5 Cuantile, percentile, cuartile și rezumatul în cinci numere ... 56
2.6 Forma distribuției: o primă privire asupra asimetriei și excesului ... 58
2.7 Reprezentările grafice ... 58
2.8 Tabelul descriptiv dintr-un articol: "Tabelul 1" ... 61
2.9 Analiza descriptivă completă a setului mg2026 ... 63
2.10 Probleme rezolvate ... 65
2.11 Exerciții propuse ... 69
Capitolul 3. Distribuțiile statistice: distribuția normală, verificarea normalității, asimetria, excesul, scorul z și transformarea datelor ... 73
3.1 Variabile aleatoare și distribuții de probabilitate ... 73
3.2 Două distribuții discrete utile în medicină ... 74

3.3 Distribuția normală ... 76
3.4 Distribuția normală standard și scorul z ... 78
3.5 Intervalele de referință în medicina de laborator ... 80
3.6 Distribuțiile de eșantionare: t, chi-pătrat și F ... 82
3.7 Teorema limitei centrale ... 83
3.8 Verificarea normalității ... 84
3.9 Transformarea datelor ... 89
3.10 Valorile extreme ... 91
3.11 Distribuții care se suprapun: de la scorul z la deciziile clinice ... 93
3.12 Probleme rezolvate ... 94
3.13 Exerciții propuse ... 97
Capitolul 4. Estimarea parametrilor și intervalele de încredere; principii de calcul al dimensiunii eșantionului ... 101
4.1 Estimarea punctuală și estimarea prin interval ... 101
4.2 Distribuția de eșantionare a mediei și eroarea standard ... 102
4.3 Intervalul pentru valori individuale și intervalul pentru medie ... 103
4.4 Intervalul de încredere pentru medie cu distribuția t ... 104
4.5 Interpretarea corectă a unui interval de încredere ... 106
4.6 Lățimea intervalului de încredere și mărimea eșantionului ... 107
4.7 Intervalul de încredere pentru o proporție ... 108
4.8 Intervalul de încredere pentru diferența a două medii și a două proporții ... 110
4.9 Intervale de încredere pentru date asimetrice ... 111
4.10 Intervale de încredere pentru rapoarte (OR, RR) și pentru aria ROC ... 111
4.11 Bootstrap, explicat simplu ... 113
4.12 Intervalul de încredere și valoarea p ... 114
4.13 Principii de calcul al dimensiunii eșantionului ... 115
4.14 Probleme rezolvate ... 118
4.15 Exerciții propuse ... 120
Capitolul 5. Bazele testării statistice: ipoteze, prag de semnificație, erori de tip I și II, decizia și interpretarea ... 124
5.1 De la întrebarea clinică la ipoteza testabilă ... 124
5.2 Logica testului: statistica testului și distribuția ei sub H0 ... 126
5.3 Valoarea p ... 126
5.4 Pragul de semnificație α și regula de decizie ... 128
5.5 Erorile de tip I și II, puterea testului ... 129
5.6 Mărimea efectului și semnificația clinică ... 132
5.7 Trei căi spre aceeași decizie: valoarea critică, valoarea p și intervalul de încredere ... 135

5.8 Algoritmul unui test statistic 135
5.9 Testele z și t pentru o singură medie 136
5.10 Comparațiile multiple și inflația erorii de tip I 138
5.11 p-hacking, HARKing și preînregistrarea 140
5.12 Echivalență și non-inferioritate, pe scurt 141
5.13 Cum se raportează un test statistic 142
5.14 Probleme rezolvate 142
5.15 Exerciții propuse 145
Capitolul 6. Teste statistice parametrice uzuale: testul t pentru eșantioane pereche și independente, analiza variației (ANOVA) 149
6.1 Familia testelor parametrice 149
6.2 Compararea unei medii cu o valoare dată: testul z și testul t 150
6.3 Testul t pentru eșantioane pereche 150
6.4 Testul t pentru eșantioane independente 153
6.5 Analiza variației (ANOVA) cu un factor 156
6.6 Testele post hoc 160
6.7 ANOVA cu măsurători repetate 162
6.8 ANOVA cu doi factori și ANCOVA, pe scurt 163
6.9 Raportarea testelor t și ANOVA 163
6.10 Probleme rezolvate 164
6.11 Exerciții propuse 167
Capitolul 7. Teste statistice neparametrice uzuale: Mann-Whitney U, Wilcoxon, Kruskal-Wallis, Friedman și testul chi-pătrat 171
7.1 Când nu mai putem folosi testele parametrice 171
7.2 Rangurile: materia primă a testelor neparametrice 173
7.3 Testul Mann-Whitney U 174
7.4 Testul Wilcoxon pentru date pereche și testul semnului 177
7.5 Testul Kruskal-Wallis 179
7.6 Testul Friedman 182
7.7 Testul chi-pătrat pentru tabele de contingență 184
7.8 Variante ale testului chi-pătrat 189
7.9 Dincolo de valoarea p: phi, V Cramér și alte măsuri ale asocierii 190
7.10 Raportarea rezultatelor testelor neparametrice și a testului chi-pătrat 191
7.11 Probleme rezolvate 192
7.12 Exerciții propuse 197
Capitolul 8. Selecția și aplicarea testelor statistice: criterii de alegere, rezolvarea de probleme și studii de caz 201
8.1 Alegerea testului este o decizie de proiectare, nu de calcul 201

8.2 Algoritmul de alegere a testului 201
8.3 Tabelul-sinteză al testelor 204
8.4 Unitatea de analiză și independența observațiilor 206
8.5 Testul se stabilește înainte de a vedea datele 206
8.6 Greșeli frecvente în alegerea și aplicarea testului 207
8.7 Raportarea unui rezultat în format standardizat (a-g) 208
8.8 Probleme rezolvate 209
8.9 Recunoașterea testului greșit: zece descrieri de tip "Metode" 218
8.10 Lectura critică a secțiunii Metode: articole ale colegilor din disciplină 219
8.11 Exerciții propuse 224
Capitolul 9. Analiza relațiilor dintre variabile: corelația și regresia liniară pentru variabile numerice și ordinale 227
9.1 Analiza bivariată: două variabile măsurate pe același pacient 227
9.2 Covarianța: primul pas spre o măsură a asocierii 229
9.3 Coeficientul de corelație Pearson 229
9.4 Corelația pentru date ordinale sau asimetrice: coeficienții Spearman și Kendall 234
9.5 Corelația nu înseamnă cauzalitate 236
9.6 Corelație și concordanță: metoda Bland-Altman 237
9.7 Regresia liniară simplă 238
9.8 Regresia liniară multiplă: introducere 242
9.9 Regresia logistică: o punte spre capitolul 10 245
9.10 Probleme rezolvate 246
9.11 Exerciții propuse 249
Capitolul 10. Aplicații statistice în epidemiologie: tipuri de studii observaționale și analiza riscului 253
10.1 Factorul de risc 253
10.2 Măsurarea frecvenței bolii 254
10.3 Tipuri de studii epidemiologice 255
10.4 Tabelul de contingență 2 x 2 259
10.5 Riscul relativ 259
10.6 Raportul de șanse 260
10.7 Intervale de încredere și testarea statistică 261
10.8 Măsuri de impact: riscul atribuibil, fracția atribuibilă, NNT și NNH 263
10.9 Confuzia, stratificarea și modificarea efectului 264
10.10 De la asociere la cauzalitate 267
10.11 Probleme rezolvate 268
10.12 Exerciții propuse 272

Capitolul 11. Metode statistice în decizia medicală: analiza bayesiană și medicina bazată pe dovezi 276
11.1 Incertitudinea în decizia clinică 276
11.2 Probabilitatea condiționată, revăzută 277
11.3 Teorema lui Bayes 279
11.4 Șanse, rapoarte de verosimilitate și forma Bayes cu șanse 281
11.5 Testarea secvențială și combinarea testelor 285
11.6 Diagnosticul diferențial cu mai multe boli 286
11.7 Pragurile de decizie: a testa, a trata 288
11.8 Screeningul în medicina de familie: exemplul FINDRISC 289
11.9 Inferența bayesiană față de inferența frecventistă 290
11.10 Medicina bazată pe dovezi 293
11.11 Erori cognitive în raționamentul clinic 294
11.12 Comunicarea riscului către pacient 295
11.13 Probleme rezolvate 295
11.14 Exerciții propuse 299
Capitolul 12. Evaluarea performanței testelor diagnostice: sensibilitate, specificitate, valori predictive și curba ROC 304
12.1 Testul de evaluat și standardul de referință 304
12.2 Tabelul 2 x 2 și indicatorii de performanță 305
12.3 Reconstituirea unui tabel 2 x 2 308
12.4 Dependența valorilor predictive de prevalență 309
12.5 Biasuri în studiile de acuratețe diagnostică 310
12.6 Teste cu rezultat continuu și alegerea pragului 311
12.7 Curba ROC 312
12.8 Acordul dintre evaluatori 316
12.9 Evaluarea clasificatorilor de inteligență artificială 317
12.10 Raportarea studiilor de acuratețe diagnostică: STARD 2015 și STARD-AI 318
12.11 Probleme rezolvate 318
12.12 Exerciții propuse 323
Capitolul 13. Studiile clinice: clasificare, proiectare, protocol, instrumente de culegere a datelor și analiza integrată a datelor biomedicale 327
13.1 Ce este un studiu clinic și ce îi cere statisticii 327
13.2 Clasificarea studiilor clinice 328
13.3 Principiile metodologice ale unui studiu randomizat 332
13.4 Protocolul, înregistrarea și cadrul etic 335
13.5 Instrumentele de culegere a datelor 337
13.6 Managementul și curățarea datelor 341

13.7 Raportarea: CONSORT 2025 ... 343
13.8 Analiza integrată a studiului mg2026: un exemplu complet ... 344
13.9 Probleme rezolvate ... 349
13.10 Exerciții propuse ... 353
Capitolul 14. Exemple recapitulative: interpretarea integrată a rezultatelor statistice ... 357
14.1 Harta disciplinei ... 357
14.2 Cum citim statistica unui articol medical ... 359
14.3 Studii de caz integrate ... 360
14.4 Cum se scrie paragraful de rezultate ... 371
14.5 Patruzeci de afirmații: adevărat sau fals ... 372
14.6 Tabelul-sinteză al formulelor ... 374
14.7 Probleme rezolvate ... 376
14.8 Exerciții propuse ... 378
Partea a II-a. Lucrări practice ... 385
Lucrarea practică 1. Introducere în biostatistică și instrumente digitale: Google Sheets, JASP și MedCalc ... 386
Lucrarea practică 2. Descrierea și organizarea datelor: chestionarul medical, tabele de frecvențe și grafice ... 394
Lucrarea practică 3. Distribuții și testarea normalității ... 401
Lucrarea practică 4. Estimare și intervale de încredere: efectul dimensiunii eșantionului ... 408
Lucrarea practică 5. Testarea ipotezelor statistice: formularea ipotezelor, erori și decizie ... 414
Lucrarea practică 6. Teste parametrice: testul t și ANOVA ... 421
Lucrarea practică 7. Teste neparametrice: Mann-Whitney, Wilcoxon, Kruskal-Wallis, Friedman și chi-pătrat ... 428
Lucrarea practică 8. Selecția testului statistic și raportarea standardizată a rezultatelor ... 435
Lucrarea practică 9. Analiza corelației și regresiei: Pearson, Spearman și regresia liniară simplă ... 441
Lucrarea practică 10. Analize epidemiologice: riscul relativ, raportul de șanse și interpretarea studiilor observaționale ... 447
Lucrarea practică 11. Analiza bayesiană: raționamentul probabilistic în diagnostic și screening 454
Lucrarea practică 12. Performanța testelor diagnostice: sensibilitate, specificitate, valori predictive și curba ROC ... 460
Lucrarea practică 13. Proiectarea unui mini-studiu clinic și analiza integrată a unui set complet de date ... 466
Lucrarea practică 14. Examenul practic: model de subiect, barem și plan de pregătire ... 473
Anexa A. Răspunsuri la întrebările de autoevaluare și la exercițiile propuse ... 482
Capitolul 1 ... 482
Capitolul 2 ... 483
Capitolul 3 ... 483

Capitolul 4 484
Capitolul 5 485
Capitolul 6 486
Capitolul 7 487
Capitolul 8 488
Capitolul 9 490
Capitolul 10 491
Capitolul 11 492
Capitolul 12 493
Capitolul 13 494
Capitolul 14 495
Anexa B. Glosar 497
Anexa C. Setul de date didactic mg2026: scenariu, dicționar de date și erorile din fișierul brut. 518
C.1 Scenariul 518
C.2 Fișierele 518
C.3 Dicționarul de date 519
C.4 Erorile din fișierul brut și corectarea lor 521
C.5 Cum a fost generat setul 523
C.6 Cum se citează setul în referate și teme 523
Anexa D. Fișa de formule 525
D.1 Capitolul 1. Probabilități 525
D.2 Capitolul 2. Descrierea datelor 526
D.3 Capitolul 3. Distribuția normală, scorul z, transformări 527
D.4 Capitolul 4. Estimare, intervale de încredere, dimensiunea eșantionului 528
D.5 Capitolul 5. Bazele testării ipotezelor 529
D.6 Capitolul 6. Teste parametrice: t și ANOVA 529
D.7 Capitolul 7. Teste neparametrice și testul chi-pătrat 531
D.8 Capitolul 8. Alegerea testului (repere numerice) 532
D.9 Capitolul 9. Corelație și regresie 532
D.10 Capitolul 10. Epidemiologie și analiza riscului 533
D.11 Capitolul 11. Bayes și medicina bazată pe dovezi 534
D.12 Capitolul 12. Teste diagnostice și curba ROC 534
D.13 Capitolul 13. Studii clinice (formule de planificare) 535
Anexa E. Tabele statistice 537
E.1 Distribuția normală standard: funcția de repartiție Φ(z) 537
E.2 Valori critice ale distribuției normale standard 538
E.3 Valori critice ale distribuției Student (t) 538

E.4 Valori critice ale distribuției chi-pătrat (χ^2) 539
E.5 Valori critice ale distribuției Fisher (F) pentru $\alpha = 0,05$ 540
E.6 Valori critice ale coeficientului de corelație Pearson (r) 541
E.7 Valori critice aproximative ale coeficientului de corelație Spearman (rho) 542
E.8 Tabel de 1000 de cifre aleatoare 543
Anexa F. Teste-grilă de antrenament pentru examen 545
Cum sunt construite întrebările și cum se punctează 545
Testul 1 545
Testul 2 553
Testul 3 561
Răspunsuri și explicații 569
Bibliografie 583

Prefață

Un medic de familie primește dimineața rezultatele a patruzeci de analize, citește la prânz un ghid nou de hipertensiune și seara un articol despre un antidiabetic. În fiecare dintre aceste momente face, fără să-și dea seama, statistică. Se întreabă dacă o glicemie de 118 mg/dL este o abatere întâmplătoare sau un semnal, dacă scăderea tensiunii raportată în studiu ar conta pentru pacienții lui și cât de mult se poate baza pe un scor de risc completat în cinci minute. Biostatistica nu este o disciplină a calculelor. Este felul în care medicina își măsoară incertitudinea.

Volumul de față este cursul de Biostatistică pentru studenții anului I de la Facultatea de Medicină a Universității de Medicină și Farmacie "Victor Babeș" din Timișoara. El continuă manualul de informatică medicală din primul semestru [1], păstrându-i forma și regulile de lucru, și urmează fișa disciplinei pentru anul universitar 2026-2027: 14 teme de curs și 14 lucrări practice, câte o pereche pe săptămână, cu examen scris de tip grilă, în sistemul folosit la rezidențiat, și examen practic în laborator. Am ales titlul "Biostatistică aplicată în medicina generală" pentru că toate exemplele vin din medicina pe care o va practica orice absolvent, indiferent de specialitatea aleasă mai târziu: cabinetul de medicină de familie, medicina internă, diabetul și bolile metabolice, hipertensiunea, bolile respiratorii, hepatologia, cardiologia, pediatria, obstetrica, urgențele și bolile infecțioase.

Prima parte a cursului, capitolele 1-4, așază fundația: ce sunt datele și de unde vin, cum se descriu, ce înseamnă o distribuție și cum se estimează un parametru al populației dintr-un eșantion. Capitolul 1 face și legătura cu informatica medicală din primul semestru, pentru că o analiză statistică nu poate fi mai bună decât datele pe care le prelucrează, iar datele de sănătate cer reguli de etică și de protecție pe care medicul trebuie să le cunoască. Capitolele 5-8 tratează testarea ipotezelor: logica testului, testele parametrice și neparametrice uzuale și, mai ales, alegerea testului potrivit, cea mai frecventă sursă de greșeli la examen și în articolele publicate. Capitolele 9-12 privesc asocierea și decizia: corelația și regresia, măsurile riscului din epidemiologie, raționamentul bayesian și evaluarea testelor diagnostice. Capitolul 13 arată cum se proiectează un studiu clinic și cum se construiește un formular de colectare a datelor, iar capitolul 14 adună totul în exemple de interpretare integrată.

Biostatistica se învață greu din definiții și mult mai ușor din exemple pe care cititorul le recunoaște. De aceea volumul folosește două surse de exemple. Prima este setul de date didactic "mg2026", un studiu randomizat simulat în cabinete de medicină de familie, cu 160 de pacienți cu risc cardiometabolic urmăriți șase luni. Același set apare în toate capitolele și în toate lucrările practice, astfel încât aceeași medie, același interval de încredere și același test să poată fi recalculate de student cu programele gratuite din laborator, iar rezultatul obținut să poată fi comparat cu cel din carte.

A doua sursă este cercetarea publicată de colegii din disciplină, de cele mai multe ori împreună cu clinicile universității. Studiile despre elastografia hepatică și ecografia de contrast [2,3], despre comportamentele alimentare și indicele de masă corporală la studenți [4-6], despre implicarea digitală a pacienților cu diabet [7], despre evoluția COVID-19 la pacienții cu diabet de tip 1 și de tip 2 [8], despre diferențele dintre femei și bărbați în urgențele cardiovasculare [9] sau despre validarea chestionarelor în limba română [10,11] sunt citite în carte ca studii de caz: ce întrebare au pus, ce design au avut, ce test au folosit, ce spun și ce nu spun cifrele lor. Studentul care învață să citească aceste articole cu creionul în mână va putea citi la fel orice articol din specialitatea pe care o va alege.

Instrumentele sunt cele din fișa disciplinei, toate gratuite sau cu licență academică: Google Forms și Google Sheets pentru colectarea și organizarea datelor, JASP și MedCalc pentru analiză, Epi Info pentru tabelele epidemiologice și pentru lucrul fără rețea. Fiecare lucrare practică indică meniurile exacte și rezultatele așteptate, astfel încât studentul să-și poată verifica singur munca.

Am păstrat regulile de lucru ale volumului de informatică medicală. Fiecare referință a fost găsită efectiv în PubMed, în CrossRef sau pe pagina instituției emitente înainte de a fi citată. Cifrele din studiile publicate sunt preluate așa cum apar în rezumatul indexat. Graficele construite pe setul didactic sunt marcate ca atare. Pacienții din exemple sunt fictivi. Acolo unde o informație nu a putut fi verificată, am preferat să nu o scriu.

Mulțumiri

Această carte datorează mult colegilor de catedră, cu care predau aceeași disciplină de ani buni, cu care am scris îndrumătoarele de lucrări practice ale disciplinei [12-14] și alături de care am lucrat la analiza statistică a multor studii clinice. Le mulțumesc, pentru colaborarea didactică și științifică din care s-a format acest curs, domnului Conf. univ. dr. Radu Moleriu, doamnei Conf. univ. dr. Anca Tudor, doamnei Șef lucrări dr. Mirela Frandeș, doamnei Șef lucrări dr. Lavinia Moleriu, domnului Asist. univ. dr. Flavius Cioca și doamnei Asist. univ. dr. Raluca Lupușoru. Multe dintre exemplele volumului provin din articole la care au contribuit, iar multe dintre problemele de laborator s-au șlefuit în discuțiile noastre de la sfârșitul orelor. Mulțumesc și studenților din ultimii ani, ale căror întrebări, uneori incomode, au hotărât ce merita explicat mai pe larg. Erorile rămase îmi aparțin.

Timișoara, octombrie 2026

Călin Muntean

Cum se citește acest volum

Volumul poate fi citit în ordine, săptămână de săptămână, dar nu este obligatoriu. Capitolele trimit unele la altele, iar harta disciplinei din capitolul 14 arată cum se leagă. Cine are nevoie repede de o temă, de exemplu de calculul raportului de șanse sau de alegerea între testul t și testul Mann-Whitney, poate porni direct de la capitolul respectiv și de la lucrarea practică pereche.

Fiecare capitol de curs este construit la fel. Începe cu obiective operaționale, formulate ca acțiuni pe care studentul le poate face la final. Conținutul trece de la noțiuni la metode, cu exemple calculate de mână și apoi pe setul didactic. Studiile de caz din cercetarea disciplinei sunt puse în casete separate. Urmează probleme rezolvate pas cu pas și exerciții propuse, o casetă cu capcanele frecvente, adică greșelile văzute repetat la laborator și la examen, apoi o casetă cu ideile de reținut. Capitolul se încheie cu zece întrebări de tip grilă, în formatul examenului: complement simplu (CS), cu un singur răspuns corect, și complement multiplu (CM), cu două până la patru răspunsuri corecte. Răspunsurile și rezolvările exercițiilor se află în anexa A.

Lucrările practice au un obiectiv, instrumentele cu versiunea folosită, pașii de lucru cu rezultatele așteptate, exerciții suplimentare, ce se predă cu un barem orientativ și o variantă pentru sălile fără rețea. Toate folosesc setul de date "mg2026", livrat în două variante: fișierul brut, cu erorile de introducere lăsate intenționat, și fișierul curat. Dicționarul de date se află în anexa C, fișa de formule în anexa D, tabelele statistice în anexa E, iar trei teste-grilă complete de antrenament în anexa F.

Pentru pregătirea curentă recomand trei pași. Întâi se citesc obiectivele. Apoi se parcurge textul cu figurile, tabelele și problemele rezolvate, refăcând calculele pe hârtie sau în foaia de calcul. În sfârșit, se rezolvă întrebările și exercițiile fără a privi răspunsurile din anexa A.

Capcanele frecvente merită citite de două ori, o dată înainte de lucrarea practică și o dată înainte de examen.

Citările sunt numerotate în stil Vancouver, în ordinea primei apariții, iar lista completă se află la finalul volumului. Abrevierile sunt reunite într-o listă separată, iar termenii principali sunt definiți pe scurt în glosarul din anexa B, cu trimitere la capitolul în care sunt explicați. În text se folosește virgula zecimală, ca în limba română (0,05; 131,13 mmHg), chiar dacă programele de analiză afișează adesea punctul zecimal.

Lista abrevierilor

Lista abrevierilor folosite în volum, în ordine alfabetică. Pentru termenii preluați din engleză, denumirea originală apare în paranteză.

Abreviere	Denumire completă
AN	adevărat negativ (engl. true negative, TN)
ANOVA	analiza variației (engl. analysis of variance)
AP	adevărat pozitiv (engl. true positive, TP)
AS	abatere standard (engl. standard deviation, SD)
AUC	aria de sub curba ROC (engl. area under the curve)
BMI	vezi IMC (engl. body mass index)
BPOC	bronhopneumopatie obstructivă cronică
CI	vezi IC (engl. confidence interval)
CM	complement multiplu (tip de întrebare grilă)
CONSORT	ghid de raportare a studiilor randomizate (engl. Consolidated Standards of Reporting Trials)
CRF	formular de raportare a cazului (engl. case report form)
CS	complement simplu (tip de întrebare grilă)
CV	coeficient de variație
DOR	raportul de șanse diagnostic (engl. diagnostic odds ratio)
ES	eroarea standard (engl. standard error, SE)
FINDRISC	scorul finlandez de risc pentru diabet (engl. Finnish Diabetes Risk Score)
FN	fals negativ
FP	fals pozitiv
GDPR	Regulamentul general privind protecția datelor, Regulamentul (UE) 2016/679
gl	grade de libertate (engl. degrees of freedom, df)
GRADE	sistemul de evaluare a certitudinii dovezilor (engl. Grading of Recommendations Assessment, Development and Evaluation)
H0, H1	ipoteza nulă, ipoteza alternativă
HbA1c	hemoglobina glicată
HR	raportul hazardurilor (engl. hazard ratio)
HTA	hipertensiune arterială
IC	interval de încredere (engl. confidence interval)
ICC	coeficient de corelație intraclasă (engl. intraclass correlation coefficient)
IMC	indicele de masă corporală
IQR	interval intercuartil (engl. interquartile range)
ITT	analiza după intenția de tratament (engl. intention to treat)
MBE	medicina bazată pe dovezi (engl. evidence-based medicine)
NNH	numărul de pacienți expuși pentru a produce un eveniment advers (engl. number needed to harm)
NNT	numărul de pacienți care trebuie tratați pentru a preveni un eveniment (engl. number needed to treat)
OR	raportul șanselor (engl. odds ratio)
PICO	populație, intervenție, comparator, rezultat (engl. population, intervention, comparison, outcome)
PRISMA	ghid de raportare a recenziilor sistematice
RAR	reducerea absolută a riscului
RCT	studiu randomizat controlat (engl. randomized controlled trial)
ROC	curba caracteristicii de operare a receptorului (engl. receiver operating characteristic)

Abreviere	**Denumire completă**
RP	raportul prevalențelor (engl. prevalence ratio)
RR	riscul relativ (engl. relative risk)
RRR	reducerea relativă a riscului
RV+, RV-	raportul de verosimilitate pozitiv și negativ (engl. likelihood ratio, LR)
Se	sensibilitate
SEE	eroarea standard a estimării în regresie (engl. standard error of the estimate)
Sp	specificitate
STARD	ghid de raportare a studiilor de acuratețe diagnostică
STROBE	ghid de raportare a studiilor observaționale
TAD	tensiunea arterială diastolică
TAS	tensiunea arterială sistolică
TTGO	testul oral de toleranță la glucoză
VPN	valoare predictivă negativă
VPP	valoare predictivă pozitivă

Partea I. Curs

Capitolul 1. Introducere în biostatistică: concepte fundamentale, rolul în cercetarea medicală și elemente de probabilități

Obiective

La finalul acestui capitol, studentul va putea:

- să explice de ce variabilitatea biologică face necesară statistica în medicină și să dea exemple din practica de cabinet;
- să definească unitatea statistică, populația, eșantionul și volumul acestora și să le identifice într-un articol publicat;
- să compare recensământul, screeningul și eșantionarea și să aleagă o metodă de eșantionare potrivită pentru o întrebare dată;
- să recunoască principalele tipuri de erori sistematice (de selecție, de informare, de confuzie) într-un scenariu clinic;
- să deosebească statistica descriptivă de cea inferențială și să folosească notațiile pentru parametri (μ, σ, π) și pentru statistici (m, s, p_i);
- să descrie etapele unui studiu statistic și să formuleze o întrebare de cercetare în format PICO;
- să enumere sursele de date biomedicale și principiile etice și legale ale prelucrării lor (GDPR, Legea 190/2018, anonimizare, pseudonimizare, consimțământ, avizul comisiei de etică);
- să calculeze probabilități simple, condiționate și totale dintr-un tabel de contingență și să verifice independența a două evenimente;
- să calculeze probabilități binomiale simple și să recunoască situațiile în care se folosește distribuția Poisson.

1.1 De ce are nevoie un medic de biostatistică

Primul contact al multor studenți cu biostatistica este o întrebare sinceră: "La ce îmi trebuie asta, dacă eu vreau să fiu chirurg?". Răspunsul nu vine din matematică, ci din ziua obișnuită a unui medic. Dimineața, medicul de familie are pe birou rezultatele de laborator ale pacienților văzuți săptămâna trecută. O glicemie a jeun de 118 mg/dL la un pacient care avea 104 mg/dL acum un an este o agravare reală sau o fluctuație fără importanță? La prânz citește rezumatul unui studiu care arată că un antidiabetic nou scade riscul de spitalizare pentru insuficiență cardiacă "cu 30%". Seara completează pentru un pacient un scor de risc și trebuie să-i spună acestuia ce înseamnă un rezultat "crescut". Fiecare dintre aceste situații cere o judecată despre incertitudine, iar biostatistica este limbajul în care se exprimă această judecată.

Medicina bazată pe dovezi a fost definită ca utilizarea conștientă, explicită și judicioasă a celor mai bune dovezi disponibile în deciziile privind îngrijirea pacientului individual, integrate cu experiența clinică a medicului și cu valorile pacientului [15]. Cuvântul "dovezi" ascunde aproape întotdeauna un calcul statistic: un interval de încredere, o valoare p, un risc relativ, o sensibilitate. Ghidurile clinice pe care le veți folosi toată viața își clasifică recomandările după calitatea studiilor din care provin, iar calitatea unui studiu se judecă în mare parte după designul și analiza lui statistică. Un medic care nu înțelege ce înseamnă "IC 95% 0,62-0,91" nu poate verifica singur dacă o recomandare se sprijină pe date solide sau pe un rezultat fragil.

Există și un motiv foarte concret, apropiat în timp: examenul de rezidențiat. Bibliografia lui conține capitole de epidemiologie, de evaluare a testelor diagnostice și de interpretare a studiilor

clinice. Întrebările despre sensibilitate și valoare predictivă, despre riscul relativ sau despre tipurile de studii revin an de an. Examenul acestei discipline este construit în același stil, cu întrebări cu complement simplu și complement multiplu, tocmai ca să vă obișnuiți devreme cu acest tip de raționament.

Ce cred studenții la medicină despre aceste discipline? Un sondaj anonim pe 121 de studenți de la universitatea noastră a cerut aprecierea mai multor cursuri pe o scală de la 1 la 10 [16]. Cursul de medicină de urgență a primit o mediană de 9 (IQR 7-10), cel de informatică biomedicală 7 (5-9), iar biostatistica 8 (5-9). Rezultatul cel mai interesant este însă altul: corelația dintre cunoștințele măsurate printr-un test și cele autoevaluate de studenți a fost practic nulă ($R = 0{,}107$, $p = 0{,}246$). Altfel spus, studenții nu își pot aprecia corect nivelul de cunoștințe în acest domeniu; unii cred că știu mai mult decât știu, alții se subestimează. Observați și cum au fost raportate rezultatele: mediană și interval intercuartilic, pentru că o notă pe o scală de la 1 la 10 este o variabilă ordinală, nu o măsurătoare propriu-zisă (capitolul 2). Despre prețul și oportunitățile predării cu ajutorul tehnologiei informației în anii pandemiei, colegii din disciplină au publicat o analiză separată [17]; multe dintre instrumentele digitale folosite în lucrările practice din acest volum s-au consolidat atunci.

Nu în ultimul rând, mulți dintre voi veți face cercetare: o lucrare de licență, un proiect în cadrul unui cerc studențesc, o teză de doctorat. Fără noțiuni de biostatistică, colaborarea cu un statistician devine un monolog. Statisticianul nu poate alege singur testul potrivit dacă nu înțelege întrebarea clinică, iar medicul nu poate formula întrebarea corect dacă nu știe ce poate și ce nu poate oferi o analiză. Cartea de față încearcă să vă pregătească pentru acest dialog, așa cum o fac și manualele de referință ale disciplinei [14,18,19].

1.1.1 Variabilitatea, materia primă a biologiei

Doi pacienți cu aceeași vârstă, același sex și același diagnostic nu au niciodată aceeași tensiune arterială. Nici același pacient nu are aceeași tensiune la două măsurători consecutive, făcute la cinci minute distanță. Această variabilitate este trăsătura de bază a fenomenelor biologice și are mai multe surse. Există o variabilitate între indivizi (genetică, vârstă, stil de viață, boli asociate), o variabilitate în timp la același individ (ritm circadian, efort, stres, alimentație) și o variabilitate a măsurării (aparatul, tehnica, observatorul, condițiile de laborator).

Dacă toate valorile ar fi identice, statistica nu ar fi necesară: o singură măsurătoare ar spune totul despre întreaga populație. Tocmai pentru că valorile diferă, avem nevoie de metode care să descrie cât de mult diferă (dispersia), în jurul cărei valori se grupează (tendința centrală) și cât de sigure sunt concluziile trase dintr-un număr limitat de observații. În setul didactic mg2026, prezentat în secțiunea 1.3, tensiunea arterială sistolică la includere variază între 102 și 169 mmHg, deși toți cei 160 de pacienți au fost recrutați după aceleași criterii. Media este 131,13 mmHg, iar abaterea standard 13,88 mmHg; vom învăța în capitolul 2 ce înseamnă aceste cifre.

La laborator, studenții întreabă adesea dacă variabilitatea este "o eroare". Răspunsul este nu. Variabilitatea biologică reală este informație: ne spune cât de diferiți sunt pacienții între ei. Eroarea de măsurare se adaugă peste ea și o putem reduce prin aparate calibrate, protocoale standardizate și măsurători repetate. Prima nu se poate elimina și nici nu trebuie. Biostatistica o acceptă și învață să lucreze cu ea.

1.1.2 Statistică și biostatistică

Statistica este știința care se ocupă cu culegerea, organizarea, descrierea, analiza și interpretarea datelor care privesc fenomene de masă, adică fenomene care se manifestă la un număr mare de indivizi și care nu pot fi înțelese dintr-un singur caz. Biostatistica este aplicarea acestor metode la datele din biologie, medicină și sănătate publică. Diferența nu ține de formule, care sunt aceleași,

ci de tipul de probleme: studii clinice, teste diagnostice, factori de risc, supraviețuire, scoruri de severitate, chestionare de calitate a vieții.

Biostatistica are câteva particularități. Datele provin de la oameni, cu consimțământul lor, și sunt protejate de reguli etice și legale stricte. Măsurătorile au adesea unități clinice (mmHg, mg/dL, kg/m²) și praguri cu semnificație medicală (140 mmHg pentru hipertensiune, 126 mg/dL pentru diabet). Multe variabile sunt asimetrice (trigliceridele, durata spitalizării, titrurile de anticorpi), iar multe rezultate sunt binare (vindecat sau nu, decedat sau nu). Mai ales, concluziile se transformă în decizii despre pacienți, ceea ce face ca greșelile să aibă un cost real.

1.2 Concepte de bază: unitate statistică, populație, eșantion

Orice studiu începe cu o întrebare simplă: "Despre cine vorbim?". Răspunsul precis la această întrebare definește câteva noțiuni pe care le veți folosi în tot restul cursului.

1.2.1 Unitatea statistică și populația

Unitatea statistică (individul statistic) este elementul asupra căruia se fac observațiile, localizat în spațiu și în timp. În cele mai multe studii medicale, unitatea statistică este pacientul. Nu este însă întotdeauna așa. Într-un studiu de prescriere, unitatea poate fi rețeta; într-un studiu de oftalmologie, ochiul; într-un studiu de stomatologie, dintele sau suprafața dentară; într-un studiu de servicii de sănătate, cabinetul sau spitalul. Alegerea unității are consecințe: doi ochi ai aceluiași pacient nu sunt observații independente, iar analiza trebuie să țină cont de acest lucru.

Populația este ansamblul tuturor unităților statistice care au cel puțin o proprietate comună și care sunt, la rândul lor, localizate în spațiu și timp. "Adulții cu prediabet din județul Timiș în anul 2026" este o populație bine definită; "pacienții cu prediabet" nu este, pentru că nu spune unde și când. Numărul de unități din populație se numește volumul populației și se notează de obicei cu N.

În cercetare se face distincția între populația țintă, la care vrem să generalizăm concluziile, și populația accesibilă (sau sursă), din care putem recruta efectiv. Pentru un studiu din cabinetele de medicină de familie din vestul țării, populația țintă poate fi "adulții cu risc cardiometabolic din România", dar populația accesibilă este "adulții înscriși pe listele celor 12 cabinete participante". Distanța dintre cele două este una dintre primele întrebări pe care trebuie să le pună cititorul unui articol.

1.2.2 Eșantionul

Eșantionul este o submulțime a populației, aleasă pentru a fi studiată în locul ei. Numărul de unități din eșantion este volumul eșantionului și se notează cu n. Eșantionul de 160 de pacienți din setul mg2026 provine dintr-o populație accesibilă de câteva mii de adulți înscriși în cabinetele participante; analizele se fac pe cei 160, dar întrebarea de interes privește întreaga populație.

Volumul eșantionului nu se alege "cât mai mare" și nici "cât se poate". El se calculează, înainte de începerea studiului, în funcție de precizia dorită sau de diferența pe care vrem să o detectăm. Vom vedea în capitolul 4 că, pentru a estima media tensiunii sistolice cu o marjă de ±3 mmHg, când abaterea standard este de aproximativ 14 mmHg, sunt suficienți 84 de pacienți, dar pentru a estima o proporție de 20% cu o marjă de ±5% sunt necesari 246. Diferența vine din natura variabilei și din precizia cerută, nu din dorința cercetătorului.

1.2.3 Recensământ, screening, eșantionare

Există trei moduri principale de a culege date despre o populație. Recensământul observă toate unitățile populației la un moment dat. Este exact (nu există eroare de eșantionare), bine localizat în timp și spațiu, dar foarte costisitor, lent și rareori posibil în medicină. Recensământul populației

României se face o dată la zece ani; un registru național complet al unei boli rare se apropie de un recensământ al cazurilor.

Screeningul examinează sistematic o populație (sau un segment larg al ei) pentru a identifica persoanele cu o anumită afecțiune, adesea încă nediagnosticată. Exemplele clasice sunt screeningul cancerului de col uterin, al cancerului colorectal, al hipoacuziei la nou-născut sau al diabetului de tip 2 cu scorul FINDRISC în cabinetul de medicină de familie. Screeningul nu cere o localizare strictă în timp (se poate desfășura pe parcursul mai multor ani), are un cost relativ ridicat și urmărește un beneficiu individual, nu doar o estimare statistică.

Eșantionarea (selecția) studiază doar o parte a populației și extinde concluziile asupra întregului. Costul este redus, datele pot fi culese mai atent, iar rezultatul este disponibil mai repede. Prețul plătit este incertitudinea: rezultatul obținut pe eșantion diferă, în mod inevitabil, de valoarea adevărată din populație. Statistica inferențială există tocmai pentru a cuantifica această incertitudine. Aproape toate studiile pe care le veți citi sunt studii pe eșantioane.

1.3 Setul didactic mg2026

De-a lungul întregului curs vom lucra pe același set de date, ca să puteți urmări cum aceleași variabile sunt descrise, estimate, comparate și corelate. Setul se numește mg2026 și este disponibil pe Moodle în două variante: mg2026_curat.csv (160 de rânduri, gata de analiză) și mg2026_brut.csv (161 de rânduri, cu erori lăsate intenționat, pe care le veți curăța în lucrarea practică 1).

Scenariul este un studiu randomizat pragmatic simulat, desfășurat în 12 cabinete de medicină de familie din vestul țării. Au fost incluși 160 de adulți cu risc cardiometabolic: persoane supraponderale, cu prediabet sau cu diabet nediagnosticat anterior. Pacienții au fost repartizați aleator în trei brațe: A, consiliere standard privind stilul de viață (53 de pacienți); B, consiliere plus o aplicație mobilă de automonitorizare (53); C, consiliere, aplicație și teleconsultație lunară cu medicul de familie (54). Măsurătorile s-au făcut la includere, la 3 luni și la 6 luni. Diabetul a fost confirmat la includere prin testul oral de toleranță la glucoză (TTGO), care este standardul de referință față de care se evaluează scorul FINDRISC. Pacienții sunt fictivi, iar valorile au fost generate prin simulare, astfel încât să semene cu datele reale din cabinet, fără să provină de la vreo persoană.

Tabelul 1.1 rezumă variabilele setului. Dicționarul complet se găsește în anexa C și va fi folosit din nou în capitolul 2, când vom discuta tipurile de variabile.

Tabelul 1.1. Variabilele setului didactic mg2026.

Variabila	Semnificația	Tipul (vezi capitolul 2)
id	codul pacientului (P001-P160)	identificator
brat	brațul de tratament (A, B, C)	nominală
sex	F, M	nominală dihotomică
varsta	vârsta la includere, ani	numerică
mediu	urban, rural	nominală dihotomică
educatie	1 gimnaziu, 2 liceu sau școală profesională, 3 universitar, 4 postuniversitar	ordinală
fumat	nefumător, fost fumător, fumător	nominală
activitate	activitate fizică autoevaluată, 1-5	ordinală
inaltime, greutate0, imc0, talie	înălțime (cm), greutate (kg), IMC (kg/m²), circumferința taliei (cm)	numerice
tas0, tad0, fc0	tensiune sistolică, diastolică (mmHg), frecvență cardiacă (bătăi/min)	numerice
colesterol, hdl, trigliceride	profil lipidic (mg/dL)	numerice

Variabila	Semnificația	Tipul (vezi capitolul 2)
glicemie0, glicemie3, glicemie6	glicemie a jeun la 0, 3 și 6 luni (mg/dL)	numerice
hba1c0	hemoglobina glicată (%)	numerică
findrisc	scorul de risc pentru diabet (0-26)	numerică discretă (scor)
diabet_ogtt	diabet confirmat prin TTGO (0/1)	nominală dihotomică
greutate6, tas6	greutatea și TAS la 6 luni	numerice
scadere5	scădere în greutate ≥ 5% la 6 luni (0/1)	nominală dihotomică
satisfactie	satisfacția față de program, 1-5 (Likert)	ordinală

Câteva cifre vă vor ajuta să vă orientați de la început. Din cei 160 de pacienți, 86 sunt femei și 74 bărbați; 95 locuiesc în mediul urban și 65 în mediul rural; 74 sunt nefumători, 33 foști fumători și 53 fumători. Vârsta medie este 56,29 ani, cu limite între 35 și 79 de ani. Diabetul a fost confirmat prin TTGO la 29 de pacienți, adică 18,1%. Aceste cifre vor reapărea, sub forme diferite, în aproape fiecare capitol.

Un detaliu de design merită reținut de acum: la includere, datele mg2026 pot fi analizate ca un studiu transversal (de exemplu, asocierea dintre obezitate și diabet), iar urmărirea la 3 și 6 luni pe brațe randomizate le transformă într-un studiu clinic controlat. Același set de date poate răspunde deci la întrebări de natură diferită, cu condiția să știm ce tip de concluzie permite fiecare parte a lui.

1.4 Eșantionarea: reprezentativitate și metode

Un eșantion este util doar dacă seamănă cu populația din care provine. Spunem că un eșantion este reprezentativ atunci când reproduce, în proporții apropiate, structura populației pentru caracteristicile care contează în studiu: vârstă, sex, mediu de proveniență, severitatea bolii și așa mai departe. Reprezentativitatea nu se poate garanta prin intuiție. Se obține prin modul în care sunt aleși indivizii.

1.4.1 Condițiile unui eșantion reprezentativ

Două condiții stau la baza unei selecții corecte. Prima este echiprobabilitatea: fiecare individ din populație trebuie să aibă aceeași probabilitate de a fi inclus în eșantion (sau, în variantele mai complexe, o probabilitate cunoscută și diferită de zero). A doua este independența: includerea unui individ nu trebuie să influențeze includerea altuia. Dacă un medic invită la studiu, pe lângă pacient, și pe toți membrii familiei acestuia, independența se pierde, pentru că membrii aceleiași familii seamănă între ei prin alimentație, obiceiuri și gene.

Orice condiție care face ca anumiți indivizi să aibă șanse mai mari de a fi selectați distorsionează eșantionul. Rezultatul este o eroare sistematică de selecție, despre care vom vorbi în secțiunea 1.5. O regulă practică: dacă nu puteți descrie exact cum a ajuns fiecare pacient în eșantion, nu puteți afirma că eșantionul este reprezentativ.

1.4.2 Eșantionarea aleatoare simplă

Eșantionarea aleatoare simplă este modelul teoretic de la care pornesc aproape toate formulele din statistică. Se pornește de la o listă completă a populației (cadrul de eșantionare), fiecare individ primește un număr, iar numerele incluse se aleg cu un generator de numere aleatoare. În Google Sheets, de exemplu, se poate adăuga o coloană cu =RAND() lângă lista pacienților, se sortează după această coloană și se păstrează primele n rânduri.

Să presupunem că un cabinet are pe listă 1 850 de adulți între 40 și 75 de ani și vrea să invite 120 dintre ei la o evaluare FINDRISC. Prin eșantionare aleatoare simplă, fiecare pacient are

aceeași probabilitate de a fi invitat, 120/1 850 = 0,065 (6,5%). Metoda este corectă, dar are două puncte slabe în practică: cere o listă completă și actualizată și poate produce, din întâmplare, un eșantion dezechilibrat (de exemplu, cu prea puțini bărbați), mai ales când n este mic.

1.4.3 Eșantionarea sistematică

În eșantionarea sistematică se calculează un pas $k = \frac{N}{n}$, se alege aleator un punct de pornire între 1 și k, apoi se ia fiecare al k-lea individ de pe listă. Pentru 1 200 de pacienți și un eșantion de 80, pasul este k = 15. Dacă numărul de pornire extras este 7, sunt incluși pacienții cu numerele 7, 22, 37, 52 și așa mai departe, până la 1 192.

Metoda este simplă și se potrivește foarte bine cu fluxul din cabinet: "fiecare al cincilea pacient care se prezintă la consultație în luna martie". Pericolul apare atunci când lista are o periodicitate care coincide cu pasul. Dacă programările sunt organizate astfel încât fiecare a cincea programare din zi este rezervată controalelor de urgență, un pas de 5 poate selecta numai astfel de pacienți. Înainte de a folosi metoda, verificați cum este ordonată lista.

1.4.4 Eșantionarea stratificată

Populația se împarte în straturi omogene după o caracteristică importantă (sex, grupă de vârstă, mediu urban sau rural), apoi se extrage un eșantion aleator din fiecare strat. În varianta proporțională, fiecare strat contribuie cu un număr de indivizi proporțional cu mărimea lui în populație. Dacă din 1 200 de pacienți 720 sunt din mediul urban (60%) și 480 din mediul rural (40%), un eșantion stratificat de 80 va conține 48 de pacienți din urban și 32 din rural.

Stratificarea crește precizia estimărilor atunci când caracteristica de stratificare influențează variabila studiată și garantează că grupurile mici nu lipsesc din eșantion. Uneori se folosește intenționat o stratificare disproporționată, cu suprareprezentarea unui grup rar (de exemplu, pacienții de peste 75 de ani), urmată de ponderarea rezultatelor la analiză.

1.4.5 Eșantionarea în cluster (pe grupuri)

Când nu există o listă a tuturor indivizilor, dar există o listă a unor grupuri naturale (cabinete, școli, sate, secții de spital), se aleg aleator câteva grupuri și se studiază toți indivizii din grupurile alese sau un eșantion din fiecare (eșantionare în două etape). Pentru un studiu național despre controlul tensiunii arteriale, este mult mai ușor să alegi aleator 40 de cabinete de medicină de familie și să incluzi câte 25 de pacienți din fiecare decât să alegi 1 000 de pacienți împrăștiați în toată țara.

Avantajul logistic are un cost statistic. Pacienții aceluiași cabinet seamănă între ei (același medic, aceeași zonă, aceleași obiceiuri de prescriere), deci 1 000 de pacienți din 40 de cabinete aduc mai puțină informație decât 1 000 de pacienți aleși individual. Analiza trebuie să țină cont de această grupare. Setul mg2026 are o structură asemănătoare, cu 12 cabinete participante, chiar dacă în acest curs îl vom analiza, pentru simplitate, ca pe un eșantion de indivizi.

1.4.6 Eșantionarea de conveniență

Eșantionul de conveniență cuprinde indivizii cel mai ușor de recrutat: pacienții care vin la consultație într-o anumită perioadă, voluntarii care răspund la un anunț, persoanele care completează un chestionar online distribuit pe rețelele sociale. Este de departe metoda cea mai folosită în cercetarea clinică, pentru că este rapidă și ieftină. Este și cea mai expusă la erori de selecție, pentru că nimic nu garantează că cei recrutați seamănă cu populația țintă.

Un eșantion de conveniență nu este automat inutil. Studiile clinice randomizate folosesc aproape întotdeauna pacienți recrutați consecutiv, deci un eșantion de conveniență, iar validitatea comparației dintre brațe vine din randomizare, nu din reprezentativitate. Problema apare atunci

când rezultatul obținut pe un eșantion de conveniență (de exemplu, o prevalență) este prezentat ca valoare valabilă pentru întreaga populație. Un medic care citește un astfel de studiu trebuie să se întrebe cine a răspuns, cine nu a răspuns și de ce.

Tabelul 1.2. Metodele de eșantionare pe scurt, cu exemple din medicina de familie.

Metoda	Principiul	Exemplu	Avantaj	Limită
Aleatoare simplă	fiecare individ are aceeași șansă	120 de pacienți extrași cu RAND() din lista cabinetului	baza teoretică a formulelor	cere listă completă; dezechilibre întâmplătoare
Sistematică	fiecare al k-lea individ, start aleator	fiecare al 15-lea pacient de pe listă, start 7	simplă, ușor de aplicat la triaj	periodicitatea listei
Stratificată	eșantion aleator în fiecare strat	48 urban + 32 rural, proporțional	precizie mai mare, grupuri mici prezente	cere informație despre straturi
În cluster	se aleg aleator grupuri	40 de cabinete, câte 25 de pacienți	logistică ușoară	observații corelate în cluster
De conveniență	cei disponibili	pacienții consultați în martie	rapidă, ieftină	risc mare de eroare de selecție

Studiu de caz: aceeași cohortă de studenți, două întrebări diferite

Două studii publicate în 2026 de colegii din disciplină au folosit același grup de 921 de studenți cu vârste între 18 și 24 de ani, aflați în primul an de facultate. În studiul transversal despre comportamentele alimentare orientate spre comoditate, 24,1% dintre participanți au fost clasificați ca supraponderali sau obezi, dintre care 19,4% supraponderali și 4,7% obezi [4]. În studiul prospectiv despre crononutriție, aceiași studenți au fost urmăriți pe parcursul primului an universitar, iar modificarea medie a IMC a fost -0,45 ± 1,07 kg/m² (IC 95% -0,52 la -0,38; $p < 0,001$); 48,2% au avut o scădere a IMC, 38,7% au rămas stabili și 13,1% au crescut, iar o creștere în greutate relevantă clinic a apărut la 7,2% din cohortă [5]. Un al treilea studiu, pe 693 de studenți de aceeași vârstă, a raportat o prevalență a prediabetului de 21,1% (diabet 1,4%) [6]. Întrebarea pe care trebuie să o pună cititorul nu este dacă cifrele sunt corecte, ci la ce populație se pot extinde. Studenții unei universități de medicină și farmacie dintr-un singur centru nu reprezintă toți tinerii de 18-24 de ani din România: au un nivel de educație mai ridicat, acces mai bun la informație medicală și un stil de viață specific. Concluziile sunt valide pentru populația sursă și, cu prudență, pentru studenții din universități asemănătoare. Ce ar trebui raportat în plus pentru a judeca reprezentativitatea: câți studenți au fost invitați, câți au acceptat (rata de participare) și cum se compară cei incluși cu cei care au refuzat, măcar după sex și facultate.

1.5 Erorile sistematice (bias)

Orice rezultat obținut pe un eșantion diferă de valoarea adevărată din populație. Diferența are două componente. Eroarea aleatoare (de eșantionare) apare pentru că am studiat doar o parte a populației; ea se micșorează când n crește și poate fi cuantificată prin intervale de încredere. Eroarea sistematică (bias, în engleză) deplasează rezultatul mereu în aceeași direcție și nu dispare când creștem eșantionul. Un cântar care arată cu 2 kg mai mult va da o greutate medie greșită cu 2 kg indiferent dacă cântărim 10 sau 10 000 de pacienți.

Distincția este fundamentală pentru lectura critică. Un studiu mare are o eroare aleatoare mică, adică intervale de încredere înguste, dar poate avea o eroare sistematică mare. Precizia nu înseamnă exactitate. Erorile sistematice se clasifică, de obicei, în trei mari familii: erori de selecție, erori de informare (de măsurare) și confuzie.

1.5.1 Eroarea de selecție

Eroarea de selecție apare atunci când modul în care participanții intră în studiu (sau rămân în el) depinde atât de expunere, cât și de rezultat, astfel încât eșantionul nu mai reflectă populația. Câteva forme frecvente:

Efectul "voluntarului sănătos": persoanele care se oferă să participe la un program de prevenție sunt, în medie, mai preocupate de sănătate, fumează mai puțin și fac mai multă mișcare decât populația generală. Un program de screening evaluat numai pe voluntari va părea mai eficient decât este în realitate.

Non-răspunsul: dacă doar 40% dintre persoanele invitate completează un chestionar despre consumul de alcool, este probabil ca marii consumatori să fie subreprezentați printre respondenți. Prevalența estimată va fi prea mică.

Autoselecția în sondajele online: un chestionar distribuit pe o rețea socială este completat de cei care folosesc acea rețea, sunt interesați de subiect și au timp. Vârsta, sexul și nivelul de educație al respondenților pot fi foarte diferite de cele ale populației.

Două studii recente ilustrează bine problema. Un sondaj online de convenienţă despre comportamentele alimentare a inclus 753 de tineri între 18 și 30 de ani, dintre care 93,5% femei; 11,6% erau subponderali, iar 29,3% supraponderali sau obezi [20]. Autorii au descris corect designul încă din titlu și au raportat prevalențele cu intervale de încredere Wilson (o metodă pe care o vom învăța în capitolul 4), dar o proporție de 93,5% femei arată că rezultatele privesc, în fapt, tinerele femei care au ales să răspundă. Un alt sondaj online, la persoane cu diabet, a obținut 329 de răspunsuri validate, iar 44,07% dintre respondenți au declarat că folosesc suplimente din plante [21]. Este foarte posibil ca pacienții interesați de suplimente să fi fost mai dispuși să completeze un chestionar despre acest subiect, ceea ce ar umfla prevalența. Prin contrast, un studiu transversal pe 2 976 de studenți din județul Timiș, cu 62,5% fete și 37,5% băieți între 18 și 25 de ani, a avut un eșantion mare, dar nici volumul mare nu garantează absența erorii de selecție: întrebarea rămâne cine a fost invitat și cine a răspuns [22].

Pierderile din urmărire (în studiile prospective): dacă pacienții care nu răspund la tratament abandonează studiul mai des decât ceilalți, cei rămași la final arată un efect mai bun decât cel real.

Eroarea de spectru (în studiile de teste diagnostice): un test evaluat pe pacienți cu boală avansată, comparați cu voluntari perfect sănătoși, pare mult mai bun decât va fi în cabinet, unde trebuie să deosebească forme ușoare de boală de alte afecțiuni asemănătoare. Vom reveni la această eroare în capitolul 12.

1.5.2 Eroarea de informare (de măsurare)

Eroarea de informare apare atunci când variabilele sunt măsurate sau clasificate greșit, iar greșeala diferă între grupuri. Eroarea de memorie (recall bias) este exemplul clasic: mamele copiilor născuți cu o malformație își amintesc mult mai detaliat medicamentele luate în sarcină decât mamele copiilor sănătoși, nu pentru că au luat mai multe, ci pentru că s-au gândit mai mult la cauze posibile. Într-un studiu caz-control, această diferență poate crea o asociere care nu există.

Alte forme sunt eroarea observatorului (medicul care știe ce tratament a primit pacientul tinde, involuntar, să-i evalueze mai favorabil evoluția), eroarea instrumentului (un tensiometru necalibrat), eroarea de raportare (pacienții subraportează fumatul sau consumul de alcool în fața medicului) și efectul de "preferință pentru cifra zero", foarte frecvent la măsurarea tensiunii: o proporție neobișnuit de mare de valori terminate în 0 (130, 140, 150) arată că măsurătorile au fost rotunjite. Orbirea (mascarea) evaluatorilor și protocoalele standardizate de măsurare sunt principalele mijloace de prevenție.

1.5.3 Confuzia

Confuzia (engl. confounding) apare atunci când asocierea dintre o expunere și un rezultat este distorsionată de o a treia variabilă, care este legată atât de expunere, cât și de rezultat, fără a fi o etapă intermediară între ele. Exemplul de manual: consumatorii de cafea au, în unele studii vechi, mai frecvent cancer pulmonar. Cafeaua nu produce cancer pulmonar; fumătorii beau mai multă cafea, iar fumatul produce cancer pulmonar. Fumatul este factorul de confuzie.

În medicina de familie, vârsta este cel mai frecvent factor de confuzie. Pacienții care primesc mai multe medicamente au mai multe internări, dar și sunt mai în vârstă și au mai multe boli. O asociere brută "mai multe medicamente, mai multe internări" nu spune nimic despre efectul medicamentelor până nu se ține cont de vârstă și de comorbidități. Confuzia se previne prin design (randomizare, restricție, împerechere) și se controlează la analiză prin stratificare sau prin modele multivariabile, pe care le vom întâlni în capitolele 9 și 10.

O observație care apare des la examen: randomizarea este singura metodă care echilibrează între grupuri atât factorii de confuzie cunoscuți, cât și pe cei necunoscuți. De aceea studiile randomizate stau în vârful ierarhiei dovezilor pentru întrebările despre efectul tratamentelor.

> **Studiu de caz: când o asociere dispare după ajustare**
>
> Într-un studiu prospectiv pe 278 de pacienți internați cu insuficiență cardiacă decompensată, urmăriți 12 luni, s-a analizat raportul dintre hemoglobină și lărgimea distribuției eritrocitare (HRR) ca marker de prognostic [23]. Mediana HRR a fost 0,89, iar evenimentul principal a apărut la 167 de pacienți (60,1%). Curbele de supraviețuire au arătat o diferență între pacienții cu HRR scăzut și cei cu HRR crescut (log-rank p = 0,027), dar după ajustarea pentru vârstă, sex, NT-proBNP, fracția de ejecție și rata de filtrare glomerulară, asocierea nu a mai fost semnificativă statistic (p = 0,240). Interpretarea corectă este că o parte importantă din asocierea brută se explica prin variabilele de ajustare: pacienții cu HRR scăzut erau, probabil, mai în vârstă, cu funcție renală și cardiacă mai proastă. HRR rămâne un marker ușor de obținut, dar informația lui de prognostic se suprapune în mare parte cu cea a markerilor clasici. Autorii au procedat corect raportând ambele analize; cititorul trebuie să rețină că o asociere brută nu este o dovadă de efect independent. Ce ar ajuta în plus: raportul de risc (HR) cu interval de încredere pentru ambele modele, ca să se vadă cât de mult s-a micșorat efectul, nu doar dacă a trecut de pragul de 0,05.

1.6 Statistica descriptivă și statistica inferențială

Biostatistica are două mari ramuri, care corespund la două întrebări diferite. Statistica descriptivă răspunde la întrebarea "Ce am observat?": organizează datele în tabele de frecvențe, le rezumă prin indicatori (medie, mediană, abatere standard, proporții) și le prezintă grafic. Ea privește doar eșantionul studiat și nu face afirmații despre populație. Capitolul 2 este în întregime dedicat ei.

Statistica inferențială răspunde la întrebarea "Ce putem spune despre populație pornind de la ce am observat?". Ea folosește teoria probabilităților pentru a estima parametrii populației (prin intervale de încredere) și pentru a testa ipoteze (prin teste statistice). Figura 1.1 prezintă schematic acest drum: din populație se extrage un eșantion, pe eșantion se calculează statistici, iar din statistici se trag concluzii, cu o anumită probabilitate de eroare, despre parametrii populației.

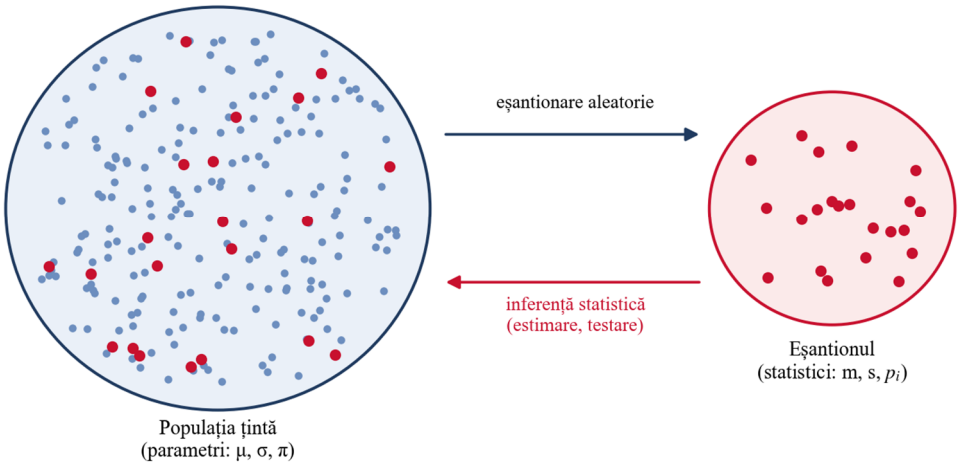


Figura 1.1. Populația, eșantionul și inferența statistică. Din populația țintă (cu parametri necunoscuți μ, σ, π) se extrage un eșantion pe care se calculează statistici (m, s, p_i); inferența merge în sens invers, de la eșantion la populație, și este însoțită întotdeauna de o măsură a incertitudinii.

1.6.1 Parametri și statistici: notațiile

O sursă constantă de confuzie la examen este amestecarea notațiilor. Caracteristicile populației se numesc parametri și se notează, prin convenție, cu litere grecești: media populației μ (miu), abaterea standard a populației σ (sigma), proporția din populație π (pi, care aici nu are legătură cu 3,14). Parametrii sunt fixi, dar de obicei necunoscuți.

Caracteristicile calculate pe eșantion se numesc statistici (sau estimatori) și se notează cu litere latine: media eșantionului m (în unele cărți $\bar{x}$), abaterea standard a eșantionului s, proporția din eșantion p_i (p cu indicele inferior i), adică proporția subiecților din clasa i a variabilei în totalul subiecților studiați: $p_i = \frac{n_i}{n}$, unde n_i este numărul subiecților din clasa i. Statisticile se pot calcula exact, dar variază de la un eșantion la altul. Tabelul 1.3 rezumă convenția, cu valorile din setul mg2026.

Tabelul 1.3. Parametri și statistici: notații și exemple pe setul mg2026.

Caracteristica	**Parametru (populație)**	**Statistică (eșantion)**	**Valoarea în mg2026 (n = 160)**
media TAS la includere	μ (necunoscută)	m	131,13 mmHg
abaterea standard a TAS	σ (necunoscută)	s	13,88 mmHg
proporția pacienților cu diabet confirmat	π (necunoscută)	p_i	$\frac{29}{160} = 0{,}181 (18{,}1\%)$
volumul	N (mii de pacienți)	n	160

Valoarea $p_i = 0{,}181$, unde clasa i este "diabet confirmat", este o estimare punctuală a lui π. Un alt eșantion de 160 de pacienți din aceleași cabinete ar fi dat, aproape sigur, o proporție puțin diferită, poate 15% sau 21%. În capitolul 4 vom vedea că intervalul de încredere de 95% pentru π, calculat prin metoda Wilson, este 12,9-24,8%. Acest interval spune ceva despre populație; cifra de 18,1% spune ceva despre eșantion.

Atenție la o capcană de vocabular: de aceea proporția din eșantion se scrie cu indice, p_i, și nu doar cu litera p. Fără indice, litera p desemnează în capitolele despre teste statistice "valoarea p", adică o probabilitate calculată de test, iar cele două noțiuni nu trebuie confundate.

1.7 Etapele unui studiu statistic și întrebarea de cercetare

Un studiu nu începe cu datele și nici cu programul statistic. Începe cu o întrebare. Figura 1.2 arată succesiunea etapelor, de la întrebarea clinică la decizie; fiecare etapă depinde de calitatea celei precedente.

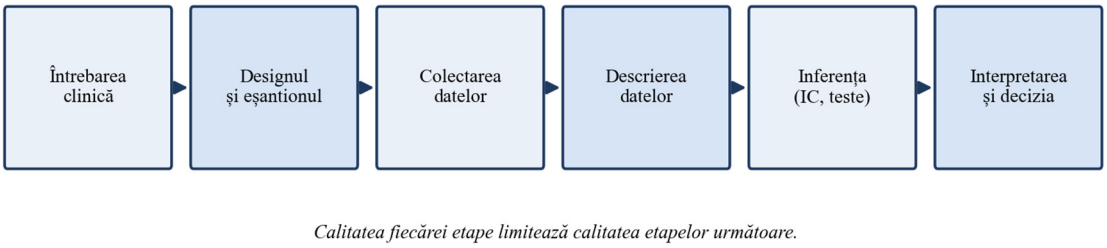


Figura 1.2. Etapele unui studiu statistic: întrebarea clinică, designul și eșantionul, colectarea datelor, descrierea datelor, inferența (intervale de încredere și teste) și interpretarea cu decizia clinică. Greșelile din primele etape nu mai pot fi corectate prin analiză.

1.7.1 Întrebarea de cercetare și formatul PICO

O întrebare de cercetare bună este precisă, relevantă clinic și se poate verifica cu date. Formatul PICO ajută la structurarea ei în patru elemente: P (populația sau pacienții), I (intervenția sau expunerea), C (comparația) și O (rezultatul, engl. outcome) [24]. Uneori se adaugă T (timpul) și S (tipul de studiu).

Pentru scenariul mg2026, o întrebare formulată vag ar fi "Ajută aplicațiile mobile pacienții cu prediabet?". În format PICO devine: "La adulții cu risc cardiometabolic din cabinetele de medicină de familie (P), adăugarea unei aplicații de automonitorizare și a teleconsultației lunare la consilierea standard (I), comparativ cu consilierea standard singură (C), crește proporția pacienților care scad cel puțin 5% din greutate în 6 luni (O)?". A doua formulare spune exact ce date trebuie culese, ce grupuri se compară și ce variabilă va fi analizată. Răspunsul, pe care îl vom calcula în capitolul 10, este că proporția a fost 29,6% în brațul C față de 11,3% în brațul A.

Formatul PICO nu este o formalitate. El obligă cercetătorul să aleagă un singur rezultat principal înainte de a vedea datele, ceea ce previne una dintre cele mai frecvente probleme din literatură: testarea a zeci de variabile și raportarea doar a celor "semnificative".

1.7.2 De la întrebare la design și la colectare

Odată formulată întrebarea, se alege designul (transversal, caz-control, cohortă, studiu clinic randomizat; vezi capitolele 10 și 13), se definește populația, se stabilesc criteriile de includere și de excludere, se calculează volumul eșantionului și se construiește instrumentul de colectare: fișa de observație, formularul electronic sau chestionarul. Fiecare variabilă primește o definiție operațională, o unitate de măsură și un mod de codificare. Abia apoi începe recrutarea.

Analiza statistică se planifică tot înainte de colectare, într-un plan de analiză statistică. Acesta precizează rezultatul principal, testele care vor fi folosite, nivelul de semnificație și modul de tratare a datelor lipsă. Analiza "după ce vedem datele" nu este interzisă, dar trebuie prezentată ca explorare, nu ca verificare a unei ipoteze. Ghidurile de raportare, cum este STROBE pentru studiile observaționale, cer ca aceste elemente să fie descrise explicit în articol [25].

Studiu de caz: etapele unui studiu prin chestionar, de la traducere la raport

Un chestionar folosit într-o limbă diferită de cea în care a fost creat nu poate fi pur și simplu tradus de un singur vorbitor bilingv. Protocolul ALSATION, publicat de o echipă din care face parte și autorul acestui curs, descrie planul de traducere în limba română a trei chestionare de alfabetizare în sănătate [26]. Faptul că traducerea unui instrument a meritat un protocol publicat separat spune mult: etapa de "construire a instrumentului" cere aceeași rigoare ca analiza. În general, adaptarea unui chestionar presupune traduceri independente, reconcilierea lor, retroversiune, evaluare de către experți și testare pe un grup mic de respondenți, iar fiecare dintre acești pași poate schimba sensul unei întrebări. Pe de altă parte, un studiu prin chestionar aplicat în 10 farmacii din mediul urban și rural a inclus 189 de persoane de 40 de ani și peste, 78,8% femei, 69,3% din mediul urban; 75,1% au declarat că știu ce este osteoporoza, dar 77,3% nu efectuaseră niciodată o osteodensitometrie DXA, iar 42,9% dintre cei care au raportat diagnosticul de osteoporoză nu urmau tratament [27]. Ambele studii arată că rezultatul unui chestionar depinde de etape care nu apar în tabelul final: cum au fost formulate întrebările, unde au fost aplicate, cine a acceptat să răspundă. În studiul din farmacii, predominanța femeilor (78,8%) și a respondenților urbani (69,3%) reflectă în parte cine intră într-o farmacie și acceptă un chestionar, deci prevalențele trebuie citite ca valabile pentru acest tip de respondenți, nu pentru toți românii de peste 40 de ani.

1.8 Datele în epoca dosarului electronic: legătura cu informatica medicală

În semestrul I, la informatica medicală, ați învățat cum se reprezintă, se stochează și se transmit datele de sănătate: dosarul electronic al pacientului, standardele de interoperabilitate, clasificările și nomenclatoarele, securitatea informației [1]. Biostatistica începe exact de unde se oprea acel curs. O analiză statistică nu poate fi mai bună decât datele pe care le prelucrează, iar majoritatea problemelor pe care le veți întâlni în practică nu țin de alegerea testului, ci de calitatea, structura și legalitatea datelor.

1.8.1 De unde vin datele biomedicale

Sursele de date se pot grupa în câteva mari categorii, fiecare cu avantaje și limite proprii [28].

Dosarul electronic de sănătate și sistemele informatice ale cabinetelor și spitalelor (engl. electronic health record, EHR) conțin diagnostice, prescripții, rezultate de laborator și consultații, înregistrate în cursul îngrijirii curente [29]. Avantajul este volumul imens și costul mic al datelor deja existente. Limita este că ele au fost colectate pentru îngrijire și decontare, nu pentru cercetare: diagnosticele pot fi codificate pentru a justifica o prescripție, unele valori lipsesc, iar definițiile diferă de la o unitate la alta. Diagnosticele sunt de obicei codificate în clasificarea internațională a bolilor [30], iar analizele de laborator pot fi identificate prin coduri standard de tip LOINC [31].

Registrele de boală (registre de cancer, de diabet, de dializă, de boli rare) și datele de supraveghere epidemiologică raportate către direcțiile de sănătate publică urmăresc sistematic toate cazurile dintr-o zonă. Sunt valoroase pentru estimarea incidenței și a tendințelor în timp, dar depind de completitudinea raportării.

Chestionarele și formularele de cercetare (pe hârtie sau electronice, de exemplu în Google Forms) colectează exact variabilele de care are nevoie studiul, cu definiții controlate. Costul este mai mare, iar calitatea depinde de formularea întrebărilor, de validarea instrumentului și de rata de răspuns.

Dispozitivele purtabile și aplicațiile mobile (ceasuri inteligente, glucometre cu monitorizare continuă, tensiometre conectate, aplicații de automonitorizare ca cea din brațul B al studiului mg2026) produc serii lungi de măsurători, cu o frecvență imposibilă în cabinet. Problemele lor sunt calibrarea, datele lipsă când pacientul nu poartă dispozitivul și faptul că utilizatorii acestor tehnologii sunt, de regulă, mai tineri și mai motivați decât media pacienților.

Datele de laborator și de imagistică provin din aparate și sisteme informatice care respectă standarde proprii; ele sunt în general precise, dar valorile de referință și metodele pot diferi între laboratoare, ceea ce contează când se combină date din mai multe centre.

1.8.2 Calitatea datelor și dicționarul de date

Calitatea datelor are mai multe dimensiuni: completitudinea (câte valori lipsesc), corectitudinea (valorile reflectă realitatea), consistența (aceeași informație este codificată la fel peste tot), plauzibilitatea (valorile sunt posibile biologic) și actualitatea. În fișierul mg2026_brut.csv veți găsi erori tipice: sexul scris o dată "F", altă dată "f" sau "Feminin"; o tensiune sistolică de 1540 mmHg (o cifră în plus); o înălțime de 1,72 introdusă în metri într-o coloană în centimetri; o greutate de 7,26 kg (virgula zecimală mutată); o glicemie scrisă "87 mg/dl", deci text în loc de număr; un rând duplicat. Fiecare dintre ele ar strica o medie, o frecvență sau un test dacă nu ar fi găsită la timp.

Instrumentul de bază pentru calitatea datelor este dicționarul de date (engl. data dictionary). Pentru fiecare variabilă, el precizează numele (scurt, fără spații și fără diacritice, de exemplu tas0), descrierea, tipul (numeric, text, dată calendaristică), unitatea de măsură, valorile permise sau intervalul plauzibil (de exemplu, TAS între 70 și 260 mmHg), codurile pentru variabilele categoriale (1 = gimnaziu, 2 = liceu...) și codul folosit pentru valorile lipsă. Dicționarul se scrie înainte de colectare, nu după; el devine apoi baza regulilor de validare din formularul electronic și din foaia de calcul. Lucrarea practică 1 vă cere să construiți un astfel de dicționar.

Regula de aur a curățării datelor este simplă: o valoare suspectă nu se corectează după intuiție, ci se verifică în documentul sursă. O tensiune de 1540 mmHg pare evident 154 mmHg cu o cifră în plus, dar poate fi la fel de bine 150 sau 140 tastată greșit, sau o valoare cu totul diferită; ipoteza cea mai probabilă trebuie totuși confirmată. Dacă documentul sursă nu este disponibil, valoarea se marchează ca lipsă, iar decizia se notează într-un jurnal al modificărilor. Datele brute originale nu se șterg niciodată; se lucrează pe o copie.

1.8.3 Etica și protecția datelor în prelucrarea datelor biomedicale

Fișa disciplinei cere, pentru nota 5, familiarizarea cu principiile eticii în prelucrarea datelor biomedicale. Nu este o cerință formală. Datele de sănătate sunt printre cele mai sensibile informații despre o persoană, iar o scurgere de date poate afecta angajarea, asigurarea sau relațiile sociale ale pacientului.

Regulamentul general privind protecția datelor (GDPR) se aplică în toată Uniunea Europeană din 2018 și include datele privind sănătatea, datele genetice și datele biometrice printre categoriile speciale de date, a căror prelucrare este în principiu interzisă, cu excepțiile prevăzute expres de regulament, printre care consimțământul explicit al persoanei, furnizarea de îngrijiri medicale și cercetarea științifică însoțită de garanții adecvate [32]. În România, Legea nr. 190/2018 stabilește măsurile de punere în aplicare a regulamentului, inclusiv reguli specifice pentru prelucrarea datelor genetice, biometrice și privind sănătatea și pentru prelucrarea codului numeric personal [33]. Regulamentul privind Spațiul european al datelor medicale, adoptat în 2025, adaugă un cadru pentru utilizarea secundară a datelor de sănătate în cercetare, prin organisme de acces autorizate [34].

Câteva principii ale GDPR se traduc direct în reguli de lucru pentru un student care face cercetare. Limitarea scopului: datele culese pentru îngrijire nu se refolosesc liber pentru alt scop. Minimizarea datelor: se culeg doar variabilele necesare întrebării de cercetare; nu se cere data nașterii completă dacă ajunge vârsta în ani, nu se notează adresa dacă ajunge mediul urban sau rural. Integritatea și confidențialitatea: fișierele se păstrează pe suporturi protejate, nu pe un stick lăsat în laborator și nici în conturi personale de stocare în cloud.

Diferența dintre anonimizare și pseudonimizare este întrebată frecvent. Pseudonimizarea înlocuiește identificatorii direcți (nume, cod numeric personal, număr de telefon) cu un cod, de exemplu P001, dar păstrează separat o listă de corespondență prin care pacientul poate fi reidentificat. Datele pseudonimizate rămân date cu caracter personal și se află în continuare sub incidența GDPR. Anonimizarea elimină ireversibil orice posibilitate rezonabilă de reidentificare, inclusiv prin combinarea mai multor variabile; datele cu adevărat anonime nu mai sunt date personale. Atenție: ștergerea numelui nu este anonimizare. Un fișier care conține data nașterii, sexul, localitatea și diagnosticul unei boli rare poate identifica un pacient chiar fără nume. Setul mg2026 are coduri P001-P160, dar fiind simulat nu corespunde niciunei persoane reale.

Pe lângă legislația privind datele, cercetarea pe oameni este guvernată de principii etice. Declarația de la Helsinki a Asociației Medicale Mondiale, revizuită în 2024, cere ca orice protocol de cercetare pe subiecți umani să fie aprobat în prealabil de o comisie de etică independentă și ca participanții să-și dea consimțământul informat, liber exprimat, după ce au primit informații clare despre scop, riscuri, beneficii și dreptul de a se retrage oricând [35]. Pentru studiile retrospective pe date existente, comisia de etică poate aproba utilizarea datelor fără consimțământ individual, în condițiile legii, dacă riscul este minim și datele sunt pseudonimizate sau anonimizate. Revistele medicale cer ca aprobarea comisiei de etică și modul de obținere a consimțământului să fie menționate în secțiunea de metode [36].

Etica prelucrării datelor nu se oprește la protecția identității. Ea cuprinde și onestitatea analizei: nu se elimină pacienți "incomozi" fără o regulă stabilită dinainte, nu se testează zeci de ipoteze raportând doar rezultatele favorabile, nu se ascund datele lipsă și nu se prezintă o asociere ca fiind o relație cauzală. Aceste reguli, discutate pe larg în literatura de etică a informaticii biomedicale [37], vor reveni în capitolele despre teste statistice și despre studiile clinice.

> **Regula care previne greșeala**
>
> Înainte de a încărca un fișier cu date de pacienți într-un program online, într-un formular, într-un instrument de inteligență artificială sau pe o platformă de stocare, puneți-vă trei întrebări: (1) Fișierul conține identificatori direcți sau combinații de variabile care pot identifica o persoană? (2) Am aprobarea (a comisiei de etică, a instituției) pentru a prelucra aceste date în acest loc? (3) Am păstrat lista de corespondență a codurilor separat de date, într-un loc protejat? Dacă răspunsul la oricare dintre întrebări este "nu știu", nu încărcați fișierul. La lucrările practice veți folosi numai setul simulat mg2026 și fișierele didactice de pe Moodle, care nu conțin date reale de identificare.

1.8.4 Datele de rutină în cercetare: exemple din regiunea noastră

Multe studii publicate de colegii din disciplină folosesc date existente, culese inițial pentru alte scopuri. Un studiu transversal al prescripțiilor electronice din mediul urban al județelor Timiș și Arad a analizat 1 498 de rețete pentru vârstnici, folosind criteriile STOPP/START [38]. Rețetele au fost emise în proporție de 57% pentru femei, vârsta medie a pacienților a fost 74,1 ± 6,95 ani, iar numărul mediu de medicamente pe rețetă 4,7 ± 1,51; statinele apăreau pe 47,4% dintre rețete. Prevalența medicamentelor potențial inadecvate a fost 18,58%, iar a omisiunilor potențiale de prescriere 49,2%. Observați că unitatea statistică este rețeta, nu pacientul: dacă un pacient a primit

mai multe rețete în perioada studiată, el contribuie de mai multe ori, iar acest lucru trebuie avut în vedere când citim "57% femei".

Un alt tip de date de rutină sunt cele din sistemele de supraveghere și din registrele centrelor de tratament. Un studiu pe hemodializă a comparat două cohorte transversale din 10 centre românești: 1 401 pacienți în 2015 și 1 698 în 2019, înainte și după programul național de tratament cu antivirale directe [39]. Prevalența infecției cu virusul hepatitic C a scăzut de la 16,9% la 10,5% ($p < 0,0001$), în timp ce prevalența infecției cu virusul hepatitic B a rămas practic aceeași (4,7% față de 4,8%, $p = 0,604$), iar mortalitatea la un an nu a diferit semnificativ (9,9% față de 10,7%, $p = 0,46$).

> **Studiu de caz: datele de supraveghere și limitele lor**
>
> Un studiu retrospectiv a folosit datele de supraveghere ale Direcției de Sănătate Publică Timiș pentru toate cazurile confirmate de COVID-19 la personalul medical, între 1 ianuarie 2021 și 31 martie 2022 [40]. Incidența la personalul auxiliar și mediu a urmărit foarte strâns evoluția din comunitate, cu coeficienți de corelație Spearman de 0,852 și 0,821 ($p < 0,001$), iar acoperirea vaccinală a depășit 70%. Este un exemplu frumos de valorificare a datelor existente, cu un cost foarte mic. Ce spune rezultatul: valurile de infecție la personalul medical au coincis în timp cu valurile din populație. Ce nu spune: corelația a două serii de timp agregate nu arată unde s-au infectat lucrătorii (în spital sau în comunitate) și nu este o dovadă de cauzalitate. Este o asociere ecologică, la nivel de grup, care poate fi explicată de faptul că ambele serii depind de același factor, circulația virusului. Studiul de hemodializă discutat mai sus are o limită înrudită: cele două cohorte, din 2015 și 2019, nu sunt formate din aceiași pacienți, deci scăderea prevalenței HCV descrie populația dializată în cele două momente, nu vindecarea unor pacienți urmăriți individual. Cohortele au diferit și prin vârstă (59,7 ± 12,92 față de 61 ± 12,93 ani, $p = 0,005$); o diferență de 1,3 ani este semnificativă statistic într-un eșantion de peste 3 000 de pacienți, dar neglijabilă clinic, o distincție pe care o vom discuta pe larg în capitolul 5. În ambele cazuri, datele de rutină permit întrebări care altfel ar fi imposibil de abordat, cu condiția ca interpretarea să respecte felul în care au fost culese.

1.9 Elemente de teoria probabilităților

Statistica inferențială se sprijină pe teoria probabilităților. Nu aveți nevoie de demonstrații riguroase, dar trebuie să stăpâniți câteva reguli simple, pentru că ele stau la baza valorilor p, a intervalelor de încredere, a riscului relativ și a valorilor predictive ale testelor diagnostice [41,42]. Toate exemplele din această secțiune folosesc setul mg2026.

1.9.1 Experiment aleator, rezultate, evenimente

Un experiment aleator este un proces al cărui rezultat nu poate fi prezis cu certitudine, deși mulțimea rezultatelor posibile este cunoscută. Aruncarea unui zar este exemplul clasic. În medicină, experimentul aleator poate fi alegerea la întâmplare a unui pacient din listă, măsurarea glicemiei la un pacient, rezultatul unui test rapid sau evoluția unui pacient după tratament.

Mulțimea tuturor rezultatelor posibile se numește spațiul rezultatelor (sau spațiul de selecție) și se notează cu Ω. Dacă alegem la întâmplare un pacient din mg2026 și notăm statutul de fumător, Ω = {nefumător, fost fumător, fumător}. Un eveniment este o submulțime a spațiului rezultatelor, de exemplu "pacientul a fumat vreodată" = {fost fumător, fumător}. Evenimentul sigur este Ω însuși (se produce întotdeauna), iar evenimentul imposibil este mulțimea vidă.

1.9.2 Definiția clasică și definiția frecvențială

Definiția clasică a probabilității se aplică atunci când toate rezultatele sunt egal posibile:

P(A) = numărul de cazuri favorabile / numărul de cazuri posibile

Dacă alegem complet aleator un pacient din cei 160 ai setului mg2026, fiecare pacient are aceeași șansă de a fi ales, iar probabilitatea ca el să fie femeie este $P(F) = \frac{86}{160} = 0{,}5375$. Probabilitatea ca el să aibă diabet confirmat prin TTGO este $P(D) = \frac{29}{160} = 0{,}18125$, adică aproximativ 18,1%.

Definiția frecvențială (statistică) pornește de la observație: probabilitatea unui eveniment este limita spre care tinde frecvența relativă a evenimentului atunci când experimentul se repetă de foarte multe ori, în aceleași condiții. Dacă din 10 000 de nou-născuți dintr-o maternitate 5 140 sunt băieți, estimăm probabilitatea nașterii unui băiat la 0,514. Legea numerelor mari garantează că, pe măsură ce numărul de repetiții crește, frecvența relativă se apropie de probabilitatea adevărată. Pe 10 nașteri, proporția de băieți poate fi 0,3 sau 0,7; pe 10 000 va fi foarte aproape de valoarea reală.

În medicină folosim aproape întotdeauna definiția frecvențială. Când spunem că riscul de diabet într-o populație este 18%, ne referim la o frecvență relativă observată pe un eșantion, folosită ca estimare a unei probabilități. Proporția din eșantion (p) estimează probabilitatea din populație (π).

1.9.3 Axiomele probabilității și evenimentul contrar

Orice probabilitate respectă trei reguli de bază (axiomele lui Kolmogorov): (1) probabilitatea oricărui eveniment este un număr între 0 și 1, $0 \leq P(A) \leq 1$; (2) probabilitatea evenimentului sigur este 1, $P(\Omega) = 1$; (3) dacă două evenimente nu se pot produce simultan (sunt mutual exclusive), probabilitatea ca unul dintre ele să se producă este suma probabilităților lor. Din aceste trei reguli decurg toate celelalte.

Evenimentul contrar lui A, notat "non-A", se produce atunci când A nu se produce. Probabilitatea lui este:

$$P(\text{non} - A) = 1 - P(A)$$

Probabilitatea ca un pacient ales aleator din mg2026 să nu aibă diabet este 1 - 0,18125 = 0,81875. Regula pare banală, dar este extrem de utilă pentru calcule de tipul "cel puțin unul": este mult mai simplu să calculezi probabilitatea ca niciun eveniment să nu se producă și să o scazi din 1.

1.9.4 Regula adunării

Pentru două evenimente mutual exclusive (care nu se pot produce împreună), probabilitatea ca unul sau celălalt să se producă este:

$$P(A \text{ sau} B) = P(A) + P(B)$$

Un pacient nu poate fi simultan fumător și fost fumător. Probabilitatea ca un pacient ales aleator să fi fumat vreodată este deci $P(\text{ fumător }) + P(\text{fost fumător }) = \frac{53}{160} + \frac{33}{160} = \frac{86}{160} = 0{,}5375$.

Dacă evenimentele se pot produce împreună, suma de mai sus ar număra de două ori indivizii care au ambele caracteristici, iar formula devine:

$$P(A \text{ sau} B) = P(A) + P(B) - P(A \text{ și } B)$$

Din cei 160 de pacienți, 53 sunt fumători, 54 sunt obezi (IMC ≥ 30 kg/m²), iar 22 sunt și fumători, și obezi. Probabilitatea ca un pacient ales aleator să fie fumător sau obez (sau ambele) este:

$$P(\text{ fumător sau obez}) = \frac{53}{160} + \frac{54}{160} - \frac{22}{160} = \frac{85}{160} = 0{,}531$$

Dacă am fi adunat simplu, am fi obținut 107/160 = 0,669, adică am fi numărat de două ori cei 22 de fumători obezi. La examen, aceasta este cea mai frecventă greșeală la problemele cu regula adunării.

1.9.5 Probabilitatea condiționată

Probabilitatea condiționată a evenimentului A, știind că s-a produs evenimentul B, se notează P(A|B) și se citește "probabilitatea lui A dat fiind B". Ea se calculează restrângând spațiul rezultatelor la indivizii care au B:

$$P(A \mid B) = \frac{P(A \text{ și } B)}{P(B)}$$

Conceptul este probabil cel mai important din tot capitolul, pentru că pe el se sprijină riscul relativ, sensibilitatea, specificitatea și valorile predictive. Să-l aplicăm pe tabelul 1.4, care încrucișează obezitatea (IMC ≥ 30 kg/m²) și diabetul confirmat prin TTGO la includere în studiul mg2026.

Tabelul 1.4. Obezitatea și diabetul confirmat prin TTGO la includere, setul mg2026 (n = 160).

	Diabet (D)	**Fără diabet**	**Total**
Obezi (O), IMC ≥ 30	18	36	54
Neobezi, IMC < 30	11	95	106
Total	29	131	160

Din tabel se pot calcula mai multe tipuri de probabilități, iar diferența dintre ele trebuie înțeleasă foarte bine.

Probabilitățile marginale privesc o singură variabilă și se calculează din totalurile de pe margine: $P(O) = \frac{54}{160} = 0{,}3375$; $P(D) = \frac{29}{160} = 0{,}181$.

Probabilitatea comună (a intersecției) privește ambele caracteristici simultan și se calculează din celula corespunzătoare raportată la totalul general: $P(O \text{ și } D) = \frac{18}{160} = 0{,}1125$.

Probabilitățile condiționate se calculează raportând celula la totalul rândului sau al coloanei care exprimă condiția. Probabilitatea de a avea diabet dacă pacientul este obez este $P(D \mid O) = \frac{18}{54} = 0{,}333 (33{,}3\%)$. Probabilitatea de a avea diabet dacă pacientul nu este obez este $P(D \mid \text{non} - O) = \frac{11}{106} = 0{,}104 (10{,}4\%)$. Se poate verifica și cu formula: $P(D \mid O) = \frac{P(O \text{ și } D)}{P(O)} = \frac{0{,}1125}{0{,}3375} = 0{,}333$.

Observați că P(D|O) nu este același lucru cu P(O|D). Probabilitatea de a fi obez dacă pacientul are diabet este $P(O \mid D) = \frac{18}{29} = 0{,}621 (62{,}1\%)$. Prima probabilitate răspunde la întrebarea "Câți dintre obezi au diabet?", a doua la întrebarea "Câți dintre diabetici sunt obezi?". Confuzia dintre ele, numită uneori eroarea condiționalei inversate, este sursa multor raționamente greșite, în medicină și nu numai. În capitolul 11 veți vedea că diferența dintre sensibilitatea unui test, P(test pozitiv|boală), și valoarea lui predictivă pozitivă, P(boală|test pozitiv), este exact diferența dintre P(O|D) și P(D|O).

Din formula probabilității condiționate rezultă regula înmulțirii, valabilă pentru orice două evenimente:

$$P(A \text{ și } B) = P(A) \times P(B \mid A) = P(B) \times P(A \mid B)$$

Pe tabelul 1.4: $P(O \text{ și } D) = P(O) \times P(D \mid O) = 0{,}3375 \times 0{,}333 = 0{,}1125$, același rezultat ca 18/160.

1.9.6 Evenimente independente

Două evenimente sunt independente dacă producerea unuia nu modifică probabilitatea celuilalt, adică P(A|B) = P(A). Pentru evenimente independente, regula înmulțirii devine mai simplă:

$P(A \text{ și } B) = P(A) \times P(B)$ (numai pentru evenimente independente)

Această relație oferă și un mod de verificare a independenței. Dacă obezitatea și diabetul ar fi independente în mg2026, probabilitatea comună ar trebui să fie $P(O) \times P(D) = 0{,}3375 \times 0{,}18125 = 0{,}0612$. Probabilitatea observată este însă 0,1125, aproape dublă. Echivalent, P(D|O) = 0,333 este foarte diferită de P(D) = 0,181 și de P(D|non-O) = 0,104. Conchidem că, în acest eșantion, cele două evenimente nu sunt independente: obezitatea și diabetul apar împreună mai des decât ar apărea întâmplător. Dacă diferența observată poate fi explicată de întâmplare (de variabilitatea de eșantionare) este o întrebare pentru testul chi-pătrat, din capitolul 7; răspunsul, anticipat aici, este că nu ($\chi^2 = 12{,}70$, $p < 0{,}001$).

Un exemplu de independență aproximativă din același set: probabilitatea de diabet la femei este 15/86 = 0,174, iar la bărbați 14/74 = 0,189, ambele foarte apropiate de probabilitatea globală de 0,181. Sexul și diabetul par independente în mg2026.

Nu confundați evenimentele independente cu evenimentele mutual exclusive. Două evenimente mutual exclusive, cu probabilități nenule, nu pot fi independente: dacă știm că s-a produs unul, știm sigur că celălalt nu s-a produs, deci probabilitatea lui a scăzut la zero. "Fumător" și "fost fumător" sunt mutual exclusive și, tocmai de aceea, puternic dependente.

Regula înmulțirii pentru evenimente independente este cea pe care o folosim când pacienții sunt aleși independent unul de altul. Dacă probabilitatea de diabet la un pacient este 0,18, probabilitatea ca doi pacienți aleși independent să aibă amândoi diabet este 0,18 x 0,18 = 0,0324, iar probabilitatea ca niciunul dintre zece pacienți să nu aibă diabet este 0,82 la puterea a zecea, adică 0,137. De aici, probabilitatea ca cel puțin unul dintre cei zece să aibă diabet este 1 - 0,137 = 0,863.

1.9.7 Formula probabilității totale și o primă privire asupra teoremei lui Bayes

Uneori cunoaștem probabilitatea unui eveniment în fiecare dintre mai multe subgrupuri care împart populația și vrem probabilitatea globală. Dacă evenimentele B1, B2,..., Bk sunt mutual exclusive și acoperă împreună toată populația, atunci:

$$P(A) = P(A \mid B_1) \times P(B_1) + P(A \mid B_2) \times P(B_2) + \cdots + P(A \mid B_k) \times P(B_k)$$

Este o medie ponderată a probabilităților condiționate, cu ponderi egale cu mărimea fiecărui subgrup. Pe tabelul 1.4:

$$P(D) = P(D \mid O) \times P(O) + P(D \mid \text{non} - O) \times P(\text{non} - O)$$
$$= 0{,}333 \times 0{,}3375 + 0{,}104 \times 0{,}6625 = 0{,}1125 + 0{,}0688 = 0{,}181$$

Regăsim, cum era de așteptat, 29/160. Formula devine cu adevărat utilă atunci când nu avem tabelul complet, ci doar probabilitățile pe subgrupuri. Să presupunem că un medic de familie lucrează cu două liste: în cabinetul din oraș sunt 60% dintre pacienți, cu o prevalență a diabetului de 15%; în punctul de lucru din comună sunt 40% dintre pacienți, cu o prevalență de 25%. Probabilitatea ca un pacient oarecare al medicului să aibă diabet este 0,15 x 0,60 + 0,25 x 0,40 = 0,09 + 0,10 = 0,19.

Teorema lui Bayes răspunde la întrebarea inversă: dacă știm că pacientul are diabet, care este probabilitatea să provină din comună? Folosind regula înmulțirii și formula probabilității totale:

$$P(\text{ comună } \mid D) = \frac{P(D \mid \text{ comună }) \times P(\text{ comună })}{P(D)} = \frac{0{,}10}{0{,}19} = 0{,}526$$

Deși doar 40% dintre pacienți sunt din comună, peste jumătate dintre diabetici (52,6%) provin de acolo, pentru că prevalența este mai mare în acel subgrup. Pe tabelul mg2026, aceeași teoremă dă $P(O \mid D) = \frac{P(D|O) \times P(O)}{P(D)} = \frac{0{,}333 \times 0{,}3375}{0{,}181} = 0{,}621$, adică valoarea calculată direct mai sus. Logica bayesiană, care actualizează o probabilitate inițială (a priori) pe baza unei informații noi și obține o probabilitate finală (a posteriori), este fundamentul interpretării testelor diagnostice și va fi tratată pe larg în capitolul 11.

1.9.8 Distribuții de probabilitate: o primă privire

O variabilă aleatoare este o variabilă al cărei rezultat depinde de un experiment aleator, de exemplu numărul de pacienți cu diabet dintre zece pacienți aleși la întâmplare. Distribuția ei de probabilitate spune ce valori poate lua și cu ce probabilitate. Distribuțiile se împart în discrete (pentru numărări: 0, 1, 2,...) și continue (pentru măsurători, ca tensiunea sau glicemia). Cea mai importantă distribuție continuă, distribuția normală, este subiectul capitolului 3. Aici prezentăm două distribuții discrete pe care le veți întâlni des.

Distribuția binomială descrie numărul de "succese" din n încercări independente, fiecare cu aceeași probabilitate de succes p. Cuvântul "succes" este convențional și desemnează doar evenimentul numărat, chiar dacă acesta este o boală. Condițiile sunt: număr fix de încercări, două rezultate posibile la fiecare încercare, probabilitate constantă, independență între încercări. Probabilitatea de a obține exact k succese este:

$$P(X = k) = C(n,\ k) \times p^k \times (1-p)^{n-k}$$

unde $C(n,\ k) = \frac{n!}{k! \times (n-k)!}$ este numărul de moduri în care se pot alege cele k succese dintre cele n încercări (combinări de n luate câte k), iar p^k înseamnă p la puterea k.

Să luăm un exemplu direct din scenariul nostru. Proporția pacienților cu diabet confirmat în mg2026 este de aproximativ 18%. Dacă un medic evaluează cu TTGO 10 pacienți cu profil similar, aleși independent, care este probabilitatea ca exact 2 dintre ei să aibă diabet? Avem n = 10, p = 0,18, k = 2:

$$C(10,\ 2) = \frac{10 \times 9}{2} = 45$$

$$P(X = 2) = 45 \times 0{,}18^2 \times 0{,}82^8 = 45 \times 0{,}0324 \times 0{,}2044 = 0{,}298$$

Tabelul 1.5 dă întreaga distribuție, iar figura 1.3A o reprezintă grafic. Media distribuției binomiale este n x p = 10 x 0,18 = 1,8 pacienți, iar abaterea standard $\sqrt{[n \times p \times (1-p)]} = \sqrt{10 \times 0{,}18 \times 0{,}82} = \sqrt{1{,}476} = 1{,}21$.

Tabelul 1.5. Distribuția binomială pentru numărul de pacienți cu diabet din 10 (p = 0,18).

k	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**≥ 7**
$P(X = k)$	0,137	0,302	0,298	0,174	0,067	0,018	0,003	< 0,001
P(X ≤ k)	0,137	0,439	0,737	0,912	0,979	0,996	1,000	1,000

Din tabel putem citi direct răspunsuri la întrebări practice. Probabilitatea ca niciun pacient să nu aibă diabet este 0,137; probabilitatea ca cel puțin unul să aibă diabet este 1 - 0,137 = 0,863; probabilitatea ca cel puțin trei să aibă diabet este $1 - P(X \leq 2) = 1 - 0{,}737 = 0{,}263$. Dacă medicul găsește 6 diabetici din 10, rezultatul ar fi foarte puțin probabil (0,003) în ipoteza că p = 0,18, ceea ce sugerează că pacienții lui au un risc mai mare decât cei din mg2026. Acest

raționament, "cât de probabil ar fi rezultatul observat dacă ipoteza noastră ar fi adevărată", este exact logica testelor statistice din capitolul 5.

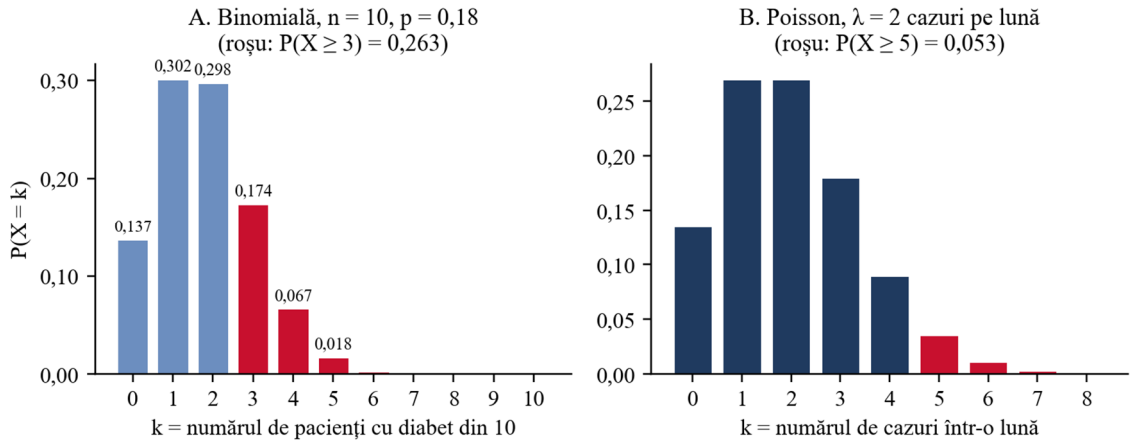


Figura 1.3. Două distribuții discrete. A: distribuția binomială a numărului de pacienți cu diabet dintre 10 pacienți evaluați, când probabilitatea individuală este 0,18; barele roșii corespund evenimentului "cel puțin 3 pacienți cu diabet" (probabilitate totală 0,263). B: distribuția Poisson pentru un eveniment rar cu o medie de 2 cazuri pe lună; barele roșii corespund evenimentului "cel puțin 5 cazuri" (probabilitate 0,053). Calcul teoretic.

Distribuția Poisson descrie numărul de evenimente rare care se produc într-un interval de timp sau într-un spațiu dat, atunci când evenimentele apar independent unele de altele, cu o rată medie constantă λ (lambda). Exemplele tipice sunt numărul de cazuri noi dintr-o boală rară raportate lunar într-un județ, numărul de reacții adverse grave la un vaccin la 100 000 de doze sau numărul de internări de urgență într-o noapte la un spital mic. Probabilitatea de a observa exact k evenimente este:

$$P(X = k) = \frac{e^{-\lambda} \times \lambda^{k}}{k!}$$

unde e ≈ 2,718 este baza logaritmilor naturali. Particularitatea distribuției Poisson este că media și varianța sunt egale, ambele cu λ. Dacă o direcție de sănătate publică primește în medie 2 raportări pe lună ale unei boli rare, probabilitatea unei luni fără niciun caz este $e^{-2} = 0{,}135$, iar probabilitatea a cel puțin 5 cazuri într-o lună este 0,053 (figura 1.3B). O lună cu 5 cazuri nu dovedește un focar, dar apare suficient de rar (aproximativ o lună din 19) pentru a justifica o verificare. Așa funcționează, în esență, sistemele de alertă epidemiologică.

Distribuția Poisson este și o aproximare bună a distribuției binomiale atunci când n este mare și p foarte mic, cu λ = n x p. Pentru o reacție adversă cu probabilitatea 1/10 000 la 20 000 de pacienți tratați, numărul de reacții urmează aproximativ o distribuție Poisson cu λ = 2.

1.10 Programele folosite în acest curs

Calculele de mână sunt indispensabile pentru a înțelege ce face o metodă, dar nimeni nu calculează de mână o abatere standard pe 160 de valori. Fișa disciplinei prevede câteva programe, alese pentru că sunt gratuite sau au o versiune gratuită, rulează pe calculatoarele din laborator și acoperă metodele din curs. Fiecare are rolul lui, iar în lucrările practice le veți folosi pe toate.

Google Sheets este foaia de calcul din suita Google, accesibilă din browser cu contul universitar. O folosim pentru introducerea și organizarea datelor, pentru reguli de validare, pentru curățarea datelor, pentru calcule descriptive cu funcții precum AVERAGE, MEDIAN, STDEV, COUNTIF, pentru tabele pivot și pentru grafice simple [43,44]. Este instrumentul cel mai apropiat de foile de calcul pe care le veți întâlni la orice loc de muncă. În configurarea regională România, zecimalele se scriu cu virgulă, iar argumentele funcțiilor se separă cu punct și virgulă.

Google Forms permite construirea rapidă a unui chestionar sau a unui formular de colectare a datelor, cu tipuri de întrebări care corespund tipurilor de variabile și cu reguli de validare a răspunsurilor; răspunsurile se salvează automat într-o foaie Google Sheets [45,46]. Îl veți folosi în lucrarea practică 2 și în lucrarea despre proiectarea studiilor clinice.

JASP este un program statistic gratuit, cu sursă deschisă, dezvoltat la Universitatea din Amsterdam [47]. Are o interfață prin meniuri, rezultatele apar imediat în tabele formatate în stil academic, iar analizele se actualizează automat când se schimbă opțiunile. Acoperă statistica descriptivă, testele parametrice și neparametrice, corelația, regresia și multe altele, inclusiv variante bayesiene ale testelor. Pentru studenți există ghiduri dedicate, de exemplu pentru statistică medicală [48,49].

MedCalc este un program statistic orientat spre cercetarea medicală, foarte folosit pentru curbe ROC, analiza testelor diagnostice, comparații de metode de măsurare și meta-analiză [50]. Versiunea completă este comercială, dar site-ul producătorului oferă gratuit calculatoare online pentru multe dintre problemele din curs: intervale de încredere pentru proporții, risc relativ, raport de șanse, sensibilitate și specificitate, numărul necesar de tratat. Le veți folosi mai ales în capitolele 10-12.

Epi Info este dezvoltat de Centrele pentru Controlul și Prevenirea Bolilor din Statele Unite și este destinat epidemiologiei de teren: construirea formularelor, introducerea datelor, tabele de frecvențe, tabele 2x2 cu risc relativ și raport de șanse, hărți [51]. Are avantajul că funcționează și fără conexiune la internet, ceea ce îl face util în anchete epidemiologice desfășurate în locuri izolate. Ultima versiune este 7.2.7; CDC a oprit dezvoltarea și suportul programului în septembrie 2025, așa că îl folosim pentru exerciții clasice de epidemiologie, știind că nu va mai primi actualizări.

Tabelul 1.6. Programele din fișa disciplinei și utilizarea lor principală.

Program	Tip	Utilizare principală în curs
Google Sheets	foaie de calcul online	introducere, validare și curățare a datelor; statistici descriptive; tabele pivot; grafice
Google Forms	formulare online	chestionare și formulare de colectare a datelor
JASP	program statistic gratuit	descriptive, normalitate, teste t, ANOVA, teste neparametrice, corelație, regresie
MedCalc	program statistic medical, calculatoare online	ROC, teste diagnostice, risc relativ, raport de șanse, intervale de încredere
Epi Info	program de epidemiologie, funcționează fără rețea	formulare, frecvențe, tabele 2x2, analiza riscului

Materialele, fișierele de date și temele se găsesc pe platforma Moodle a universității (Emediqual), unde se încarcă și lucrările predate [52]. Un ultim sfat practic: indiferent de program, verificați întotdeauna numărul de observații (n) raportat în fiecare tabel de rezultate. Dacă vă așteptați la 160 și programul raportează 158, înseamnă că două valori lipsesc sau au fost citite ca text, iar acest lucru trebuie lămurit înainte de orice interpretare.

1.11 Probleme rezolvate

Problema rezolvată 1.1. Fumatul și obezitatea în setul mg2026

Enunț. Tabelul 1.7 încrucișează statutul de fumător și obezitatea (IMC ≥ 30 kg/m²) la includere, pentru cei 160 de pacienți ai setului mg2026. Se alege la întâmplare un pacient. Calculați: (a) probabilitatea ca pacientul să fie fumător; (b) probabilitatea ca pacientul să fie fumător și obez; (c)

probabilitatea ca pacientul să fie fumător sau obez; (d) probabilitatea de obezitate la fumători, la foștii fumători și la nefumători; (e) verificați dacă evenimentele "fumător" și "obez" sunt independente.

Tabelul 1.7. Statutul de fumător și obezitatea la includere, setul mg2026.

	Obezi	**Neobezi**	**Total**
Fumători	22	31	53
Foști fumători	9	24	33
Nefumători	23	51	74
Total	54	106	160

Rezolvare. Toți pacienții au aceeași probabilitate de a fi aleși, deci folosim definiția clasică, cu 160 de cazuri posibile.

(a) $P(\text{ fumător }) = \frac{53}{160} = 0{,}331$.

(b) Probabilitatea comună se citește din celula de la intersecție: $P(\text{ fumător și obez}) = \frac{22}{160} = 0{,}1375$.

(c) Evenimentele nu sunt mutual exclusive (există 22 de pacienți care au ambele caracteristici), deci folosim regula generală a adunării: $P(\text{ fumător sau obez}) = \frac{53}{160} + \frac{54}{160} - \frac{22}{160} = \frac{85}{160} = 0{,}531$.

(d) Probabilitățile condiționate se calculează pe rânduri: $P(\text{obez} \mid \text{ fumător }) = \frac{22}{53} = 0{,}415$; $P(\text{obez} \mid \text{fost fumător }) = \frac{9}{33} = 0{,}273$; $P(\text{obez} \mid \text{ nefumător }) = \frac{23}{74} = 0{,}311$.

(e) Dacă cele două evenimente ar fi independente, am avea $P(\text{ fumător și obez}) = P(\text{ fumător }) \times P(\text{obez}) = 0{,}331 \times 0{,}3375 = 0{,}112$. Valoarea observată este 0,1375, ceva mai mare. Echivalent, P(obez|fumător) = 0,415 este mai mare decât P(obez) = 0,3375. Evenimentele nu sunt perfect independente în eșantion.

Interpretare. În setul mg2026, obezitatea apare ceva mai frecvent la fumători (41,5%) decât la nefumători (31,1%) și la foștii fumători (27,3%). Diferențele sunt însă mult mai mici decât în cazul obezității și diabetului, iar pe un eșantion de 160 de pacienți ele pot fi produse și de întâmplare. Faptul că P(A și B) diferă de P(A) x P(B) într-un eșantion nu dovedește dependența în populație; pentru asta avem nevoie de un test statistic (capitolul 7). Din punct de vedere clinic, asocierea dintre fumat și obezitate nu este una cauzală directă; ambele pot fi legate de alți factori ai stilului de viață.

Problema rezolvată 1.2. Probabilitatea binomială: pacienții hipertensivi din sala de așteptare

Enunț. În setul mg2026, 41 dintre cei 160 de pacienți (25,6%) au la includere o tensiune sistolică ≥ 140 mmHg. Presupunem că probabilitatea ca un pacient cu profil similar să aibă TAS ≥ 140 mmHg este p = 0,256 și că pacienții sunt independenți. În sala de așteptare a cabinetului sunt 5 astfel de pacienți. Care este probabilitatea ca: (a) niciunul să nu aibă TAS ≥ 140; (b) cel puțin unul să aibă TAS ≥ 140; (c) exact doi să aibă TAS ≥ 140; (d) cel puțin doi să aibă TAS ≥ 140?

Rezolvare. Verificăm condițiile distribuției binomiale: număr fix de pacienți (n = 5), două rezultate posibile pentru fiecare (TAS ≥ 140 sau nu), probabilitate constantă (0,256), independență (presupusă). Sunt îndeplinite.

(a) $P(X = 0) = (1 - 0{,}256)^5 = 0{,}744^5$. Calculăm pas cu pas: 0,744² = 0,5535; 0,744^4 = 0,5535² = 0,3064; $0{,}744^5 = 0{,}3064 \times 0{,}744 = 0{,}228$.

(b) Folosim evenimentul contrar: $P(X \geq 1) = 1 - P(X = 0) = 1 - 0{,}228 = 0{,}772$.

(c) $C(5,\ 2) = \frac{5\times4}{2} = 10.\ P(X = 2) = 10 \times 0{,}256^2 \times 0{,}744^3 = 10 \times 0{,}0655 \times 0{,}4118 = 0{,}270.$

(d) Avem nevoie și de $P(X = 1) = 5 \times 0{,}256 \times 0{,}744^4 = 5 \times 0{,}256 \times 0{,}3064 = 0{,}392.$ Apoi $P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - 0{,}228 - 0{,}392 = 0{,}380.$

Interpretare. Cu o probabilitate de aproximativ 77%, cel puțin unul dintre cei cinci pacienți va avea o tensiune sistolică de cel puțin 140 mmHg la măsurare, iar cu o probabilitate de 38%, cel puțin doi. Practic, într-o zi obișnuită, medicul trebuie să se aștepte la valori crescute la aproape fiecare grup de cinci pacienți cu risc cardiometabolic. Atenție: o singură măsurătoare ≥ 140 mmHg nu înseamnă diagnosticul de hipertensiune, care cere măsurători repetate după un protocol standardizat.

Problema rezolvată 1.3. Proiectarea unui eșantion într-un cabinet

Enunț. Un cabinet de medicină de familie are pe listă 1 600 de adulți între 40 și 75 de ani: 1 000 din oraș și 600 din satele arondate. Medicul vrea să evalueze riscul de diabet cu scorul FINDRISC la un eșantion de 80 de pacienți. (a) Descrieți o eșantionare sistematică. (b) Descrieți o eșantionare stratificată proporțională după mediu. (c) Ce s-ar întâmpla dacă medicul ar include primii 80 de pacienți care vin la cabinet în luna ianuarie?

Rezolvare. (a) Pasul de eșantionare este k = N/n = 1 600/80 = 20. Se extrage aleator un număr între 1 și 20, de exemplu 13 (în Google Sheets, =RANDBETWEEN(1;20)). Sunt incluși pacienții cu numerele de ordine 13, 33, 53, 73 și așa mai departe, adică 13 + 20 x (i - 1) pentru i = 1,..., 80. Ultimul inclus este pacientul cu numărul 13 + 20 x 79 = 1 593. Înainte de aplicare, medicul trebuie să verifice că lista nu este ordonată după un criteriu periodic.

(b) Proporțiile în populație sunt 1 000/1 600 = 62,5% pentru oraș și 600/1 600 = 37,5% pentru sate. Eșantionul de 80 va cuprinde 80 x 0,625 = 50 de pacienți din oraș și 80 x 0,375 = 30 din sate, aleși prin eșantionare aleatoare simplă în fiecare strat. Fracția de eșantionare este aceeași în ambele straturi: 50/1 000 = 30/600 = 5%.

(c) Primii 80 de pacienți care vin la cabinet formează un eșantion de convenință. Pacienții care vin des la medic sunt, în general, mai bolnavi sau mai în vârstă (mai multe boli cronice, rețete lunare), iar în ianuarie predomină infecțiile respiratorii. Pacienții sănătoși, care nu au venit la cabinet de luni de zile, ar fi subreprezentați. Prevalența riscului crescut ar fi, cel mai probabil, supraestimată. Este o eroare de selecție.

Interpretare. Eșantionarea stratificată garantează că sătenii, care au acces mai dificil la cabinet și poate un profil de risc diferit, sunt prezenți exact în proporția lor din populație. Varianta sistematică este mai simplă, dar proporția sătenilor din eșantion ar putea să varieze întâmplător în jurul valorii de 37,5%.

Problema rezolvată 1.4. Recunoașterea erorilor sistematice

Enunț. Pentru fiecare dintre situațiile de mai jos, identificați tipul de eroare sistematică și propuneți o măsură de prevenire. (1) Într-un studiu caz-control despre consumul de antiinflamatoare nesteroidiene și hemoragia digestivă, cazurile sunt intervievate în spital, la câteva zile după episod, iar martorii sunt intervievați telefonic. (2) Un program de educație pentru diabet este evaluat comparând participanții la program cu pacienții care au refuzat să participe. (3) Un studiu arată că persoanele care poartă brichetă au mai des boală coronariană. (4) Asistenta care măsoară tensiunea știe care pacienți sunt în brațul cu teleconsultație și rotunjește valorile la cea mai apropiată valoare terminată în 0.

Rezolvare. (1) Eroare de informare, mai precis eroare de memorie: cazurile, preocupate de cauza bolii, își amintesc mai bine medicamentele luate decât martorii; în plus, modul diferit de

intervievare (față în față față de telefon) poate produce diferențe de raportare. Prevenire: același mod de intervievare pentru cazuri și martori, chestionar structurat, verificarea consumului în dosarul medical sau în baza de date a prescripțiilor.

(2) Eroare de selecție (autoselecție, efectul voluntarului sănătos): pacienții care acceptă un program educațional sunt, de obicei, mai motivați și mai aderenți la tratament. Diferența de control glicemic ar putea fi explicată de motivație, nu de program. Prevenire: randomizarea pacienților eligibili în grupul cu program și grupul fără program.

(3) Confuzie: purtatul unei brichete este asociat cu fumatul, iar fumatul produce boală coronariană. Bricheta nu are niciun efect propriu. Prevenire: stratificarea pe statutul de fumător sau ajustarea pentru fumat într-un model multivariabil.

(4) Eroare de informare: eroarea observatorului (cunoașterea brațului poate influența involuntar măsurarea) și preferința pentru cifra zero. Prevenire: evaluatori care nu cunosc brațul pacientului (orbire), tensiometre automate, protocol standardizat de măsurare, cu înregistrarea valorii exacte afișate.

Interpretare. Eroarea de selecție privește cine intră în studiu, eroarea de informare privește cum sunt măsurate variabilele, iar confuzia privește interpretarea unei asocieri în prezența unei a treia variabile. Toate trei deplasează rezultatul sistematic și nu dispar prin creșterea eșantionului.

Problema rezolvată 1.5. Probabilitatea totală și teorema lui Bayes cu trei cabinete

Enunț. O rețea de trei cabinete de medicină de familie participă la un program de screening pentru diabet. Cabinetul 1 contribuie cu 50% dintre pacienții evaluați, cabinetul 2 cu 30%, iar cabinetul 3 cu 20%. Proporția pacienților cu diabet nediagnosticat anterior este 12% în cabinetul 1, 20% în cabinetul 2 și 30% în cabinetul 3. (a) Care este probabilitatea ca un pacient oarecare din program să aibă diabet? (b) Un pacient are diabet confirmat; care este probabilitatea să provină din cabinetul 3?

Rezolvare. Notăm cu C1, C2, C3 evenimentele "pacientul provine din cabinetul 1, 2, 3"; ele sunt mutual exclusive și acoperă toată populația programului.

(a) Formula probabilității totale:

$$P(D) = P(D \mid C_1) \times P(C_1) + P(D \mid C_2) \times P(C_2) + P(D \mid C_3) \times P(C_3)\, P(D)$$
$$= 0{,}12 \times 0{,}50 + 0{,}20 \times 0{,}30 + 0{,}30 \times 0{,}20 = 0{,}06 + 0{,}06 + 0{,}06 = 0{,}18$$

(b) Teorema lui Bayes:

$$P(C_3 \mid D) = \frac{P(D \mid C_3) \times P(C_3)}{P(D)} = \frac{0{,}06}{0{,}18} = 0{,}333$$

Interpretare. Prevalența globală a diabetului în program este 18%, o medie ponderată a celor trei prevalențe, nu media lor simplă (care $\frac{\text{ar fi} \quad (12+20+30)}{3} = 20{,}7\%$). Deși cabinetul 3 contribuie doar cu 20% dintre pacienți, el dă o treime dintre cazurile de diabet, pentru că prevalența la el este cea mai mare. Observați simetria: fiecare cabinet contribuie cu 6 puncte procentuale la prevalența globală, deci fiecare dă o treime dintre cazuri. Planificarea resurselor pentru confirmarea diagnosticului (TTGO, consultații) ar trebui să țină cont de această distribuție, nu de numărul de pacienți evaluați în fiecare cabinet.

Problema rezolvată 1.6. Distribuția Poisson într-un sistem de alertă

Enunț. Direcția de sănătate publică a unui județ primește, în medie, 3 raportări pe lună de hepatită virală A. Presupunem că raportările sunt independente și au o rată constantă. (a) Care este probabilitatea ca într-o lună să nu fie raportat niciun caz? (b) Care este probabilitatea a exact 3 cazuri? (c) Într-o lună sunt raportate 7 cazuri; cât de neobișnuit este acest număr?

Rezolvare. Numărul de cazuri lunare urmează aproximativ o distribuție Poisson cu λ = 3.

(a) $P(X = 0) = \frac{e^{-3} \times 3^0}{0!} = e^{-3} = 0{,}0498$. Aproximativ o lună din 20 nu va avea niciun caz.

(b) $P(X = 3) = \frac{e^{-3} \times 3^3}{3!} = \frac{0{,}0498 \times 27}{6} = 0{,}224$.

(c) Calculăm probabilitatea a cel puțin 7 cazuri, folosind evenimentul contrar. Probabilitățile pentru k = 0, 1,..., 6 sunt 0,0498; 0,1494; 0,2240; 0,2240; 0,1680; 0,1008; 0,0504, care însumate dau 0,966. Deci $P(X \geq 7) = 1 - 0{,}9665 = 0{,}034$.

Interpretare. O lună cu 7 cazuri sau mai multe apare, în condiții obișnuite, cu o probabilitate de aproximativ 3,4%, adică mai rar decât o dată la doi ani. Rezultatul nu dovedește existența unui focar, dar este suficient de neobișnuit pentru a justifica o anchetă epidemiologică: cazurile au o sursă comună, provin din aceeași localitate, din aceeași colectivitate de copii? Așa se stabilesc, în principiu, pragurile de alertă în supravegherea bolilor transmisibile.

1.12 Exerciții propuse

1. În setul mg2026, 19 dintre cei 95 de pacienți din mediul urban și 10 dintre cei 65 de pacienți din mediul rural au diabet confirmat prin TTGO. Construiți tabelul 2x2 și calculați: P(D|urban), P(D|rural), P(urban și D), P(urban) x P(D). Sunt evenimentele "urban" și "diabet" independente în eșantion?
2. În mg2026 sunt 86 de femei și 53 de fumători, iar 30 dintre femei sunt fumătoare. Calculați P(fumător|femeie), P(femeie|fumător) și P(femeie sau fumător). Explicați de ce primele două probabilități sunt diferite.
3. Prevalența hipertensiunii arteriale într-o populație de adulți de peste 40 de ani este 25%. Se aleg la întâmplare 8 persoane. Calculați probabilitatea ca: (a) niciuna să nu fie hipertensivă; (b) exact una să fie hipertensivă; (c) cel puțin două să fie hipertensive.
4. Un spital are în registrul de externări 2 400 de pacienți cu insuficiență cardiacă din ultimii trei ani. Se dorește un eșantion sistematic de 120 de dosare. Care este pasul de eșantionare? Dacă numărul de pornire extras este 9, scrieți primele cinci numere de dosar incluse și ultimul.
5. Un cabinet are 1 500 de adulți pe listă: 600 între 18 și 39 de ani, 600 între 40 și 64 de ani și 300 de 65 de ani și peste. Câți pacienți din fiecare grupă de vârstă trebuie incluși într-un eșantion stratificat proporțional de 90 de pacienți? Care este fracția de eșantionare?
6. Clasificați eroarea sistematică (selecție, informare sau confuzie) din fiecare situație: (a) un chestionar despre consumul de alcool, completat în prezența medicului de familie; (b) un studiu despre efectul vitaminei D asupra fracturilor, în care persoanele care iau suplimente sunt și cele care fac mai multă mișcare; (c) un studiu despre eficacitatea unui tratament pentru BPOC în care 30% dintre pacienții din grupul tratat abandonează studiul din cauza efectelor adverse și nu mai sunt evaluați; (d) un test nou pentru cancerul de prostată evaluat pe pacienți cu metastaze și pe studenți voluntari sănătoși.
7. Citiți descrierea studiului despre prescrierea la vârstnici din secțiunea 1.8.4 [38]. Precizați: unitatea statistică, volumul eșantionului, populația țintă probabilă, tipul de studiu, o variabilă numerică și o variabilă calitativă. Ce problemă poate apărea dacă același pacient are mai multe rețete în bază?
8. Un serviciu de urgență pediatrică primește, în medie, 1,5 cazuri pe săptămână de intoxicație accidentală cu medicamente. Presupunând o distribuție Poisson, calculați probabilitatea ca într-o săptămână să nu fie niciun caz și probabilitatea a cel puțin 4 cazuri.

9. Într-un cabinet, 30% dintre adulți sunt fumători. Prevalența BPOC este 15% la fumători și 3% la nefumători. Calculați prevalența BPOC în populația cabinetului și probabilitatea ca un pacient cu BPOC să fie fumător.
10.
    (a) Formulați în format PICO întrebarea: "Teleconsultația ajută la scăderea tensiunii arteriale?", folosind scenariul mg2026 și variabilele tas0 și tas6. (b) Un fișier conține coloanele: cod pacient (P001...), vârstă în ani, sex, IMC, glicemie; lista care leagă codurile de numele pacienților este păstrată de medicul investigator într-un dosar separat, protejat. Datele sunt anonimizate sau pseudonimizate? Ce regim legal li se aplică?

Capcane frecvente

- Confundarea lui P(A|B) cu P(B|A): probabilitatea de diabet la obezi (33,3%) nu este probabilitatea de obezitate la diabetici (62,1%).
- Aplicarea regulii simple a adunării, P(A) + P(B), la evenimente care se pot produce împreună; rezultatul poate depăși chiar 1.
- Confundarea evenimentelor independente cu cele mutual exclusive.
- Credința că un eșantion mare este automat reprezentativ; volumul mare reduce eroarea aleatoare, nu eroarea sistematică.
- Folosirea literelor grecești (μ, σ) pentru valori calculate pe eșantion; în eșantion se scrie m și s.
- Corectarea "după intuiție" a unei valori aberante în locul verificării în documentul sursă.
- Considerarea unui fișier ca anonim doar pentru că s-au șters numele pacienților.

De reținut

- Variabilitatea biologică este motivul pentru care medicina are nevoie de statistică; ea se descrie, nu se elimină.
- Populația se definește în spațiu și timp; eșantionul este o parte a ei, aleasă astfel încât să fie reprezentativă (echiprobabilitate și independență).
- Eșantionarea aleatoare simplă, sistematică, stratificată și în cluster sunt metode probabilistice; eșantionul de conveniență este frecvent, dar expus erorii de selecție.
- Erorile sistematice (selecție, informare, confuzie) nu scad când crește eșantionul; se previn prin design.
- Statistica descriptivă rezumă eșantionul; statistica inferențială generalizează la populație, cu o incertitudine cuantificată.
- Parametrii populației (μ, σ, π) sunt necunoscuți; statisticile eșantionului (m, s, p_i) îi estimează.
- Probabilitatea condiționată $P(A \mid B) = \frac{P(A \text{ și } B)}{P(B)}$ stă la baza riscului relativ și a testelor diagnostice.
- Evenimentele independente satisfac $P(A \text{ și } B) = P(A) \times P(B)$; verificarea pe eșantion nu înlocuiește testul statistic.
- Datele de sănătate sunt date cu caracter special (GDPR, Legea 190/2018); datele pseudonimizate rămân date personale; cercetarea pe oameni cere avizul comisiei de etică și, de regulă, consimțământ informat.

Întrebări de autoevaluare

1. (CS) Într-un studiu care analizează 1 498 de rețete electronice emise pentru vârstnici, unitatea statistică este:
    a) pacientul vârstnic
    b) rețeta
    c) farmacia
    d) medicul prescriptor

e) județul

2. (CS) Notația corectă pentru media unei variabile în populație este:
   a) m
   b) s
   c) μ
   d) p
   e) σ
3. (CS) În tabelul 1.4 (obezitate și diabet, mg2026), probabilitatea ca un pacient obez să aibă diabet este:
   a) 0,113
   b) 0,181
   c) 0,333
   d) 0,621
   e) 0,104
4. (CS) Pentru o eșantionare sistematică de 60 de pacienți dintr-o listă de 900, pasul de eșantionare este:
   a) 60
   b) 15
   c) 9
   d) 6
   e) 90
5. (CS) Dacă probabilitatea ca un pacient să aibă diabet este 0,18, iar pacienții sunt independenți, probabilitatea ca niciunul dintre 10 pacienți să nu aibă diabet este aproximativ:
   a) 0,18
   b) 0,82
   c) 0,137
   d) 0,863
   e) 0,298
6. (CM) Despre pseudonimizarea datelor sunt corecte afirmațiile:
   a) înlocuiește identificatorii direcți cu coduri
   b) datele pseudonimizate rămân date cu caracter personal în sensul GDPR
   c) este ireversibilă, pacientul nu mai poate fi reidentificat în niciun fel
   d) lista de corespondență dintre coduri și identitate se păstrează separat de date
   e) scutește studiul de avizul comisiei de etică
7. (CM) Despre erorile sistematice sunt corecte afirmațiile:
   a) scad pe măsură ce crește volumul eșantionului
   b) eroarea de memorie este o eroare de informare
   c) randomizarea echilibrează între grupuri factorii de confuzie cunoscuți și necunoscuți
   d) efectul voluntarului sănătos este o eroare de selecție
   e) un interval de încredere îngust garantează absența erorilor sistematice
8. (CM) Fie A și B două evenimente mutual exclusive, cu $P(A) > 0$ și $P(B) > 0$. Sunt corecte afirmațiile:

a) $P(A \text{ și } B) = 0$
b) $P(A \text{ sau} B) = P(A) + P(B)$
c) A și B sunt independente
d) $P(A \mid B) = 0$
e) P(A) + P(B) poate fi mai mare decât 1

9. (CM) Statistica inferențială:
   a) estimează parametrii populației prin intervale de încredere
   b) testează ipoteze despre populație
   c) se referă exclusiv la indivizii din eșantion
   d) se sprijină pe teoria probabilităților
   e) constă în construirea tabelelor de frecvențe
10. (CM) Despre eșantionul de conveniență sunt corecte afirmațiile:

a) este format din indivizii cel mai ușor de recrutat
b) garantează reprezentativitatea dacă volumul lui depășește 500
c) poate fi folosit într-un studiu randomizat, unde validitatea comparației dintre brațe provine din randomizare
d) un sondaj online distribuit pe rețelele sociale este un exemplu
e) elimină eroarea de selecție

Capitolul 2. Descrierea și organizarea datelor: tipuri de variabile, indicatori descriptivi și reprezentări grafice

Obiective

La finalul acestui capitol, studentul va putea:

- să clasifice o variabilă după tipul ei (nominală, ordinală, numerică discretă sau continuă) și după scala de măsură (nominală, ordinală, de interval, de raport);
- să organizeze datele unui studiu într-un tabel cu un rând pentru fiecare pacient și o coloană pentru fiecare variabilă, însoțit de un dicționar de date, și să recunoască erorile frecvente de introducere;
- să construiască un tabel de frecvențe absolute, relative și cumulate, inclusiv pentru date grupate în clase, și să reconstituie valorile lipsă dintr-un astfel de tabel;
- să calculeze de mână și cu programul media, mediana, modul, media ponderată și media geometrică și să aleagă indicatorul potrivit tipului de date și formei distribuției;
- să calculeze amplitudinea, abaterea medie absolută, varianța, abaterea standard, coeficientul de variație și intervalul intercuartilic;
- să explice diferența dintre abaterea standard și eroarea standard a mediei;
- să determine cuartilele, percentilele și rezumatul în cinci numere și să identifice valorile extreme cu regula 1,5 x IQR;
- să aleagă graficul potrivit pentru fiecare tip de variabilă și să recunoască graficele care induc în eroare;
- să citească și să construiască un tabel descriptiv de tipul "Tabelul 1" dintr-un articol medical.

2.1 De la pacient la tabelul de date

Un pacient care intră în cabinet nu este o înșiruire de cifre. Pentru a putea fi analizat statistic, însă, el trebuie descris printr-un set de caracteristici măsurate sau observate, aceleași pentru toți pacienții din studiu. Fiecare caracteristică se numește variabilă, pentru că ia valori diferite de la un pacient la altul. Vârsta, sexul, tensiunea arterială, statutul de fumător, nivelul de educație, satisfacția față de tratament sunt variabile. Valoarea unei variabile la un anumit pacient se numește observație.

Tipul unei variabile determină aproape tot ce urmează: ce indicatori descriptivi se pot calcula, ce grafic se potrivește, ce test statistic se poate aplica. De aceea, primul pas al oricărei analize, la examen ca și în cercetare, este clasificarea corectă a variabilelor. Studenții care sar acest pas ajung să calculeze media sexului sau abaterea standard a brațului de tratament, rezultate pe care programul le afișează fără protest, dar care nu au niciun sens.

2.1.1 Variabile calitative și cantitative

Figura 2.1 rezumă clasificarea de bază. Variabilele calitative (categoriale) încadrează fiecare individ într-o categorie. Variabilele cantitative (numerice) exprimă o mărime prin număr, de regulă însoțit de o unitate de măsură.

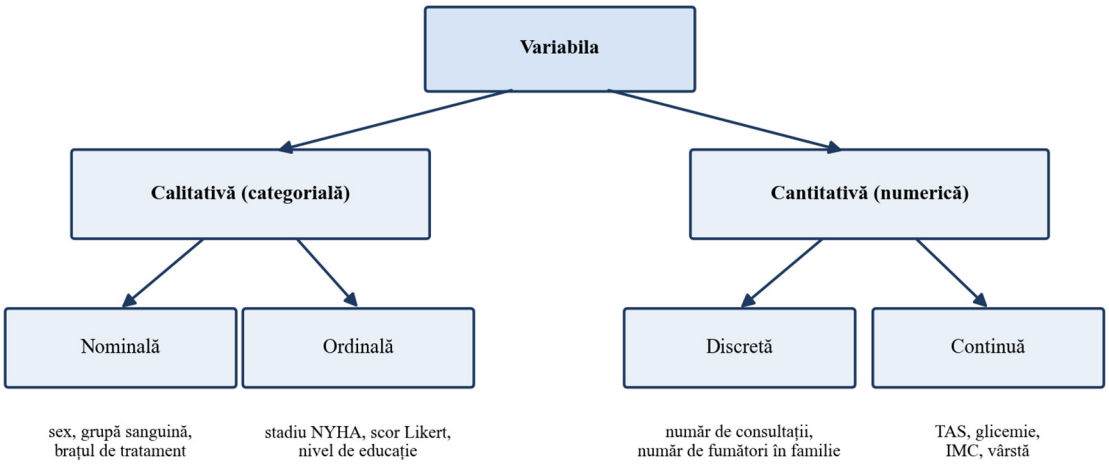


Figura 2.1. Clasificarea variabilelor statistice. Variabilele calitative pot fi nominale (categorii fără ordine) sau ordinale (categorii ordonate); variabilele cantitative pot fi discrete (numărări) sau continue (măsurători). Exemplele sunt din medicina generală și din setul mg2026.

Variabilele nominale au categorii între care nu există o ordine naturală: sexul (F, M), grupa sanguină (0, A, B, AB), brațul de tratament (A, B, C), statutul de fumător (nefumător, fost fumător, fumător), mediul de proveniență (urban, rural). Chiar dacă le codificăm cu cifre, de exemplu 1 pentru femei și 2 pentru bărbați, cifrele sunt doar etichete. Media codurilor nu are sens. O variabilă nominală cu doar două categorii se numește dihotomică (sau binară): diabet confirmat da/nu, decedat da/nu, scădere în greutate de cel puțin 5% da/nu.

Variabilele ordinale au categorii ordonate, dar distanțele dintre ele nu sunt egale sau nu sunt cunoscute. Stadiile NYHA ale insuficienței cardiace, stadiile de fibroză hepatică F0-F4, gradele de steatoză S0-S3, nivelul de educație, scalele Likert ("total dezacord"... "total de acord"), scorul de durere de la 0 la 10 sunt ordinale. Știm că stadiul NYHA III este mai sever decât II, dar nu putem spune că diferența dintre I și II este egală cu diferența dintre III și IV. În setul mg2026, educația (1-4), activitatea fizică autoevaluată (1-5) și satisfacția (1-5) sunt ordinale.

Variabilele cantitative discrete iau valori izolate, de obicei numere întregi rezultate din numărare: numărul de consultații pe an, numărul de medicamente pe rețetă, numărul de sarcini, numărul de dinți cariați. Variabilele cantitative continue pot lua, cel puțin teoretic, orice valoare dintr-un interval, limitată doar de precizia instrumentului: tensiunea arterială, glicemia, greutatea, înălțimea, vârsta, concentrația unui medicament. Faptul că o variabilă continuă este rotunjită la întreg (TAS de 131 mmHg, vârsta de 56 de ani) nu o face discretă.

2.1.2 Scalele de măsură

O clasificare mai fină, foarte folosită în manualele de metodologie, distinge patru scale de măsură, în ordinea crescătoare a informației pe care o conțin [19,53].

Scala nominală permite doar verificarea egalității: doi pacienți au sau nu aceeași grupă sanguină. Scala ordinală permite, în plus, ordonarea: un pacient are un stadiu mai avansat decât altul. Scala de interval permite, în plus, compararea diferențelor: diferența dintre 36 °C și 38 °C este egală cu diferența dintre 38 °C și 40 °C. Zeroul ei este însă convențional, deci rapoartele nu au sens; o temperatură de 40 °C nu este "de două ori mai mare" decât 20 °C. Scala de raport (sau de proporție) are un zero absolut, care înseamnă absența mărimii măsurate, și permite toate operațiile, inclusiv rapoarte: o glicemie de 200 mg/dL este dublul uneia de 100 mg/dL, iar un pacient de 90 kg cântărește de 1,5 ori mai mult decât unul de 60 kg.

Cele mai multe variabile cantitative din medicină sunt pe scală de raport: greutatea, înălțimea, glicemia, colesterolul, tensiunea arterială, frecvența cardiacă, durata spitalizării. Distincția interval-raport contează rar în practică pentru alegerea testului, dar are o consecință pe care o veți întâlni în secțiunea 2.4: coeficientul de variație are sens doar pentru variabilele pe scală de raport.

Tabelul 2.1. Scalele de măsură și operațiile permise.

Scala	Operații cu sens	Exemple	Tendința centrală potrivită
Nominală	=, ≠	sex, grupă sanguină, brat de tratament, fumat	modul (moda), proporții
Ordinală	=, ≠, <, >	stadiu NYHA, educație, scală Likert, scor de durere	mediana, modul (moda)
De interval	în plus, diferențe	temperatura în °C, anul nașterii	media, mediana
De raport	în plus, rapoarte	glicemie, greutate, TAS, vârstă, durata spitalizării	media, mediana, media geometrică

2.1.3 Variabile derivate și scoruri

Multe variabile din analiză nu sunt măsurate direct, ci calculate din altele. Indicele de masă corporală este exemplul tipic:

IMC = greutate (kg) / [înălțime (m)]²

Pentru pacientul P003 din mg2026, cu greutatea de 127,4 kg și înălțimea de 183 cm (1,83 m), $\text{IMC} = \frac{127{,}4}{1{,}83^2} = \frac{127{,}4}{3{,}3489} = \frac{38{,}0\ \text{kg}}{m^2}$. Alte variabile derivate frecvente sunt vârsta (calculată din data nașterii și data includerii), rata de filtrare glomerulară estimată (din creatinină, vârstă și sex), modificarea unei variabile în timp (glicemie0 - glicemie6) sau scăderea procentuală în greutate.

Variabilele derivate se calculează în foaia de calcul cu o formulă, nu se introduc de mână. Dacă o înălțime este introdusă greșit, formula va produce un IMC absurd, ușor de observat; dacă IMC-ul ar fi tastat separat, eroarea s-ar ascunde. În lucrarea practică 1 veți vedea că o înălțime de 1,72 introdusă într-o coloană în centimetri dă, la recalcularea IMC cu formula, o valoare de peste 260 000 kg/m², care sare imediat în ochi. Invers, o variabilă derivată păstrată în fișier poate ajuta la recuperarea unei valori greșite: din IMC și greutate se poate recalcula înălțimea.

Scorurile clinice (FINDRISC, scorul Apgar, scorul Glasgow, scorurile de risc cardiovascular) sunt sume de puncte acordate pentru diferite caracteristici. Formal sunt variabile discrete, cu un număr mare de valori posibile; în practică sunt adesea tratate ca numerice, dar cu prudență, pentru că un punct în plus nu are aceeași semnificație clinică în toate zonele scalei. Scorul FINDRISC din mg2026 ia valori între 0 și 25, cu media 10,37 și mediana 10.

2.1.4 Dihotomizarea și costul ei

Medicii iubesc pragurile: hipertensiv sau nu (140 mmHg), obez sau nu (IMC 30 kg/m²), diabetic sau nu (glicemie a jeun 126 mg/dL). Transformarea unei variabile continue într-una dihotomică se numește dihotomizare. Este utilă pentru decizii clinice și pentru comunicarea cu pacientul, dar în analiza statistică are un cost.

Primul cost este pierderea de informație. După dihotomizare, un pacient cu IMC 30,0 și unul cu IMC 48,3 sunt la fel de "obezi", în timp ce un pacient cu IMC 29,9 este "neobez", deși diferă de primul cu doar 0,1 kg/m². În setul mg2026, tensiunea sistolică medie este 136,87 mmHg la cei 54 de pacienți obezi și 128,21 mmHg la cei 106 neobezi, dar această comparație ignoră faptul că, între pacienții obezi, TAS crește în continuare cu IMC-ul. Regresia din capitolul 9 va arăta că fiecare kg/m² în plus se asociază, în medie, cu 1,435 mmHg în plus, informație pe care dihotomizarea o pierde. Al doilea cost este scăderea puterii statistice: pentru a detecta aceeași

asociere cu o variabilă dihotomizată sunt necesari mai mulți pacienți. Al treilea este arbitrarul pragului: dacă pragul nu este stabilit dinainte, cercetătorul poate fi tentat să-l aleagă pe cel care dă rezultatul "cel mai frumos".

Un exemplu bun de analiză care păstrează ambele perspective vine din studiul prospectiv pe 921 de studenți [5]. Modificarea IMC a fost analizată mai întâi ca variabilă continuă (medie -0,45 ± 1,07 kg/m²), cu regresie liniară, apoi, separat, a fost definită o creștere în greutate relevantă clinic, apărută la 7,2% din cohortă, analizată cu regresie logistică. Prima analiză folosește toată informația; a doua răspunde la o întrebare clinică ("cine ia semnificativ în greutate?"). Concluzia practică: păstrați variabila continuă pentru analiza principală și folosiți pragurile clinice consacrate, stabilite dinainte, pentru analize secundare.

2.1.5 Codificarea și tabelul de date

Datele unui studiu se organizează într-un tabel dreptunghiular, după o regulă simplă: un rând pentru fiecare unitate statistică (pacient) și o coloană pentru fiecare variabilă. Primul rând conține numele variabilelor. Fiecare celulă conține o singură valoare. Toate programele statistice (JASP, MedCalc, Epi Info) presupun această structură, numită uneori format "lung" sau "ordonat" (engl. tidy data).

Câteva reguli practice previn majoritatea problemelor. Numele variabilelor sunt scurte, fără spații, fără diacritice și fără caractere speciale (tas0, nu "TA sistolică inițială (mmHg)"); descrierea completă stă în dicționarul de date. O coloană conține un singur tip de date: în coloana glicemiei se scriu doar numere, fără unitate ("87", nu "87 mg/dl"). Categoriile se scriu întotdeauna la fel ("F", nu o dată "F" și altă dată "f" sau "Feminin"). Valorile lipsă se lasă necompletate sau se marchează cu un cod unic, declarat în dicționar; nu se scrie 0 pentru o valoare lipsă, pentru că 0 este o valoare reală pentru multe variabile. Măsurătorile repetate (glicemia la 0, 3 și 6 luni) se pun în coloane separate, cu nume care indică momentul.

Tabelul 2.2 prezintă un fragment din dicționarul setului mg2026. Dicționarul complet se află în anexa C.

Tabelul 2.2. Fragment din dicționarul de date al setului mg2026.

Nume	Descriere	Tip	Unitate / coduri	Interval plauzibil
id	codul pacientului	text	P001-P160	unic
sex	sexul	nominal	F, M	-
varsta	vârsta la includere	numeric întreg	ani	18-90
educatie	nivelul de educație	ordinal	1 gimnaziu; 2 liceu sau profesională; 3 universitar; 4 postuniversitar	1-4
inaltime	înălțimea	numeric	cm	130-210
tas0	tensiunea sistolică la includere	numeric întreg	mmHg	70-260
glicemie0	glicemia a jeun la includere	numeric	mg/dL	40-400
hba1c0	hemoglobina glicată	numeric, o zecimală	%	3,5-15
diabet_ogtt	diabet confirmat prin TTGO	dihotomic	0 nu, 1 da	0, 1
satisfactie	satisfacția la 6 luni	ordinal	1 foarte nemulțumit... 5 foarte mulțumit	1-5

2.1.6 Erori frecvente de introducere a datelor

Fișierul mg2026_brut.csv conține intenționat câteva dintre erorile pe care le veți întâlni în orice bază de date reală. Le prezentăm aici pe tipuri, iar în lucrarea practică 1 le veți găsi și corecta.

Tabelul 2.3. Tipuri de erori de introducere, cu exemple din mg2026_brut.csv.

Tipul erorii	Exemplu din fișierul brut	Cum se detectează
Codificare inconsecventă	sex "f" și "Feminin" în loc de "F"; mediu "Urban" în loc de "urban"	tabel de frecvențe al variabilei (apar categorii în plus)
Spații invizibile	"Fumator " (cu spațiu la final)	frecvențe; funcția TRIM
Cifră în plus	TAS 1540 mmHg	minim și maxim; interval plauzibil
Unitate sau zecimală greșită	înălțime 1,72 (metri) într-o coloană în cm; greutate 7,26 kg în loc de 72,6	minim și maxim; variabila derivată IMC
Text în coloană numerică	glicemie "87 mg/dl"; HbA1c "5,2" scrisă cu virgulă într-un fișier cu punct zecimal, deci citită ca text	numărul de valori numerice (n) mai mic decât numărul de rânduri
Valori lipsă	vârstă lipsă; glicemie la 6 luni lipsă	COUNTBLANK, n raportat de program
Rând duplicat	același pacient introdus de două ori	numărarea codurilor unice; sortare după id

Ordinea de lucru recomandată este: mai întâi frecvențele pentru variabilele calitative (apar imediat categoriile în plus), apoi minimul, maximul și numărul de valori pentru variabilele numerice, apoi verificarea codurilor unice. O jumătate de oră de curățare economisește zile de analiză refăcută.

> **Studiu de caz: aceeași variabilă, trei moduri de raportare**
>
> Felul în care un articol raportează o variabilă spune cum au înțeles autorii tipul ei. Într-un sondaj pe 121 de studenți la medicină, aprecierile pentru cursuri, pe o scală de la 1 la 10, au fost raportate ca mediană și interval intercuartilic: 9 (7-10) pentru medicina de urgență, 7 (5-9) pentru informatica biomedicală și 8 (5-9) pentru biostatistică [16]. Alegerea este corectă: o notă dată de student este ordinală, iar o mediană spune "jumătate dintre studenți au dat cel puțin 8", fără a presupune că distanța dintre 7 și 8 este egală cu cea dintre 9 și 10. Într-un studiu despre prescrierea la vârstnici, numărul de medicamente pe rețetă, o variabilă discretă, a fost raportat ca medie ± abatere standard, 4,7 ± 1,51 [38]; pentru o numărare cu multe valori posibile și o distribuție relativ simetrică, media este acceptabilă și ușor de interpretat. Într-un studiu despre acceptabilitatea tehnologiei mobile la 103 pacienți cu diabet, vârsta a fost raportată ca mediană cu amplitudine: 37 de ani, cu limite între 18 și 65 [54]. Amplitudinea arată spectrul de vârstă al participanților, dar depinde doar de doi indivizi extremi; intervalul intercuartilic ar fi descris mai bine unde se află majoritatea pacienților. Niciuna dintre aceste alegeri nu este "greșită", dar fiecare transmite altceva, iar cititorul trebuie să știe ce anume.

2.2 Tabelele de frecvențe

Primul rezumat al unei variabile este tabelul de frecvențe: lista valorilor (sau a claselor de valori) împreună cu numărul de apariții al fiecăreia. Pare un instrument modest, dar este cel mai sigur mod de a vedea cum arată datele, de a descoperi erorile de codificare și de a pregăti graficele.

2.2.1 Frecvențe absolute, relative şi cumulate

Frecvența absolută (ni) a unei valori este numărul de observații egale cu acea valoare. Suma frecvențelor absolute este volumul eșantionului, n. Frecvența relativă (fi) este frecvența absolută raportată la n, exprimată ca fracție sau ca procent:

$$f_i = \frac{n_i}{n} \text{ (sau } f_i\% = \frac{n_i}{n} \times 100)$$

Frecvența cumulată (absolută sau relativă) a unei valori este suma frecvențelor tuturor valorilor mai mici sau egale cu ea. Ea are sens doar dacă valorile au o ordine, deci pentru variabilele ordinale și numerice, nu pentru cele nominale. Frecvența cumulată relativă a ultimei clase este întotdeauna 100% (sau 99,9-100,1%, din cauza rotunjirilor).

Tabelul 2.4 prezintă satisfacția față de program la 6 luni, o variabilă ordinală măsurată pe o scală Likert de la 1 (foarte nemulțumit) la 5 (foarte mulțumit).

Tabelul 2.4. Satisfacția față de program la 6 luni, setul mg2026 (n = 160).

Satisfacția	Frecvența absolută	Frecvența relativă (%)	Frecvența cumulată absolută	Frecvența cumulată relativă (%)
1 foarte nemulțumit	4	2,5	4	2,5
2 nemulțumit	19	11,9	23	14,4
3 neutru	54	33,8	77	48,1
4 mulțumit	55	34,4	132	82,5
5 foarte mulțumit	28	17,5	160	100,0
Total	160	100,0		

Tabelul se citește în mai multe feluri. Cei mai mulți pacienți au ales 4 (55 de pacienți, 34,4%), aproape la egalitate cu 3 (54). Doar 14,4% dintre pacienți au fost nemulțumiți sau foarte nemulțumiți (frecvența cumulată la valoarea 2), iar 51,9% au fost mulțumiți sau foarte mulțumiți (34,4 + 17,5). Frecvența cumulată trece de 50% la valoarea 4, deci mediana satisfacției este 4: cel puțin jumătate dintre pacienți au dat nota 4 sau 5.

Pentru variabilele nominale, tabelul se limitează la frecvențele absolute și relative. În mg2026, statutul de fumător are distribuția: nefumători 74 (46,3%), foști fumători 33 (20,6%), fumători 53 (33,1%). Mediul de proveniență: urban 95 (59,4%), rural 65 (40,6%). Sexul: femei 86 (53,8%), bărbați 74 (46,3%); suma procentelor rotunjite este 100,1%, ceea ce este normal și se poate menționa într-o notă sub tabel.

2.2.2 Gruparea datelor continue în clase

O variabilă continuă precum tensiunea sistolică are, în mg2026, zeci de valori distincte, de la 102 la 169 mmHg. Un tabel cu fiecare valoare ar avea peste 50 de rânduri, cele mai multe cu frecvența 1 sau 2. De aceea valorile se grupează în clase (intervale) de lățime egală.

Numărul de clase se alege astfel încât tabelul să fie lizibil și să păstreze forma distribuției. O regulă orientativă este regula lui Sturges:

$$k = 1 + 3{,}322 \times \log_{10}(n)$$

Pentru n = 160, log10(160) = 2,204, deci k = 1 + 3,322 x 2,204 = 8,32, adică aproximativ 8 clase. Lățimea clasei (pasul) se obține împărțind amplitudinea la numărul de clase:

$$h = \frac{x_{\max} - x_{\min}}{k}$$

Pentru TAS, amplitudinea este 169 - 102 = 67 mmHg, iar $h = \frac{67}{8} = 8{,}4$ mmHg. În practică, lățimea se rotunjește la o valoare comodă pentru cititor, aici 10 mmHg, iar limitele claselor se aleg la valori rotunde. Cu clase de 10 mmHg începând de la 100 rezultă 7 clase, ceea ce este perfect acceptabil: regula lui Sturges este un punct de plecare, nu o lege.

Limitele claselor trebuie să fie fără ambiguitate: fiecare valoare trebuie să aparțină unei singure clase. Pentru valori întregi, clasele se scriu 100-109, 110-119 și așa mai departe. Pentru valori cu zecimale se folosesc intervale de forma [100; 110), unde paranteza dreaptă arată că 110 nu aparține clasei, sau limite cu o zecimală în plus față de date (99,5-109,5), ca în tabelele din culegerile clasice de probleme. Mijlocul (centrul) clasei este media limitelor reale: pentru clasa 100-109, cu limitele reale 99,5 și 109,5, centrul este 104,5.

Tabelul 2.5. Tensiunea arterială sistolică la includere, grupată în clase de 10 mmHg, setul mg2026 (n = 160).

TAS (mmHg)	Centrul clasei	ni	fi (%)	Frecvența cumulată	Frecvența cumulată (%)
100-109	104,5	11	6,9	11	6,9
110-119	114,5	21	13,1	32	20,0
120-129	124,5	35	21,9	67	41,9
130-139	134,5	52	32,5	119	74,4
140-149	144,5	25	15,6	144	90,0
150-159	154,5	13	8,1	157	98,1
160-169	164,5	3	1,9	160	100,0
Total		160	100,0		

Din tabel se pot citi rapid informații utile clinic. Clasa cea mai frecventă (clasa modală) este 130-139 mmHg. Proporția pacienților cu TAS sub 140 mmHg este frecvența cumulată a clasei 130-139, adică 74,4%; prin urmare 25,6% au TAS ≥ 140 mmHg, exact proporția calculată din datele individuale (41 de pacienți). Proporția pacienților cu TAS de cel puțin 150 mmHg este 8,1 + 1,9 = 10,0%, adică 16 pacienți.

Gruparea are un preț: se pierde valoarea exactă a fiecărei observații. Dacă avem doar tabelul, media se poate estima presupunând că toate valorile unei clase sunt egale cu centrul ei:

$$m \approx \frac{\sum(n_i \times c_i)}{n}$$

unde ci este centrul clasei i. Pentru tabelul 2.5:

$$\frac{11\times104{,}5+21\times114{,}5+35\times124{,}5+52\times134{,}5+25\times144{,}5+13\times154{,}5+3\times164{,}5}{160} = \frac{21\,020}{160} = 131{,}38 \text{ mmHg.}$$

Media calculată din datele individuale este 131,13 mmHg; eroarea de grupare este mică, de un sfert de mmHg.

2.2.3 Tabelele de frecvențe cu valori lipsă

Un tip clasic de problemă, prezent în toate culegerile disciplinei și la examen, dă un tabel de frecvențe din care lipsesc câteva celule și cere reconstituirea lor. Nu este nevoie de nicio formulă nouă, ci doar de cele trei relații de mai sus, folosite în ordinea potrivită:

- frecvența absolută = frecvența relativă x n (rotunjită la întreg, pentru că numărăm pacienți);
- frecvența relativă = frecvența absolută / n x 100;
- frecvența cumulată a unei clase = frecvența cumulată a clasei anterioare + frecvența clasei; invers, frecvența unei clase = diferența dintre două frecvențe cumulate consecutive;
- suma frecvențelor absolute = n; suma frecvențelor relative = 100%.

Strategia de lucru este să începeți cu celulele care se pot calcula direct, apoi să le folosiți pe acestea pentru celelalte, și să verificați la final că totalurile se închid. Problemele rezolvate 2.1 și 2.2 arată două astfel de tabele rezolvate complet, iar lucrarea practică 2 conține încă două exerciții.

2.3 Indicatorii tendinței centrale

Tabelele de frecvențe arată întreaga distribuție. Adesea avem însă nevoie de o singură valoare care să rezume "centrul" datelor, valoarea tipică. Există mai mulți astfel de indicatori, iar alegerea dintre ei depinde de tipul variabilei și de forma distribuției.

2.3.1 Media aritmetică

Media aritmetică este suma valorilor împărțită la numărul lor:

$$m = \frac{x_1 + x_2 + \cdots + \text{xn}}{n} = \frac{\sum x_i}{n}$$

Este indicatorul cel mai cunoscut și cel mai folosit, are proprietăți matematice excelente și stă la baza majorității testelor statistice. Pentru tensiunea sistolică din mg2026, m = 131,13 mmHg. Media are o interpretare fizică simplă: este "centrul de greutate" al datelor, punctul în care suma abaterilor pozitive este egală cu suma abaterilor negative, $\sum(x_i - m) = 0$.

Media are și un defect important: este sensibilă la valorile extreme. O singură valoare foarte mare o trage în sus, o singură valoare foarte mică o trage în jos. De aceea media este indicatorul potrivit pentru variabilele numerice cu distribuție aproximativ simetrică, dar poate fi înșelătoare pentru distribuțiile asimetrice.

Media ponderată se folosește atunci când valorile au importanțe (ponderi) diferite, de exemplu când combinăm medii calculate pe grupuri de mărimi diferite:

$$\text{mp} = \frac{\sum(w_i \times x_i)}{\sum w_i}$$

Media TAS în cele trei brațe ale studiului mg2026 este 131,53 mmHg (n = 53), 132,79 mmHg (n = 53) și 129,11 mmHg (n = 54). Media globală se obține ponderând cu mărimea grupurilor: $\frac{53\times131{,}53+53\times132{,}79+54\times129{,}11}{160} = \frac{6\,971{,}09+7\,037{,}87+6\,971{,}94}{160} = \frac{20\,980{,}90}{160} = 131{,}13$ mmHg, adică exact media celor 160 de pacienți. Aici grupurile au mărimi aproape egale, iar media simplă a celor trei medii (131,14) diferă foarte puțin. Diferența devine importantă când grupurile sunt foarte inegale, cum este cazul celor trei cabinete din problema rezolvată 1.5. Estimarea mediei din tabelul de frecvențe, din secțiunea 2.2.2, este tot o medie ponderată, cu ponderile ni.

2.3.2 Mediana

Mediana este valoarea care împarte seria ordonată a observațiilor în două jumătăți egale ca număr: jumătate dintre observații sunt mai mici sau egale cu ea, jumătate mai mari sau egale. Pentru a o calcula, se ordonează valorile crescător. Dacă n este impar, mediana este valoarea din poziția (n + 1)/2. Dacă n este par, mediana este media celor două valori din mijloc, din pozițiile n/2 și n/2 + 1.

Pentru vârstele primilor 20 de pacienți din mg2026, ordonate crescător (37, 40, 45, 47, 50, 52, 53, 54, 55, 56, 56, 57, 57, 60, 62, 62, 65, 66, 69, 70), n = 20 este par, iar valorile din pozițiile 10 și 11 sunt 56 și 56; mediana este 56 de ani. Pentru întregul set, mediana vârstei este tot 56 de ani, iar mediana TAS este 131 mmHg.

Mediana este robustă: nu este influențată de valorile extreme, ci doar de poziția lor. Dacă cel mai vârstnic dintre cei 20 de pacienți ar avea 99 de ani în loc de 70, mediana ar rămâne 56. De aceea mediana este indicatorul preferat pentru variabilele asimetrice (trigliceride, durata spitalizării, valori ale unor markeri biologici, timpi de așteptare) și pentru variabilele ordinale, unde media nu are o interpretare corectă.

2.3.3 Modul (moda)

Modul (valoarea modală) întâlinita sub denumirea de modă este valoarea cu frecvența cea mai mare. Este singurul indicator al tendinței centrale care se poate folosi pentru variabile nominale: modul statutului de fumător în mg2026 este "nefumător" (74 de pacienți). Pentru variabilele numerice, modul este mai puțin util, pentru că depinde mult de întâmplare și de rotunjire. În mg2026, valoarea TAS cea mai frecventă este 138 mmHg (10 pacienți), vârsta cea mai frecventă este 51 de ani (10 pacienți), iar scorul FINDRISC cel mai frecvent este 10 (23 de pacienți).

O distribuție poate avea mai multe moduri. O distribuție bimodală, cu două vârfuri, sugerează adesea că eșantionul amestecă două populații diferite, de exemplu înălțimea într-un grup cu femei și bărbați (în mg2026, înălțimea medie este 162,6 cm la femei și 175,97 cm la bărbați). Într-un

astfel de caz, media globală descrie un "pacient mediu" care nu există, iar descrierea separată pe subgrupuri este mai informativă.

2.3.4 Media geometrică

Media geometrică a n valori pozitive este rădăcina de ordinul n a produsului lor. Se calculează mai comod prin logaritmare: se calculează media logaritmilor, apoi se revine la scara inițială prin funcția exponențială:

$$\text{mg} = \exp\left[\frac{\ln x_1 + \ln x_2 + \cdots + \ln \text{xn}}{n}\right]$$

Media geometrică este potrivită pentru variabilele pozitive cu asimetrie spre dreapta, a căror distribuție devine aproximativ simetrică după logaritmare: concentrațiile unor markeri biologici, titrurile de anticorpi, încărcătura virală, trigliceridele. În mg2026, media aritmetică a trigliceridelor este 161,18 mg/dL, mediana 151,5 mg/dL, iar media logaritmilor naturali este 4,97; media geometrică este $\exp(4{,}97) \approx 144{,}0$ mg/dL. Media geometrică este mai mică decât media aritmetică (este întotdeauna așa, cu excepția cazului în care toate valorile sunt egale) și apropiată de mediană, pentru că nu este trasă în sus de cele câteva valori foarte mari (maximul este 442 mg/dL). Capitolul 3 va reveni la transformarea logaritmică, iar capitolul 4 va arăta cum se calculează un interval de încredere pentru media geometrică.

Un exemplu simplu arată logica. Pentru trei titruri de anticorpi de 1/10, 1/100 și 1/1 000, media aritmetică a numitorilor este 370, o valoare care nu seamănă cu niciuna dintre cele trei. Media geometrică este rădăcina cubică din 10 x 100 x 1 000 = 1 000 000, adică 100, valoarea din mijloc pe scara logaritmică pe care se gândește de fapt laboratorul.

2.3.5 Alegerea indicatorului și efectul valorilor extreme

Relația dintre medie și mediană spune ceva despre forma distribuției. Într-o distribuție simetrică, media și mediana sunt practic egale: pentru TAS, 131,13 și 131 mmHg; pentru vârstă, 56,29 și 56 de ani. Într-o distribuție cu asimetrie pozitivă (coadă lungă spre dreapta), media este mai mare decât mediana. Pentru IMC, media este 29,04 kg/m², iar mediana 28,25 kg/m². Figura 2.2 arată de ce: câțiva pacienți cu obezitate severă, până la 48,3 kg/m², trag media spre dreapta, dar nu mută mediana.

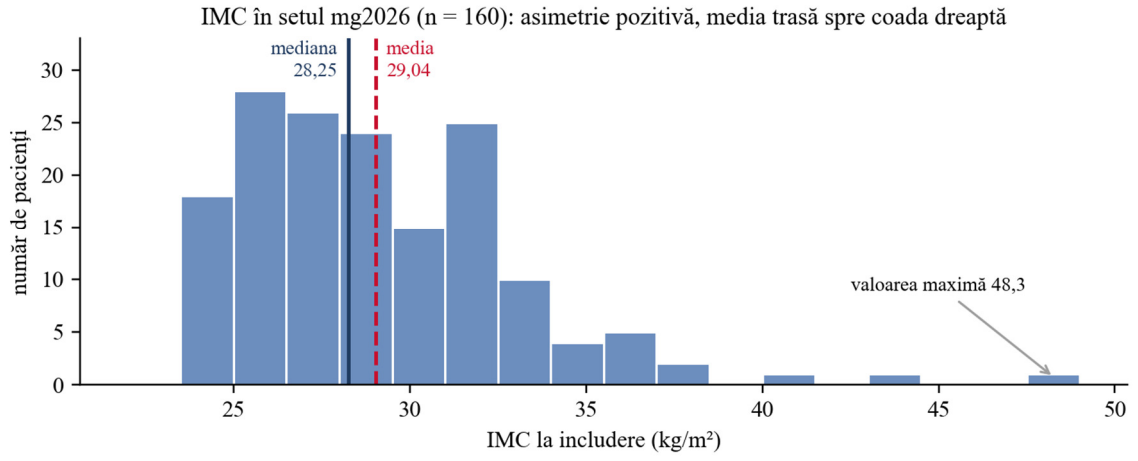


Figura 2.2. Distribuția IMC la includere în setul mg2026 (n = 160). Histograma are o coadă spre dreapta, iar media (29,04 kg/m², linia roșie întreruptă) este mai mare decât mediana (28,25 kg/m², linia albastră). Valoarea maximă, 48,3 kg/m², influențează media, dar nu și mediana. Date sintetice (setul mg2026).

Experimentul este ușor de făcut. Dacă eliminăm pacientul cu IMC 48,3 kg/m², media celor 159 de pacienți rămași scade de la 29,04 la 28,92 kg/m², iar abaterea standard de la 3,91 la 3,61

kg/m²; mediana trece de la 28,25 la 28,2 kg/m², o diferență neglijabilă. O singură observație din 160 a schimbat media cu 0,12 kg/m² și abaterea standard cu aproape 8%. Pe un eșantion de 20 de pacienți, același pacient ar fi schimbat media cu peste 1 kg/m².

Asta nu înseamnă că pacientul cu IMC 48,3 kg/m² trebuie eliminat. Valoarea este plauzibilă biologic (obezitate de gradul III), a fost verificată, iar pacientul face parte din populația studiată. Valorile extreme se elimină doar dacă sunt erori demonstrate; altfel, se aleg metode care le tolerează (mediana, intervalul intercuartilic, testele neparametrice) sau se raportează analiza cu și fără ele.

Tabelul 2.6. Alegerea indicatorului tendinței centrale.

Tipul datelor	Indicator recomandat	Observații
Nominale	modul; proporțiile fiecărei categorii	media codurilor nu are sens
Ordinale	mediana (și modul)	media se raportează uneori pentru scoruri cu multe valori, cu prudență
Numerice, aproximativ simetrice	media (și mediana)	TAS, vârstă, colesterol, înălțime în mg2026
Numerice, asimetrice	mediana; media geometrică pentru date log-normale	trigliceride, IMC, durata spitalizării, markeri inflamatori
Date cu valori extreme plauzibile	mediana	se raportează și influența lor asupra mediei

2.4 Indicatorii de dispersie

Două grupuri de pacienți pot avea aceeași medie a tensiunii sistolice, 131 mmHg, și totuși să fie foarte diferite: într-unul toți pacienții au între 125 și 137 mmHg, în celălalt valorile se întind de la 100 la 165 mmHg. Tendința centrală nu spune nimic despre această împrăștiere. Pentru ea avem nevoie de indicatorii de dispersie. Ca să-i calculăm de mână, vom folosi un exemplu mic.

Exemplu de lucru. Opt pacienți au la un control glicemiile a jeun (mg/dL): 88, 94, 97, 102, 105, 110, 116, 128. Suma lor este 840, deci media este $m = \frac{840}{8} = 105$ mg/dL. Valorile sunt deja ordonate; mediana este media valorilor din pozițiile 4 și 5, (102 + 105)/2 = 103,5 mg/dL.

2.4.1 Amplitudinea

Amplitudinea (engl. range) este diferența dintre valoarea maximă și cea minimă: 128 - 88 = 40 mg/dL în exemplul nostru. Pentru TAS în mg2026, amplitudinea este 169 - 102 = 67 mmHg. Este simplă și intuitivă, dar depinde doar de două observații, cele extreme, și crește aproape întotdeauna cu volumul eșantionului: cu cât studiem mai mulți pacienți, cu atât e mai probabil să întâlnim valori neobișnuite. În articole se raportează de obicei sub forma "minim-maxim", nu ca o diferență.

2.4.2 Abaterea medie absolută

Abaterea (deviația) unei observații de la medie este xi - m. Pentru exemplul nostru, abaterile sunt -17, -11, -8, -3, 0, 5, 11, 23. Suma lor este zero, ca întotdeauna, deci media abaterilor nu poate măsura dispersia. O soluție este să luăm valorile absolute. Abaterea medie absolută este media distanțelor față de medie:

$$\text{AMA} = \frac{\sum |x_i - m|}{n} = \frac{17 + 11 + 8 + 3 + 0 + 5 + 11 + 23}{8} = \frac{78}{8} = 9{,}75 \text{ mg/dL}$$

Pacienții se află, în medie, la aproape 10 mg/dL de media grupului. Indicatorul este ușor de înțeles, dar are proprietăți matematice incomode, așa că în statistica inferențială se folosește mult mai rar decât abaterea standard. Pentru TAS în mg2026, AMA este 10,91 mmHg.

2.4.3 Varianța și abaterea standard

A doua soluție la problema semnelor este ridicarea abaterilor la pătrat. Varianța eșantionului este suma pătratelor abaterilor împărțită la n - 1:

$$s^2 = \frac{\sum(x_i - m)^2}{n - 1}$$

Abaterea standard (deviația standard, engl. standard deviation, notată s sau AS) este rădăcina pătrată a variației și are aceeași unitate de măsură ca datele:

$$s = \sqrt{s^2}$$

Tabelul 2.7 arată calculul complet pentru cele opt glicemii.

Tabelul 2.7. Calculul de mână al variației și al abaterii standard pentru opt glicemii.

x_i (mg/dL)	**xi - m**	$(x_i - m)^2$
88	-17	289
94	-11	121
97	-8	64
102	-3	9
105	0	0
110	5	25
116	11	121
128	23	529
Σ = 840	Σ = 0	Σ = 1 158

$$s^2 = \frac{1\,158}{8 - 1} = \frac{1\,158}{7} = 165{,}43 \text{ (mg/dL)}^2$$

$$s = \sqrt{165{,}43} = 12{,}86 \text{ mg/dL}$$

De ce n - 1 și nu **n**? Abaterile sunt calculate față de media eșantionului, nu față de media necunoscută a populației, iar media eșantionului este, prin construcție, punctul față de care suma pătratelor este cea mai mică. Împărțirea la n ar subestima sistematic varianța populației. Împărțirea la n - 1 corectează această deplasare; numărul n - 1 se numește număr de grade de libertate, pentru că, odată cunoscută media, doar n - 1 abateri pot varia liber (ultima este determinată de condiția ca suma să fie zero). În Google Sheets, funcția STDEV (sau STDEV.S) împarte la n - 1, iar STDEVP (STDEV.P) împarte la n; pentru datele unui eșantion se folosește întotdeauna prima. Pentru n mare, diferența este neglijabilă.

Abaterea standard este cel mai folosit indicator de dispersie. Ea are o interpretare foarte utilă pentru distribuțiile aproximativ normale, pe care o vom detalia în capitolul 3: aproximativ 68% dintre valori se află în intervalul $m \pm s$ și aproximativ 95% în intervalul $m \pm 2s$. Pentru TAS din mg2026, s = 13,88 mmHg, iar intervalul $m \pm s$ este 117,3-145,0 mmHg. Varianța, în schimb, are unitatea de măsură la pătrat (mmHg², (mg/dL)²), greu de interpretat clinic; pentru TAS, $s^2 = 192{,}66$ mmHg^2. Ea este însă indispensabilă în calcule, de exemplu în analiza variației din capitolul 6.

2.4.4 Coeficientul de variație

Abaterea standard depinde de unitatea de măsură și de ordinul de mărime al datelor. O abatere standard de 13,88 mmHg pentru TAS și una de 0,45% pentru HbA1c nu pot fi comparate direct. Coeficientul de variație exprimă abaterea standard ca procent din medie:

$$\text{CV} = \frac{s}{m} \times 100$$

Pentru cele opt glicemii, $\text{CV} = \frac{12{,}86}{105} \times 100 = 12{,}2\%$. În mg2026, CV este 10,6% pentru TAS, 15,6% pentru vârstă, 13,5% pentru IMC, 8,0% pentru HbA1c, 17,6% pentru colesterol și 47,3% pentru trigliceride. Trigliceridele sunt, relativ la media lor, de aproape cinci ori mai variabile decât tensiunea sistolică. Un CV mare (peste 30-50%) la o variabilă care nu poate fi negativă este și un semnal de asimetrie: dacă abaterea standard ar fi aproape egală cu media, intervalul m - 2s ar conține valori negative, imposibile biologic.

Coeficientul de variație are două utilizări frecvente în medicină: compararea variabilității unor mărimi diferite și evaluarea preciziei metodelor de laborator (un CV analitic mic arată o metodă reproductibilă). El are sens doar pentru variabilele pe scală de raport, cu zero absolut și medie pozitivă. Pentru temperatura în grade Celsius, de exemplu, CV-ul s-ar schimba dacă am trece la grade Fahrenheit, deci nu are interpretare.

2.4.5 Abaterea standard și eroarea standard: o confuzie frecventă

Eroarea standard a mediei (engl. standard error of the mean, ES sau SEM) se calculează din abaterea standard și din volumul eșantionului:

$$\text{ES} = \frac{s}{\sqrt{n}}$$

Pentru cele opt glicemii, $\text{ES} = \frac{12{,}86}{\sqrt{8}} = \frac{12{,}86}{2{,}83} = 4{,}55$ mg/dL. Pentru TAS în mg2026, $\text{ES} = \frac{13{,}88}{\sqrt{160}} = \frac{13{,}88}{12{,}65} = 1{,}097$ mmHg.

Cele două mărimi răspund la întrebări diferite. Abaterea standard descrie cât de mult variază pacienții între ei; este o caracteristică a populației și nu scade când studiem mai mulți pacienți. Eroarea standard descrie cât de precis estimează media eșantionului media populației; ea scade pe măsură ce n crește. Dacă am studia 640 de pacienți în loc de 160, abaterea standard a TAS ar rămâne în jur de 14 mmHg, dar eroarea standard s-ar înjumătăți, la aproximativ 0,55 mmHg.

Confuzia are consecințe practice. Unele articole raportează "media ± ES" în tabelele descriptive, fără să precizeze. Pentru TAS, "131,13 ± 13,88" și "131,13 ± 1,10" descriu aceleași date, dar a doua formă sugerează, pentru un cititor neatent, o populație extrem de omogenă. Regula de raportare este simplă: pentru a descrie pacienții se folosește abaterea standard; pentru a descrie precizia unei estimări se folosește intervalul de încredere (capitolul 4), care se construiește cu eroarea standard. Orice tabel trebuie să precizeze ce reprezintă cifra de după semnul ±.

2.5 Cuantile, percentile, cuartile și rezumatul în cinci numere

Mediana este cazul particular al unei familii mai largi de indicatori de poziție, numiți cuantile. O cuantilă de ordinul q (între 0 și 1) este valoarea sub care se află o fracție q din observații. Percentilele împart seria ordonată în 100 de părți: percentila 10 (P10) este valoarea sub care se află 10% din observații, P90 valoarea sub care se află 90%. Cuartilele împart seria în patru părți egale: prima cuartilă Q1 = P25, a doua Q2 = P50 = mediana, a treia Q3 = P75.

Percentilele sunt familiare oricărui medic pediatru, din curbele de creștere: un copil "pe percentila 90 la greutate" cântărește mai mult decât 90% dintre copiii de aceeași vârstă și sex din populația de referință. În mg2026, P10 și P90 pentru TAS sunt 112 și 148,2 mmHg, iar pentru glicemia a jeun 86 și 126 mg/dL. Altfel spus, 10% dintre pacienți au o glicemie de cel puțin 126 mg/dL, pragul diagnostic pentru diabet, ceea ce se potrivește cu numărarea directă: 18 pacienți din 160 (11,3%).

2.5.1 Calculul cuartilelor

Pentru serii mici, cuartilele se pot calcula de mână prin metoda "medianelor jumătăților": se găsește mediana, apoi Q1 este mediana jumătății inferioare, iar Q3 mediana jumătății superioare. Pentru cele opt glicemii (88, 94, 97, 102 | 105, 110, 116, 128), jumătatea inferioară are mediana (94 + 97)/2 = 95,5, iar cea superioară (110 + 116)/2 = 113. Deci $Q_1 = 95{,}5$ mg/dL și $Q_3 = 113$ mg/dL.

Programele folosesc formule de interpolare ușor diferite. Google Sheets (funcția QUARTILE) și JASP folosesc, implicit, o interpolare liniară care, pentru aceleași opt valori, dă Q1 = 96,25 și $Q_3 = 111{,}5$ mg/dL. Pe eșantioane mari diferențele devin neglijabile, iar pentru seriile mici de la examen se acceptă oricare metodă, cu condiția să fie aplicată corect. Nu vă speriați dacă MedCalc și JASP dau cuartile ușor diferite pentru aceleași date.

2.5.2 Intervalul intercuartilic şi rezumatul în cinci numere

Intervalul intercuartilic (IQR, engl. interquartile range) este diferența Q3 - Q1 și cuprinde jumătatea centrală a observațiilor. Este indicatorul de dispersie care însoțește în mod natural mediana și, ca și ea, este robust la valorile extreme. În articole se raportează de obicei ca interval, "mediană (Q1-Q3)", de exemplu IMC 28,25 (26,2-31,4) kg/m², nu ca o singură diferență.

Rezumatul în cinci numere (minim, Q1, mediană, Q3, maxim) descrie compact o distribuție și stă la baza diagramei box-plot. Pentru variabilele principale din mg2026:

Tabelul 2.8. Rezumatul în cinci numere pentru câteva variabile din mg2026 (n = 160).

Variabila	Minim	Q1	Mediana	Q3	Maxim	IQR
Vârsta (ani)	35	50,75	56	62	79	11,25
IMC $\left(\frac{\text{kg}}{m^2}\right)$	23,5	26,2	28,25	31,4	48,3	5,2
TAS (mmHg)	102	122	131	140	169	18
Glicemie a jeun (mg/dL)	75	93	104,5	116	154	23
Trigliceride (mg/dL)	44	107	151,5	203,25	442	96,25

Un rezumat în cinci numere se citește ca o hartă. Pentru TAS, distanțele de la mediană la Q1 (9 mmHg) și la Q3 (9 mmHg) sunt egale, semn de simetrie. Pentru trigliceride, distanța de la mediană la Q3 (51,75) este mai mare decât cea până la Q1 (44,5), iar maximul (442) este foarte departe de Q3: distribuția are o coadă lungă spre dreapta.

2.5.3 Valorile extreme: regula 1,5 x IQR

O regulă practică, propusă de Tukey, consideră "valori extreme" (engl. outliers) observațiile situate la mai mult de 1,5 x IQR sub Q1 sau peste Q3. Limitele (numite uneori "garduri") sunt:

limita inferioară = Q1 - 1,5 x IQR; limita superioară = Q3 + 1,5 x IQR

Pentru TAS: IQR = 18, deci limitele sunt 122 - 27 = 95 și 140 + 27 = 167 mmHg. Două valori (168 și 169 mmHg) depășesc limita superioară. Pentru IMC: IQR = 5,2, limitele sunt 26,2 - 7,8 = 18,4 și 31,4 + 7,8 = 39,2 kg/m²; trei valori depășesc limita superioară (40,0; 43,4 și 48,3). Pentru trigliceride: limita superioară este $203{,}25 + 1{,}5 \times 96{,}25 = 347{,}6$ mg/dL, depășită de cinci valori (364, 364, 373, 405 și 442 mg/dL).

Regula identifică valori care merită privite, nu valori care trebuie șterse. O tensiune de 169 mmHg sau o trigliceridemie de 442 mg/dL sunt întâlnite zilnic în cabinet. Întrebarea corectă este întotdeauna "Este valoarea posibilă și a fost verificată?", nu "Strică media?". În schimb, o tensiune de 1540 mmHg, cea din fișierul brut, nu este o valoare extremă, ci o eroare: este imposibilă biologic.

2.6 Forma distribuției: o primă privire asupra asimetriei și excesului

Doi indicatori numerici descriu forma unei distribuții și vor fi studiați detaliat în capitolul 3. Coeficientul de asimetrie (engl. skewness) este aproximativ zero pentru o distribuție simetrică, pozitiv pentru o distribuție cu coada spre dreapta (valori mari rare) și negativ pentru o coadă spre stânga. Coeficientul de exces (engl. excess kurtosis, aplatizare) compară "greutatea cozilor" cu cea a distribuției normale: este zero pentru distribuția normală, pozitiv pentru distribuții cu cozi grele și vârf ascuțit, negativ pentru distribuții mai turtite.

În mg2026, TAS are asimetria 0,01 și excesul -0,10, practic valorile unei distribuții normale. IMC are asimetria 1,42 și excesul 3,68, deci o coadă dreaptă pronunțată și câteva valori foarte mari. Trigliceridele au asimetria 0,95 și excesul 1,12; după logaritmare, asimetria devine -0,28, aproape de simetrie. Glicemia a jeun este ușor asimetrică (0,37). În practică, o asimetrie între -0,5 și 0,5 este considerată mică, una peste 1 în valoare absolută este importantă. Pentru variabilele biologice care nu pot fi negative și au medii mici față de abaterea standard, asimetria pozitivă este regula, nu excepția.

Un indiciu simplu, fără niciun calcul: dacă media este sensibil mai mare decât mediana și abaterea standard este comparabilă cu media, distribuția este aproape sigur asimetrică spre dreapta.

2.7 Reprezentările grafice

Un grafic bun arată în câteva secunde ceea ce un tabel arată în câteva minute. Alegerea graficului depinde, ca și alegerea indicatorilor, de tipul variabilei. Figura 2.3 prezintă patru dintre cele mai folosite grafice, construite pe setul mg2026.

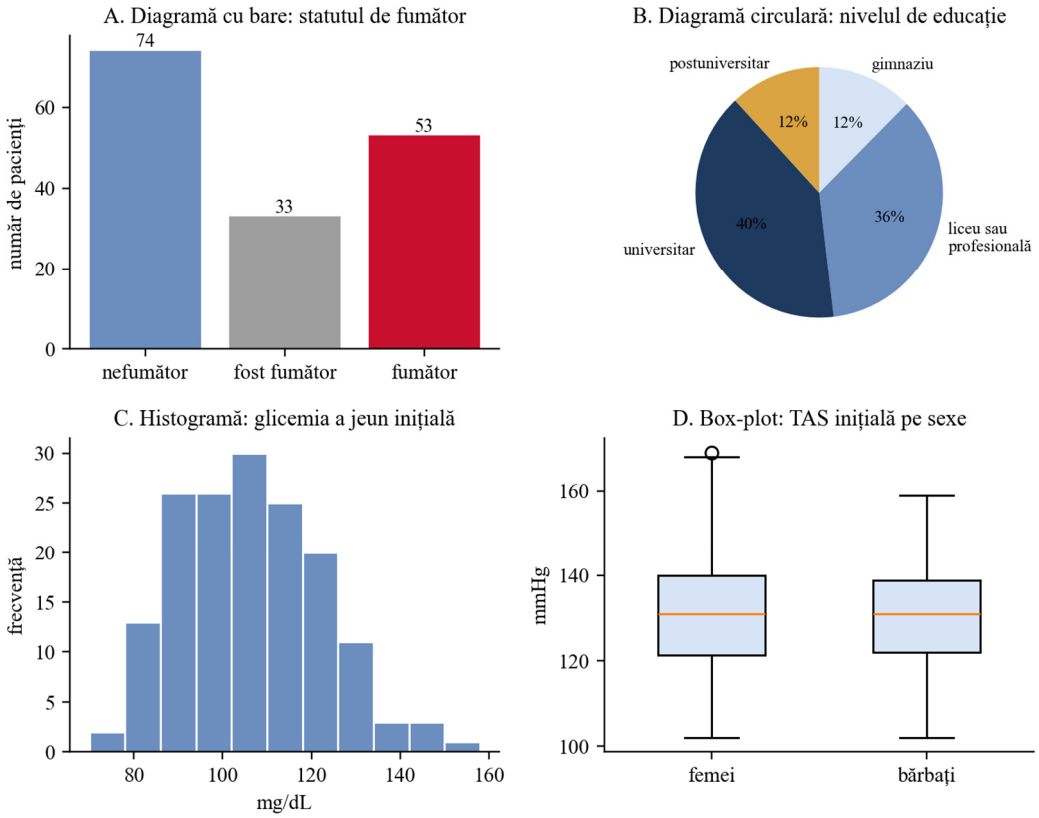


Figura 2.3. Grafice uzuale pentru date biomedicale, setul mg2026 (n = 160). A: diagramă cu bare pentru statutul de fumător (variabilă nominală). B: diagramă circulară pentru nivelul de educație (variabilă ordinală cu patru categorii). C: histograma glicemiei a jeun la includere (variabilă continuă). D: diagrame box-plot ale tensiunii sistolice la femei și la bărbați. Date sintetice.

2.7.1 Grafice pentru variabile calitative

Diagrama cu bare (coloane) reprezintă fiecare categorie printr-o bară cu înălțimea proporțională cu frecvența (absolută sau relativă). Barele sunt separate între ele, pentru că categoriile sunt distincte. Pentru variabilele nominale, ordinea barelor se poate alege liber (de exemplu, descrescător după frecvență); pentru cele ordinale se păstrează ordinea naturală a categoriilor. Figura 2.3A arată că nefumătorii sunt cel mai numeros grup (74), urmați de fumători (53) și de foștii fumători (33). Diagramele cu bare grupate sau stivuite permit compararea unei variabile între grupuri, de exemplu statutul de fumător pe sexe.

Diagrama circulară (engl. pie chart) împarte un cerc în sectoare proporționale cu frecvențele relative; unghiul fiecărui sector este fi x 360°. Ea arată bine structura unui întreg cu puține categorii (două până la cinci). Pentru nivelul de educație din mg2026 (figura 2.3B), sectoarele corespund proporțiilor de 12,5% (gimnaziu), 35,6% (liceu sau profesională), 40,0% (universitar) și 11,9% (postuniversitar); sectorul universitar are un unghi de 0,40 x 360 = 144°. Ochiul uman compară însă greu unghiurile apropiate, așa că, pentru mai mult de cinci categorii sau pentru comparații între grupuri, diagrama cu bare este aproape întotdeauna mai bună.

2.7.2 Grafice pentru variabile cantitative

Histograma este reprezentarea grafică a tabelului de frecvențe pentru date grupate în clase. Pe axa orizontală sunt clasele, pe cea verticală frecvențele, iar barele sunt lipite unele de altele, pentru că variabila este continuă. Histograma arată forma distribuției: simetrică sau asimetrică, cu un vârf sau cu două, cu valori izolate. Figura 2.3C arată o distribuție a glicemiei cu un vârf în jurul valorilor de 95-115 mg/dL și o coadă ușoară spre dreapta. Aspectul histogramei depinde de lățimea claselor: clase prea late ascund forma distribuției, clase prea înguste produc un grafic zimțat. Merită încercate două-trei variante.

Diagrama box-plot (cutie cu mustăți) desenează rezumatul în cinci numere. Cutia se întinde de la Q1 la Q3, o linie în interiorul ei marchează mediana, iar mustățile se prelungesc până la cea mai mică și cea mai mare valoare situate în limitele 1,5 x IQR; valorile din afara acestor limite apar ca puncte separate. Box-plotul este ideal pentru compararea mai multor grupuri. În figura 2.3D, mediana TAS este 131 mmHg atât la femei, cât și la bărbați, iar cutiile se suprapun aproape complet (femei: Q1 121,25, Q3 140; bărbați: Q1 122, Q3 138,75). Graficul anticipează rezultatul testului din capitolul 6: nu există o diferență de TAS între sexe în acest eșantion.

Diagrama tulpină-frunză (engl. stem-and-leaf) este o histogramă "scrisă", care păstrează valorile individuale. Fiecare valoare se desparte într-o tulpină (cifrele din față) și o frunză (ultima cifră). Pentru vârstele primilor 20 de pacienți din mg2026:

Tabelul 2.9. Diagrama tulpină-frunză pentru vârstele primilor 20 de pacienți (tulpina = zecile, frunza = unitățile).

Tulpina	Frunzele
3	7
4	0 5 7
5	0 2 3 4 5 6 6 7 7
6	0 2 2 5 6 9
7	0

Diagrama arată imediat că majoritatea pacienților au între 50 și 59 de ani și că distribuția este aproximativ simetrică. Din ea se pot citi direct minimul (37), maximul (70) și mediana (media valorilor din pozițiile 10 și 11, ambele 56). Este utilă pentru seturi mici, scrise de mână.

Diagrama cu puncte (engl. dot plot) reprezintă fiecare observație printr-un punct pe o axă; valorile egale se așază una peste alta. Pentru eșantioane mici (sub 30-40 de observații) este mai sinceră decât o diagramă cu bare care arată doar media, pentru că lasă cititorul să vadă fiecare

pacient. Multe reviste recomandă astăzi, pentru grupuri mici, grafice care arată punctele individuale peste box-plot.

Diagrama de dispersie (norul de puncte, engl. scatter plot) reprezintă relația dintre două variabile numerice: fiecare pacient este un punct cu coordonatele (x, y). O vom folosi pe larg în capitolul 9 pentru corelație și regresie; de exemplu, între IMC și TAS în mg2026 norul de puncte urcă moderat, cu un coeficient de corelație r = 0,405. Diagrama de dispersie ajută și la judecarea unor rezultate publicate. Într-un studiu pe 867 de pacienți cu cancer de colon, corelația dintre vârstă și durata intervenției a fost r = -0,088, semnificativă statistic (p = 0,01) [55]. Un nor de puncte cu r = -0,088 arată practic ca o pată fără nicio direcție; semnificația statistică vine din volumul mare al eșantionului, nu dintr-o relație importantă clinic.

Graficul liniar (engl. line plot) unește prin linii valori ordonate în timp: media glicemiei la 0, 3 și 6 luni, incidența lunară a unei boli, evoluția temperaturii unui pacient pe foaia de observație. Liniile au sens doar când axa orizontală are o ordine naturală, de regulă timpul.

Tabelul 2.10. Alegerea graficului în funcție de tipul variabilei.

Variabila (variabilele)	**Grafic recomandat**	**De evitat**
O variabilă nominală	bare; circulară pentru 2-5 categorii	circulară cu multe categorii sau 3D
O variabilă ordinală	bare în ordinea categoriilor; bare stivuite pentru scale Likert	circulară care amestecă ordinea
O variabilă numerică	histogramă; box-plot; tulpină-frunză sau puncte pentru n mic	bare cu media singură
Numerică pe grupuri	box-plot alăturate; puncte individuale pentru n mic	bare cu "media ± ES" fără precizare
Două variabile numerice	diagramă de dispersie	linii între puncte fără ordine în timp
Evoluție în timp	grafic liniar	bare 3D

2.7.3 Reguli pentru un grafic corect

Un grafic poate fi corect din punct de vedere numeric și totuși înșelător. Figura 2.4 arată cea mai frecventă capcană: axa trunchiată. Glicemia medie a scăzut în mg2026 de la 105,4 mg/dL la includere la 99,4 mg/dL la 3 luni și la 95,2 mg/dL la 6 luni, adică cu aproximativ 10%. Dacă axa verticală începe la 94 mg/dL (panoul A), bara de la 6 luni pare de aproape zece ori mai mică decât cea de la includere, ca și cum glicemia s-ar fi prăbușit. Cu axa pornind de la zero (panoul B), proporțiile dintre bare sunt cele reale.

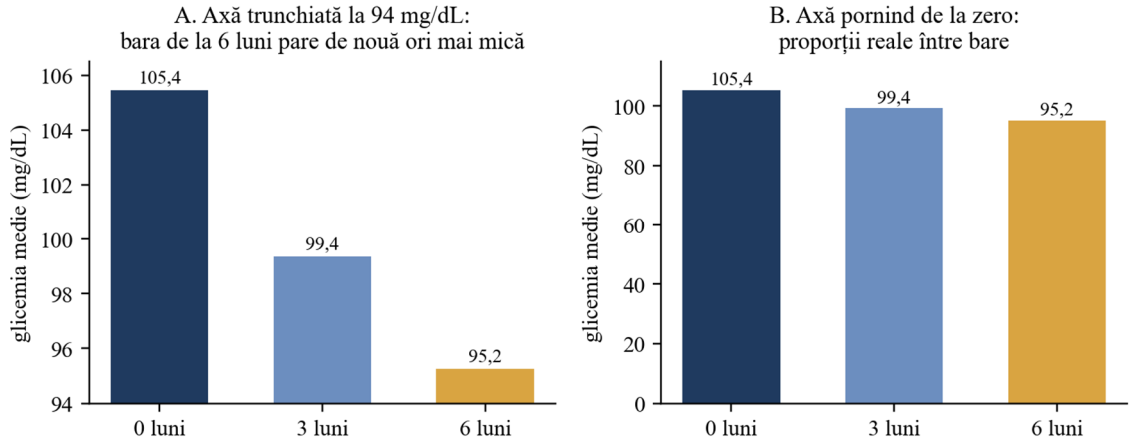


Figura 2.4. Același rezultat, două impresii diferite. Glicemia medie la 0, 3 și 6 luni în setul mg2026 (105,4; 99,4; 95,2 mg/dL). A: axa verticală trunchiată la 94 mg/dL exagerează vizual scăderea. B: axa pornind de la zero păstrează proporțiile reale dintre bare. Date sintetice (setul mg2026).

Regula este simplă: la diagramele cu bare, axa valorilor începe de la zero, pentru că lungimea barei este cea care transmite mărimea. La graficele liniare și la diagramele de dispersie, axa poate începe de la o valoare diferită de zero, pentru că acolo contează poziția punctelor, nu lungimea barelor; în acest caz, întreruperea axei trebuie marcată vizibil.

Alte reguli, adunate din experiența corectării lucrărilor:

- Fără efecte tridimensionale. O diagramă circulară 3D înclinată face ca sectoarele din față să pară mai mari decât sunt.
- Fără două axe verticale cu scări diferite pe același grafic, decât dacă este absolut necesar; alegerea scărilor poate face ca două curbe să pară că se urmăresc sau se despart, după dorința autorului.
- Axele au titlu și unitate de măsură (TAS, mmHg; glicemie, mg/dL).
- Legenda sau titlul figurii precizează n, perioada și sursa datelor; figura trebuie să poată fi înțeleasă fără text.
- Barele de eroare sunt explicate: abatere standard, eroare standard sau interval de încredere de 95%. Fără această precizare, barele nu pot fi interpretate.
- Culorile se aleg astfel încât graficul să poată fi citit și tipărit alb-negru și de persoanele cu deficit de vedere cromatică.
- Un grafic, un mesaj. Dacă graficul are nevoie de un paragraf de explicații ca să fie înțeles, probabil trebuie împărțit în două.

2.8 Tabelul descriptiv dintr-un articol: "Tabelul 1"

Aproape orice articol de cercetare clinică începe secțiunea de rezultate cu un tabel care descrie participanții, numit informal "Tabelul 1". El prezintă caracteristicile demografice și clinice ale eșantionului, adesea separat pe grupurile comparate. Convențiile sunt aceleași în toate revistele: variabilele numerice simetrice se raportează ca medie ± abatere standard; variabilele numerice asimetrice sau ordinale, ca mediană (Q1-Q3); variabilele calitative, ca număr de pacienți și procent, n (%). Ghidurile de raportare cer ca tabelul să precizeze, pentru fiecare variabilă, câte valori lipsesc [25]. Tabelul 2.11 este "Tabelul 1" al studiului simulat mg2026.

Tabelul 2.11. Caracteristicile pacienților la includere, pe brațe de tratament, setul mg2026.

Caracteristica	Braț A (n = 53)	Braț B (n = 53)	Braț C (n = 54)	Total (n = 160)
Vârsta, ani, medie ± AS	57,53 ± 9,19	56,87 ± 9,42	54,52 ± 7,47	56,29 ± 8,77
Sex feminin, n (%)	32 (60,4)	26 (49,1)	28 (51,9)	86 (53,8)
Mediu urban, n (%)	25 (47,2)	36 (67,9)	34 (63,0)	95 (59,4)
Fumători activi, n (%)	18 (34,0)	18 (34,0)	17 (31,5)	53 (33,1)
IMC, kg/m², mediană (Q1-Q3)	29,2 (26,6-31,1)	27,9 (26,1-31,4)	27,7 (25,95-31,4)	28,25 (26,2-31,4)
TAS, mmHg, medie ± AS	131,53 ± 13,84	132,79 ± 13,79	129,11 ± 14,01	131,13 ± 13,88
Glicemie a jeun, mg/dL, medie ± AS	107,79 ± 15,65	106,28 ± 14,70	102,30 ± 16,64	105,44 ± 15,76
Trigliceride, mg/dL, mediană (Q1-Q3)	143 (95-204)	144 (112-185)	166,5 (107-213)	151,5 (107-203,25)
Scor FINDRISC, mediană (Q1-Q3)	10	10	10	10 (8-13)
Diabet confirmat prin TTGO, n (%)	10 (18,9)	9 (17,0)	10 (18,5)	29 (18,1)

Notă: AS, abatere standard; Q1, Q3, prima și a treia cuartilă. Nu există valori lipsă. Procentele sunt calculate din numărul de pacienți al fiecărui braț.

Câteva observații despre citirea unui astfel de tabel. Primul lucru de verificat este dacă numerele se închid: 32 + 26 + 28 = 86 de femei; 53 + 53 + 54 = 160. Al doilea este dacă formatul fiecărei variabile se potrivește cu distribuția ei: IMC și trigliceridele, asimetrice, sunt raportate ca mediană (Q1-Q3), iar TAS și vârsta, simetrice, ca medie ± AS. Al treilea, specific studiilor randomizate, este compararea brațelor. Brațul A are mai multe femei (60,4%) și mai puțini pacienți din mediul urban (47,2%) decât celelalte. Aceste diferențe sunt efectul întâmplării, pentru că repartizarea a fost aleatoare; de aceea ghidurile pentru studiile randomizate recomandă să nu se calculeze valori p pentru comparațiile din Tabelul 1, ci să se judece dacă dezechilibrele sunt suficient de mari încât să conteze clinic [56].

Articolele colegilor din disciplină oferă exemple variate de raportare descriptivă. Un studiu de acuratețe diagnostică a ecografiei cu contrast pentru ciroză a raportat vârsta ca medie ± AS, 58,98 ± 8,90 ani, și sexul ca procent, 56,3% bărbați, pentru 64 de participanți (37 cu ciroză, 27 voluntari sănătoși); parametrul ecografic a fost 0,83 ± 0,09 la pacienții cu ciroză față de 0,49 ± 0,11 la voluntari [2]. Un studiu despre diagnosticul obstrucției biliare maligne, pe 160 de pacienți, a raportat 85 de bărbați (53,1%) și o vârstă medie de 69,31 ± 10,96 ani [57]. În studiul de hemodializă din capitolul 1, vârstele celor două cohorte au fost raportate ca 59,7 ± 12,92 și 61 ± 12,93 ani [39]. Toate aceste raportări respectă convenția "medie ± AS" pentru o vârstă aproximativ simetrică.

Merită să verificați întotdeauna și coerența internă a cifrelor. Într-un studiu despre cuantificarea ecografică a steatozei hepatice, gradele de steatoză au fost raportate ca S0 48/176 (27,2%), S1 56/176 (31,6%), S2 14/176 (7,3%) și S3 59/176 (33,9%), cu valori medii ale parametrului UGAP care cresc de la $198{,}3 \pm \frac{25{,}7 d_B}{m}$ în S0 la $270{,}8 \pm \frac{31{,}62 d_B}{m}$ în S3 [58]. Dacă adunați numărătorii, obțineți 177, nu 176, iar 14/176 înseamnă 8,0%, nu 7,3%. Cel mai probabil este o greșeală de redactare a rezumatului, fără consecințe asupra concluziilor, dar exercițiul arată de ce un cititor atent refăce întotdeauna sumele. Observați și numărul inegal de zecimale (198,3 față de 216,86): o regulă bună este ca toate valorile aceleiași variabile să fie raportate cu același număr de zecimale, iar acesta să nu depășească precizia măsurării.

Pentru variabilele calitative din chestionare, raportarea constă aproape exclusiv în procente. În sondajul despre osteoporoză aplicat în farmacii, din 189 de respondenți, 78,8% au fost femei, grupele de vârstă cele mai numeroase au fost 60-69 de ani (31,7%) și 50-59 de ani (30,7%), iar 69,3% proveneau din mediul urban [27]. Un tabel de frecvențe complet ar adăuga, pentru fiecare procent, și numărul absolut, ca cititorul să poată reface calculele.

Studiu de caz: medii, mediane și scoruri cu efect de podea

Un studiu retrospectiv pe 200 de pacienți cu diabet de tip 2 tratați cu inhibitori SGLT-2 a raportat valorile înainte și după un an de tratament [59]. Glicemia a scăzut de la 180,00 mg% (IQR 154,50-207,00) la 130,00 mg% (IQR 117,50-150,00), HbA1c de la 8,40% (IQR 7,98-9,15%) la 7,30% (IQR 6,90-7,95%), greutatea de la 90,50 kg (82,00-106,50) la 89,00 kg (77,50-100,00), tensiunea sistolică de la 140,00 mmHg (130,00-160,00) la 120,00 mmHg (115,50-130,00), toate cu $p < 0{,}001$, cu excepția greutății ($p = 0{,}018$). Circumferința abdominală a fost raportată altfel, ca 107,05 (± 16,39) și $102{,}50 (\pm 15{,}11)$ ($p = 0{,}042$). Prezența IQR arată că primele valori sunt mediane, deși în rezumat sunt numite "medii"; este o neconcordanță de terminologie frecventă, pe care cititorul trebuie să o recunoască după forma raportării. Faptul că unele variabile sunt raportate ca mediană (IQR) și altele ca medie ± AS este corect dacă alegerea a urmat distribuția fiecărei variabile, dar metodele ar trebui să spună explicit acest lucru și ce test pereche a fost folosit. Un al doilea exemplu: într-un studiu transversal pe 304 adulți cu diabet, un scor de implicare digitală a fost raportat ca mediană [Q1-Q3], 1,71 [1,71-2,39] față de 3,86 [3,86-4,43] între grupuri, iar 35,5% dintre participanți au avut o implicare digitală ridicată [7]. Observați că în ambele grupuri mediana coincide cu

> prima cuartilă. Acest lucru se întâmplă când un scor are puține valori posibile și multe observații egale cu cea mai mică valoare a unui grup: cel puțin un sfert dintre pacienți, și chiar până la jumătate, au exact aceeași valoare. Fenomenul este înrudit cu efectul de podea, descris explicit într-un studiu de validare a unui chestionar despre povara menopauzei, pe 308 femei, unde 37,0% dintre respondente au avut scorul minim la o subscală și 23,4% la alta [60]. Când o proporție mare de pacienți au scorul minim, media și abaterea standard descriu prost distribuția, iar instrumentul nu poate distinge între pacienții din partea de jos a scalei. Ce ar trebui raportat în plus: proporția observațiilor cu valoarea minimă și maximă și, pentru scorurile mărginite, o histogramă.

2.9 Analiza descriptivă completă a setului mg2026

Încheiem partea teoretică cu descrierea completă a setului de date pe care îl veți folosi tot semestrul. Valorile din tabelele 2.12 și 2.13 sunt cele pe care trebuie să le regăsiți în JASP și în Google Sheets la lucrările practice 1 și 2; dacă rezultatul vostru diferă, căutați cauza la import sau la curățare.

Tabelul 2.12. Statistici descriptive pentru variabilele numerice la includere, setul mg2026 (n = 160).

Variabila	Media	AS	Mediana	Q1-Q3	Min-max	CV (%)	Asimetrie
Vârsta (ani)	56,29	8,77	56	50,75-62	35-79	15,6	0,01
Înălțimea (cm)	168,79	9,05	168	162-177	145-192	5,4	0,10
Greutatea (kg)	83,05	15,09	80,2	72,05-90,93	58,1-159,9	18,2	1,29
$\mathrm{IMC}\left(\frac{\mathrm{kg}}{m^2}\right)$	29,04	3,91	28,25	26,2-31,4	23,5-48,3	13,5	1,42
Talia (cm)	88,54	12,11	86	80,75-95	65-140	13,7	0,86
TAS (mmHg)	131,13	13,88	131	122-140	102-169	10,6	0,01
TAD (mmHg)	79,23	8,63	80	74-85	58-103	10,9	-0,09
Frecvența cardiacă (bătăi/min)	74,71	9,30	74,5	69-80	51-102	12,4	0,05
Colesterol total (mg/dL)	209,39	36,91	207	183-235	119-306	17,6	0,02
HDL-colesterol (mg/dL)	49,52	11,78	49	41-59	20-78	23,8	0,04
Trigliceride (mg/dL)	161,18	76,30	151,5	107-203,25	44-442	47,3	0,95
Glicemie a jeun (mg/dL)	105,44	15,76	104,5	93-116	75-154	15,0	0,37
HbA1c (%)	5,58	0,45	5,6	5,3-5,9	4,5-7,0	8,0	0,14
Scor FINDRISC	10,37	4,27	10	8-13	0-25	41,2	0,39

Tabelul 2.13. Distribuția variabilelor calitative, setul mg2026 (n = 160).

Variabila	**Categoria**	**n**	**%**
Sex	femei / bărbați	86 / 74	53,8 / 46,3
Mediu	$\frac{\text{urban}}{\text{rural}}$	95 / 65	59,4 / 40,6
Fumat	nefumător / fost fumător / fumător	74 / 33 / 53	46,3 / 20,6 / 33,1
Educație	gimnaziu / liceu sau profesională / universitar / postuniversitar	20 / 57 / 64 / 19	12,5 / 35,6 / 40,0 / 11,9
Activitate fizică (1-5)	1 / 2 / 3 / 4 / 5	21 / 42 / 60 / 34 / 3	13,1 / 26,3 / 37,5 / 21,3 / 1,9
Satisfacție la 6 luni (1-5)	1 / 2 / 3 / 4 / 5	4 / 19 / 54 / 55 / 28	2,5 / 11,9 / 33,8 / 34,4 / 17,5
Brațul de tratament	$\frac{A}{\frac{B}{C}}$	53 / 53 / 54	33,1 / 33,1 / 33,8
Diabet confirmat prin TTGO	da / nu	29 / 131	18,1 / 81,9

Notă: sumele procentelor pot fi 100,1 din cauza rotunjirii.

Cum se scrie interpretarea acestor tabele? Iată un model, în stilul secțiunii de rezultate a unui articol. "Au fost incluși 160 de pacienți (86 de femei, 53,8%), cu vârsta medie de 56,29 ± 8,77 ani (limite 35-79). Majoritatea proveneau din mediul urban (59,4%), iar o treime erau fumători activi (33,1%). Mediana IMC a fost 28,25 kg/m² (Q1-Q3 26,2-31,4), cu 54 de pacienți (33,8%) obezi. Tensiunea sistolică medie a fost 131,13 ± 13,88 mmHg, iar 41 de pacienți (25,6%) aveau TAS ≥ 140 mmHg la includere. Glicemia a jeun medie a fost 105,44 ± 15,76 mg/dL, HbA1c medie 5,58 ± 0,45%, iar diabetul a fost confirmat prin TTGO la 29 de pacienți (18,1%). Trigliceridele au avut o distribuție asimetrică, cu mediana 151,5 mg/dL (Q1-Q3 107-203,25)."

Câteva remarci despre alegerile din acest text. Vârsta, TAS, glicemia, HbA1c, colesterolul și HDL au asimetrie mică și medii practic egale cu medianele, deci se raportează ca medie ± AS. IMC, greutatea, talia și trigliceridele au asimetrie pozitivă importantă (peste 0,8) și se raportează mai corect ca mediană (Q1-Q3). Scorul FINDRISC este o variabilă discretă cu multe valori posibile; se poate raporta în ambele moduri, dar mediana cu IQR este mai prudentă. Educația, activitatea fizică și satisfacția sunt ordinale: se raportează prin frecvențe și, eventual, prin mediană (3, 3 și respectiv 4), niciodată prin medie ± AS. Pentru variabilele dihotomice, procentul complet descrie variabila; abaterea standard nu se raportează.

Studiu de caz: când abaterea standard este aproape cât media

Abaterea standard raportată într-un articol poate fi folosită ca test rapid pentru forma distribuției. Într-un studiu retrospectiv pe 486 de pacienți cu COVID-19 (200 fără diabet, 62 cu diabet de tip 1, 224 cu diabet de tip 2), durata spitalizării a fost 11,6 ± 7,0 zile în diabetul de tip 2 și 9,1 ± 5,8 zile în diabetul de tip 1, iar glicemia la internare 178,3 ± 34,7 mg/dL față de 164,8 ± 39,6 mg/dL [8]. Pentru glicemie, coeficientul de variație este de aproximativ 20%, compatibil cu o distribuție rezonabil de simetrică. Pentru durata spitalizării, CV este de aproximativ 60% (7,0/11,6), iar m - 2s = 11,6 - 14,0 = -2,4 zile, o valoare imposibilă. Distribuția este, aproape sigur, asimetrică spre dreapta, cu câțiva pacienți internați foarte mult, iar mediana cu IQR ar fi descris-o mai bine. Situația este și mai clară într-un studiu de diagnostic al accidentului vascular cerebral în urgență, pe 85 de pacienți: proteina GFAP a fost $142{,}91 \pm 102{,}19$ pg/mL la pacienții cu accident vascular și $37{,}76 \pm 19{,}92$ pg/mL la cei fără, iar UCH-L1 1307,68 ± 967,54 față de $189{,}81 \pm 92{,}69$ pg/mL[61]. O abatere standard de peste 70% din medie semnalează o asimetrie puternică; pentru o astfel de variabilă, comparația ar

trebui făcută cu un test neparametric (capitolul 7) sau după logaritmare (capitolul 3). În schimb, un alt studiu din același serviciu de urgență, despre tromboliza în accidentul vascular ischemic, a raportat corect timpii, variabile tipic asimetrice, prin mediane: o mediană de 100 de minute de la debut la sosirea în urgență și 15,5 minute de la sosire la tomografie, pe 258 de pacienți eligibili dintre cei 355 evaluați inițial, dintre care 46 au primit tromboliză (17,83%) [62]. Concluzia pentru cititor: înainte de a accepta o comparație de medii, comparați abaterea standard cu media; dacă s este mai mare decât aproximativ jumătate din m la o variabilă care nu poate fi negativă, distribuția este asimetrică și media nu este valoarea tipică.

2.10 Probleme rezolvate

Problema rezolvată 2.1. Tabel de frecvențe cu valori lipsă: tensiunea sistolică într-un ambulatoriu de cardiologie

Enunț. Într-un ambulatoriu de cardiologie s-a măsurat tensiunea arterială sistolică la 50 de pacienți hipertensivi la prima prezentare. Rezultatele au fost grupate în tabelul de mai jos, din care lipsesc câteva valori. (a) Completați valorile A, B, C și D. (b) Care este probabilitatea ca un pacient ales la întâmplare din acest grup să aibă TAS de cel puțin 160 mmHg? (c) În ce clasă se află mediana? Estimați mediana și media din tabel.

Tabelul 2.14. TAS la 50 de pacienți hipertensivi (date grupate).

TAS (mmHg)	Frecvența absolută	Frecvența relativă (%)	Frecvența cumulată (%)
100-119	4	8,0	8,0
120-139	11	22,0	30,0
140-159	A	34,0	C
160-179	12	B	88,0
180-199	D	8,0	96,0
200-219	2	4,0	100,0
Total	50	100,0	

Rezolvare. (a) A se obține din frecvența relativă: A = 34,0% x 50 = 0,34 x 50 = 17 pacienți. B se obține din frecvența absolută: $B = \frac{12}{50} \times 100 = 24{,}0\%$. C este frecvența cumulată a clasei a treia: C = 30,0 + 34,0 = 64,0%. Verificare: 64,0 + 24,0 = 88,0, exact valoarea din rândul următor. D se obține din frecvența relativă: D = 0,08 x 50 = 4 pacienți. Verificare finală: 4 + 11 + 17 + 12 + 4 + 2 = 50, iar 8 + 22 + 34 + 24 + 8 + 4 = 100%. Tabelul se închide.

(b) TAS ≥ 160 mmHg înseamnă ultimele trei clase: 24,0 + 8,0 + 4,0 = 36,0%. Echivalent, 100 - 64,0 = 36,0% (complementul frecvenței cumulate până la 159 mmHg). Probabilitatea este 0,36, adică 18 pacienți din 50.

(c) Mediana este valoarea care împarte cei 50 de pacienți în două jumătăți, deci se află între a 25-a și a 26-a valoare ordonată. Frecvența cumulată este 30,0% (15 pacienți) la sfârșitul clasei 120-139 și 64,0% (32 de pacienți) la sfârșitul clasei 140-159; a 25-a valoare este deci în clasa 140-159, care este clasa mediană. Dacă presupunem că cei 17 pacienți din această clasă sunt distribuiți uniform, mediana se estimează prin interpolare, folosind limitele reale ale clasei (139,5-159,5, lățimea h = 20):

$$\text{Mediana} \approx 139{,}5 + \frac{25-15}{17} \times 20 = 139{,}5 + 11{,}8 = 151{,}3 \text{ mmHg}$$

Media se estimează cu centrele claselor (109,5; 129,5; 149,5; 169,5; 189,5; 209,5):

$$m \approx \frac{4 \times 109{,}5 + 11 \times 129{,}5 + 17 \times 149{,}5 + 12 \times 169{,}5 + 4 \times 189{,}5 + 2 \times 209{,}5}{50}$$

$$= \frac{438 + 1\,424{,}5 + 2\,541{,}5 + 2\,034 + 758 + 419}{50} = \frac{7\,615}{50} = 152{,}3 \text{ mmHg}$$

Interpretare. Peste o treime dintre pacienții hipertensivi (36%) se prezintă cu valori de cel puțin 160 mmHg, adică în zona hipertensiunii de gradul 2 sau mai mult. Media și mediana estimate sunt apropiate (152,3 față de 151,3 mmHg), deci distribuția este aproximativ simetrică, cu o ușoară coadă spre valorile mari. Atenție: ambele valori sunt estimări din date grupate; cu datele individuale am obține valori ușor diferite.

Problema rezolvată 2.2. Tabel de frecvențe pentru o variabilă discretă: consultațiile pe an

Enunț. Un medic de familie a numărat consultațiile din ultimul an pentru 80 de pacienți cu diabet de tip 2 aleși aleator de pe lista sa. (a) Completați valorile A, B, C și D. (b) Calculați modul, mediana și media numărului de consultații. (c) Ce proporție dintre pacienți au avut cel puțin 4 consultații?

Tabelul 2.15. Numărul de consultații pe an la 80 de pacienți cu diabet de tip 2.

Consultații pe an	Frecvența absolută	Frecvența relativă (%)	Frecvența cumulată (%)
0	6	7,5	7,5
1	14	17,5	25,0
2	A	25,0	50,0
3	18	B	72,5
4	C	12,5	D
5	8	10,0	95,0
6	4	5,0	100,0
Total	80	100,0	

Rezolvare. (a) A = 0,25 x 80 = 20 de pacienți (sau din frecvențele cumulate: 50,0 - 25,0 = 25,0%, deci tot 20). $B = \frac{18}{80} \times 100 = 22{,}5\%$ (verificare: 50,0 + 22,5 = 72,5). C = 0,125 x 80 = 10 pacienți. D = 72,5 + 12,5 = 85,0% (verificare: 85,0 + 10,0 = 95,0). Suma frecvențelor absolute: 6 + 14 + 20 + 18 + 10 + 8 + 4 = 80.

(b) Modul este valoarea cu frecvența cea mai mare: 2 consultații (20 de pacienți). Pentru mediană, n = 80 este par, deci mediana este media valorilor din pozițiile 40 și 41. Frecvența cumulată este exact 50,0% (40 de pacienți) la valoarea 2: a 40-a valoare este 2, iar a 41-a este prima valoare de 3. Mediana este (2 + 3)/2 = 2,5 consultații. Media este o medie ponderată cu frecvențele:

$$m = \frac{0\times6+1\times14+2\times20+3\times18+4\times10+5\times8+6\times4}{80} = \frac{0+14+40+54+40+40+24}{80} = \frac{212}{80} = 2{,}65 \text{ consultații}$$

(c) Cel puțin 4 consultații: 12,5 + 10,0 + 5,0 = 27,5%, adică 22 de pacienți. Echivalent, 100 - 72,5 = 27,5%.

Interpretare. Pacientul "tipic" are două sau trei consultații pe an, iar un sfert dintre pacienți (25,0%, frecvența cumulată la valoarea 1) au cel mult o consultație. Capcana problemei este mediana: când frecvența cumulată atinge exact 50% la o valoare, mediana nu este acea valoare, ci media dintre ea și valoarea următoare. La examen, mulți studenți răspund "2". Din punct de vedere clinic, cei 6 pacienți (7,5%) fără nicio consultație într-un an sunt cei care ar trebui chemați activ la control.

Problema rezolvată 2.3. Clasificarea variabilelor dintr-o fișă de observație

Enunț. Fișa unui studiu despre pacienții internați cu pneumonie conține următoarele câmpuri: (1) codul pacientului; (2) vârsta (ani); (3) sexul; (4) scorul CURB-65 (0-5); (5) saturația în oxigen (%); (6) numărul de zile de spitalizare; (7) severitatea percepută a dispneei (absentă, ușoară, moderată, severă); (8) germenul izolat (pneumococ, Haemophilus, Mycoplasma, altul, neizolat); (9) decesul în spital (da/nu); (10) temperatura la internare (°C). Pentru fiecare variabilă precizați tipul, indicatorul descriptiv potrivit și un grafic potrivit. Rezolvare. Tabelul 2.16 sintetizează răspunsul.

Tabelul 2.16. Clasificarea variabilelor din fișa studiului despre pneumonie.

Variabila	Tipul	Indicator descriptiv	Grafic
Codul pacientului	identificator, nu se analizează	-	-
Vârsta	numerică continuă, scală de raport	medie ± AS (sau mediană, dacă e asimetrică)	histogramă, box-plot
Sexul	nominală dihotomică	n (%)	bare
Scorul CURB-65	ordinală (scor cu 6 valori)	mediană (Q1-Q3), frecvențe	bare în ordinea scorului
Saturația în O2	numerică continuă, mărginită superior la 100%	mediană (Q1-Q3), pentru că este de obicei asimetrică spre stânga	histogramă, box-plot
Zilele de spitalizare	numerică discretă, asimetrică	mediană (Q1-Q3)	histogramă, box-plot
Severitatea dispneei	ordinală	frecvențe, mediană	bare în ordinea categoriilor
Germenul izolat	nominală cu 5 categorii	n (%)	bare (descrescător)
Decesul în spital	nominală dihotomică	n (%)	bare
Temperatura	numerică continuă, scală de interval	medie ± AS	histogramă, box-plot

Interpretare. Două detalii merită atenție. Scorul CURB-65 este numeric ca formă, dar ordinal ca natură: diferența dintre 0 și 1 punct nu are aceeași semnificație clinică precum cea dintre 3 și 4. Saturația în oxigen nu poate depăși 100%, iar la pacienții cu pneumonie valorile se strâng spre partea de sus a scalei, cu o coadă spre valori mici; mediana este mai potrivită decât media. Temperatura este exemplul clasic de scală de interval: nu are sens să spunem că 40 °C este de două ori 20 °C, deci nici coeficientul de variație nu are sens.

Problema rezolvată 2.4. Descrierea completă a unei serii mici: HbA1c la zece pacienți

Enunț. Valorile HbA1c (%) ale celor 10 pacienți cu diabet de tip 2 consultați într-o zi sunt: 6,2; 6,8; 7,1; 7,4; 7,5; 7,9; 8,3; 8,8; 9,6; 11,4. Calculați: media, mediana, abaterea standard, coeficientul de variație, eroarea standard, cuartilele, intervalul intercuartilic și verificați dacă există valori extreme după regula 1,5 x IQR. Ce proporție dintre pacienți au HbA1c ≥ 7%?

Rezolvare. Valorile sunt deja ordonate, n = 10.

Media: suma valorilor este 81,0, deci $m = \frac{81,0}{10} = 8,1\%$.

Mediana: n par, valorile din pozițiile 5 și 6 sunt 7,5 și 7,9; mediana = (7,5 + 7,9)/2 = 7,7%.

Abaterea standard: abaterile de la medie sunt -1,9; -1,3; -1,0; -0,7; -0,6; -0,2; 0,2; 0,7; 1,5; 3,3. Pătratele lor sunt 3,61; 1,69; 1,00; 0,49; 0,36; 0,04; 0,04; 0,49; 2,25; 10,89, cu suma 20,86.

$s^2 = \frac{20,86}{9} = 2,318$; $s = \sqrt{2,318} = 1,52\%$

Coeficientul de variație: $\text{CV} = \frac{1,52}{8,1} \times 100 = 18,8\%$. Eroarea standard: $\text{ES} = \frac{1,52}{\sqrt{10}} = \frac{1,52}{3,162} = 0,48\%$.

Cuartilele (metoda medianelor jumătăților): jumătatea inferioară 6,2; 6,8; 7,1; 7,4; 7,5 are mediana Q1 = 7,1%; jumătatea superioară 7,9; 8,3; 8,8; 9,6; 11,4 are mediana Q3 = 8,8%. IQR = 8,8 - 7,1 = 1,7%. (Google Sheets și JASP dau, prin interpolare, Q1 = 7,175 și Q3 = 8,675.)

Valori extreme: limitele sunt 7,1 - 1,5 x 1,7 = 7,1 - 2,55 = 4,55% și 8,8 + 2,55 = 11,35%. Valoarea 11,4% depășește ușor limita superioară și apare ca valoare extremă.

Proporția cu HbA1c ≥ 7%: 8 pacienți din 10, adică 80%.

Interpretare. Media (8,1%) este mai mare decât mediana (7,7%), semn că o valoare mare, 11,4%, trage media în sus. Fără această valoare, media celor nouă pacienți $\frac{\text{ar fi}}{9} \quad \frac{69,6}{9} = 7,73\%$, practic egală cu mediana. Pacientul cu 11,4% nu este o eroare, ci un pacient cu diabet foarte prost

controlat, care are nevoie urgentă de intensificarea tratamentului. Pentru o serie atât de mică, raportarea corectă este mediana 7,7% (Q1-Q3 7,1-8,8), eventual completată de medie ± AS, 8,1 ± 1,52%. Din punct de vedere clinic, 80% dintre pacienți nu ating ținta de HbA1c sub 7%, iar acesta este, de fapt, mesajul principal al seriei.

Problema rezolvată 2.5. Abaterea standard sau eroarea standard?

Enunț.

(a) Un articol raportează pentru 100 de pacienți hipertensivi "TAS $158 \pm 2{,}1$ mmHg", fără a preciza ce reprezintă 2,1. Ce este mai plauzibil: o abatere standard sau o eroare standard? (b) Un alt articol raportează durata spitalizării pentru 81 de pacienți ca 6,5 ± 5,9 zile (medie ± AS). Ce puteți spune despre forma distribuției? Calculați eroarea standard.

Rezolvare.

(a) Dacă 2,1 ar fi abaterea standard, aproximativ 95% dintre pacienți ar avea TAS între 158 - 4,2 = 153,8 și 158 + 4,2 = 162,2 mmHg, adică toți pacienții hipertensivi ar avea tensiuni aproape identice. Este biologic neplauzibil: în mg2026, abaterea standard a TAS este 13,88 mmHg. Dacă 2,1 este eroarea standard, abaterea standard ar fi $s = \text{ES} \times \sqrt{n} = 2{,}1 \times \sqrt{100} = 2{,}1 \times 10 = 21$ mmHg, o valoare perfect plauzibilă pentru un grup de hipertensivi. Concluzia: aproape sigur este vorba despre eroarea standard, iar autorul ar fi trebuit să precizeze acest lucru.

(b) Coeficientul de variație este $\frac{5{,}9}{6{,}5} \times 100 = 91\%$, iar m - 2s = 6,5 - 11,8 = -5,3 zile, imposibil. Distribuția este puternic asimetrică spre dreapta: cei mai mulți pacienți stau câteva zile, iar câțiva stau foarte mult. Mediana și IQR ar fi fost mai informative. Eroarea standard este $\text{ES} = \frac{5{,}9}{\sqrt{81}} = \frac{5{,}9}{9} = 0{,}66$ zile.

Interpretare. Prima întrebare pe care trebuie să o puneți când vedeți semnul ± este "ce reprezintă?". A doua este "este plauzibil?". Un raport s/m foarte mic la o variabilă biologică sugerează o eroare standard; un raport s/m apropiat de 1 la o variabilă pozitivă semnalează asimetrie. Eroarea standard de 0,66 zile arată că media de 6,5 zile este estimată destul de precis, deși pacienții individuali variază foarte mult; sunt două lucruri diferite.

Problema rezolvată 2.6. Media ponderată și media geometrică

Enunț. (a) Trei cabinete raportează media HbA1c a pacienților lor cu diabet: 7,2% (40 de pacienți), 7,8% (25 de pacienți) și 6,9% (85 de pacienți). Care este media HbA1c pentru toți cei 150 de pacienți? (b) La un pacient, proteina C reactivă (PCR) a avut, la trei determinări succesive, valorile 2, 8 și 32 mg/L. Calculați media aritmetică și media geometrică. Care descrie mai bine valoarea tipică?

Rezolvare. (a) Media ponderată cu numărul de pacienți:

$$m = \frac{40 \times 7{,}2 + 25 \times 7{,}8 + 85 \times 6{,}9}{40 + 25 + 85} = \frac{288 + 195 + 586{,}5}{150} = \frac{1\,069{,}5}{150} = 7{,}13\%$$

Media simplă a celor trei medii ar fi $\frac{(7{,}2+7{,}8+6{,}9)}{3} = 7{,}3\%$, o valoare greșită, pentru că acordă cabinetului cu 25 de pacienți aceeași importanță ca celui cu 85.

(b) Media aritmetică: $\frac{2+8+32}{3} = \frac{42}{3} = \frac{14\text{mg}}{L}$. Media geometrică: rădăcina cubică din 2 x 8 x 32 = 512, adică 8 mg/L. Echivalent, cu logaritmi zecimali: log 2 = 0,301; log 8 = 0,903; log 32 = 1,505; media lor = 2,709/3 = 0,903; 10 la puterea 0,903 = 8.

Interpretare. Valorile PCR cresc de patru ori de la o determinare la alta; pe scara logaritmică ele sunt egal distanțate, iar valoarea din mijloc, 8 mg/L, este cea tipică. Media aritmetică, 14 mg/L,

este trasă în sus de ultima valoare. Pentru markerii care variază multiplicativ (PCR, titruri, încărcătură virală, trigliceride), media geometrică este alegerea naturală.

2.11 Exerciții propuse

1. Clasificați următoarele variabile (nominală, ordinală, numerică discretă, numerică continuă) și precizați scala de măsură: grupa sanguină; scorul Apgar la 5 minute; numărul de sarcini anterioare; temperatura corporală; clasa NYHA; HbA1c; acordul cu afirmația "Sunt mulțumit de tratament" (scală Likert în 5 trepte); codul numeric al pacientului; durata spitalizării în zile; statutul de fumător (da/nu).
2. Completați tabelul de frecvențe al glicemiei a jeun la 40 de pacienți: clasa 70-89 mg/dL, 6 pacienți, 15,0%, cumulat 15,0%; clasa 90-109, A pacienți, 35,0%, cumulat B; clasa 110-125, 10 pacienți, C%, cumulat 75,0%; clasa 126-139, D pacienți, 12,5%, cumulat 87,5%; clasa ≥ 140, E pacienți, cumulat 100,0%. Ce proporție dintre pacienți au glicemia în intervalul diagnostic pentru diabet (≥ 126 mg/dL)?
3. Pentru valorile TAS 118, 124, 130, 136, 142, 150 mmHg, calculați de mână media, mediana, abaterea medie absolută, varianța, abaterea standard, coeficientul de variație și eroarea standard.
4. Durata spitalizării (zile) la 10 pacienți operați de colecistită a fost: 3, 4, 4, 5, 5, 6, 7, 8, 12, 31. Calculați media, mediana și modul. Determinați cuartilele prin metoda medianelor jumătăților și verificați valorile extreme cu regula 1,5 x IQR. Cum se schimbă media fără ultima valoare? Ce indicatori ați raporta?
5. Un studiu are 250 de valori ale trigliceridelor, între 68 și 412 mg/dL. Câte clase recomandă regula lui Sturges? Ce lățime de clasă rezultă și la ce valoare comodă ați rotunji-o?
6. Un articol raportează, pentru 64 de pacienți, glicemia a jeun ca 112 ± 3 mg/dL. Dacă 3 este eroarea standard, care este abaterea standard? Care interpretare (AS sau ES) este mai plauzibilă și de ce? Calculați eroarea standard a TAS pentru femeile din mg2026 (AS 14,47 mmHg, n = 86).
7. Vârsta medie a pacienților din trei secții ale unui spital este 45,2 ani (120 de pacienți), 61,8 ani (60 de pacienți) și 70,4 ani (20 de pacienți). Calculați vârsta medie a tuturor celor 200 de pacienți și comparați-o cu media simplă a celor trei medii.
8. Folosind tabelul 2.11: (a) verificați procentul femeilor din brațul B; (b) identificați variabilele raportate ca mediană (Q1-Q3) și explicați alegerea; (c) explicați de ce într-un studiu randomizat nu se calculează valori p pentru comparațiile din Tabelul 1.
9. Alegeți graficul potrivit pentru: (a) distribuția grupelor sanguine la 500 de donatori; (b) compararea glicemiei între trei brațe de tratament; (c) relația dintre vârstă și tensiunea sistolică; (d) evoluția lunară a cazurilor de gripă într-un județ pe parcursul a doi ani; (e) distribuția greutății la naștere pentru 1 200 de nou-născuți; (f) răspunsurile la o întrebare Likert în 5 trepte, comparate între femei și bărbați.
10. Încărcătura virală HIV la patru pacienți este 1 000, 10 000, 100 000 și 1 000 000 de copii/mL. Calculați media aritmetică și media geometrică (folosiți logaritmii zecimali). De ce laboratoarele și clinicienii exprimă aceste valori pe scară logaritmică?

Capcane frecvente

- Calcularea mediei pentru variabile nominale codificate cu cifre (de exemplu, "media sexului = 1,46").
- Raportarea mediei ± AS pentru variabile ordinale (scoruri Likert, stadii) sau pentru variabile foarte asimetrice (durata spitalizării, markeri inflamatori).
- Confundarea abaterii standard cu eroarea standard; "media ± ES" face datele să pară mult mai omogene.
- Împărțirea la n în loc de n - 1 la calculul variației unui eșantion (STDEVP în loc de STDEV).
- Alegerea valorii ca mediană când frecvența cumulată atinge exact 50% la acea valoare; corect este media dintre ea și valoarea următoare.
- Eliminarea valorilor extreme doar pentru că "strică media", fără a demonstra că sunt erori.
- Diagrame cu bare cu axa trunchiată, diagrame circulare 3D sau cu prea multe categorii.
- Coeficientul de variație calculat pentru variabile pe scală de interval (temperatura în °C).
- Procente care nu se pot verifica pentru că lipsesc numerele absolute.

De reținut

- Tipul variabilei (nominală, ordinală, discretă, continuă) determină indicatorii, graficele și testele permise.
- Datele se organizează un rând pe pacient și o coloană pe variabilă, cu un dicționar de date scris înainte de colectare; erorile se verifică în documentul sursă.
- Frecvențele relative și cumulate se pot reconstitui unele din altele; frecvența cumulată are sens doar pentru date ordonate.
- Media este potrivită pentru date numerice simetrice; mediana pentru date asimetrice sau ordinale; modul pentru date nominale; media geometrică pentru date pozitive cu variație multiplicativă.
- Abaterea standard descrie variabilitatea pacienților; eroarea standard $\left(\frac{s}{\sqrt{n}}\right)$ descrie precizia mediei și scade când n crește.
- Rezumatul în cinci numere și box-plotul arată centrul, dispersia, asimetria și valorile extreme; regula 1,5 x IQR semnalează valori de verificat, nu de șters.
- Un s apropiat de m la o variabilă pozitivă semnalează asimetrie; în acest caz se raportează mediana (Q1-Q3).
- În "Tabelul 1": medie ± AS pentru variabile simetrice, mediană (Q1-Q3) pentru cele asimetrice, n (%) pentru cele calitative; numerele trebuie să se închidă.
- Graficele cu bare încep de la zero, fără efecte 3D; fiecare grafic are axe etichetate, unități și n.

Întrebări de autoevaluare

1. (CS) Stadiul fibrozei hepatice (F0, F1, F2, F3, F4) este o variabilă:
    a) nominală
    b) ordinală
    c) numerică discretă pe scală de raport
    d) numerică continuă
    e) dihotomică
2. (CS) Pentru variabila "durata spitalizării", cu media 11,6 zile și abaterea standard 7,0 zile, indicatorul de tendință centrală cel mai potrivit este:
    a) media aritmetică
    b) modul
    c) mediana

   d) amplitudinea
   e) coeficientul de variație

3. (CS) Într-un tabel de frecvențe pentru 40 de pacienți, o clasă are frecvența relativă 22,5%. Frecvența ei absolută este:
   a) 22
   b) 9
   c) 18
   d) 10
   e) 4
4. (CS) Abaterea standard a TAS într-un eșantion de 100 de pacienți este 15 mmHg. Eroarea standard a mediei este:
   a) 0,15 mmHg
   b) 1,5 mmHg
   c) 15 mmHg
   d) 150 mmHg
   e) 3,87 mmHg
5. (CS) Pentru seria 2, 3, 3, 4, 5, 6, 7, 8, mediana este:
   a) 4
   b) 4,5
   c) 4,75
   d) 3
   e) 5
6. (CM) Despre mediană sunt corecte afirmațiile:
   a) este egală cu a doua cuartilă și cu percentila 50
   b) este puțin influențată de valorile extreme
   c) este indicatorul potrivit pentru variabilele ordinale
   d) este întotdeauna mai mare decât media
   e) se poate calcula pentru variabile nominale
7. (CM) Despre abaterea standard sunt corecte afirmațiile:
   a) are aceeași unitate de măsură ca datele
   b) este rădăcina pătrată a varianței
   c) scade pe măsură ce volumul eșantionului crește, la fel ca eroarea standard
   d) la eșantion se calculează cu n - 1 la numitor
   e) nu poate fi negativă
8. (CM) Într-o distribuție cu asimetrie pozitivă (coadă spre dreapta):
   a) media este, de regulă, mai mare decât mediana
   b) coeficientul de asimetrie este pozitiv
   c) valorile mari rare trag media în sus
   d) media și mediana sunt egale
   e) mediana cu IQR descrie mai bine datele decât media ± AS
9. (CM) Sunt grafice potrivite pentru o variabilă numerică continuă:
   a) histograma

b) diagrama box-plot

c) diagrama tulpină-frunză

d) diagrama circulară

e) diagrama cu puncte (dot plot) pentru eșantioane mici

10. (CM) În setul mg2026, IMC are media 29,04 kg/m², mediana 28,25 kg/m², Q1 26,2, Q3 31,4 și maximul 48,3 kg/m². Sunt corecte afirmațiile:

a) intervalul intercuartilic este 5,2 kg/m²

b) limita superioară după regula 1,5 x IQR este 39,2 kg/m²

c) valoarea 48,3 kg/m² trebuie eliminată din analiză pentru că este extremă

d) distribuția este asimetrică spre dreapta

e) raportarea corectă în Tabelul 1 este 28,25 (26,2-31,4) kg/m²

Capitolul 3. Distribuțiile statistice: distribuția normală, verificarea normalității, asimetria, excesul, scorul z și transformarea datelor

Obiective

La finalul acestui capitol, studentul va putea:

- să deosebească o variabilă aleatoare discretă de una continuă și să explice ce descriu funcția de probabilitate, densitatea și funcția de repartiție;
- să calculeze probabilități simple cu distribuția binomială și cu distribuția Poisson, pornind de la exemple din cabinetul de medicină de familie;
- să enumere proprietățile distribuției normale și să aplice regula 68,3-95,4-99,7 pe date reale;
- să calculeze un scor z, să găsească percentila corespunzătoare în tabelul din anexa E și în Google Sheets și să interpreteze rezultatul clinic;
- să construiască un interval de referință de tip $m \pm 1{,}96s$ și să explice de ce 5% dintre persoanele sănătoase ies "în afara normalului";
- să verifice normalitatea unei variabile prin grafice (histogramă, Q-Q, box-plot), prin indicatori (asimetrie, exces și erorile lor standard) și prin teste (Shapiro-Wilk, Kolmogorov-Smirnov cu corecția Lilliefors, Anderson-Darling), cunoscând limitele fiecărei metode;
- să aleagă o transformare potrivită (logaritmică, radical, inversă, Box-Cox) și să retransforme corect rezultatele, inclusiv media geometrică și intervalul ei de încredere;
- să recunoască valorile extreme prin scorul z și prin regula 1,5 x IQR și să decidă argumentat ce face cu ele;
- să citească critic modul în care articolele publicate raportează variabilele simetrice și pe cele asimetrice.

3.1 Variabile aleatoare și distribuții de probabilitate

În capitolul 2 am descris datele așa cum sunt: tabele de frecvențe, medii, mediane, histograme. Acum facem un pas înapoi și ne întrebăm de unde vin aceste forme. De ce histograma tensiunii sistolice din setul mg2026 seamănă cu un clopot, iar cea a trigliceridelor are o "coadă" lungă spre dreapta? Răspunsul ține de noțiunea de distribuție, adică de legea după care se repartizează valorile unei variabile într-o populație. Aproape tot ce urmează în această carte, de la intervalele de încredere la testele statistice, se sprijină pe presupuneri despre distribuții, așa că merită să le înțelegem bine de la început [19,63].

3.1.1 De la frecvența observată la probabilitate

Când spunem că 29 dintre cei 160 de pacienți din mg2026 au diabet confirmat prin TTGO, descriem eșantionul: frecvența relativă este 29/160 = 18,1%. Dacă ne imaginăm că alegem la întâmplare un pacient din aceeași populație de adulți cu risc cardiometabolic, probabilitatea ca el să aibă diabet este aproximativ aceeași, în jur de 0,18. Probabilitatea este, într-un sens practic, frecvența relativă spre care ne apropiem atunci când numărul de observații crește foarte mult. O monedă aruncată de 10 ori poate da 7 steme; aruncată de 10.000 de ori, va da foarte aproape de 50%.

Probabilitatea ia valori între 0 (eveniment imposibil) și 1 (eveniment sigur). Două reguli ne ajung pentru acest capitol. Probabilitatea ca un eveniment să NU se producă este 1 minus

probabilitatea ca el să se producă. Pentru evenimente independente, probabilitatea ca ambele să se producă este produsul probabilităților: dacă doi pacienți sunt aleși independent, probabilitatea ca amândoi să aibă diabet este 0,18 x 0,18 = 0,0324.

3.1.2 Variabile aleatoare discrete şi continue

O variabilă aleatoare este o mărime al cărei rezultat nu îl cunoaștem dinainte, dar despre care știm ce valori poate lua și cu ce probabilități. Numărul de pacienți cu diabet dintr-un grup de 10 este o variabilă aleatoare discretă: poate fi 0, 1, 2 și așa mai departe până la 10, dar nu poate fi 2,7. La fel, numărul de reacții adverse raportate într-o lună într-un cabinet, numărul de nașteri pe zi într-o maternitate sau numărul de colonii pe o placă Petri sunt variabile discrete (numărători).

Tensiunea arterială, glicemia, greutatea sau timpul până la tromboliză sunt variabile continue: între oricare două valori posibile mai există, teoretic, o infinitate de valori. Faptul că tensiometrul afișează numere întregi ține de instrument, nu de fenomen. Această distincție contează, pentru că modul în care calculăm probabilitățile diferă: la variabilele discrete adunăm probabilitățile unor valori punctuale, la cele continue calculăm arii sub o curbă.

3.1.3 Funcţia de probabilitate, densitatea şi funcţia de repartiţie

Pentru o variabilă discretă, distribuția se descrie printr-o funcție de probabilitate (engl. probability mass function): o listă care spune, pentru fiecare valoare posibilă, care este probabilitatea ei. Suma tuturor acestor probabilități este 1. Graficul ei este o diagramă cu bare, ca în figura 3.1.

Pentru o variabilă continuă, probabilitatea unei valori exacte este zero. Nu are sens să întrebăm care este probabilitatea ca un pacient să aibă TAS exact 131,000... mmHg; are sens să întrebăm care este probabilitatea să aibă TAS între 130 și 140 mmHg. De aceea folosim densitatea de probabilitate (engl. probability density function), o curbă f(x) pentru care aria de sub curbă între două valori a și b este probabilitatea ca variabila să cadă în intervalul [a, b]. Aria totală de sub curbă este 1. Histograma unui eșantion mare, desenată cu frecvențe relative și cu clase din ce în ce mai înguste, se apropie de densitate.

Funcția de repartiție (engl. cumulative distribution function), notată F(x), dă probabilitatea ca variabila să fie mai mică sau egală cu x:

$$F(x) = P(X \leq x)$$

Ea crește de la 0 la 1 și corespunde frecvențelor cumulate din tabelele de frecvență din capitolul 2. Când vom citi tabelul distribuției normale din anexa E, vom citi de fapt valori ale funcției de repartiție. Tot de aici vine noțiunea de percentilă: percentila 90 a unei variabile este valoarea x pentru care F(x) = 0,90.

3.2 Două distribuţii discrete utile în medicină

3.2.1 Distribuţia binomială

Să presupunem că în populația de adulți cu risc cardiometabolic din care provine setul mg2026 prevalența diabetului nediagnosticat este 18%. Un medic de familie trimite la TTGO 10 pacienți aleși la întâmplare din această populație. Câți vor avea diabet? Nu putem ști sigur, dar putem calcula probabilitatea fiecărui rezultat posibil.

Situația îndeplinește condițiile distribuției binomiale: un număr fix de încercări (n = 10 pacienți), fiecare cu două rezultate posibile (diabet sau nu), aceeași probabilitate de "succes" la fiecare încercare (p = 0,18) și independență între încercări. Probabilitatea de a obține exact k cazuri este:

$$P(X = k) = C(n,\ k) \times p^k \times (1-p)^{n-k}$$

unde $C(n,\ k) = \frac{n!}{k! \times (n-k)!}$ este numărul de moduri în care putem alege cei k pacienți din n. Media și varianța distribuției binomiale au formule foarte simple:

medie = n x p varianță = n x p x (1 - p)

Pentru exemplul nostru, media este 10 x 0,18 = 1,8 pacienți, varianța 10 x 0,18 x 0,82 = 1,476, iar abaterea standard √1,476 = 1,21 pacienți. Probabilitățile individuale sunt date în tabelul 3.1.

Tabelul 3.1. Distribuția binomială pentru n = 10 pacienți și prevalența p = 0,18.

k (pacienți cu diabet)	**0**	**1**	**2**	**3**	**4**	**5**	**6 sau mai mulți**
$P(X = k)$	0,1374	0,3017	0,2980	0,1745	0,0670	0,0177	0,0037
P(X ≤ k), cumulat	0,1374	0,4392	0,7372	0,9117	0,9787	0,9963	1,0000

Să verificăm prima valoare de mână: $P(X = 0) = 0{,}82^{10} = 0{,}1374$. Cu alte cuvinte, în aproape un grup din șapte de câte 10 pacienți nu vom găsi niciun diabetic, deși prevalența este 18%. Probabilitatea de a găsi cel puțin un caz este 1 - 0,1374 = 0,8626, iar probabilitatea de a găsi trei sau mai multe cazuri este 1 - 0,7372 = 0,2628. Aceste cifre explică o observație pe care o face orice medic tânăr: în grupuri mici, frecvențele observate "sar" mult în jurul valorii așteptate, fără ca acest lucru să însemne ceva special.

Să aplicăm aceeași logică întregului eșantion mg2026. Dacă prevalența reală ar fi 18%, numărul așteptat de diabetici în 160 de pacienți ar fi 160 x 0,18 = 28,8, cu o abatere standard de $\sqrt{160 \times 0{,}18 \times 0{,}82} = 4{,}86$. Am observat 29, adică practic exact valoarea așteptată. Probabilitatea de a observa 35 sau mai mulți ar fi 0,122, deci nici o asemenea valoare nu ar fi fost surprinzătoare. Remarcați un lucru important pentru capitolele următoare: când n este mare, distribuția binomială seamănă tot mai mult cu o distribuție normală cu media n x p și abaterea standard $\sqrt{n \times p \times (1-p)}$. Calculul aproximativ cu distribuția normală (cu o mică corecție de continuitate) dă pentru aceeași întrebare 0,120, foarte aproape de valoarea exactă. Pe această aproximare se bazează intervalul de încredere Wald pentru o proporție, pe care îl vom discuta în capitolul 4.

3.2.2 Distribuția Poisson

Într-un cabinet de medicină de familie se raportează, în medie, două reacții adverse la medicamente pe lună. Într-o lună oarecare pot fi zero, una, trei sau șase. Aici nu avem un număr fix de încercări: numărul pacienților expuși este mare, iar probabilitatea ca un anumit pacient să facă o reacție adversă într-o anumită lună este foarte mică. Pentru astfel de evenimente rare, numărate într-un interval de timp sau de spațiu, se folosește distribuția Poisson:

$$P(X = k) = \frac{e^{-\lambda} \times \lambda^k}{k!}$$

unde λ (lambda) este numărul mediu de evenimente pe interval, iar e = 2,718... este baza logaritmului natural. Distribuția Poisson are o proprietate neobișnuită: media și varianța sunt egale, ambele fiind λ. Pentru λ = 2, abaterea standard este √2 = 1,41.

Tabelul 3.2. Distribuția Poisson pentru λ = 2 reacții adverse pe lună.

k (reacții într-o lună)	**0**	**1**	**2**	**3**	**4**	**5 sau mai multe**
$P(X = k)$	0,1353	0,2707	0,2707	0,1804	0,0902	0,0527

Probabilitatea unei luni fără nicio reacție adversă este $e^{-2} = 0{,}1353$, adică aproximativ o lună din șapte. Probabilitatea de a avea patru sau mai multe reacții este 0,0902 + 0,0527 = 0,1429, iar cinci sau mai multe apar în 5,3% din luni. Dacă medicul ar trage un semnal de alarmă de

fiecare dată când numără patru reacții într-o lună, ar declanșa alarme false în aproape o lună din șapte, chiar dacă nimic nu s-a schimbat. Pragul de cinci reacții dă alarme false în aproximativ o lună din douăzeci. Este exact tipul de raționament pe care îl folosesc sistemele de farmacovigilență și de supraveghere epidemiologică: un număr observat devine "semnal" doar când este improbabil sub regimul obișnuit.

Distribuția Poisson este, de fapt, cazul limită al distribuției binomiale când n este foarte mare și p foarte mic, cu n x p = λ. Dacă 500 de pacienți primesc un medicament, fiecare cu o probabilitate de 0,004 de a face o anumită reacție într-o lună, binomiala dă P(X = 0) = 0,1348, iar Poisson cu λ = 2 dă 0,1353. Diferența este neglijabilă. Pe un an întreg, cu λ = 24, probabilitatea de a număra 30 sau mai multe reacții este 0,132; observați că, pentru λ mare, distribuția Poisson devine și ea aproape simetrică și se apropie de o normală. Figura 3.1 compară cele două distribuții discrete.

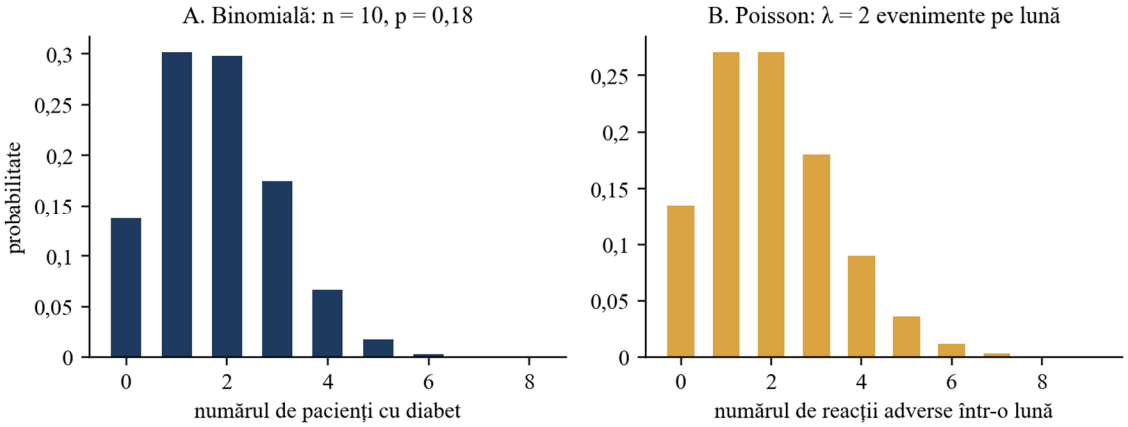


Figura 3.1. Distribuții discrete. A: distribuția binomială a numărului de pacienți cu diabet într-un grup de 10, la o prevalență de 18%; media este 1,8. B: distribuția Poisson a numărului lunar de reacții adverse într-un cabinet, cu λ = 2. Ambele sunt asimetrice spre dreapta. Valori teoretice.

3.3 Distribuția normală

3.3.1 De ce apare atât de des

Distribuția normală, numită și distribuția Gauss sau Laplace-Gauss, descrie variabile continue care rezultă din însumarea multor influențe mici și independente. Înălțimea unui adult depinde de zeci de gene, de alimentația din copilărie, de boli, de mediul social; tensiunea arterială depinde de factori genetici, de greutate, de sare, de stres, de momentul măsurării, de tehnica asistentei. Când multe cauze mici se adună, rezultatul tinde să se distribuie simetric, în formă de clopot. Explicația matematică este teorema limitei centrale, pe care o vom întâlni în secțiunea 3.7.

Trebuie spus de la început că multe variabile biologice NU sunt normale. Concentrațiile serice (trigliceride, enzime hepatice, CRP, insulină), duratele (spitalizare, timpi de intervenție), titrurile de anticorpi, numărul de zile de concediu medical au de regulă asimetrie pozitivă. Nu există o lege a naturii care să oblige datele medicale să fie normale. Distribuția normală este un model, adesea bun, uneori prost, și tocmai de aceea trebuie verificat de fiecare dată [42,53].

3.3.2 Proprietățile distribuției normale

Densitatea distribuției normale are formula:

$$f(x) = \left(\frac{1}{\sigma \times \sqrt{2\pi}}\right) \times e^{-\frac{(x-\mu)^2}{2\,\sigma^2}}$$

Nu trebuie să o memorați. Trebuie însă să rețineți ce consecințe are:

- Curba este simetrică față de media μ. Jumătate din arie se află la stânga mediei, jumătate la dreapta.
- Media, mediana și modul coincid. Într-un eșantion real ele vor fi doar apropiate: pentru TAS din mg2026, media este 131,13 mmHg, iar mediana 131 mmHg.
- Distribuția este complet determinată de doi parametri: media μ, care dă poziția, și abaterea standard σ, care dă lățimea. Două normale cu aceeași medie și σ diferite au același centru, dar una este înaltă și îngustă, cealaltă joasă și lată.
- Aria totală de sub curbă este 1.
- "Cozile" curbei se apropie de axa orizontală pe măsură ce ne depărtăm de medie, dar nu o ating niciodată. Teoretic, orice valoare este posibilă, oricât de mare sau de mică, doar că valorile foarte depărtate de medie au probabilități extrem de mici.
- Punctele de inflexiune ale curbei se află la μ - σ și μ + σ.

Ultima proprietate a cozilor are o consecință practică pe care studenții o descoperă singuri când aplică modelul normal unor variabile care nu pot fi negative. O normală cu media 11,6 zile și abaterea standard 7,0 zile ar prezice că aproximativ 5% dintre pacienți stau în spital un număr negativ de zile, iar 2,3% chiar mai puțin de -2,4 zile, ceea ce este absurd. Vom reveni la acest exemplu, care provine dintr-un articol publicat, în secțiunea 3.8.

3.3.3 Regula 68,3%-95,4%-99,7%

Pentru orice distribuție normală, indiferent de μ și σ, proporția valorilor aflate la o anumită distanță de medie, măsurată în abateri standard, este aceeași (figura 3.2):

- între μ - σ și μ + σ se află 68,3% din valori;
- între μ - 2σ și μ + 2σ se află 95,4% din valori;
- între μ - 3σ și μ + 3σ se află 99,7% din valori.

Două valori derivate se folosesc la fel de des: între μ - 1,96σ și μ + 1,96σ se află exact 95% din valori, iar între μ - 2,576σ și μ + 2,576σ se află 99%. Diferența dintre 2 și 1,96 pare o subtilitate, dar o veți întâlni la fiecare interval de încredere de 95%.

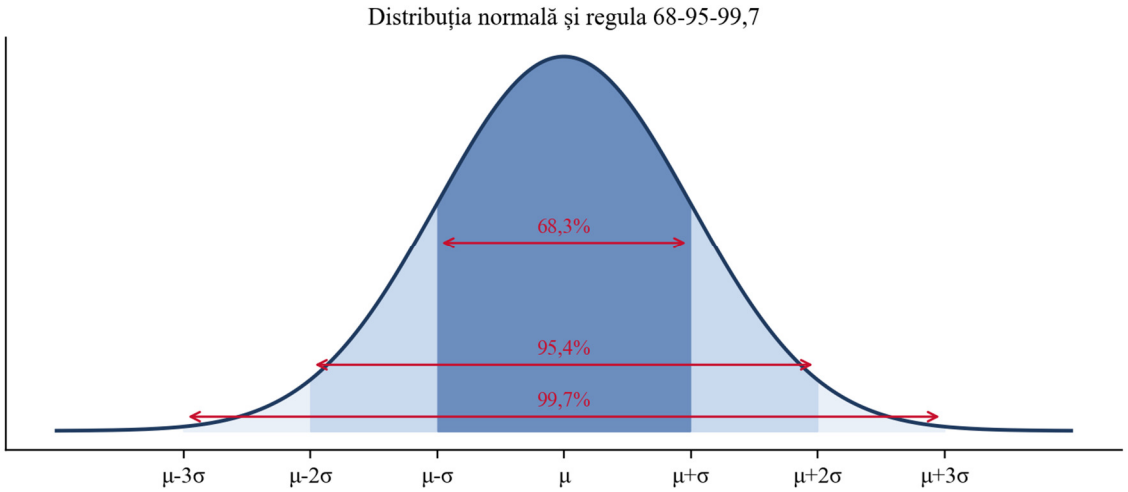


Figura 3.2. Distribuția normală și regula 68-95-99,7: aria cuprinsă între μ - σ și μ + σ este 68,3% din aria totală, între μ - 2σ și μ + 2σ este 95,4%, iar între μ - 3σ și μ + 3σ este 99,7%. Curbă teoretică.

Să verificăm regula pe setul mg2026. Tensiunea sistolică inițială are media 131,13 mmHg și abaterea standard 13,88 mmHg. Intervalul $m \pm s$ este 117,3-145,0 mmHg și conține 70,0% dintre cei 160 de pacienți (teoretic 68,3%). Intervalul $m \pm 2s$, adică 103,4-158,9 mmHg, conține 93,8% (teoretic 95,4%). Intervalul $m \pm 3s$, 89,5-172,8 mmHg, îi conține pe toți (teoretic 99,7%).

Potrivirea nu este perfectă și nici nu are cum să fie, pentru că 160 de pacienți reprezintă un eșantion, nu populația. Este însă suficient de bună ca să putem folosi modelul normal pentru TAS. Atenție la notație: μ și σ sunt parametrii populației, pe care nu îi cunoaștem; m și s sunt estimările lor din eșantion. În practică, înlocuim parametrii cu estimările.

3.3.4 Proporția observată și proporția prezisă de modelul normal

Un test simplu și foarte instructiv al modelului normal este să comparăm proporția observată a pacienților peste un prag clinic cu proporția prezisă de o normală cu media și abaterea standard din eșantion. Pentru TAS ≥ 140 mmHg, pragul clasic al hipertensiunii, scorul z este (140 - 131,13) / 13,88 = 0,64. Din tabelul E.1, Φ(0,64) = 0,7389, deci proporția teoretică peste prag este 1 - 0,7389 = 0,261, adică 26,1%. În eșantion, 41 dintre 160 de pacienți au TAS ≥ 140, adică 25,6%. Diferența de jumătate de punct procentual este neglijabilă.

Același calcul, repetat pentru alte variabile, ne spune multe despre forma distribuției (tabelul 3.3).

Tabelul 3.3. Proporții observate și prezise de modelul normal (setul mg2026, n = 160).

Variabila și pragul	Scorul z al pragului	Proporție teoretică	Proporție observată	Comentariu
TAS ≥ 140 mmHg	0,64	26,1%	25,6%	potrivire foarte bună
TAS < 120 mmHg	-0,80	21,1%	20,0%	potrivire foarte bună
Vârsta ≥ 65 ani	0,99	16,0%	16,3%	potrivire foarte bună
Colesterol ≥ 240 mg/dL	0,83	20,3%	23,8%	acceptabilă
Glicemie ≥ 126 mg/dL	1,30	9,6%	11,3%	coada dreaptă puțin mai groasă
IMC ≥ $\frac{30\ \text{kg}}{m^2}$	0,25	40,3%	33,8%	model nepotrivit
IMC ≥ $\frac{40\ \text{kg}}{m^2}$	2,80	0,3%	1,9%	model nepotrivit

Pentru IMC, modelul normal greșește în ambele direcții: supraestimează proporția obezilor (40,3% față de 33,8%), pentru că media este trasă în sus de câteva valori foarte mari, și subestimează sever proporția celor cu obezitate de gradul III (0,3% față de 1,9%, adică 3 pacienți observați față de mai puțin de 1 prezis). Este semnătura tipică a unei distribuții asimetrice spre dreapta, cu coadă lungă. Pentru tensiune și vârstă, modelul funcționează excelent. Vom vedea în secțiunea 3.8 că testele formale de normalitate ajung exact la aceeași concluzie.

3.4 Distribuția normală standard și scorul z

3.4.1 Standardizarea

Există o infinitate de distribuții normale, câte una pentru fiecare pereche (μ, σ). Ar fi imposibil să avem câte un tabel pentru fiecare. Soluția este să le aducem pe toate la una singură, distribuția normală standard, cu media 0 și abaterea standard 1, notată N(0, 1). Transformarea se numește standardizare și produce scorul z:

$$z = \frac{x - \mu}{\sigma}$$

sau, cu estimările din eșantion, $z = \frac{x-m}{s}$. Scorul z spune câte abateri standard separă o valoare de medie și în ce direcție. Un z pozitiv înseamnă o valoare peste medie, un z negativ, sub medie. Scorul z nu are unitate de măsură: mmHg se simplifică cu mmHg. De aceea putem compara

scoruri z ale unor variabile complet diferite. Un pacient cu z = 2,1 la glicemie și z = 0,3 la colesterol este "mai neobișnuit" la glicemie decât la colesterol, chiar dacă cele două se măsoară pe scale diferite.

Dacă o variabilă este normală, scorurile ei z urmează distribuția normală standard. Media scorurilor z ale unui eșantion este întotdeauna 0, iar abaterea lor standard este 1, chiar dacă variabila nu este normală. Standardizarea nu schimbă însă forma distribuției: o variabilă asimetrică rămâne la fel de asimetrică după transformarea în scoruri z. Este o greșeală frecventă să se creadă că "standardizarea normalizează datele".

3.4.2 Un exemplu complet: TAS de 160 mmHg

Un pacient din mg2026 are TAS 160 mmHg. Cât de neobișnuită este această valoare în eșantion?

$$z = \frac{160 - 131{,}13}{13{,}88} = \frac{28{,}87}{13{,}88} = 2{,}08$$

Valoarea se află la 2,08 abateri standard deasupra mediei (figura 3.3). În tabelul E.1 din anexa E căutăm rândul 2,0 și coloana 0,08 și citim Φ(2,08) = 0,9812. Asta înseamnă că, sub modelul normal, 98,1% dintre pacienți au TAS mai mică sau egală cu 160 mmHg: pacientul se află la percentila 98. Proporția celor cu valori mai mari este 1 - 0,9812 = 0,0188, adică 1,9%, iar aria dintre medie și 160 mmHg este 0,9812 - 0,5 = 0,4812.

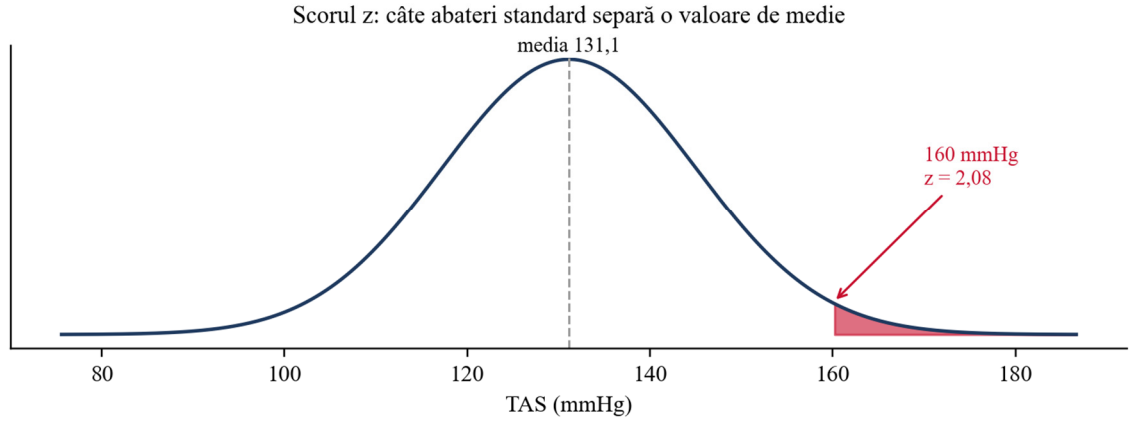


Figura 3.3. Scorul z al unei tensiuni sistolice de 160 mmHg în setul mg2026 (media 131,1 mmHg, abaterea standard 13,9 mmHg) este 2,08. Aria colorată de la dreapta valorii reprezintă proporția teoretică a pacienților cu TAS mai mare, aproximativ 1,9%. Date sintetice (setul mg2026).

Pentru valori sub medie, folosim simetria curbei. Un pacient cu TAS 108 mmHg are $z = \frac{108-131{,}13}{13{,}88} = -1{,}67$. Tabelul dă Φ(1,67) = 0,9525, iar prin simetrie Φ(-1,67) = 1 - 0,9525 = 0,0475. Pacientul se află aproximativ la percentila 5. În eșantion, 6,9% dintre pacienți au TAS ≤ 108 mmHg, din nou aproape de valoarea teoretică.

3.4.3 Problema inversă: de la procent la valoare

Uneori întrebarea este inversă: care este valoarea TAS sub care se află 90% dintre pacienți? Căutăm în tabel z pentru care Φ(z) = 0,90 și găsim z = 1,28 (mai precis 1,2816, valoare dată în tabelul E.2). Apoi refacem drumul invers:

$$x = m + z \times s = 131{,}13 + 1{,}28 \times 13{,}88 = 148{,}9 \text{ mmHg}$$

Percentila 90 observată în eșantion este 148,2 mmHg, iar cea teoretică 148,9 mmHg. Valorile z pe care merită să le știți pe de rost sunt puține: 1,28 (percentila 90), 1,645 (percentila 95), 1,96 (percentila 97,5), 2,33 (percentila 99) și 2,576 (percentila 99,5). Le veți folosi la intervalele de încredere, la testele unilaterale și bilaterale și la calculul dimensiunii eșantionului.

3.4.4 Distribuția normală în Google Sheets

În foile de calcul nu mai este nevoie de tabel. Google Sheets oferă patru funcții pentru distribuția normală, plus funcția de standardizare [14,43]. Cu configurarea regională România, separatorul de argumente este ";" și virgula este separator zecimal.

Tabelul 3.4. Funcțiile Google Sheets pentru distribuția normală, cu exemple pe TAS din mg2026.

Funcția	Ce calculează	Exemplu	Rezultat
=STANDARDIZE(x; medie; AS)	scorul z	=STANDARDIZE(160; 131,13; 13,88)	2,08
=NORM.DIST(x; medie; AS; TRUE)	aria la stânga lui x (funcția de repartiție)	=NORM.DIST(160; 131,13; 13,88; TRUE)	0,981
=NORM.DIST(x; medie; AS; FALSE)	înălțimea curbei (densitatea), rar folosită	=NORM.DIST(131; 131,13; 13,88; FALSE)	0,029
=NORM.S.DIST(z; TRUE)	aria la stânga lui z în normala standard	=NORM.S.DIST(2,08; TRUE)	0,9812
=NORM.INV(p; medie; AS)	valoarea x sub care se află proporția p	=NORM.INV(0,9; 131,13; 13,88)	148,9
=NORM.S.INV(p)	valoarea z sub care se află proporția p	=NORM.S.INV(0,975)	1,96

În practică, în locul valorilor numerice se pun referințe: =STANDARDIZE(M2; AVERAGE($M$2:$M$161); STDEV($M$2:$M$161)) calculează scorul z al primului pacient și poate fi copiată pe toată coloana. Ultimul argument al funcției NORM.DIST este o capcană clasică la laborator: cu TRUE obținem probabilitatea cumulată (care ne interesează aproape întotdeauna), cu FALSE obținem densitatea, un număr care nu este o probabilitate și care nu are interpretare directă. Funcțiile NORMDIST și NORMSDIST din versiunile vechi de Excel există și ele, cu aceeași logică.

3.4.5 Scorul z în clinică

Scorurile z nu sunt doar un exercițiu de seminar. Curbele de creștere pediatrice exprimă greutatea și înălțimea copilului ca scor z (sau percentilă) față de o populație de referință de aceeași vârstă și același sex; un copil cu înălțimea la z = -2 este la percentila 2,3. Densitometria osoasă raportează scorul T, care este un scor z calculat față de media adulților tineri, și scorul Z, calculat față de persoane de aceeași vârstă. Ecocardiografia pediatrică exprimă dimensiunile cavităților și ale vaselor ca scoruri z față de suprafața corporală. În toate aceste situații, medicul compară pacientul cu o distribuție de referință, iar întrebarea este mereu aceeași: cât de departe de centru se află valoarea lui, măsurat în abateri standard?

3.5 Intervalele de referință în medicina de laborator

3.5.1 De unde vin "valorile normale"

Pe buletinul de analize, lângă fiecare rezultat, apare un interval de referință, popular numit "valori normale". El se construiește, de regulă, pe un eșantion de persoane sănătoase, ca interval care cuprinde 95% central din valorile lor [64]. Dacă variabila este aproximativ normală, intervalul este:

$$m - 1{,}96 \times s;\ m + 1{,}96 \times s$$

Dacă variabila nu este normală, se folosesc direct percentilele 2,5 și 97,5 ale eșantionului de referință (metoda neparametrică) sau se transformă datele, se calculează intervalul pe scala transformată și se retransformă limitele [42,53].

Consecința logică a acestei construcții surprinde pe mulți: prin definiție, 5% dintre persoanele perfect sănătoase au o valoare "în afara normalului", 2,5% sub limita inferioară și 2,5% peste cea superioară. Un rezultat ușor în afara intervalului nu înseamnă boală. Înseamnă doar că valoarea este neobișnuită comparativ cu populația de referință.

Problema se agravează când se cer multe analize. Dacă cele k analize ar fi independente, probabilitatea ca o persoană sănătoasă să aibă toate rezultatele în interval ar fi $0{,}95^k$. Pentru 5 analize, 0,95^5 = 0,774, deci 22,6% dintre persoanele sănătoase ar avea cel puțin un rezultat "anormal". Pentru 12 analize, proporția crește la 46%, iar pentru 20 de analize la 64%. Analizele reale sunt corelate între ele, așa că cifrele exacte diferă, dar ideea rămâne: un "screening" cu zeci de parametri va găsi aproape la fiecare om sănătos ceva de investigat. Acesta este unul dintre motivele pentru care analizele se cer cu o întrebare clinică în minte, nu "la grămadă".

3.5.2 Un interval de 95% pe setul mg2026

Să construim intervalul $m \pm 1{,}96s$ pentru TAS din mg2026:

limita inferioară = 131,13 - 1,96 x 13,88 = 131,13 - 27,20 = 103,9 mmHg limita superioară = 131,13 + 1,96 x 13,88 = 131,13 + 27,20 = 158,3 mmHg

Intervalul 103,9-158,3 mmHg conține 150 dintre cei 160 de pacienți, adică 93,8%, aproape de 95%. Percentilele 2,5 și 97,5 calculate direct din date sunt 103,0 și 156,1 mmHg, foarte apropiate de limitele obținute cu modelul normal. Când variabila este normală, cele două metode dau rezultate similare, iar metoda parametrică are avantajul că folosește toate datele și nu doar cele câteva valori extreme.

Trebuie făcută o precizare de fond: pacienții din mg2026 au risc cardiometabolic, deci nu formează o populație de referință sănătoasă. Intervalul calculat descrie acest eșantion și ne servește ca exercițiu, nu ca interval de referință pentru tensiunea arterială. Valorile-țintă pentru TA nu se definesc statistic, ci pe baza riscului cardiovascular asociat, prin ghiduri. Distincția dintre "normal statistic" (frecvent) și "normal clinic" (fără risc crescut) merită ținută minte: o colesterolemie de 220 mg/dL poate fi foarte frecventă într-o populație și totuși asociată cu risc crescut.

3.5.3 Ce se întâmplă cu o variabilă asimetrică

Pentru trigliceride (media 161,18 mg/dL, abaterea standard 76,30 mg/dL), formula $m \pm 1{,}96s$ dă 11,6-310,7 mg/dL. Limita inferioară este fără sens clinic: nimeni din eșantion nu are trigliceride sub 44 mg/dL, iar 11,6 mg/dL este o valoare practic inexistentă la adult. În partea dreaptă, 6 pacienți depășesc limita superioară, toți în coada lungă. Intervalul conține 96,3% din date, dar este deplasat: prea larg jos, prea strâmt sus.

Dacă logaritmăm valorile (secțiunea 3.9), media logaritmilor este 4,970 și abaterea lor standard 0,488. Intervalul pe scala logaritmică este $4{,}970 \pm 1{,}96 \times 0{,}488$, iar după retransformare cu funcția exponențială obținem 55,3-374,9 mg/dL. Percentilele 2,5 și 97,5 observate sunt 53,8 și 364,0 mg/dL, aproape identice cu limitele obținute pe scala logaritmică. Intervalul retransformat nu mai este simetric față de medie, iar asta este corect: reflectă forma reală a distribuției.

> **Studiu de caz: praguri pe distribuții continue în elastografia hepatică**
>
> În diagnosticul neinvaziv al bolilor hepatice, parametrii măsurați sunt continui, iar decizia clinică cere praguri. Moga și colaboratorii au evaluat elastografia time-harmonic față de elastografia tranzitorie la 165 de pacienți și au propus două praguri pentru ciroză: un prag de excludere (engl. rule-out) sub 1,61 m/s (7,77 kPa), cu sensibilitate 90,7% și specificitate 66,6%, și un prag de confirmare (engl. rule-in) peste 1,83 m/s (10 kPa), cu sensibilitate 65,1%

> și specificitate 96,7%; aria de sub curba ROC a fost 0,90 (IC 95% 0,82-0,93) [3]. Două praguri în loc de unul nu sunt o complicație inutilă: ele recunosc faptul că distribuțiile valorilor la pacienții cu și fără ciroză se suprapun. Sub primul prag se află mai ales pacienți fără ciroză, peste al doilea mai ales pacienți cu ciroză, iar între ele rămâne o "zonă gri" unde decizia cere alte investigații. Nica și colaboratorii au procedat asemănător pentru fracția de grăsime măsurată ecografic (USFF), la 169 de pacienți, cu praguri de peste 5% pentru steatoza cel puțin ușoară (AUROC 0,90), peste 7,5% pentru cea moderată (AUROC 0,92) și peste 10% pentru cea severă (AUROC 0,91), folosind ca referință CAP, cu pragurile 248, 268 și 280 dB/m [65]. Ce nu ne spun rezumatele: forma distribuțiilor în fiecare grup și proporția pacienților din zona gri, două informații utile pentru a judeca aplicabilitatea pragurilor în alt centru. Le vom relua în capitolul 12, la curbele ROC.

3.6 Distribuțiile de eșantionare: t, chi-pătrat și F

Până acum am vorbit despre distribuția valorilor individuale: cum se repartizează tensiunile pacienților. Statistica inferențială folosește însă un alt tip de distribuție, la fel de important: distribuția unei statistici calculate din eșantion, de exemplu media, dacă am repeta studiul de foarte multe ori. Aceasta se numește distribuție de eșantionare (engl. sampling distribution). Trei distribuții de eșantionare apar continuu în biostatistică și le prezentăm acum doar ca formă, pentru a le recunoaște în capitolele următoare (figura 3.4).

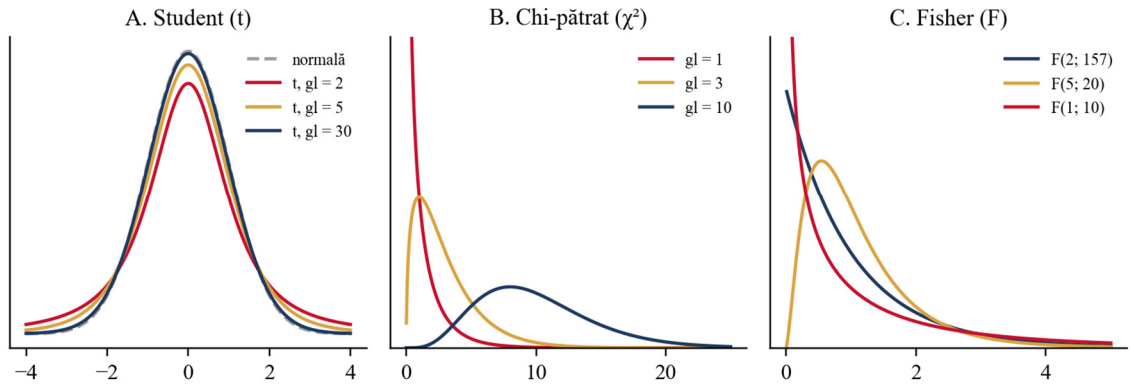


Figura 3.4. Trei distribuții de eșantionare. A: distribuția Student (t) pentru 2, 5 și 30 de grade de libertate, comparată cu normala standard (linie întreruptă); cu cât gl sunt mai puține, cu atât cozile sunt mai groase. B: distribuția chi-pătrat pentru 1, 3 și 10 grade de libertate, asimetrică spre dreapta. C: distribuția Fisher (F) pentru câteva perechi de grade de libertate. Curbe teoretice.

Distribuția Student (t) a fost publicată de W. S. Gosset sub pseudonimul "Student". Ea apare când standardizăm o medie folosind abaterea standard estimată din eșantion, s, în locul abaterii standard reale σ, pe care nu o cunoaștem. Are formă de clopot, este simetrică față de 0, dar are cozile mai groase decât normala: estimarea lui σ din puține date adaugă incertitudine. Forma depinde de un singur parametru, numărul gradelor de libertate (gl), care pentru o medie este n - 1. Când gl cresc, distribuția t se apropie de normala standard. Valoarea critică bilaterală pentru 95% este 2,776 la 4 gl, 2,262 la 9 gl, 2,045 la 29 gl, 1,975 la 159 gl și 1,96 la infinit (anexa E, tabelul E.3). Aceste valori vor înlocui pe 1,96 în intervalele de încredere pentru eșantioane mici (capitolul 4) și în testele t (capitolul 6).

Gradele de libertate merită o explicație intuitivă. Dacă știm media a 10 valori și cunoaștem 9 dintre ele, a zecea este determinată: nu mai este "liberă". Când calculăm abaterea standard în jurul mediei eșantionului, folosim deja o informație extrasă din date (media), așa că rămân n - 1 valori libere. De aici vine numitorul n - 1 din formula abaterii standard, despre care vom vorbi în capitolul 4.

Distribuția chi-pătrat (χ^2) apare când adunăm pătratele unor variabile normale standard. Ia doar valori pozitive, este asimetrică spre dreapta, iar asimetria scade când gl cresc. O vom folosi la testul chi-pătrat pentru tabele de contingență (capitolul 7), unde gl = (r - 1) x (c - 1), și la intervalul de încredere pentru o varianță. Valoarea critică pentru α = 0,05 este 3,84 la 1 gl, 5,99 la 2 gl și 7,81 la 3 gl.

Distribuția Fisher (F) este raportul a două variabile chi-pătrat, fiecare împărțită la gradele ei de libertate. Are două numere de grade de libertate, unul pentru numărător și unul pentru numitor, ia doar valori pozitive și este asimetrică spre dreapta. O vom folosi la analiza de varianță (ANOVA, capitolul 6), la testul Levene pentru egalitatea variațelor și la regresie. De exemplu, pentru ANOVA cu trei brațe și 160 de pacienți, gl sunt 2 și 157, iar valoarea critică pentru α = 0,05 este 3,05.

Nu trebuie să calculați aceste distribuții de mână. Trebuie să știți ce este fiecare, la ce test apare și că valorile critice scad pe măsură ce gradele de libertate cresc.

3.7 Teorema limitei centrale

3.7.1 Enunțul, fără formule

Teorema limitei centrale este, probabil, cel mai important rezultat al statisticii aplicate. În cuvinte simple: dacă extragem eșantioane de mărime n dintr-o populație cu media μ și abaterea standard σ, oricare ar fi forma distribuției populației, mediile acestor eșantioane au o distribuție care:

- este centrată pe μ (mediile eșantioanelor nu sunt sistematic nici prea mari, nici prea mici);
- are abaterea standard $\frac{\sigma}{\sqrt{n}}$, numită eroare standard a mediei;
- se apropie de o distribuție normală pe măsură ce n crește.

Ultimul punct este partea surprinzătoare. Populația poate fi asimetrică, bimodală, oricum, iar mediile eșantioanelor tot vor tinde spre o curbă în formă de clopot. Cu cât populația este mai asimetrică, cu atât este nevoie de un n mai mare. Pentru distribuții moderat asimetrice, 30-50 de observații ajung de obicei; pentru distribuții foarte asimetrice (de exemplu costuri de spitalizare, unde câțiva pacienți costă de zeci de ori mai mult decât restul), poate fi nevoie de sute.

3.7.2 O simulare pe trigliceride

Ca să vedem teorema la lucru, am tratat cele 160 de valori ale trigliceridelor din mg2026 ca pe o "populație" și am extras din ea, cu calculatorul, câte 10.000 de eșantioane aleatoare (cu întoarcere) de mărime 2, 5, 10, 30 și 50. Pentru fiecare eșantion am calculat media, apoi am privit distribuția celor 10.000 de medii (figura 3.5 și tabelul 3.5).

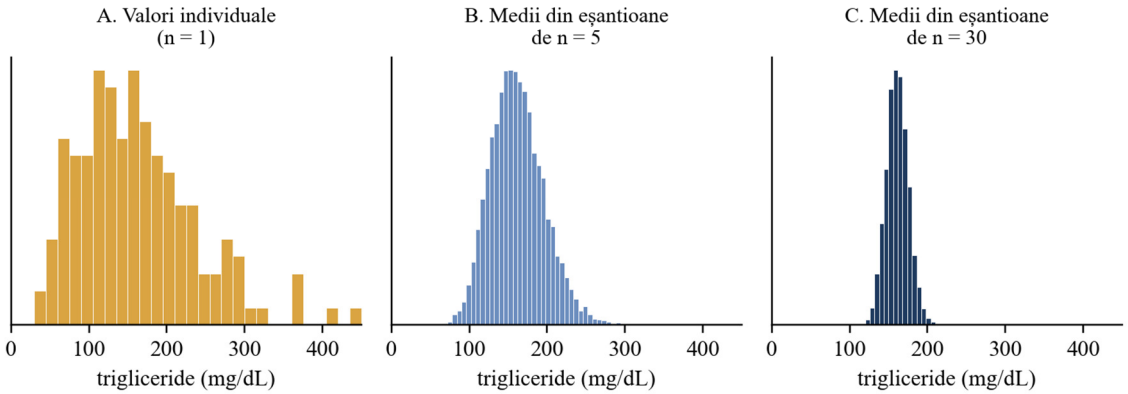


Figura 3.5. Teorema limitei centrale pe trigliceride. A: distribuția valorilor individuale, asimetrică spre dreapta (asimetrie 0,95). B: distribuția mediilor a 10.000 de eșantioane de câte 5 pacienți, mai simetrică. C: distribuția mediilor eșantioanelor de câte 30 de pacienți, aproape normală și mult mai îngustă. Simulare pe setul mg2026.

Tabelul 3.5. Simularea teoremei limitei centrale: 10.000 de eșantioane din valorile trigliceridelor (mg2026).

Mărimea eșantionului n	**Media celor 10.000 de medii (mg/dL)**	**Abaterea standard a mediilor (mg/dL)**	$\frac{\sigma}{\sqrt{n}}$ teoretic (mg/dL)	**Asimetria mediilor**
1 (valori individuale)	161,2	76,3	-	0,95
2	162,2	54,3	53,8	0,65
5	160,6	33,7	34,0	0,39
10	161,1	24,0	24,1	0,25
30	161,3	14,1	13,9	0,21
50	161,2	10,8	10,8	0,19

Trei lucruri se văd imediat. Media mediilor rămâne în jur de 161 mg/dL, adică media "populației", oricare ar fi n. Abaterea standard a mediilor scade exact cum prezice formula $\frac{\sigma}{\sqrt{n}}$: la n = 30 este 14,1 mg/dL, față de 76,3 mg/dL pentru valorile individuale. Asimetria scade de la 0,95 la aproximativ 0,2. Mediile eșantioanelor de 30 de pacienți sunt deja aproape normale, deși trigliceridele individuale nu sunt.

3.7.3 De ce contează pentru medic

Teorema limitei centrale explică de ce testul t și intervalul de încredere pentru medie funcționează bine și când datele nu sunt perfect normale, cu condiția ca eșantionul să fie suficient de mare. Ea explică și de ce o medie calculată pe 160 de pacienți este mult mai stabilă decât o valoare individuală. În capitolul 4 vom transforma aceste observații în intervale de încredere. Rețineți însă și limita: teorema vorbește despre medii, nu despre valori individuale. Un interval de referință pentru pacienți (secțiunea 3.5) cere ca valorile individuale să fie normale; teorema limitei centrale nu ne ajută acolo.

Simularea o puteți repeta și voi, în Google Sheets, fără niciun program special: în lucrarea practică 3 se extrag eșantioane cu funcția RAND și se calculează mediile lor, iar în lucrarea practică 4 se urmărește cum se schimbă lățimea intervalelor de încredere.

3.8 Verificarea normalității

3.8.1 De ce verificăm și ce anume verificăm

Multe metode din capitolele următoare (intervalul de încredere cu distribuția t, testul t, ANOVA, corelația Pearson, regresia liniară) presupun că datele sau reziduurile provin dintr-o distribuție normală. Dacă presupunerea este grav încălcată, mai ales în eșantioane mici, valorile p și

intervalele de încredere pot fi înșelătoare. De aceea, înainte de a alege un test, privim distribuția. Verificarea are trei componente, care se folosesc împreună: metodele grafice, indicatorii numerici și testele formale [49,53].

O precizare care scutește multe confuzii: ce trebuie să fie normal depinde de test. La testul t pentru eșantioane independente ne interesează distribuția variabilei în fiecare grup, nu în eșantionul total. La testul t pereche ne interesează distribuția diferențelor, nu a valorilor inițiale și finale. La regresie ne interesează distribuția reziduurilor. Testarea normalității pe "toată coloana" este uneori irelevantă pentru testul pe care urmează să-l aplicăm.

3.8.2 Metodele grafice

Histograma este primul lucru la care ne uităm. Pentru o variabilă normală, ea arată un singur vârf, aproximativ la mijloc, și două "umeri" care coboară simetric. Histograma TAS din figura 3.6A se potrivește acestei descrieri, cea a trigliceridelor din figura 3.6B are vârful deplasat spre stânga și o coadă lungă spre dreapta. Histograma are însă un defect: aspectul ei depinde de numărul și lățimea claselor, iar în eșantioane mici (sub 30-40 de valori) poate arăta neregulat chiar dacă datele vin dintr-o populație perfect normală.

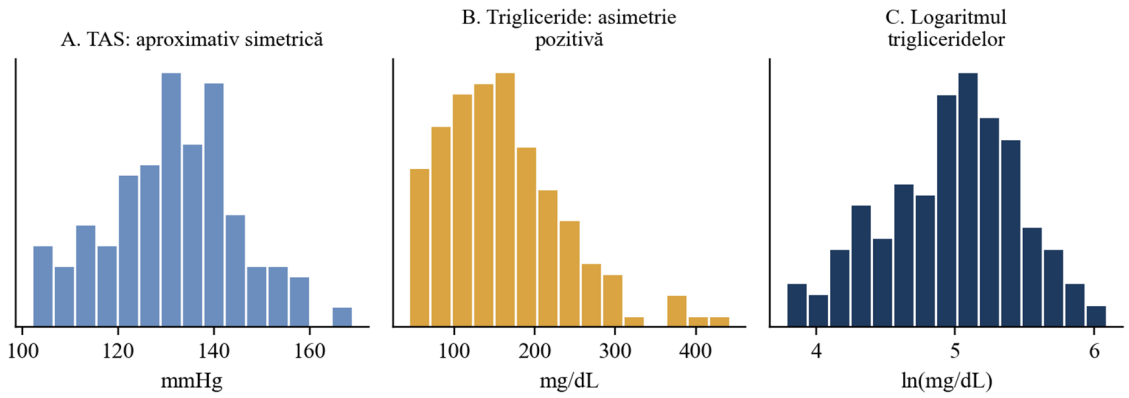


Figura 3.6. Forma distribuției în setul mg2026. A: tensiunea sistolică inițială, aproximativ simetrică (asimetrie 0,01). B: trigliceridele, cu asimetrie pozitivă (0,95) și coadă lungă spre dreapta. C: logaritmul natural al trigliceridelor, aproximativ simetric (asimetrie -0,28). Date sintetice (setul mg2026).

Graficul Q-Q (engl. quantile-quantile plot) compară cuantilele observate cu cuantilele pe care le-am aștepta dacă datele ar fi normale. Fiecare punct este o observație: pe axa orizontală, valoarea pe care ar trebui să o aibă dacă distribuția ar fi normală, pe axa verticală, valoarea reală. Dacă datele sunt normale, punctele se așază pe o linie dreaptă. Abaterile au tipare ușor de recunoscut:

- punctele se curbează în sus la capătul din dreapta și stau deasupra liniei la capătul din stânga (formă de "banană" sau de "U"): asimetrie pozitivă, ca la trigliceride;
- curbarea inversă (formă de "n"): asimetrie negativă;
- punctele coboară sub linie la stânga și urcă peste linie la dreapta, ca un "S" culcat: cozi mai groase decât normala (exces pozitiv);
- un singur punct departe de linie, la un capăt: o valoare extremă.

În figura 3.7, punctele pentru TAS urmează linia aproape perfect, cu mici abateri la capete, normale pentru 160 de observații. Punctele pentru trigliceride formează clar o curbă: valorile mari sunt mai mari decât ar prezice modelul normal, iar valorile mici sunt mai puțin mici decât ar trebui.

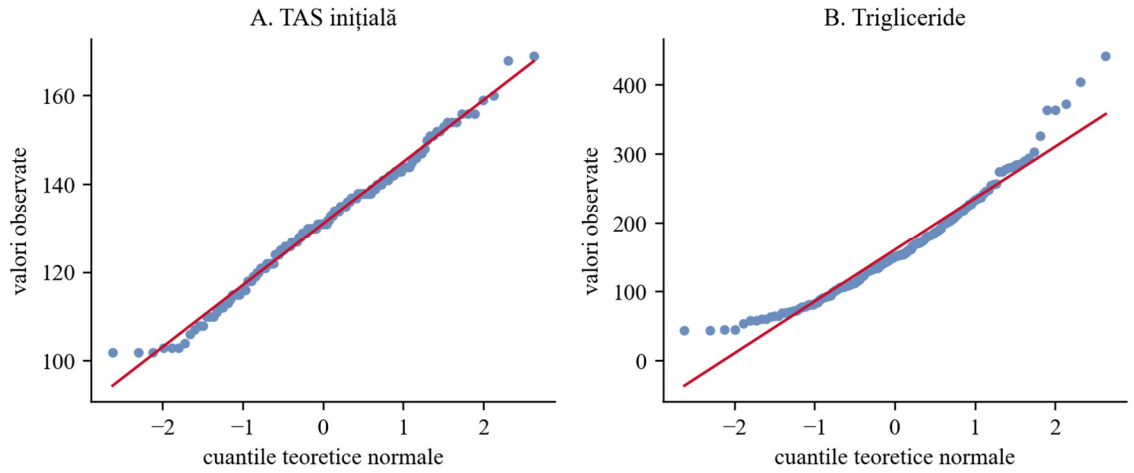


Figura 3.7. Graficele Q-Q pentru tensiunea sistolică inițială (A), unde punctele urmează linia, și pentru trigliceride (B), unde punctele formează o curbă caracteristică asimetriei pozitive. Date sintetice (setul mg2026).

Box-plot-ul, descris în capitolul 2, arată asimetria prin poziția medianei în cutie (mai aproape de Q1 când asimetria este pozitivă), prin lungimea inegală a "mustăților" și prin punctele individuale din afara lor. Este excelent pentru a compara mai multe grupuri și pentru a vedea valorile extreme, dar spune puțin despre forma exactă a distribuției.

3.8.3 Indicatorii numerici: asimetria și excesul

Asimetria (engl. skewness) măsoară cât de dezechilibrată este distribuția față de medie. Pentru o distribuție simetrică, asimetria este 0. Valorile pozitive indică o coadă mai lungă spre dreapta (valori mari rare, dar foarte depărtate), valorile negative, o coadă spre stânga. Excesul (engl. excess kurtosis), numit adesea simplu "kurtosis" în programe, măsoară cât de groase sunt cozile comparativ cu o normală: normala are excesul 0, o distribuție cu cozi groase și vârf ascuțit are exces pozitiv, una "turtită", cu cozi subțiri, are exces negativ. Google Sheets (funcțiile SKEW și KURT), JASP și MedCalc raportează versiunile corectate pentru eșantion ale acestor indicatori, deci dau aceleași valori pe aceleași date.

Cât de mare trebuie să fie asimetria ca să ne îngrijoreze? Răspunsul depinde de n, pentru că în eșantioane mici asimetria variază mult de la un eșantion la altul chiar dacă populația este normală. De aceea, asimetria și excesul se raportează la erorile lor standard. Pentru n mare, aproximativ:

$$\mathrm{ES(\ asimetrie\ )} \approx \sqrt{\frac{6}{n}} \qquad \mathrm{ES(\ exces\ )} \approx \sqrt{\frac{24}{n}}$$

Formulele exacte, folosite de programe, dau pentru n = 160 o eroare standard de 0,192 pentru asimetrie și 0,381 pentru exces (aproximările dau 0,194 și 0,387). Regula practică spune că un raport |asimetrie / ES| sau |exces / ES| mai mare de 2 (unii autori folosesc 1,96 sau, pentru eșantioane mari, 3) semnalează o abatere de la normalitate care merită luată în seamă.

Alți doi indicatori simpli completează imaginea. Primul este compararea mediei cu mediana: într-o distribuție simetrică sunt apropiate, într-una cu asimetrie pozitivă media este mai mare decât mediana (trasă de coada dreaptă). Al doilea este testul "mediei minus două abateri standard": pentru o variabilă care nu poate fi negativă, dacă m - 2s este sub zero, distribuția nu poate fi normală, pentru că modelul normal ar prezice cel puțin 2,3% valori negative, imposibile [53].

3.8.4 Testele formale de normalitate

Testele de normalitate formulează ipotezele astfel:

H0: datele provin dintr-o populație cu distribuție normală; H1: datele provin dintr-o populație cu distribuție care nu este normală.

O valoare p mică (sub 0,05) ne face să respingem H0 și să spunem că distribuția se abate semnificativ de la normalitate. O valoare p mare înseamnă că nu avem dovezi suficiente împotriva normalității. Observați formularea: nu "datele sunt normale", ci "nu am găsit dovezi că nu ar fi". Diferența va deveni centrală în capitolul 5.

Testul Shapiro-Wilk [66] calculează o statistică W, între 0 și 1, care măsoară cât de bine se aliniază valorile ordonate cu cuantilele normale așteptate; este în esență o corelație pe graficul Q-Q. Valori ale lui W apropiate de 1 indică normalitate. Este testul cu cea mai bună putere pentru majoritatea abaterilor de la normalitate și este testul implicit în JASP. Este recomandat mai ales pentru eșantioane mici și medii.

Testul Kolmogorov-Smirnov compară funcția de repartiție observată (frecvențele cumulate ale datelor) cu funcția de repartiție a normalei și folosește drept statistică D cea mai mare distanță verticală dintre cele două curbe. Varianta clasică presupune că media și abaterea standard ale populației sunt cunoscute dinainte. Când le estimăm din aceleași date, cum se întâmplă aproape întotdeauna, testul clasic devine prea "îngăduitor": dă valori p prea mari. Corecția Lilliefors rezolvă problema. În mg2026, testul Kolmogorov-Smirnov clasic aplicat IMC dă p = 0,082, ceea ce ar sugera, greșit, normalitatea, în timp ce varianta cu corecția Lilliefors dă p = 0,002, iar Shapiro-Wilk p < 0,001. Unele programe afișează testul clasic fără să o spună explicit; merită verificat în documentație.

Testul Anderson-Darling este o variantă a ideii Kolmogorov-Smirnov care dă mai multă greutate cozilor distribuției. Statistica A^2 se compară cu o valoare critică (pentru n = 160 și α = 0,05, aproximativ 0,75). Îl veți întâlni în MedCalc și în programele de control al calității din laborator.

Tabelul 3.6 adună toate aceste informații pentru variabilele principale din mg2026.

Tabelul 3.6. Verificarea normalității pentru câteva variabile din setul mg2026 (n = 160; ES asimetrie 0,192; ES exces 0,381).

Variabila	Media	Mediana	Asimetria (raport la ES)	Excesul (raport la ES)	Shapiro-Wilk W; p	Lilliefors p	Anderson-Darling A^2
Vârsta (ani)	56,29	56	0,01 (0,06)	-0,09 (-0,24)	0,993; 0,687	0,835	0,26
TAS0 (mmHg)	131,13	131	0,01 (0,03)	-0,10 (-0,26)	0,990; 0,303	0,534	0,38
Glicemie_0(n	105,44	104,5	0,37 (1,92)	-0,23 (-0,60)	0,982; 0,037	0,127	0,63
$\text{IMC}_0\left(\frac{\text{kg}}{m^2}\right)$	29,04	28,25	1,42 (7,42)	3,68 (9,64)	0,904; < 0,001	0,002	2,80
Trigliceride (mg/dL)	161,18	151,5	0,95 (4,97)	1,12 (2,94)	0,945; < 0,001	0,016	1,78
ln(trigliceride)	4,97	5,02	-0,28 (-1,45)	-0,28 (-0,73)	0,988; 0,200	0,364	0,50

Citind tabelul pe rânduri, concluziile sunt clare pentru patru dintre variabile. Vârsta și TAS sunt compatibile cu normalitatea pe toate criteriile: asimetrie și exces practic nule, medie egală cu mediana, toate testele nesemnificative (Shapiro-Wilk p = 0,687 pentru vârstă și p = 0,303 pentru TAS; Kolmogorov-Smirnov clasic pentru TAS p = 0,823). IMC este clar asimetric: asimetria de 1,42 este de peste 7 ori mai mare decât eroarea ei standard, excesul de 3,68 arată cozi groase (câțiva pacienți cu obezitate severă), media depășește mediana cu 0,8 kg/m², iar toate testele

resping normalitatea. Trigliceridele sunt și ele asimetrice, dar logaritmul lor se comportă ca o variabilă normală: spunem că trigliceridele au o distribuție log-normală.

3.8.5 Cazul la limită: glicemia a jeun

Glicemia inițială merită o discuție separată, pentru că aici criteriile nu mai sunt de acord. Shapiro-Wilk dă p = 0,037, deci, strict formal, respingem normalitatea la pragul de 0,05. În schimb, testul Lilliefors (p = 0,127) și testul Anderson-Darling ($A^2 = 0{,}63$, sub valoarea critică de 0,75) nu o resping. Asimetria este 0,37, cu un raport la eroarea standard de 1,92, chiar sub pragul de 2. Excesul este neglijabil. Media (105,44 mg/dL) și mediana (104,5 mg/dL) diferă cu sub 1 mg/dL. Histograma glicemiei din figura 2.3C (capitolul 2) arată o ușoară alungire spre dreapta.

Ce facem? Mai întâi înțelegem de unde vine abaterea. Eșantionul conține 29 de pacienți cu diabet nediagnosticat, la care glicemia este mai mare; ei formează o "umflătură" în coada dreaptă, peste o distribuție altfel simetrică. Este o explicație biologică plauzibilă, nu un artefact. Apoi ne întrebăm ce analiză urmează. Pentru un interval de încredere al mediei sau un test t cu 160 de pacienți, o asimetrie de 0,37 este nesemnificativă practic: teorema limitei centrale garantează că media este aproape normal distribuită (secțiunea 3.7). Pentru un interval de referință al valorilor individuale, abaterea ar conta mai mult, pentru că afectează exact cozile. În fine, raportăm transparent: "glicemia a avut o ușoară asimetrie pozitivă (0,37; Shapiro-Wilk p = 0,037); am folosit teste parametrice, iar rezultatele au fost confirmate cu teste neparametrice". În capitolul 6 veți vedea că, pentru comparația glicemiei între sexe, testul t și testul Mann-Whitney duc la aceeași concluzie.

Lecția generală este că decizia "normal sau nu" nu se ia mecanic, după o singură valoare p, ci prin cântărirea tuturor indiciilor, ținând cont de mărimea eșantionului și de analiza care urmează [18,42].

3.8.6 Capcanele interpretării testelor de normalitate

Testele de normalitate au o problemă de fond: puterea lor depinde puternic de mărimea eșantionului, exact în sens invers față de nevoile noastre.

În eșantioane mari (sute sau mii de observații), testele devin extrem de sensibile și resping normalitatea pentru abateri minuscule, fără importanță practică. Într-o simulare cu câte 2000 de valori extrase dintr-o distribuție cu asimetrie de doar 0,2, testul Shapiro-Wilk a respins normalitatea în peste 90% din repetări, deși un test t ar fi funcționat perfect pe asemenea date. Tocmai în eșantioanele mari, unde teorema limitei centrale ne protejează, testul ne alarmează cel mai des.

În eșantioane mici (sub 20-30 de observații), testele au putere mică și nu detectează nici abateri importante. Am extras din mg2026 trei eșantioane aleatoare de câte 12 valori ale trigliceridelor, o variabilă pe care o știm asimetrică. Shapiro-Wilk a dat p = 0,084, p = 0,313 și p = 0,259: niciun eșantion nu a "picat" testul, deși unul dintre ele avea asimetria 1,40. Tocmai în eșantioanele mici, unde normalitatea contează cel mai mult, testul ne liniștește pe nedrept.

Mai rău, o valoare p > 0,05 nu dovedește normalitatea. Ea spune doar că, cu datele disponibile, nu putem respinge ipoteza. Formulări ca "testul Shapiro-Wilk a confirmat normalitatea" sunt greșite și apar frecvent în lucrări de licență și chiar în articole. Formularea corectă este "testul Shapiro-Wilk nu a indicat o abatere semnificativă de la normalitate (p = 0,303)".

Concluzia practică: folosiți testele ca pe un indiciu, nu ca pe un verdict. Priviți întotdeauna graficele. Pentru n mare, judecați mai ales după grafic, asimetrie și diferența medie-mediană; pentru n mic, fiți prudenți chiar și atunci când testul nu este semnificativ și luați în considerare ce

ştiți despre variabilă din alte studii (de exemplu, că duratele de spitalizare sunt aproape întotdeauna asimetrice).

> **Studiu de caz: abaterea standard care "trădează" asimetria**
>
> Cioca și colaboratorii au comparat evoluția COVID-19 la 486 de pacienți: 200 fără diabet, 62 cu diabet de tip 1 și 224 cu diabet de tip 2 [8]. Durata spitalizării a fost raportată ca medie ± AS: 11,6 ± 7,0 zile pentru diabetul de tip 2 și 9,1 ± 5,8 zile pentru tipul 1. Aplicați testul "m - 2s": 11,6 - 2 x 7,0 = -2,4 zile și 9,1 - 2 x 5,8 = -2,5 zile. Dacă duratele ar fi normale, aproximativ 5% dintre pacienți ar fi stat în spital un număr negativ de zile, ceea ce este imposibil. Prin urmare, distribuția este asimetrică spre dreapta, cum se întâmplă aproape întotdeauna cu duratele de spitalizare: majoritatea pacienților stau o perioadă tipică, iar câțiva, cu complicații, foarte mult. Mediana și intervalul intercuartilic ar fi descris mai fidel datele, iar comparația ar fi trebuit făcută cu un test neparametric sau pe scală logaritmică. Același tipar apare în studiul lui Popa și colaboratorii asupra biomarkerilor în accidentul vascular cerebral, la 85 de pacienți [61]: GFAP a fost $142{,}91 \pm 102{,}19$ pg/mL la pacienții cu AVC, iar UCH $-L_1$ $1307{,}68 \pm 967{,}54$ pg/mL, abateri standard de 70-75% din medie, cu m - 2s mult sub zero. Pentru biomarkerii serici, o abatere standard apropiată de medie este aproape întotdeauna semnul unei distribuții log-normale. Nu este vorba de o greșeală de calcul, ci de o alegere de raportare care merită discutată: media ± AS descrie bine doar datele simetrice.

3.9 Transformarea datelor

3.9.1 Ideea transformării

Când o variabilă este asimetrică, avem trei opțiuni: folosim metode care nu presupun normalitatea (metodele neparametrice, capitolul 7), transformăm variabila astfel încât să devină aproximativ normală și aplicăm apoi metodele parametrice, sau folosim modele mai avansate care acceptă direct alte distribuții. Transformarea înseamnă să aplicăm aceeași funcție matematică fiecărei valori: logaritm, radical, inversul. Ordinea valorilor se păstrează (cine avea trigliceride mai mari le are în continuare mai mari), dar distanțele dintre ele se modifică: valorile mari sunt "comprimate" mai mult decât cele mici, iar coada lungă se scurtează [42,53].

3.9.2 Transformarea logaritmică

Transformarea logaritmică este cea mai folosită în medicină, pentru că multe procese biologice sunt multiplicative: o concentrație crește "de două ori", nu "cu 50 de unități". Se poate folosi logaritmul natural (ln, baza e) sau cel zecimal (log10); forma distribuției rezultate este aceeași, doar scala diferă. În Google Sheets: =LN(R2) sau =LOG10(R2). Logaritmul cere valori strict pozitive; pentru variabile care pot fi zero (de exemplu numărul de zile de concediu medical) se folosește uneori ln(x + 1), cu prudență la interpretare.

Pentru trigliceridele din mg2026, logaritmarea a redus asimetria de la 0,95 la -0,28 și excesul de la 1,12 la -0,28; Shapiro-Wilk trece de la $p < 0{,}001$ la $p = 0{,}200$ (figura 3.6C). Pe scala logaritmică nu mai există nici valori extreme: niciun pacient nu are un scor z peste 2,3 în modul.

3.9.3 Alte transformări și "scara puterilor"

Transformările uzuale pot fi privite ca trepte pe o scară, de la cea mai blândă la cea mai puternică: radical $\left(\sqrt{x}\right)$, logaritm (ln x), inversul radicalului $\left(\frac{1}{\sqrt{x}}\right)$, inversul (1/x). Cu cât asimetria pozitivă este mai mare, cu atât urcăm mai sus pe scară. Pentru asimetrie negativă (rară în medicină, de exemplu scorurile unui test ușor, unde majoritatea obțin aproape punctajul maxim) se transformă mai întâi variabila "reflectată" (valoarea maximă + 1 - x). Transformarea radical este folosită tradițional pentru numărători (de exemplu numărul de colonii, numărul de leziuni), iar

transformarea arcsinus pentru proporții, deși astăzi acestea se analizează de obicei cu modele dedicate.

Tabelul 3.7. Efectul transformărilor asupra trigliceridelor (mg2026, n = 160).

Transformarea	Asimetria	Excesul	Shapiro-Wilk p
fără transformare	0,95	1,12	< 0,001
radical $\sqrt{x}$	0,32	-0,10	0,181
logaritm ln x	-0,28	-0,28	0,200
Box-Cox, λ = 0,22	-0,01	-	0,710
inversul 1/x	1,48	2,30	< 0,001

Radicalul și logaritmul sunt amândouă acceptabile, cu asimetrii de semne opuse și de mărime similară. Inversul este prea puternic: transformă o asimetrie pozitivă moderată într-una și mai mare (inversul inversează ordinea, iar valorile mici, de 44-60 mg/dL, devin acum coada lungă). Transformarea Box-Cox generalizează scara: caută automat exponentul λ care face datele cât mai apropiate de normalitate (λ = 1 înseamnă fără transformare, λ = 0,5 radical, λ = 0 logaritm, λ = -1 inversul). Pentru trigliceride, λ optim este 0,22, între radical și logaritm. În practică, alegem o transformare cu interpretare simplă (de obicei logaritmul) în locul unui exponent "exotic" ca 0,22, chiar dacă acesta dă un p ceva mai mare; un rezultat exprimat în "trigliceride la puterea 0,22" nu poate fi comunicat nimănui.

Transformarea nu rezolvă întotdeauna problema. IMC din mg2026 rămâne asimetric după logaritmare (asimetrie 0,89, Shapiro-Wilk p < 0,001), iar Box-Cox cere un exponent de -2,5, adică o transformare foarte puternică și fără interpretare clinică, după care Shapiro-Wilk dă p = 0,076. Asimetria IMC vine în bună parte din câțiva pacienți cu obezitate severă, nu dintr-un proces multiplicativ. Aici, alegerea mai bună este o metodă neparametrică sau raportarea medianei și a IQR.

3.9.4 Retransformarea: media geometrică

După ce am analizat datele pe scala logaritmică, rezultatele trebuie aduse înapoi pe scala originală, ca să le poată înțelege un clinician. Media logaritmilor trigliceridelor este 4,970. Funcția inversă logaritmului natural este exponențiala:

$$e^{4,970} = 144,0 \text{ mg/dL}$$

Această valoare nu este media aritmetică a trigliceridelor (161,18 mg/dL). Este media geometrică, adică rădăcina de ordinul n a produsului celor n valori. Pentru o distribuție log-normală, media geometrică este apropiată de mediană (151,5 mg/dL) și descrie mai bine valoarea "tipică" decât media aritmetică, trasă în sus de coada dreaptă.

Abaterea standard pe scala logaritmică (0,488) nu se retransformă într-o abatere standard. Exponențiala ei, $e^{0,488} = 1,63$, este un factor multiplicativ, numit abatere standard geometrică: aproximativ două treimi dintre pacienți au trigliceride între 144,0 / 1,63 și 144,0 x 1,63, adică între 88 și 235 mg/dL. Observați asimetria intervalului față de 144: 56 mg/dL sub, 91 mg/dL peste.

Intervalul de încredere pentru media geometrică se calculează pe scala logaritmică și se retransformă. Eroarea standard a mediei logaritmilor este 0,488 / √160 = 0,0386, iar valoarea critică t pentru 159 gl este 1,975, deci:

$4,9699 \pm 1,975 \times 0,03859 = 4,9699 \pm 0,0762$, adică 4,8937 până la 5,0461

Retransformând limitele, $e^{4,8937} = 133,44$ și $e^{5,0461} = 155,42$ mg/dL, deci IC 95% pentru media geometrică este 133,44-155,42 mg/dL. Am păstrat intenționat patru zecimale pe scala logaritmică: exponențiala amplifică micile erori, iar o rotunjire la două zecimale înainte de retransformare ar deplasa fiecare limită cu aproape 0,5 mg/dL. Regula practică este că pe scala

logaritmică se rotunjește doar la final, după retransformare. Intervalul retransformat nu este simetric în jurul lui 144,0, iar asta este firesc. Vom relua calculul în capitolul 4.

O ultimă precizare: o diferență pe scala logaritmică devine, după retransformare, un raport. Dacă media logaritmilor trigliceridelor ar fi cu 0,10 mai mare într-un grup decât în altul, retransformarea dă $e^{0,10} = 1,105$: media geometrică este cu aproximativ 10,5% mai mare, nu "cu 0,10 mg/dL mai mare". Acesta este motivul pentru care studiile care analizează biomarkeri pe scală logaritmică raportează rapoarte de medii geometrice.

3.9.5 Când alegem metode neparametrice în loc de transformare

Transformarea este utilă când există o transformare naturală, ușor de interpretat (logaritmul pentru concentrații), când vrem să estimăm parametri cu intervale de încredere și când urmează analize multivariate (regresie). Metodele neparametrice, bazate pe ranguri, sunt preferabile când nicio transformare simplă nu normalizează datele (IMC din mg2026), când variabila este ordinală (scoruri Likert, stadii), când eșantionul este mic și nu putem judeca forma distribuției, când există valori extreme pe care nu le putem exclude sau când datele au un "plafon" sau un "prag" (multe valori egale cu minimul sau maximul scalei). Prețul metodelor neparametrice este o putere ceva mai mică atunci când datele sunt totuși normale și faptul că ele compară distribuții sau mediane, nu medii. Algoritmul complet de alegere a testului este prezentat în capitolul 8.

> **Studiu de caz: date asimetrice raportate ca mediană (IQR)**
>
> Gherbon și colaboratorii au urmărit timp de un an 200 de pacienți cu diabet de tip 2 tratați cu inhibitori SGLT-2 [59]. Glicemia a scăzut de la 180,00 mg% (IQR 154,50-207,00) la 130,00 mg% (IQR 117,50-150,00), HbA1c de la 8,40% (IQR 7,98-9,15%) la 7,30% (IQR 6,90-7,95%), iar TAS de la 140,00 mmHg (130,00-160,00) la 120,00 mmHg (115,50-130,00), toate cu $p < 0,001$. Observați că autorii au ales mediana și IQR pentru variabilele asimetrice, dar media ± AS pentru circumferința abdominală ($107,05 \pm 16,39$ cm, apoi $102,50 \pm 15,11$ cm; $p = 0,042$), probabil distribuită aproape simetric. Este alegerea corectă: indicatorul se potrivește distribuției fiecărei variabile. Uitați-vă la quartilele glicemiei inițiale: 180 - 154,5 = 25,5 mg% sub mediană, 207 - 180 = 27 mg% peste, o ușoară asimetrie. Pentru TAS, Q1 = 130 se află cu 10 mmHg sub mediana de 140, iar Q3 = 160 cu 20 mmHg peste ea: distribuția este întinsă spre dreapta. Un singur detaliu de lectură critică: rezumatul numește aceste valori "medii", deși sunt mediane; o asemenea scăpare terminologică apare des și este bine să o recunoașteți. Frandeș și colaboratorii au procedat la fel într-un studiu transversal pe 304 adulți cu diabet, raportând scorurile de implicare digitală ca mediană [Q1-Q3]: 1,71 [1,71-2,39] față de 3,86 [3,86-4,43] [7]. Faptul că mediana coincide cu Q1 în ambele grupuri arată că cel puțin un sfert dintre participanți au exact valoarea minimă a grupului lor, o distribuție cu "podea" pentru care media ±AS ar fi fost înșelătoare; autorii au folosit, coerent, regresia beta, potrivită pentru scoruri mărginite. Tot cu mediane au fost comparate valorile indicelui TyG în trei fenotipuri endocrin-metabolice, la 387 de pacienți evaluați pentru disfuncție tiroidiană: 8,24, 8,74 și 8,87, $p < 0,001$ [67].

3.10 Valorile extreme

3.10.1 Cum le recunoaștem

O valoare extremă (engl. outlier) este o observație care se află neobișnuit de departe de restul datelor. Ea poate fi o eroare (de tastare, de unitate de măsură, de aparat), o valoare reală dar rară (un pacient cu obezitate severă într-un studiu de prevenție) sau semnul că distribuția are, de fapt, o coadă lungă. Primul pas este întotdeauna să o identificăm; al doilea, să înțelegem de unde vine. În

mg2026_brut.csv, valoarea TAS 1540 și înălțimea 1,72 sunt erori evidente (un zero în plus, metri în loc de centimetri), corectate în capitolul 2. Ce facem însă cu valorile plauzibile, dar depărtate?

Trei criterii simple sunt folosite curent:

- Scorul z: valorile cu |z| > 3 sunt suspecte. Criteriul presupune o distribuție aproximativ normală; într-o normală, doar 0,3% dintre valori depășesc 3 abateri standard. Are un defect: valoarea extremă însăși umflă media și abaterea standard, deci își "ascunde" parțial propria extremitate.
- Regula 1,5 x IQR (Tukey): se calculează "gardurile" Q1 - 1,5 x IQR și Q3 + 1,5 x IQR; valorile din afara lor sunt marcate ca extreme (punctele individuale din box-plot). Pentru valori foarte extreme se folosește 3 x IQR. Avantajul este că se bazează pe quartile, care nu sunt influențate de valorile extreme.
- Testul Grubbs: un test formal pentru o singură valoare extremă într-un eșantion presupus normal, cu statistica $G = \frac{|x-m|}{s}$ comparată cu o valoare critică ce depinde de n (pentru n = 160 și α = 0,05, aproximativ 3,54). Există și variante pentru mai multe valori extreme. Testul se aplică pentru datele aproape normale; pe o variabilă asimetrică va găsi "valori extreme" care sunt de fapt coada naturală a distribuției.

Tabelul 3.8. Valori extreme în setul mg2026 după criteriile scorului z și IQR.

Variabila	**Q1-Q3**	**Gardurile 1,5 x IQR**	**Valori în afara gardurilor**	**Valori cu z în modul > 3**	**Valoarea maximă și scorul ei z**
TAS0 (mmHg)	122-140	95-167	2 (168, 169)	0	169; 2,73
Vârsta (ani)	50,75-62	33,9-78,9	1 (79)	0	79; 2,59
Glicemie_0(mg/dl	93-116	58,5-150,5	1 (154)	1	154; 3,08
$\text{IMC}_0\left(\frac{\text{kg}}{m^2}\right)$	26,2-31,4	18,4-39,2	3 (40,0; 43,4; 48,3)	2	48,3; 4,92
Trigliceride (mg/dL)	107-203,25	-37,4-347,6	5 (364; 364; 373; 405; 442)	2	442; 3,68
ln(trigliceride)	4,67-5,31	3,71-6,28	0	0	6,09; 2,30

Remarcați că regula IQR marchează una-două valori chiar și la variabile normale, ca TAS și vârsta. Este de așteptat: și într-o distribuție perfect normală, aproximativ 0,7% dintre valori ies din gardurile Tukey, adică unul-doi pacienți din 160. Pentru IMC, statistica Grubbs a valorii 48,3 este 4,92, mult peste valoarea critică, iar pentru trigliceride valoarea 442 are G = 3,68, chiar peste prag. După logaritmare însă, trigliceridele nu mai au nicio valoare extremă: "valorile extreme" erau coada normală a unei distribuții log-normale.

3.10.2 Ce facem cu ele

Regula de aur: valorile extreme nu se șterg niciodată în tăcere. Ordinea pașilor este următoarea. Verificăm mai întâi dacă valoarea este o eroare, comparând cu documentul sursă (fișa pacientului, buletinul de laborator). Dacă este eroare și valoarea corectă poate fi recuperată, o corectăm și notăm corecția. Dacă este eroare și nu poate fi recuperată, o tratăm ca valoare lipsă și raportăm. Dacă valoarea este plauzibilă biologic, o păstrăm. Un IMC de 48,3 kg/m² este rar, dar există; un pacient cu obezitate de gradul III este exact tipul de pacient care face parte din populația studiată.

Pentru valorile plauzibile, alegem metode care le tolerează: mediană și IQR în loc de medie și AS, teste neparametrice, transformare logaritmică, metode robuste. Facem apoi o analiză de sensibilitate: repetăm analiza cu și fără valorile extreme și raportăm dacă concluzia se schimbă. Dacă excludem pacientul cu IMC 48,3, media IMC scade de la 29,04 la 28,92 kg/m², abaterea standard de la 3,91 la 3,61, iar asimetria de la 1,42 la 0,95; Shapiro-Wilk rămâne totuși semnificativ (p < 0,001). Excluderea unei singure valori nu "repară" distribuția, iar a continua să excludem valori până când testul devine nesemnificativ este o formă de manipulare a datelor. În

capitolul 9 veți vedea cât de mult poate schimba o singură valoare extremă un coeficient de corelație.

Orice excludere trebuie stabilită dinainte în protocol (de exemplu "valorile glicemiei peste 600 mg/dL vor fi verificate în fișa de laborator") și raportată în articol, cu numărul de observații excluse și motivul.

> **Studiu de caz: timpi, scoruri și teste bazate pe ranguri**
>
> Variabilele de timp din medicina de urgență sunt aproape întotdeauna asimetrice: nu pot fi negative, majoritatea pacienților se încadrează într-un interval tipic, iar câțiva întârzie foarte mult. Popa și colaboratorii au analizat 258 de pacienți eligibili cu AVC ischemic acut (din 355 evaluați), dintre care doar 46 au primit tromboliză intravenoasă, o rată de 17,83% [62]. Timpii au fost descriși prin mediane: o mediană de 100 de minute de la debut la sosirea în urgență ($p < 0,001$) și un timp ușă-CT de 15,5 minute ($p = 0,009$). Alegerea medianei este corectă aici: o medie ar fi fost trasă în sus de cei câțiva pacienți ajunși după multe ore. În boli infecțioase, Marin și colaboratorii au comparat markerii imunologici între trei tipuri de răspuns la terapia antiretrovirală, la 462 de adulți cu HIV, folosind testul Kruskal-Wallis, cu $H = 153,62$ ($p < 0,001$, $\varepsilon^2 = 0,33$) și $H = 115,10$ ($p < 0,001$, $\varepsilon^2 = 0,25$) [68]. Raportul CD4/CD8, un raport de două numărători celulare, are prin natura lui o distribuție asimetrică, iar testul bazat pe ranguri, însoțit de o mărime a efectului (epsilon pătrat), este alegerea firească. În același spirit, Gherbon și colaboratorii au folosit testul Kruskal-Wallis ($p = 0,003$) pentru scorurile de distres diabetic la 453 de adulți cu diabet de tip 2, dintre care 19 (4,2%) aveau distres sever [69]: scorurile unui chestionar sunt ordinale și adesea asimetrice, iar un test care compară mediile ar fi fost greu de justificat. Ce merită observat în toate trei studiile: alegerea metodei urmează forma distribuției și tipul variabilei, nu obișnuința.

3.11 Distribuții care se suprapun: de la scorul z la deciziile clinice

Până acum am privit o singură distribuție. În diagnostic, avem de obicei două: distribuția unui marker la pacienții cu boală și distribuția lui la cei fără boală. Cât de bine separă markerul cele două grupuri depinde de distanța dintre medii, măsurată în abateri standard, adică exact de ideea scorului z.

Lupușoru și colaboratorii au evaluat ecografia cu contrast cu imagistică parametrică (CEUS-PAT) pentru diagnosticul cirozei, la 64 de persoane: 37 de pacienți cu ciroză și 27 de voluntari sănătoși [2]. Indicele măsurat a fost 0,83 ± 0,09 la pacienții cu ciroză și 0,49 ± 0,11 la voluntari ($p < 0,0001$). Raportarea ca medie ± AS este potrivită aici, pentru că intervalele $m \pm 2s$ rămân în domeniul plauzibil al indicelui. Diferența dintre medii, 0,34, reprezintă peste trei abateri standard: cele două distribuții aproape nu se suprapun. Nu e de mirare că pragul > 0,7 a dat o arie sub curba ROC de 0,98, sensibilitate 89,19% și specificitate 100%. Vom verifica în problema rezolvată 3.6 că modelul normal, aplicat celor două medii și abateri standard, prezice cifre foarte apropiate. Ce nu spune studiul este cât de bine ar funcționa indicele la pacienți cu fibroză intermediară: comparația între bolnavi clari și voluntari sănătoși maximizează distanța dintre distribuții, fenomen numit bias de spectru, pe care îl vom discuta în capitolul 11.

Bende și colaboratorii au măsurat parametrul de atenuare UGAP la 179 de pacienți, comparat cu CAP [58]. Mediile au crescut de la un grad de steatoză la altul: S_0 198,3 $\pm \frac{25,7 d_B}{m}$, S1 216,86 ± 26,3, S2 237,79 ± 26,3, S_3 270,8 $\pm$ 31,62($p < 0,001$). Diferența dintre S0 și S1 este de doar 18,6 dB/m, mai puțin decât o abatere standard: distribuțiile celor două grade se suprapun mult. Diferența dintre S0 și S3 este de peste 70 dB/m, adică în jur de 2,5 abateri standard. Ariile ROC raportate urmează exact această logică: 0,83 pentru pragul S1 (192,5 dB/m), 0,90 pentru S2 (231

dB/m) și 0,91 pentru S3 (248 dB/m). Cu cât distribuțiile sunt mai depărtate, cu atât un prag le separă mai bine.

> **Studiu de caz: distribuția unei diferențe**
>
> Muntean și colaboratorii au urmărit 921 de studenți în primul an de facultate și au măsurat modificarea IMC între începutul și sfârșitul anului [5]. Diferența medie a fost -0,45 ± 1,07 kg/m² (IC 95% -0,52 până la -0,38; p < 0,001). Uitați-vă la cele două cifre: abaterea standard (1,07) este de peste două ori mai mare decât media în valoare absolută (0,45). Asta înseamnă că modificările individuale au fost foarte variate și s-au distribuit de ambele părți ale lui zero, deși media scade clar. Rezumatul confirmă: 48,2% dintre studenți au avut o scădere a IMC, 38,7% au rămas stabili, iar 13,1% au crescut în greutate, cu o creștere semnificativă clinic la 7,2%. O medie a diferențelor descrie grupul, nu individul; un studiu care ar raporta doar "IMC a scăzut în medie cu 0,45 kg/m²" ar ascunde faptul că unul din opt studenți a luat în greutate. Distribuția diferențelor, nu a valorilor inițiale sau finale, este cea care contează pentru testul t pereche (capitolul 6). Iar faptul că autorii au analizat aceeași variabilă și ca diferență continuă (regresie liniară), și dihotomizată ca "creștere semnificativă clinic" (regresie logistică) arată că o distribuție continuă poate fi privită din două unghiuri, fiecare cu întrebarea lui.

3.12 Probleme rezolvate

Problema rezolvată 3.1. Procente sub curba normală pentru TAS

Enunț. În setul mg2026, TAS inițială are media 131,13 mmHg și abaterea standard 13,88 mmHg, iar distribuția este compatibilă cu normalitatea (Shapiro-Wilk p = 0,303). (a) Ce proporție de pacienți ar trebui, teoretic, să aibă TAS între 110 și 150 mmHg? (b) Comparați cu proporția observată, 83,8%. (c) Ce valoare a TAS separă cei 10% pacienți cu valorile cele mai mici?

Rezolvare. Verificăm întâi condiția: modelul normal se poate folosi, pentru că asimetria (0,01) și excesul (-0,10) sunt practic nule, iar testele de normalitate nu sunt semnificative.

(a) Calculăm scorurile z ale celor două limite:

$$z_1 = \frac{110 - 131{,}13}{13{,}88} = -\frac{21{,}13}{13{,}88} = -1{,}52 z_2 = \frac{150 - 131{,}13}{13{,}88} = \frac{18{,}87}{13{,}88} = 1{,}36$$

Din tabelul E.1: Φ(1,36) = 0,9131 și Φ(1,52) = 0,9357, deci Φ(-1,52) = 1 - 0,9357 = 0,0643. Proporția dintre cele două limite este:

$$P(110 \leq \text{TAS} \leq 150) = \Phi(1{,}36) - \Phi(-1{,}52) = 0{,}9131 - 0{,}0643 = 0{,}8488$$

Teoretic, aproximativ 84,9% dintre pacienți au TAS între 110 și 150 mmHg. În Google Sheets: =NORM.DIST(150; 131,13; 13,88; TRUE) - NORM.DIST(110; 131,13; 13,88; TRUE).

(b) Proporția observată este 83,8% (134 de pacienți din 160). Diferența de 1,1 puncte procentuale este de mărimea fluctuațiilor de eșantionare așteptate pentru n = 160. Modelul normal descrie bine datele.

(c) Căutăm z pentru care Φ(z) = 0,10. Prin simetrie, este opusul valorii pentru care Φ(z) = 0,90, adică z = -1,28. Atunci x = 131,13 - 1,28 x 13,88 = 131,13 - 17,77 = 113,4 mmHg. În Sheets: =NORM.INV(0,1; 131,13; 13,88) dă 113,3 mmHg (diferența vine din rotunjirea lui z). Percentila 10 observată în eșantion este 112,0 mmHg.

Interpretare clinică. Aproape cinci din șase pacienți cu risc cardiometabolic din acest eșantion au TAS între 110 și 150 mmHg, iar unul din zece are sub 113 mmHg. Cifrele descriu eșantionul și nu înlocuiesc pragurile clinice ale ghidurilor.

Problema rezolvată 3.2. Distribuția binomială la screeningul diabetului

Enunț. Prevalența diabetului nediagnosticat într-o populație de adulți cu risc cardiometabolic este 18%. Un medic de familie trimite la TTGO 20 de pacienți aleși la întâmplare. (a) Care este numărul așteptat de cazuri și abaterea standard? (b) Care este probabilitatea să nu găsească niciun caz? (c) Dar probabilitatea să găsească 5 sau mai multe? (d) Medicul găsește 8 cazuri. Este un rezultat surprinzător?

Rezolvare. Condițiile binomialei sunt îndeplinite: n = 20 încercări independente, două rezultate posibile, p = 0,18 constant.

(a) Media = n x p = 20 x 0,18 = 3,6 cazuri. Varianța = 20 x 0,18 x 0,82 = 2,952, iar abaterea standard = √2,952 = 1,72 cazuri.

(b) $P(X = 0) = 0{,}82^{20}$. Calculăm pas cu pas: 0,82² = 0,6724; 0,82^4 = 0,4521; 0,82^8 = 0,2044; 0,82^16 = 0,0418; $0{,}82^{20} = 0{,}0418 \times 0{,}4521 = 0{,}0189$. Probabilitatea de a nu găsi niciun caz este 1,9%.

(c) P(X ≥ 5) = 1 - P(X ≤ 4). Calculul de mână al celor cinci termeni este laborios; în Google Sheets folosim =1 - BINOM.DIST(4; 20; 0,18; TRUE), care dă 0,2849. Aproximativ în 28% din grupurile de câte 20 de pacienți vom găsi cel puțin 5 cazuri.

(d) P(X ≥ 8) = 1 - BINOM.DIST(7; 20; 0,18; TRUE) = 0,0177. Un asemenea rezultat apare în mai puțin de 2% din grupuri dacă prevalența reală este 18%. Este deci neobișnuit și sugerează fie că pacienții acestui medic au un risc mai mare decât populația de referință (de exemplu, a selectat pacienți mai vârstnici sau mai obezi), fie o întâmplare rară. Observați că am făcut, fără să-i spunem pe nume, un test statistic: am calculat probabilitatea unui rezultat cel puțin la fel de extrem sub o ipoteză dată. Este exact logica valorii p din capitolul 5.

Problema rezolvată 3.3. Distribuția Poisson și pragul de alarmă

Enunț. Într-un centru de vaccinare se raportează în medie 1,5 reacții adverse minore pe lună. (a) Care este probabilitatea unei luni fără nicio reacție? (b) Echipa vrea un prag de alarmă astfel încât alarmele false să apară în cel mult 5% din luni. Care este cel mai mic număr de reacții pentru care se declanșează alarma?

Rezolvare. Evenimente rare, numărate pe un interval de timp, cu medie cunoscută: distribuția Poisson cu λ = 1,5.

(a) $P(X = 0) = e^{-1{,}5} = 0{,}2231$. Aproximativ o lună din patru-cinci nu are nicio reacție raportată.

(b) Calculăm probabilitățile succesive cu formula $P(X = k) = \frac{e^{-\lambda} \times \lambda^k}{k!}$:

$$P(1) = 0{,}2231 \times 1{,}5 = 0{,}3347 P(2) = \frac{0{,}2231 \times 1{,}5^2}{2} = 0{,}2510 P(3) = \frac{0{,}2231 \times 1{,}5^3}{6}$$
$$= 0{,}1255 P(4) = \frac{0{,}2231 \times 1{,}5^4}{24} = 0{,}0471$$

P(X ≤ 3) = 0,2231 + 0,3347 + 0,2510 + 0,1255 = 0,9343, deci P(X ≥ 4) = 0,0657 (6,6%), peste limita de 5%. P(X ≤ 4) = 0,9814, deci P(X ≥ 5) = 0,0186 (1,9%). Cel mai mic prag cu alarme false sub 5% este 5 reacții pe lună. În Sheets: =1 - POISSON.DIST(4; 1,5; TRUE) = 0,0186.

Interpretare. Un prag de 4 reacții ar declanșa o alarmă falsă în aproximativ o lună din cincisprezece; pragul de 5, în aproximativ o lună din cincizeci. Alegerea pragului este un compromis: un prag mai mic detectează mai repede o problemă reală, dar produce mai multe

alarme false. Aceeași dilemă o vom regăsi la testele de diagnostic (sensibilitate față de specificitate).

Problema rezolvată 3.4. Un interval de referință și analizele multiple

Enunț. Într-un eșantion de 400 de femei sănătoase, hemoglobina are media 13,8 g/dL și abaterea standard 1,0 g/dL, cu distribuție aproximativ normală (date fictive). (a) Construiți intervalul de referință de 95%. (b) O pacientă are 11,6 g/dL. Calculați scorul z și percentila. (c) O femeie sănătoasă face 12 analize independente, fiecare cu interval de referință de 95%. Care este probabilitatea ca toate să fie "în limite normale"?

Rezolvare. $(a)\ m \pm 1{,}96s = 13{,}8 \pm 1{,}96 \times 1{,}0 = 13{,}8 \pm 1{,}96$, adică 11,8-15,8 g/dL (rotunjit la o zecimală).

(b) $z = \frac{11{,}6-13{,}8}{1{,}0} = -2{,}20$. Din tabel, Φ(2,20) = 0,9861, deci Φ(-2,20) = 1 - 0,9861 = 0,0139. Pacienta se află la percentila 1,4: doar 1,4% dintre femeile sănătoase din populația de referință au o valoare atât de mică sau mai mică. Valoarea este neobișnuită și justifică investigarea unei anemii, dar statistica singură nu pune diagnosticul.

(c) $P(\text{ toate } 12 \text{ normale }) = 0{,}95^{12} = 0{,}540$. Aproape jumătate (46%) dintre femeile sănătoase care fac 12 analize independente vor avea cel puțin un rezultat în afara intervalului de referință. De aici sfatul practic: un rezultat izolat, ușor în afara intervalului, la o persoană fără simptome, se repetă și se interpretează în context înainte de orice concluzie.

Problema rezolvată 3.5. Alegerea unei transformări pentru trigliceride

Enunț. Trigliceridele din mg2026 au media 161,18 mg/dL, mediana 151,5 mg/dL, abaterea standard 76,30 mg/dL, asimetria 0,95, excesul 1,12 și Shapiro-Wilk $p < 0{,}001$. (a) Argumentați de ce distribuția nu este normală. (b) Alegeți o transformare pe baza tabelului 3.7. (c) Calculați media geometrică și un interval în care se află aproximativ 95% dintre pacienți.

Rezolvare. (a) Patru indicii converg. Asimetria împărțită la eroarea ei standard este 0,95 / 0,192 = 4,95, mult peste 2. Excesul raportat la eroarea lui standard este 1,12 / 0,381 = 2,94, tot peste 2. Media depășește mediana cu aproape 10 mg/dL. Testul Shapiro-Wilk respinge normalitatea. Graficul Q-Q (figura 3.7B) are forma curbată tipică asimetriei pozitive.

(b) Din tabelul 3.7, radicalul (asimetrie 0,32; p = 0,181) și logaritmul (asimetrie -0,28; p = 0,200) normalizează ambele suficient. Inversul este prea puternic. Box-Cox dă cel mai bun rezultat numeric (λ = 0,22), dar fără interpretare clinică. Alegem logaritmul: este la fel de eficient ca radicalul, are sens biologic (concentrațiile variază multiplicativ) și permite retransformarea în medie geometrică și în rapoarte.

(c) Media logaritmilor este 4,970, abaterea lor standard 0,488. Media geometrică este $e^{4{,}970} = 144{,}0$ mg/dL. Intervalul de 95% pe scala logaritmică este $4{,}970 \pm 1{,}96 \times 0{,}488 = 4{,}970 \pm 0{,}957$, adică 4,013-5,927. Retransformat: $e^{4{,}013} = 55{,}3$ și $e^{5{,}927} = 374{,}9$ mg/dL. Aproximativ 95% dintre pacienți au trigliceride între 55 și 375 mg/dL; în eșantion, 95,6% dintre valori se află în acest interval, iar percentilele 2,5 și 97,5 observate sunt 53,8 și 364,0 mg/dL.

Interpretare. Pe scala originală, intervalul nu este simetric în jurul mediei geometrice (89 mg/dL sub, 231 mg/dL peste), exact cum arată histograma. Intervalul calculat naiv, 11,6-310,7 mg/dL, ar fi fost greșit la ambele capete.

Problema rezolvată 3.6. Modelul normal aplicat unor date publicate

Enunț. Într-un studiu de diagnostic al cirozei [2], indicele CEUS-PAT a fost 0,83 ± 0,09 la 37 de pacienți cu ciroză și 0,49 ± 0,11 la 27 de voluntari sănătoși. Autorii au raportat, la pragul > 0,7, o sensibilitate de 89,19% și o specificitate de 100%. Presupunând că indicele este normal distribuit în fiecare grup, ce sensibilitate și specificitate ar prezice modelul?

Rezolvare. Sensibilitatea este proporția bolnavilor cu valoare peste prag. La pacienții cu ciroză:

$z = \frac{0,7-0,83}{0,09} = -1,44$; Φ(-1,44) = 1 - 0,9251 = 0,0749

Proporția peste prag este 1 - 0,0749 = 0,925, deci sensibilitatea prezisă este 92,5%.

Specificitatea este proporția celor fără boală cu valoare sub prag. La voluntari:

$z = \frac{0,7-0,49}{0,11} = 1,91$; Φ(1,91) = 0,9719

Specificitatea prezisă este 97,2%.

Interpretare. Modelul normal prezice 92,5% și 97,2%; studiul a observat 89,19% și 100%. Concordanța este bună, ținând cont de eșantioanele mici: 100% din 27 de voluntari înseamnă că niciunul nu a depășit pragul, iar modelul spune că ne-am aștepta la 0,8 voluntari, adică la zero sau unul. Exercițiul arată că, dacă știm media și abaterea standard în două grupuri și distribuțiile sunt normale, putem anticipa performanța unui prag. Arată și limita: rezultatul depinde de presupunerea de normalitate și de compoziția grupurilor (aici, bolnavi certi față de voluntari sănătoși, ceea ce exagerează separarea).

Problema rezolvată 3.7. Decizia de normalitate pentru HbA1c

Enunț. Pentru HbA1c inițială din mg2026: n = 160, media 5,58%, abaterea standard 0,45%, mediana 5,6%, asimetria 0,14, excesul 0,16, Shapiro-Wilk W = 0,990, p = 0,304. (a) Este rezonabil modelul normal? (b) Câți pacienți ar trebui să aibă, teoretic, HbA1c ≥ 6,5%? În eșantion sunt 4.

Rezolvare. (a) Raportul asimetrie / ES $= \frac{0,14}{0,192} = 0,73$; raportul exces / ES $= \frac{0,16}{0,381} = 0,42$; ambele mult sub 2. Media și mediana sunt practic egale. Testul Shapiro-Wilk nu indică o abatere semnificativă de la normalitate. Cu 160 de observații, testul are o putere bună, deci rezultatul nesemnificativ este informativ. Modelul normal este rezonabil.

(b) $z = \frac{6,5-5,58}{0,45} = 2,04$; Φ(2,04) = 0,9793; proporția teoretică peste prag este 2,1% (cu z exact, 1,9%), adică aproximativ 3 pacienți din 160. Observăm 4 (2,5%). Concordanță foarte bună.

Interpretare. Faptul că doar 4 pacienți au HbA1c ≥ 6,5%, deși 29 au diabet confirmat prin TTGO, nu este o contradicție statistică, ci una clinică: HbA1c și glicemia a jeun identifică persoane parțial diferite, iar diabetul nediagnosticat este adesea în stadii incipiente. Corelația dintre glicemie și HbA1c va fi studiată în capitolul 9.

3.13 Exerciții propuse

1. TAS din mg2026 are media 131,13 mmHg și AS 13,88 mmHg. Calculați scorurile z pentru valorile 125 și 150 mmHg și percentilele corespunzătoare, folosind tabelul E.1 și apoi funcția NORM.S.DIST.
2. Colesterolul total are media 209,39 mg/dL și AS 36,91 mg/dL (Shapiro-Wilk p = 0,927). Calculați proporția teoretică a pacienților cu colesterol ≥ 240 mg/dL și cu colesterol < 190 mg/dL. În eșantion, proporțiile observate sunt 23,8% și 30,0%. Comentați.

3. HDL are media 49,52 mg/dL și AS 11,78 mg/dL. Construiți intervalul $m \pm 1{,}96s$. Ce proporție de pacienți ar trebui să conțină? Calculați proporția teoretică a celor cu HDL < 40 mg/dL (observat: 19,4%).
4. În populația din care provine mg2026, proporția fumătorilor activi este aproximativ 0,33. Pentru un grup de 6 pacienți aleși la întâmplare, calculați: media și varianța numărului de fumători, probabilitatea să nu existe niciun fumător și probabilitatea să existe cel puțin 2.
5. Într-o secție se înregistrează în medie 0,5 evenimente adverse grave pe lună. Calculați probabilitatea ca într-o lună să apară cel puțin un eveniment și probabilitatea să apară 3 sau mai multe.
6. Într-o populație de referință sănătoasă, potasemia are media 4,3 mmol/L și AS 0,35 mmol/L, cu distribuție normală (date fictive). Construiți intervalul de referință de 95%. Un pacient are 5,1 mmol/L: calculați scorul z și percentila.
7. Într-un articol, durata spitalizării la pacienții cu diabet de tip 1 este raportată ca 9,1 ± 5,8 zile. Fără alte informații, ce puteți spune despre forma distribuției? Ce indicatori ar fi fost mai potriviți?
8. Pentru scorul FINDRISC din mg2026 (n = 160): media 10,37, mediana 10, asimetria 0,39, excesul 0,74, Shapiro-Wilk p = 0,034. Calculați rapoartele asimetrie / ES și exces / ES (ES 0,192 și 0,381) și argumentați ce indicatori descriptivi și ce tip de test ați alege, ținând cont că scorul ia doar valori întregi între 0 și 26.
9. Media logaritmilor naturali ai unei variabile este 4,97, iar abaterea lor standard 0,49. Calculați media geometrică, abaterea standard geometrică (ca factor) și intervalul în care se află aproximativ 95% din valori pe scala originală.
10. Glicemia inițială din mg2026 are Q1 = 93 și $Q_3 = 116$ mg/dL, media 105,44 și AS 15,76 mg/dL; valoarea maximă este 154 mg/dL. Este 154 o valoare extremă după regula 1,5 x IQR? Dar după criteriul |z| > 3? Ce ați face cu această valoare?
11. Descrieți cum ar arăta graficul Q-Q pentru: (a) o variabilă normală; (b) durata spitalizării; (c) o variabilă cu cozi mai groase decât normala; (d) o variabilă normală cu o singură eroare de tastare (un zero în plus).

Capcane frecvente

- A crede că un test de normalitate nesemnificativ "dovedește" normalitatea. Înseamnă doar că nu am găsit dovezi împotriva ei, iar în eșantioane mici testul are putere redusă.
- A respinge automat metodele parametrice într-un eșantion mare doar pentru că Shapiro-Wilk dă p < 0,05. Priviți graficul, mărimea asimetriei și analiza care urmează.
- A testa normalitatea pe tot eșantionul când testul cere normalitate în fiecare grup sau normalitatea diferențelor.
- A confunda funcția de repartiție cu densitatea: în Google Sheets, NORM.DIST cu argumentul FALSE nu dă o probabilitate.
- A crede că standardizarea (scorurile z) transformă o distribuție asimetrică într-una normală.
- A raporta media ± AS pentru variabile la care m - 2s este negativ, deși variabila nu poate fi negativă.
- A retransforma abaterea standard calculată pe scală logaritmică ca și cum ar fi o abatere standard pe scala originală, sau a rotunji limitele intervalului pe scala logaritmică înainte de retransformare.
- A șterge valorile extreme fără verificare și fără raportare, sau a exclude valori până când testul de normalitate devine nesemnificativ.
- A folosi testul Kolmogorov-Smirnov clasic, fără corecția Lilliefors, când media și abaterea standard sunt estimate din aceleași date.
- A interpreta un rezultat de laborator ușor în afara intervalului de referință ca dovadă de boală.

De reținut

- Distribuția normală este determinată complet de μ și σ; este simetrică, media, mediana și modul (moda) coincid, iar 68,3%, 95,4% și 99,7% dintre valori se află la 1, 2 și 3 abateri standard de medie (95% la 1,96 abateri standard).
- Scorul z = (x - m) / s exprimă distanța față de medie în abateri standard; tabelul E.1 sau funcția NORM.S.DIST îl transformă în percentilă. În mg2026, TAS de 160 mmHg are z = 2,08 (percentila 98).
- Distribuția binomială (n încercări, probabilitatea p, medie n x p) și distribuția Poisson (evenimente rare, medie = varianță = λ) descriu numărători.
- Intervalul de referință $m \pm 1{,}96s$ lasă, prin construcție, 5% dintre persoanele sănătoase în afara lui; pentru variabilele asimetrice se folosesc percentile sau scala logaritmică.
- Distribuțiile t, χ^2 și F sunt distribuții de eșantionare, cu forme care depind de gradele de libertate; teorema limitei centrale face ca mediile să fie aproape normale în eșantioane suficient de mari.
- Normalitatea se judecă prin grafice, indicatori (asimetria și excesul raportate la erorile lor standard, media față de mediană) și teste (Shapiro-Wilk, Lilliefors, Anderson-Darling), împreună.
- Transformarea logaritmică normalizează adesea concentrațiile serice; media retransformată este media geometrică (trigliceride: 144,0 mg/dL), iar diferențele devin rapoarte.
- Valorile extreme se identifică (|z| > 3, 1,5 x IQR, Grubbs), se verifică la sursă și se raportează; nu se șterg în tăcere.

Întrebări de autoevaluare

1. (CS) Ce proporție din valorile unei distribuții normale se află între μ - 2σ și μ + 2σ?
    a) 68,3%
    b) 90,0%
    c) 95,4%
    d) 99,7%
    e) 50,0%
2. (CS) În setul mg2026, TAS are media 131,13 mmHg și abaterea standard 13,88 mmHg. Scorul z al unei valori de 160 mmHg este:
    a) 1,64
    b) 2,08
    c) 2,58
    d) 28,87
    e) -2,08
3. (CS) Care dintre următoarele teste este recomandat în primul rând pentru verificarea normalității în eșantioane mici și medii?
    a) testul chi-pătrat de asociere
    b) testul Shapiro-Wilk
    c) testul t pentru eșantioane independente
    d) testul Levene
    e) testul Mann-Whitney
4. (CS) Un articol raportează durata spitalizării ca 11,6 ± 7,0 zile. Cea mai probabilă formă a distribuției este:
    a) normală, deoarece este raportată media
    b) asimetrică spre dreapta (asimetrie pozitivă)
    c) asimetrică spre stânga (asimetrie negativă)

d) uniformă

e) cu exces negativ și fără asimetrie

5. (CS) Dacă se calculează media logaritmilor naturali ai unei variabile și apoi se aplică funcția exponențială, se obține:
   a) media aritmetică
   b) mediana exactă
   c) media geometrică
   d) media armonică
   e) modul
6. (CM) Despre distribuția normală sunt adevărate afirmațiile:
   a) este simetrică față de medie
   b) media, mediana și modul coincid
   c) este determinată complet de medie și de abaterea standard
   d) cozile curbei ating axa orizontală la $\mu \pm 3\sigma$
   e) aria totală de sub curbă este egală cu 1
7. (CM) Pe un eșantion de 12 pacienți, testul Shapiro-Wilk dă p = 0,30. Sunt corecte afirmațiile:
   a) testul dovedește că datele sunt normale
   b) nu se respinge ipoteza de normalitate
   c) testul poate avea putere mică la un eșantion atât de mic
   d) este obligatoriu să se folosească un test parametric
   e) trebuie examinate și graficele (histogramă, Q-Q)
8. (CM) Care dintre următoarele indică o asimetrie pozitivă?
   a) media este mai mare decât mediana
   b) raportul asimetrie / eroare standard este mai mare de 2
   c) pe graficul Q-Q, punctele se curbează în sus la capătul din dreapta
   d) pentru o variabilă care nu poate fi negativă, m - 2s este negativ
   e) excesul este egal cu zero
9. (CM) Despre valorile extreme sunt corecte afirmațiile:
   a) se șterg automat dacă scorul lor z depășește 3 în modul
   b) se verifică în documentul sursă înainte de orice decizie
   c) se recomandă o analiză de sensibilitate, cu și fără ele
   d) regula 1,5 x IQR se bazează pe quartile
   e) testul Grubbs presupune o distribuție aproximativ normală
10. (CM) Despre un interval de referință construit ca $m \pm 1{,}96s$ pe o populație sănătoasă, sunt corecte afirmațiile:

a) conține aproximativ 95% dintre persoanele sănătoase

b) aproximativ 5% dintre persoanele sănătoase au valori în afara lui

c) o valoare în afara intervalului dovedește boala

d) pentru variabilele asimetrice se pot folosi percentilele 2,5 și 97,5 sau scala logaritmică

e) la 20 de analize independente, aproximativ 64% dintre persoanele sănătoase vor avea cel puțin un rezultat în afara intervalului

Capitolul 4. Estimarea parametrilor și intervalele de încredere; principii de calcul al dimensiunii eșantionului

Obiective

La finalul acestui capitol, studentul va putea:

- să deosebească estimarea punctuală de estimarea prin interval și să explice proprietățile unui estimator bun (nedeplasare, consistență, eficiență);
- să calculeze eroarea standard a mediei și să nu o confunde cu abaterea standard;
- să construiască, pe aceleași date, intervalele de 68,3%, 95,4% și 99,7% pentru valorile individuale și pentru medie și să explice de ce diferă atât de mult;
- să calculeze de mână un interval de încredere pentru medie cu distribuția t și să-l verifice în JASP, MedCalc și Google Sheets;
- să interpreteze corect un interval de încredere și să recunoască interpretările greșite;
- să calculeze intervalul de încredere pentru o proporție prin metodele Wald și Wilson și să aplice regula lui trei pentru zero evenimente;
- să construiască intervale de încredere pentru diferența a două medii și a două proporții și să citească intervalele pentru OR, RR, AUC și pentru date asimetrice;
- să calculeze dimensiunea eșantionului pentru estimarea unei medii, a unei proporții și pentru compararea a două medii, ținând cont de putere, de mărimea efectului și de pierderile din studiu;
- să citească critic intervalele de încredere raportate în articolele publicate.

4.1 Estimarea punctuală și estimarea prin interval

4.1.1 Parametri, estimatori, estimări

În capitolul 1 am deosebit populația de eșantion. Populația are parametri, valori fixe, dar necunoscute: media tensiunii sistolice μ a tuturor adulților cu risc cardiometabolic din vestul țării, proporția π a celor cu diabet nediagnosticat. Eșantionul ne dă statistici, calculate din date: media m, proporția p. Când folosim o statistică pentru a aproxima un parametru, o numim estimator, iar valoarea concretă obținută pe eșantionul nostru se numește estimare.

Estimarea punctuală dă un singur număr: "media TAS în populație este estimată la 131,13 mmHg". Este simplu, dar incomplet. Dacă am repeta studiul pe alți 160 de pacienți, am obține aproape sigur altă medie, poate 129,8 sau 132,5 mmHg. Un singur număr nu spune nimic despre cât de mult ne putem baza pe el. Estimarea prin interval completează informația: "media TAS în populație se află, cu o încredere de 95%, între 128,96 și 133,30 mmHg". Intervalul exprimă precizia estimării, iar lățimea lui ne spune cât de mult știm, de fapt [42,53].

Ghidurile actuale de raportare a studiilor cer ca rezultatele principale să fie prezentate ca estimări cu intervale de încredere, nu doar ca valori p. Motivul va deveni clar până la sfârșitul capitolului: intervalul de încredere conține tot ce conține valoarea p și, în plus, mărimea efectului și precizia ei [70].

4.1.2 Proprietățile unui estimator bun

Un estimator este nedeplasat (engl. unbiased) dacă, în medie, pe toate eșantioanele posibile, nimerește parametrul. Media eșantionului este un estimator nedeplasat al mediei populației: unele

eșantioane o supraestimează, altele o subestimează, dar fără tendință sistematică. Simularea din capitolul 3 (tabelul 3.5) a arătat exact acest lucru: media celor 10.000 de medii a rămas în jurul valorii de 161 mg/dL pentru orice n.

Aici apare explicația unui detaliu care intrigă studenții de la primul seminar: de ce împărțim la n - 1, și nu la n, în formula abaterii standard?

$$s^2 = \frac{\sum(x_i - m)^2}{n - 1}$$

Răspunsul ține de nedeplasare. Abaterile se calculează față de media eșantionului m, nu față de media reală μ, pe care nu o cunoaștem. Media eșantionului este, prin construcție, punctul față de care suma pătratelor abaterilor este minimă, deci abaterile față de m sunt sistematic puțin mai mici decât cele față de μ. Împărțirea la n ar subestima varianța populației. Am verificat prin simulare: din 20.000 de eșantioane de câte 5 valori extrase dintr-o populație normală cu σ = 14 mmHg (deci σ² = 196), varianța calculată cu n - 1 a avut în medie 196,2, iar cea calculată cu n doar 157,0, adică exact 196 x 4/5. Corecția n - 1 (corecția Bessel) elimină această deplasare. Pentru n mare diferența devine neglijabilă: pentru TAS din mg2026, abaterea standard cu n - 1 este 13,88 mmHg, iar cu n ar fi 13,84 mmHg. În Google Sheets, STDEV (sau STDEV.S) împarte la n - 1, iar STDEVP (STDEV.P) la n; pentru eșantioane se folosește întotdeauna prima.

Un estimator este consistent dacă, pe măsură ce n crește, se apropie tot mai mult de parametru. Media eșantionului este consistentă: eroarea ei standard $\frac{\sigma}{\sqrt{n}}$ scade spre zero când n crește. Un eșantion mai mare nu garantează o estimare corectă dacă eșantionul este deplasat (de exemplu, doar pacienți dintr-un spital universitar), dar garantează o estimare tot mai precisă a ceea ce măsoară.

Un estimator este eficient dacă, dintre estimatorii nedeplasați, are cea mai mică variabilitate. Pentru date normale, atât media, cât și mediana estimează centrul distribuției, dar media este mai eficientă: eroarea standard a medianei este aproximativ $\frac{1{,}25 \times \sigma}{\sqrt{n}}$. Pentru TAS din mg2026, aceasta ar fi 1,25 x 1,097 = 1,37 mmHg, față de 1,10 mmHg pentru medie. Ca să obținem aceeași precizie cu mediana, ne-ar trebui cu aproximativ 57% mai mulți pacienți. Pentru date asimetrice sau cu valori extreme, situația se poate inversa, iar mediana devine estimatorul mai stabil. Alegerea estimatorului depinde deci de forma distribuției, adică de tot ce am discutat în capitolul 3.

4.2 Distribuția de eșantionare a mediei și eroarea standard

4.2.1 Eroarea standard a mediei

Teorema limitei centrale (secțiunea 3.7) ne-a spus că mediile eșantioanelor de mărime n au o distribuție aproximativ normală, centrată pe μ, cu abaterea standard $\frac{\sigma}{\sqrt{n}}$. Această abatere standard a mediilor se numește eroarea standard a mediei (engl. standard error of the mean), notată ES sau SEM. Cum σ este necunoscut, o estimăm cu s:

$$\text{ES} = \frac{s}{\sqrt{n}}$$

Pentru TAS din mg2026:

$$\text{ES} = \frac{13{,}88}{\sqrt{160}} = \frac{13{,}88}{12{,}649} = 1{,}097 \text{ mmHg}$$

Eroarea standard este de aproape 13 ori mai mică decât abaterea standard, pentru că √160 ≈ 12,65. Ea nu descrie împrăștierea pacienților, ci împrăștierea mediilor pe care le-am obține dacă am repeta studiul de multe ori cu câte 160 de pacienți. În Google Sheets se calculează cu

=STDEV(M2:M161)/SQRT(COUNT(M2:M161)); JASP o afișează în Descriptives, opțiunea "Std. error of mean", iar MedCalc o raportează automat în Summary statistics.

4.2.2 Abaterea standard sau eroarea standard?

Este probabil cea mai frecventă confuzie din literatura medicală, iar la examen apare an de an. Cele două mărimi răspund la întrebări diferite.

Abaterea standard (AS) descrie variabilitatea datelor: cât de diferiți sunt pacienții între ei. Nu scade când crește eșantionul; cu mai mulți pacienți o estimăm doar mai precis. Se raportează în partea descriptivă (tabelul 1 al unui articol), de forma "TAS 131,13 ± 13,88 mmHg".

Eroarea standard (ES) descrie precizia mediei: cât de mult ar varia media de la un eșantion la altul. Scade odată cu creșterea lui n. Nu se raportează de obicei singură, ci servește la construirea intervalului de încredere, care este mult mai ușor de interpretat.

Raportarea "media ± ES" în locul "media ± AS" face datele să pară mai puțin variabile decât sunt, iar barele de eroare dintr-un grafic par mai scurte. Când citiți un articol, verificați întotdeauna ce reprezintă semnul ±. Dacă nu se precizează și n este cunoscut, puteți deduce: o "AS" de 1,1 mmHg pentru tensiunea sistolică într-un eșantion de 160 de pacienți este, aproape sigur, o eroare standard. Regula de raportare recomandată este simplă: media ± AS pentru a descrie datele, media cu IC 95% pentru a exprima precizia estimării.

4.3 Intervalul pentru valori individuale și intervalul pentru medie

În tradiția disciplinei noastre, primele exerciții de estimare se formulează astfel: "estimați, cu o încredere de 95,4%, valorile înălțimii studenților" și "estimați, cu o încredere de 95,4%, media înălțimii la nivelul populației studenților". Formulările seamănă, dar cer lucruri diferite, iar diferența este exact cea dintre AS și ES.

Intervalul pentru valori individuale (numit și interval de predicție sau, în laborator, interval de referință) folosește abaterea standard:

$m \pm k \times s$, cu k = 1 (68,3%), 2 (95,4%) sau 3 (99,7%)

El răspunde la întrebarea: unde se află majoritatea pacienților? Presupune ca valorile individuale să fie aproximativ normale (capitolul 3).

Intervalul pentru medie folosește eroarea standard:

$m \pm k \times$ ES, cu k = 1 (68,3%), 2 (95,4%) sau 3 (99,7%)

El răspunde la întrebarea: unde se află, probabil, media populației? Datorită teoremei limitei centrale, funcționează și pentru date moderat asimetrice, dacă n este suficient de mare.

Tabelul 4.1. Intervalele pentru valorile individuale și pentru media TAS (mg2026: m = 131,13 mmHg; s = 13,88 mmHg; ES = 1,097 mmHg; n = 160).

Nivelul de încredere	Pentru valori individuale: $m \pm k \times s$	Lățimea	Pentru medie: $m \pm k \times$ ES	Lățimea
68,3% ($k = 1$)	117,3-145,0 mmHg	27,8 mmHg	130,03-132,23 mmHg	2,2 mmHg
95,4% ($k = 2$)	103,4-158,9 mmHg	55,5 mmHg	128,94-133,33 mmHg	4,4 mmHg
99,7% ($k = 3$)	89,5-172,8 mmHg	83,3 mmHg	127,84-134,42 mmHg	6,6 mmHg

Diferența dintre cele două coloane este spectaculoasă (figura 4.1). Intervalul de 95,4% pentru valori are peste 55 mmHg lățime și cuprinde pacienți de la tensiune normală-joasă până la hipertensiune de gradul II. Intervalul de 95,4% pentru medie are sub 4,5 mmHg: știm destul de

precis unde se află media populației, chiar dacă pacienții individuali diferă foarte mult între ei. Raportul dintre lățimi este √160 ≈ 12,6.

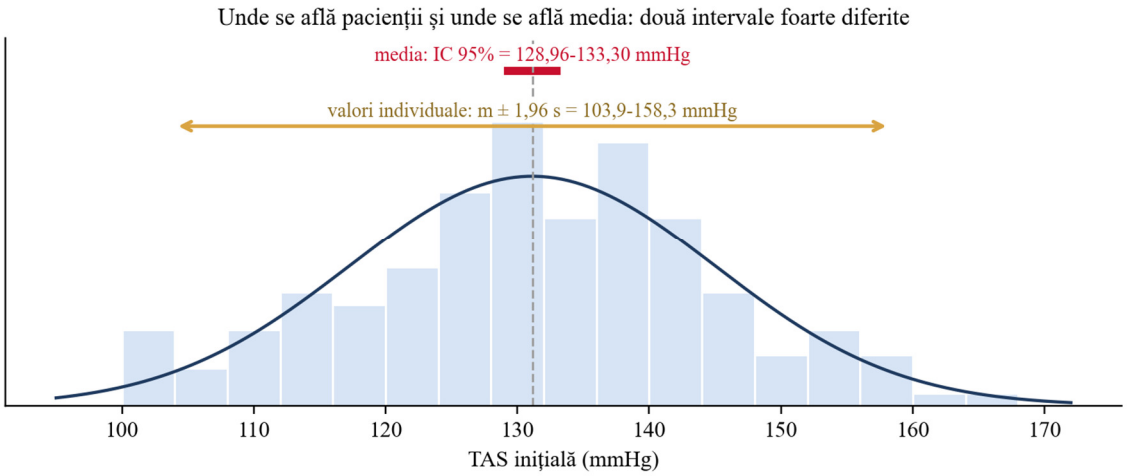


Figura 4.1. Histograma TAS inițiale din setul mg2026, cu curba normală ajustată. Săgeata galbenă arată intervalul m ± 1,96 s, în care se află aproximativ 95% dintre pacienți (103,9-158,3 mmHg). Bara roșie scurtă arată intervalul de încredere de 95% pentru media populației (128,96-133,30 mmHg). Date sintetice (setul mg2026).

Pentru intervalul exact de 95% se folosește 1,96 în loc de 2. Pentru valori individuale: 103,9-158,3 mmHg (cap. 3); pentru medie, cu z: $131{,}13 \pm 1{,}96 \times 1{,}097 = 128{,}98 - 133{,}28$ mmHg. Programele folosesc însă pentru medie distribuția t, nu distribuția normală, iar diferența apare în a doua zecimală, cum vom vedea imediat.

4.4 Intervalul de încredere pentru medie cu distribuția t

4.4.1 De ce t și nu z

Formula $m \pm 1{,}96 \times \text{ES}$ presupune că știm σ. În realitate, îl estimăm cu s, iar această estimare are și ea o incertitudine, mai mare în eșantioanele mici. Distribuția t (secțiunea 3.6), cu cozile ei mai groase, ține cont de această incertitudine suplimentară. Intervalul de încredere pentru medie devine:

$$\text{IC} = m \pm \frac{t\left(\text{gl};\, 1-\frac{\alpha}{2}\right) \times s}{\sqrt{n}}, \text{ cu gl = n - 1}$$

unde t(gl; 1 - α/2) este valoarea critică din tabelul E.3 al anexei E. Pentru un IC 95%, α = 0,05 și căutăm coloana 0,05 bilateral (sau 0,975 cumulat).

Tabelul 4.2. Valori critice t pentru intervale de încredere de 95% și 99%.

n	gl = $n - 1$	t pentru IC 95%	t pentru IC 99%
5	4	2,776	4,604
10	9	2,262	3,250
20	19	2,093	2,861
30	29	2,045	2,756
40	39	2,023	2,708
60	59	2,001	2,662
100	99	1,984	2,626
160	159	1,975	2,607
foarte mare	infinit	1,960	2,576

Pentru n = 160, diferența dintre t (1,975) și z (1,96) este mică. Pentru n = 10, valoarea t de 2,262 lărgește intervalul cu 15% față de cel calculat cu z. Pentru n = 5, cu 42%. Vechea regulă

"pentru n ≥ 30 folosim z, pentru n < 30 folosim t" provine din vremea tabelelor tipărite; astăzi programele folosesc t întotdeauna, pentru că este corect pentru orice n.

4.4.2 Calculul complet pentru TAS din mg2026

Pas 1. Statisticile descriptive: n = 160, m = 131,13 mmHg, s = 13,88 mmHg.

Pas 2. Eroarea standard: ES $= \frac{13,88}{\sqrt{160}} = 1,097$ mmHg.

Pas 3. Valoarea critică: gl = 159; t(159; 0,975) = 1,975.

Pas 4. Marja de eroare: E = 1,975 x 1,097 = 2,167 mmHg.

Pas 5. Intervalul: 131,13 - 2,167 = 128,96 mmHg și 131,13 + 2,167 = 133,30 mmHg.

IC 95% pentru media TAS: 128,96-133,30 mmHg.

Pentru IC 99%, t(159; 0,995) = 2,607, marja este 2,607 x 1,097 = 2,861 mmHg, iar intervalul este 128,27-133,99 mmHg. Pentru IC 90%, t(159; 0,95) = 1,654, iar intervalul este 129,32-132,95 mmHg. Observați regula: cu cât cerem o încredere mai mare, cu atât intervalul este mai larg. Nu există un interval care să fie în același timp foarte sigur și foarte îngust, decât dacă mărim eșantionul.

În Google Sheets, marja de eroare se obține direct cu =CONFIDENCE.T(0,05; STDEV(M2:M161); COUNT(M2:M161)), care dă 2,167. Funcția CONFIDENCE.NORM (sau vechea CONFIDENCE) folosește z în loc de t și dă 2,151. Limitele se calculează apoi ca =AVERAGE(M2:M161) - CONFIDENCE.T(...) și =AVERAGE(M2:M161) + CONFIDENCE.T(...). În JASP, intervalul apare în T-Tests > One Sample T-Test, cu opțiunea "Location parameter" și "Confidence interval" bifate [48,49]; în MedCalc, în Statistics > Summary statistics, la rândul "95% CI for the mean" [50].

Pentru glicemia inițială, cu $m = 105,44$ mg/dL, $s = 15,76$ mg/dL și n = 160, ES = 1,246 mg/dL, iar IC 95% este $105,44 \pm 1,975 \times 1,246$, adică 102,98-107,90 mg/dL. Asimetria ușoară a glicemiei (capitolul 3) nu afectează practic acest interval, pentru că, la n = 160, distribuția mediei este aproape normală.

4.4.3 Un eșantion mic: 10 pacienți

Să presupunem că am fi avut doar 10 pacienți, extrași aleatoriu din mg2026. Valorile lor TAS, ordonate, sunt: 102, 103, 110, 130, 131, 136, 137, 147, 150, 154 mmHg.

Media: suma este 1300, deci m = 130,0 mmHg.

Abaterea standard: abaterile față de medie sunt -28, -27, -20, 0, 1, 6, 7, 17, 20, 24; pătratele lor sunt 784, 729, 400, 0, 1, 36, 49, 289, 400, 576, cu suma 3264. Varianța $s^2 = \frac{3264}{9} = 362,67$, iar s = 19,04 mmHg.

Eroarea standard: ES $= \frac{19,04}{\sqrt{10}} = \frac{19,04}{3,162} = 6,022$ mmHg.

Valoarea critică: gl = 9, t = 2,262.

Marja: 2,262 x 6,022 = 13,62 mmHg.

IC 95%: 130,0 ± 13,62, adică 116,38-143,62 mmHg, cu o lățime de 27,2 mmHg.

Cu z în loc de t, am fi obținut 118,20-141,80 mmHg, un interval prea îngust, care ar fi exagerat precizia. Comparați cu intervalul pe 160 de pacienți, lățime 4,3 mmHg. Cu 10 pacienți, media populației ar putea fi oriunde între "normal" și "hipertensiune de gradul I"; cu 160, o știm cu o precizie de ±2 mmHg.

4.5 Interpretarea corectă a unui interval de încredere

4.5.1 Ce înseamnă "95%"

Interpretarea corectă a unui IC 95% este una de tip frecvenționist și se referă la procedură, nu la un interval anume: dacă am repeta studiul de foarte multe ori, de fiecare dată cu un eșantion nou și calculând intervalul în același fel, aproximativ 95% dintre intervalele obținute ar conține media reală a populației. Cele 5% rămase ar rata-o, din cauza fluctuațiilor de eșantionare.

Figura 4.2 face această idee vizibilă. Am simulat 50 de studii, fiecare pe 25 de pacienți extrași dintr-o populație cu media TAS cunoscută (131 mmHg) și abaterea standard 14 mmHg, și am calculat câte un IC 95% pentru fiecare. Intervalele au poziții și lățimi diferite, pentru că fiecare eșantion are altă medie și altă abatere standard. 48 dintre ele conțin media reală; 2 (marcate cu roșu) nu o conțin. Proporția de 96% este foarte aproape de 95%, cum era de așteptat. În viața reală facem un singur studiu și nu știm dacă intervalul nostru este unul dintre cele 95% "bune" sau dintre cele 5% "ratate".

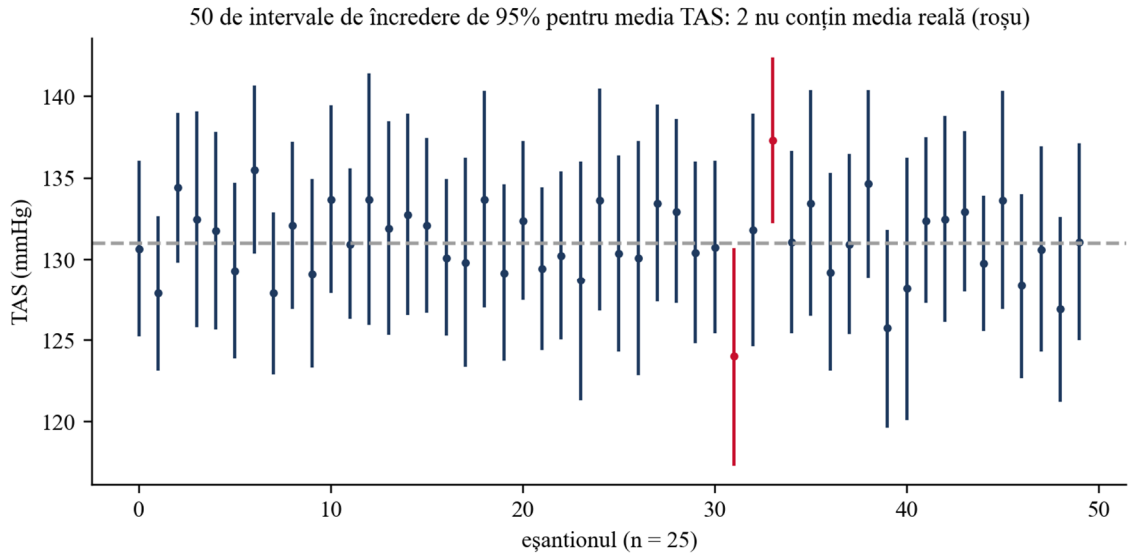


Figura 4.2. Cincizeci de intervale de încredere de 95% pentru media TAS, calculate pe 50 de eșantioane simulate de câte 25 de pacienți dintr-o populație cu media 131 mmHg (linia întreruptă). Două intervale, marcate cu roșu, nu conțin media reală. Simulare.

În limbaj obișnuit, putem spune: "suntem 95% încrezători că media populației se află între 128,96 și 133,30 mmHg", cu înțelegerea că "încrederea" se referă la metoda de construcție, care reușește în 95% din cazuri. Abordarea bayesiană, care nu face obiectul acestui curs, permite afirmații directe de probabilitate despre parametru, dar cere informații suplimentare (o distribuție a priori).

4.5.2 Interpretări greșite

Aceste formulări apar frecvent în lucrări și la examen și sunt greșite:

- "Există o probabilitate de 95% ca media populației să fie între 128,96 și 133,30 mmHg." În abordarea frecvenționistă, media populației este o constantă: fie este în interval, fie nu. Probabilitatea de 95% se referă la procedură.
- "95% dintre pacienți au TAS între 128,96 și 133,30 mmHg." Este confuzia dintre intervalul pentru medie și intervalul pentru valori. În realitate, doar 25 dintre cei 160 de pacienți (15,6%) au TAS în acest interval îngust.

- "Dacă repetăm studiul, media noului eșantion va fi cu 95% probabilitate în acest interval." Nu: noul eșantion va avea propria variabilitate, iar probabilitatea ca media lui să cadă în intervalul nostru este, în medie, în jur de 83%, nu 95%.
- "Valorile din centrul intervalului sunt sigure, cele de la capete imposibile." Valorile din apropierea estimării punctuale sunt mai compatibile cu datele decât cele de la capete, dar nu există o graniță bruscă la limitele intervalului [70].
- "Un interval mai larg înseamnă un efect mai mare." Lățimea exprimă precizia (dependentă de n și de variabilitate), nu mărimea efectului.

4.6 Lățimea intervalului de încredere și mărimea eșantionului

Lățimea unui IC pentru medie este $\frac{2 \times t \times s}{\sqrt{n}}$. Ea depinde de trei lucruri: nivelul de încredere (prin t), variabilitatea datelor (prin s) și mărimea eșantionului (prin $\sqrt{n}$). Primul îl alegem noi, al doilea ține de biologie și de calitatea măsurării, al treilea ține de proiectarea studiului. Cel mai important mesaj este că lățimea scade cu rădăcina pătrată a lui n, nu cu n: pentru a înjumătăți lățimea intervalului, trebuie să mărim eșantionul de patru ori (figura 4.3).

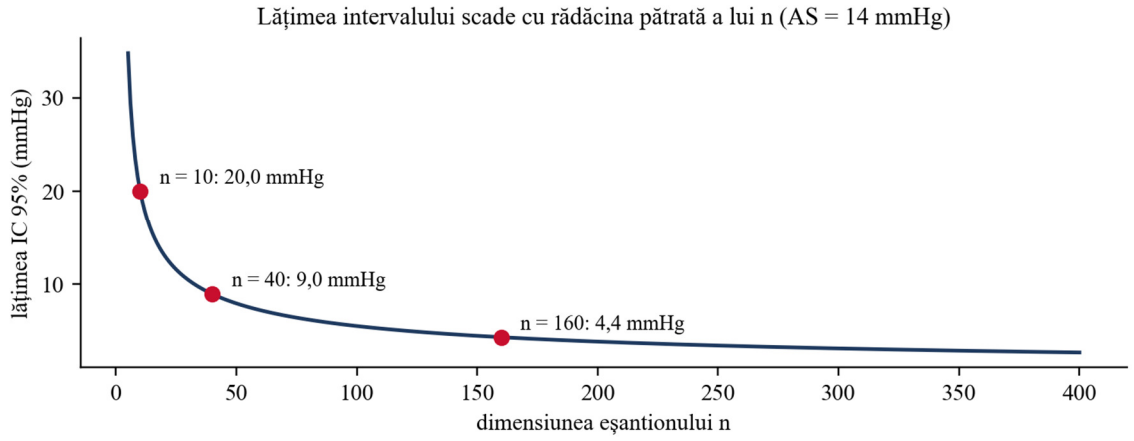


Figura 4.3. Lățimea intervalului de încredere de 95% pentru media TAS, în funcție de mărimea eșantionului, pentru o abatere standard de 14 mmHg. Curba scade repede la început și tot mai lent apoi: dublarea preciziei cere de patru ori mai mulți pacienți. Curbă teoretică.

Pe datele reale din mg2026, am extras aleatoriu un subeșantion de 10 pacienți și unul de 40 de pacienți și am calculat IC 95% pentru media TAS (tabelul 4.3).

Tabelul 4.3. Intervalul de încredere de 95% pentru media TAS în funcție de mărimea eșantionului (setul mg2026).

n	Media (mmHg)	IC 95% (mmHg)	Lățimea (mmHg)	Lățimea teoretică pentru s = 14 mmHg
10	130,0	116,38-143,62	27,2	20,0
40	128,4	123,08-133,67	10,6	9,0
160	131,1	128,96-133,30	4,3	4,4

Lățimile observate diferă de cele teoretice, pentru că fiecare subeșantion are propria abatere standard: subeșantionul de 10 pacienți a avut, din întâmplare, s = 19,04 mmHg, mai mare decât cea a eșantionului complet, iar cel de 40, s = 16,54 mmHg. Tendința este însă clară: de la 10 la 40 de pacienți (de patru ori mai mulți), lățimea scade de la 27,2 la 10,6 mmHg; de la 40 la 160 (din nou de patru ori), de la 10,6 la 4,3 mmHg. Observați și că toate cele trei intervale conțin media eșantionului complet, iar intervalele mai mici o "prind" cu o marjă mai largă. În lucrarea practică

4 veți repeta această simulare de zeci de ori în Google Sheets și veți vedea cum variază intervalele de la un eșantion la altul.

Consecința practică pentru cercetare este importantă: câștigul de precizie are randament descrescător. Trecerea de la 40 la 160 de pacienți câștigă 6,3 mmHg în lățime, trecerea de la 160 la 640 ar câștiga doar 2,2 mmHg. De aceea calculul dimensiunii eșantionului (secțiunea 4.13) pornește de la întrebarea "ce precizie ne trebuie?", nu de la "câți pacienți putem aduna?".

> **Studiu de caz: intervalul de încredere al unei modificări medii**
>
> Muntean și colaboratorii au urmărit 921 de studenți în primul an de facultate și au raportat o modificare medie a IMC de -0,45 ± 1,07 kg/m², cu IC 95% de la -0,52 la -0,38 kg/m² ($p < 0,001$) [5]. Putem verifica de mână intervalul. Eroarea standard a diferenței medii este 1,07 / √921 = 1,07 / 30,35 = 0,0353 kg/m². Marja de eroare este $1,96 \times 0,0353 = \frac{0,069\ \text{kg}}{m^2}$ (pentru 920 de grade de libertate, t = 1,963, practic 1,96). Intervalul este -0,45 ± 0,069, adică de la -0,519 la -0,381, care, rotunjit, coincide cu cel publicat. Lecția cifrelor: abaterea standard a modificărilor (1,07) este mare față de medie, dar cu 921 de participanți eroarea standard este de 30 de ori mai mică, iar intervalul este foarte îngust. Știm foarte precis că IMC mediu a scăzut, cu aproximativ o jumătate de unitate; în același timp, 13,1% dintre studenți au crescut în greutate. Precizia unei medii nu spune nimic despre eterogenitatea indivizilor. Un al doilea detaliu merită observat: intervalul exclude zero, iar $p < 0,001$; vom vedea în secțiunea 4.12 că cele două informații sunt legate.

4.7 Intervalul de încredere pentru o proporție

4.7.1 Metoda Wald

O proporție calculată pe un eșantion este și ea o estimare, cu propria eroare standard. Dacă în eșantion k din n subiecți au caracteristica studiată, proporția observată este $p_i = \frac{k}{n}$, iar eroarea ei standard este:

$$\text{ES}(p_i) = \sqrt{\frac{p_i \times (1 - p_i)}{n}}$$

Cea mai simplă metodă de calcul al intervalului, numită metoda Wald, se bazează pe aproximarea normală a distribuției binomiale (secțiunea 3.2.1):

$$\text{IC } 95\% = p_i \pm 1,96 \times \sqrt{\frac{p_i \times (1 - p_i)}{n}}$$

În mg2026, 29 dintre cei 160 de pacienți au diabet confirmat prin TTGO: $p_i = \frac{29}{160} = 0,181 (18,1\%)$. Eroarea standard este $\sqrt{\frac{0,181 \times 0,819}{160}} = \sqrt{0,000927} = 0,0305$. Marja de eroare: 1,96 x 0,0305 = 0,0597. Intervalul Wald: 0,181 ± 0,060, adică 12,2%-24,1%.

4.7.2 Metoda Wilson și de ce este preferată

Metoda Wald este ușor de calculat de mână, dar are defecte cunoscute. Intervalul este întotdeauna simetric, deși distribuția unei proporții mici nu este simetrică (nu poate coborî sub 0, dar poate urca mult). Pentru proporții apropiate de 0 sau de 1 și pentru eșantioane mici, intervalul Wald poate coborî sub 0% sau urca peste 100% și, mai grav, acoperă parametrul real mai rar decât promite: un "IC 95%" Wald poate avea, în realitate, o acoperire de 85-90%. De aceea, ghidurile și

programele moderne recomandă metoda Wilson (scorul Wilson) sau metoda exactă Clopper-Pearson [42,53].

Formula Wilson arată intimidant, dar se poate calcula pas cu pas:

centrul = ($p_i + z^2 / (2n)$) / ($1 + z^2 / n$) marja = $\left(\frac{z}{1+\frac{z^2}{n}}\right) \times \sqrt{\frac{p_i \times (1-p_i)}{n} + \frac{z^2}{4n^2}}$

Pentru 29/160, cu $z = 1{,}96 (z^2 = 3{,}8416)$:

$\frac{z^2}{n} = \frac{3{,}8416}{160} = 0{,}0240$; $\frac{z^2}{2n} = 0{,}0120$ centrul = (0,1813 + 0,0120) / 1,0240 = 0,1887 sub radical: $\frac{0{,}1813 \times 0{,}8187}{160} + \frac{3{,}8416}{102400} = 0{,}000927 + 0{,}0000375 = 0{,}000965$; radicalul = 0,0311 marja = $\left(\frac{1{,}96}{1{,}0240}\right) \times 0{,}0311 = 1{,}914 \times 0{,}0311 = 0{,}0595$; IC 95% Wilson: 0,1887 ± 0,0595, adică 12,9%-24,8%

Observați două lucruri. Centrul intervalului Wilson (18,9%) este ușor deplasat spre 50% față de proporția observată (18,1%), iar intervalul nu mai este simetric în jurul lui 18,1%: 5,2 puncte procentuale sub și 6,7 peste. Pentru n = 160 și p_i = 18%, diferența față de Wald este mică. Metoda exactă Clopper-Pearson dă 12,5%-25,0%, ceva mai larg, pentru că este conservatoare.

Diferențele devin importante în eșantioane mici. Dacă din 20 de pacienți 2 au o complicație (10%), intervalul Wald este 10% ± 13,1%, adică de la -3,1% la 23,1%, o limită inferioară imposibilă, pe care programele o "taie" la 0. Intervalul Wilson este 2,8%-30,1%, iar cel exact 1,2%-31,7%. Ambele sunt asimetrice și realiste. Concluzia practică: pentru proporții raportați intervalul Wilson (sau exact); metoda Wald este utilă doar pentru înțelegerea ideii și pentru calcule rapide de mână, cu n mare și p_i departe de 0 și 1. În MedCalc, intervalul exact apare în Statistics > Summary statistics pentru variabile dihotomice sau în calculatorul pentru o proporție; în JASP, în Frequencies > Binomial Test, care afișează intervalul exact Clopper-Pearson. În Google Sheets nu există o funcție dedicată, dar formula Wilson se poate scrie într-o celulă, cum vom face în lucrarea practică 4.

4.7.3 Zero evenimente: regula lui trei

O situație specială, foarte frecventă în medicină: un studiu pe 50 de pacienți tratați cu un medicament nou nu observă nicio reacție adversă gravă. Proporția observată este 0/50 = 0%, iar intervalul Wald este 0% ± 0, adică "exact zero", ceea ce este evident absurd. Nimeni nu ar susține că riscul real este zero doar pentru că nu l-am văzut la 50 de pacienți.

Regula lui trei dă o soluție simplă: când într-un eșantion de n subiecți nu se observă niciun eveniment, limita superioară a IC 95% pentru proporția reală este aproximativ 3 / n. Pentru 50 de pacienți: 3/50 = 6%. Riscul real ar putea fi, deci, de până la aproximativ 6%, adică până la 1 pacient din 17. Metoda exactă dă 7,1%; regula funcționează cel mai bine pentru n mai mare (pentru 100 de pacienți, 3% față de 3,6% exact). Pentru 30 de pacienți fără evenimente, limita este 3/30 = 10%. Regula explică de ce reacțiile adverse rare (1 la 10.000) nu pot fi detectate în studiile clinice de fază III cu câteva mii de pacienți și de ce farmacovigilența de după autorizare este indispensabilă.

> **Studiu de caz: intervale Wilson pentru prevalențe**
>
> Moleriu și colaboratorii au studiat 136 de pacienți dintr-o clinică de gastroenterologie, dintre care 52 cu un fenotip de tulburare a interacțiunii intestin-creier și 84 de controale asimptomatice [71]. Doar 4,4% dintre ei (IC 95% 2,0-9,3%) aveau un diagnostic formal de colon iritabil, în timp ce 47,8% (IC 95% 39,6-56,1%) îndeplineau criteriile unui grup de simptome compatibil, de unde un raport de subdiagnostic de 10,8. Uitați-vă la primul interval: 4,4% se află la 2,4 puncte de limita inferioară și la 4,9 puncte de cea superioară. Este asimetria

> tipică a unui interval Wilson sau exact pentru o proporție mică. Putem verifica: 4,4% din 136 înseamnă 6 pacienți, iar intervalul Wilson pentru 6/136 este exact 2,0-9,3%; pentru 65/136 (47,8%), intervalul Wilson este 39,6-56,1%, din nou identic cu cel publicat. Un interval Wald pentru 6/136 ar fi dat 1,0-7,9%, simetric și prea îngust în partea de sus. Al doilea interval, pentru o proporție apropiată de 50%, este aproape simetric, pentru că acolo metodele dau rezultate similare. Mischie și colaboratorii au raportat explicit prevalențe cu intervale Wilson într-un sondaj online pe 753 de tineri de 18-30 de ani, cu 11,6% subponderali și 29,3% supraponderali sau obezi [20]. Ca exercițiu, intervalul Wilson calculat de noi pentru 29,3% din 753 este aproximativ 26,2-32,7%. Precizia este bună datorită volumului mare, dar intervalul nu corectează un alt neajuns, pe care autorii îl recunosc chiar în titlu: eșantionul este de conveniență, cu 93,5% femei. Intervalul de încredere măsoară doar eroarea aleatoare de eșantionare, nu și eroarea sistematică de selecție.

4.8 Intervalul de încredere pentru diferența a două medii și a două proporții

4.8.1 Diferența a două medii

Când comparăm două grupuri independente, parametrul care ne interesează este diferența dintre mediile populațiilor, μ1 - μ2. Estimarea punctuală este m1 - m2, iar eroarea ei standard combină erorile standard ale celor două medii:

$$\text{ES}(m_1 - m_2) = \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}$$

Intervalul de încredere este $(m_1 - m_2) \pm t \times \text{ES}(m_1 - m_2)$, cu un număr de grade de libertate calculat prin formula Welch (programele îl calculează automat) sau, dacă variațele sunt egale, n1 + n2 - 2.

În mg2026, TAS este 130,85 ± 13,25 mmHg la cei 74 de bărbați și 131,37 ± 14,47 mmHg la cele 86 de femei. Diferența bărbați minus femei este -0,52 mmHg. Eroarea standard:

$$\text{ES} = \sqrt{\frac{13{,}25^2}{74} + \frac{14{,}47^2}{86}} = \sqrt{2{,}372 + 2{,}435} = \sqrt{4{,}807} = 2{,}193 \text{ mmHg}$$

Cu 157,4 grade de libertate (Welch), t = 1,975, iar marja este 1,975 x 2,193 = 4,33 mmHg. IC 95% pentru diferență: -0,52 ± 4,33, adică de la -4,9 la 3,8 mmHg.

Intervalul conține zero: datele sunt compatibile atât cu o TAS puțin mai mare la bărbați (până la 3,8 mmHg), cât și cu una mai mare la femei (până la 4,9 mmHg), și desigur cu nicio diferență. Intervalul spune însă și ceva în plus față de "diferența nu este semnificativă" (p = 0,81, capitolul 6): o diferență reală de 10 mmHg între sexe, în această populație, este foarte puțin probabilă, pentru că se află mult în afara intervalului. Un rezultat nesemnificativ cu un interval îngust este informativ; unul cu un interval foarte larg nu ne spune aproape nimic.

4.8.2 Diferența a două proporții

Pentru două proporții independente $p_1 = \frac{k_1}{n_1}$ și $p_2 = \frac{k_2}{n_2}$:

$$\text{ES}(p_1 - p_2) = \sqrt{\frac{p_1 \times (1 - p_1)}{n_1} + \frac{p_2 \times (1 - p_2)}{n_2}}$$

În studiul simulat mg2026, o scădere în greutate de cel puțin 5% la 6 luni s-a obținut la 16 din 54 de pacienți din brațul C (29,6%) și la 6 din 53 din brațul A (11,3%). Diferența absolută este 18,3 puncte procentuale. Eroarea standard:

$$\text{ES} = \sqrt{\frac{0{,}296 \times 0{,}704}{54} + \frac{0{,}113 \times 0{,}887}{53}} = \sqrt{0{,}00386 + 0{,}00189} = \sqrt{0{,}00576} = 0{,}0759$$

IC 95%: $0{,}183 \pm 1{,}96 \times 0{,}0759 = 0{,}183 \pm 0{,}149$, adică de la 3,4 la 33,2 puncte procentuale. Intervalul exclude zero, deci diferența este semnificativă statistic. Este însă foarte larg: beneficiul real al teleconsultației ar putea fi modest (3 pacienți în plus din 100) sau foarte mare (33 din 100). Din această diferență se calculează numărul necesar de tratat, $\text{NNT} = \frac{1}{0{,}183} = 5{,}5$, cu un interval de 3,0 până la 29,1 pacienți, pe care îl vom discuta în capitolul 10.

4.9 Intervale de încredere pentru date asimetrice

4.9.1 Scala logaritmică și media geometrică

Pentru o variabilă log-normală, ca trigliceridele din mg2026, intervalul de încredere se calculează pe scala logaritmică și se retransformă (secțiunea 3.9.4). Media logaritmilor este 4,970, abaterea lor standard 0,488, eroarea standard 0,488 / √160 = 0,0386, iar IC 95% pe scala logaritmică este $4{,}9699 \pm 1{,}975 \times 0{,}0386$, adică 4,8937 până la 5,0461. Retransformat, obținem $e^{4{,}8937} = 133{,}44$ și $e^{5{,}0461} = 155{,}42$ mg/dL, în jurul mediei geometrice de 144,0 mg/dL (limitele de pe scala logaritmică se păstrează cu toate zecimalele până la retransformare). Intervalul pentru media aritmetică, calculat direct cu formula t, ar fi 149,26-173,09 mg/dL. Cele două intervale nu se contrazic: estimează parametri diferiți. Primul estimează centrul "tipic" al unei distribuții asimetrice, al doilea media aritmetică, trasă în sus de coada dreaptă. Când raportați, spuneți explicit ce ați estimat.

4.9.2 Intervalul de încredere pentru mediană

Pentru variabile pentru care nu există o transformare potrivită, se poate estima direct intervalul de încredere al medianei. Metoda clasică, bazată pe distribuția binomială, nu presupune nicio formă a distribuției. Se ordonează valorile crescător și se calculează rangurile:

rangul inferior $\approx \frac{n}{2} - \frac{1{,}96 \times \sqrt{n}}{2}$; rangul superior $\approx \frac{n}{2} + \frac{1{,}96 \times \sqrt{n}}{2}$ (rotunjite spre exterior)

Pentru n = 160: $80 \pm \frac{1{,}96 \times 12{,}65}{2} = 80 \pm 12{,}4$, deci rangurile 67 și 93. Valorile de pe aceste poziții dau intervalul. Pentru trigliceride (mediana 151,5 mg/dL), a 67-a și a 93-a valoare ordonată sunt 134 și 162 mg/dL: IC 95% pentru mediană 134-162 mg/dL. Pentru TAS (mediana 131 mmHg), intervalul este 129-134 mmHg, puțin mai larg decât cel al mediei (128,96-133,30 mmHg, lățime 4,3 față de 5), ilustrând eficiența mai mică a medianei pentru date normale. MedCalc calculează automat intervalul de încredere al medianei în Summary statistics.

4.10 Intervale de încredere pentru rapoarte (OR, RR) și pentru aria ROC

4.10.1 De ce sunt asimetrice intervalele pentru OR și RR

Raportul șanselor (OR) și riscul relativ (RR), pe care le vom studia în capitolul 10, sunt rapoarte: iau valori între 0 și infinit, cu valoarea 1 pentru "nicio asociere". Distribuția lor de eșantionare este asimetrică, dar logaritmul lor are o distribuție aproximativ normală. Prin urmare, intervalul se calculează pe scala logaritmică și se retransformă, la fel ca pentru media geometrică.

Pentru asocierea obezitate-diabet din mg2026 (tabel 2 x 2 cu a = 18, b = 36, c = 11, d = 95), $\mathrm{OR} = \frac{18\times95}{36\times11} = 4{,}32$. Pe scala logaritmică:

$$\ln(\mathrm{OR}) = \ln(4{,}32) = 1{,}463;\ \mathrm{ES}(\ln \mathrm{OR}) = \sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}} = \sqrt{\frac{1}{18}+\frac{1}{36}+\frac{1}{11}+\frac{1}{95}} =$$

$\sqrt{0{,}1848} = 0{,}430 \ln(\mathrm{OR}) \pm 1{,}96 \times 0{,}430 = 1{,}463 \pm 0{,}843$, adică 0,620 până la 2,306 IC 95% pentru OR: $e^{0{,}620} = 1{,}86$ până la $e^{2{,}306} = 10{,}03$

Intervalul este puternic asimetric în jurul lui 4,32: 2,46 unități sub, 5,71 peste. Asimetria nu este un defect, ci consecința scalei multiplicative. O verificare utilă: estimarea punctuală este media geometrică a limitelor, $\sqrt{1{,}86 \times 10{,}03} = 4{,}32$. Pentru RR al scăderii ponderale în brațul C față de A (16/54 față de 6/53), RR = 2,62, iar IC 95%, calculat la fel, este 1,11-6,17.

4.10.2 Intervalul pentru aria de sub curba ROC

Aria de sub curba ROC (AUC), pe care o vom studia în capitolul 12, ia valori între 0,5 (test inutil) și 1 (test perfect). Și intervalul ei de încredere este adesea asimetric când AUC este aproape de 1, pentru că limita superioară nu poate depăși 1. În mg2026, scorul FINDRISC are AUC 0,852 (IC 95% 0,762-0,943) pentru diabetul confirmat prin TTGO. Gherbon și colaboratorii raportează pentru vârstă, ca predictor al distresului diabetic sever, AUC 0,726 (IC 95% 0,602-0,827), cu doar 19 cazuri, ceea ce explică lățimea mare a intervalului [69]. Regula de lectură este aceeași ca la orice estimare: ne uităm întâi la limita inferioară. O AUC de 0,726 cu limita inferioară 0,602 descrie un predictor de la "aproape inutil" la "acceptabil".

Moga și colaboratorii raportează pentru elastografia time-harmonic, în diagnosticul cirozei, AUROC 0,90 (IC 95% 0,82-0,93) [3]: estimarea este mult mai aproape de limita superioară decât de cea inferioară (0,08 sub, 0,03 peste), o asimetrie firească pentru o mărime mărginită la 1. Același studiu raportează, cu intervale de încredere, și reproductibilitatea măsurătorilor, pe un subgrup de 27 de pacienți: acordul global între examinatori 0,94 (IC 95% 0,89-0,97) și coeficienți de corelație intraclasă de 0,92 (0,82-0,96), 0,94 (0,87-0,97) și 0,97 (0,95-0,99). Intervalul pentru 0,92 se întinde pe 14 sutimi, ceea ce reflectă mărimea mică a subgrupului; o reproductibilitate "excelentă" estimată pe 27 de pacienți rămâne compatibilă și cu una doar "bună".

Rațiu și colaboratorii au comparat trei metode de diagnostic al obstrucției biliare maligne la 160 de pacienți, cu AUROC 0,88 (IC 95% 0,75-0,90) pentru colangio-RM, 0,94 (0,85-0,99) pentru ERCP și 0,80 (0,70-0,82) pentru histologie [57]. Două observații de lectură critică. Intervalele se suprapun parțial, dar suprapunerea a două intervale nu înseamnă automat că diferența nu este semnificativă; comparația a două AUC măsurate pe aceiași pacienți cere un test dedicat, care ține cont de corelația dintre ele (capitolul 12). Apoi, pentru colangio-RM și histologie, estimarea punctuală se află foarte aproape de limita superioară (0,88 într-un interval care urcă doar până la 0,90; 0,80 într-unul care urcă până la 0,82). Tiparul este neobișnuit și poate proveni din metoda de calcul sau dintr-o scăpare de redactare; merită verificat în textul complet înainte ca cifrele să fie folosite mai departe.

> **Studiu de caz: rapoarte de șanse cu intervale de încredere, citite corect**
>
> Trei studii ale colegilor din disciplină pe studenți și pacienți din Timișoara ilustrează lectura intervalelor pentru rapoarte. Moleriu și colaboratorii au analizat 921 de studenți, dintre care 24,1% supraponderali sau obezi, și au raportat rapoarte ajustate ale șanselor de 7,20 (IC 95% 4,26-12,17) și 1,98 (IC 95% 1,26-3,10), plus un OR de 1,43 pentru fiecare punct al unui scor de risc alimentar (IC 95% 1,23-1,66) [4]. Verificați proprietatea mediei geometrice: $\sqrt{4{,}26 \times 12{,}17} = 7{,}20$, $\sqrt{1{,}26 \times 3{,}10} = 1{,}98$, $\sqrt{1{,}23 \times 1{,}66} = 1{,}43$. Toate cele trei intervale exclud 1, deci asocierile sunt semnificative, dar primul este larg: datele sunt compatibile cu o asociere de 4 ori sau de 12 ori mai mare a șanselor. Piroș și colaboratorii au găsit, pe 693 de

> studenți, o prevalență a prediabetului de 21,1% și un OR de 1,78 pe categorie de tipar alimentar (IC 95% 1,38-2,30; p < 0,001) [6], un interval mai îngust, care exclude confortabil valoarea 1. La polul opus, Popa și colaboratorii, comparând 100 de femei și 113 bărbați tratați în urgență pentru infecții bacteriene suspectate, au raportat OR 0,71 (IC 95% 0,34-1,49) pentru un rezultat clinic înregistrat la 14 femei (14,0%) și 21 de bărbați (18,6%) și OR 0,87 (IC 95% 0,31-2,43) pentru mortalitatea intraspitalicească, cu 7 și 9 decese [9]. Aceste intervale includ 1 și sunt foarte largi: pentru mortalitate, datele sunt compatibile atât cu șanse de trei ori mai mici, cât și cu șanse de peste două ori mai mari la femei. Concluzia corectă nu este "nu există diferențe între sexe", ci "studiul nu are suficiente evenimente pentru a răspunde la întrebare". Același studiu a găsit, pentru sursa urinară a infecției, OR 2,14 (IC 95% 1,20-3,81). Prin contrast, coeficienții de regresie liniară au intervale simetrice, pentru că se află pe o scală aditivă: Cîrstea și colaboratorii raportează B = 0,076 (IC 95% 0,017-0,135; p = 0,011) pentru asocierea dintre satisfacția față de serviciile farmaceutice și aderența la tratament, la 220 de pacienți dermatologici [72]; estimarea este exact la mijloc, 0,076 ± 0,059.

4.11 Bootstrap, explicat simplu

Pentru multe statistici (mediana, raportul a două mediane, coeficientul de corelație Spearman, AUC) nu există formule simple pentru eroarea standard, sau formulele cer presupuneri pe care datele nu le îndeplinesc. Metoda bootstrap, propusă în anii '70, ocolește problema printr-o idee aproape naivă: dacă eșantionul nostru este reprezentativ pentru populație, putem simula "repetarea studiului" extrăgând noi eșantioane din eșantionul însuși.

Procedura este următoarea. Din cele 160 de valori ale trigliceridelor, extragem la întâmplare 160 de valori cu întoarcere: unele valori vor apărea de două sau trei ori, altele deloc. Calculăm mediana acestui "eșantion bootstrap". Repetăm de 10.000 de ori. Obținem 10.000 de mediane, a căror distribuție aproximează distribuția de eșantionare a medianei. Percentilele 2,5 și 97,5 ale acestei distribuții formează intervalul de încredere bootstrap de 95% (metoda percentilelor).

Pentru mediana trigliceridelor din mg2026, rezultatul este 137-162 mg/dL (figura 4.4), foarte apropiat de intervalul obținut prin metoda rangurilor (134-162 mg/dL). Distribuția medianelor bootstrap este neregulată, cu "salturi", pentru că mediana poate lua doar valori din setul de date sau medii a două valori vecine. Pentru media TAS, bootstrap-ul dă 128,99-133,26 mmHg, aproape identic cu intervalul t clasic (128,96-133,30 mmHg), iar abaterea standard a celor 10.000 de medii bootstrap este 1,09 mmHg, practic egală cu eroarea standard calculată cu formula (1,097 mmHg). Când presupunerile metodei clasice sunt îndeplinite, cele două metode concordă; când nu sunt, bootstrap-ul rămâne valid.

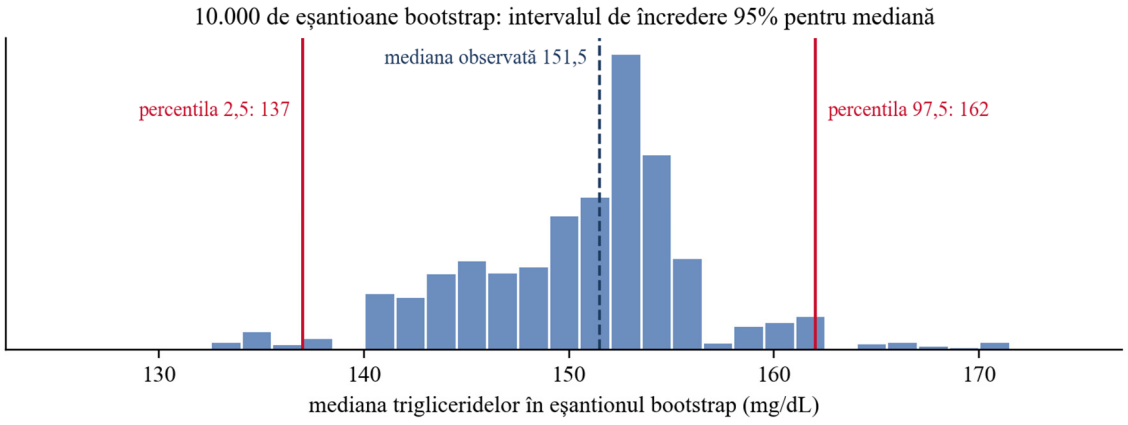


Figura 4.4. Distribuția medianelor trigliceridelor în 10.000 de eșantioane bootstrap extrase din setul mg2026. Liniile roșii marchează percentilele 2,5 (137 mg/dL) și 97,5 (162 mg/dL), care formează intervalul de încredere bootstrap de 95% pentru mediană; linia întreruptă este mediana observată (151,5 mg/dL). Simulare pe setul mg2026.

Bootstrap-ul nu este o soluție magică: nu poate corecta un eșantion nereprezentativ sau prea mic (cu 10 valori, eșantioanele bootstrap repetă la nesfârșit aceleași 10 numere). Este însă folosit tot mai des în articolele medicale, de exemplu pentru intervalele AUC sau pentru modelele de mediere. Dicu și colaboratorii au folosit 5000 de reeșantionări bootstrap în validarea unui instrument de evaluare a consumatorilor, pe 504 participanți [73]. JASP oferă opțiunea "Bootstrap" în multe analize, iar MedCalc o folosește pentru unele intervale.

4.12 Intervalul de încredere și valoarea p

Capitolul 5 este dedicat testării ipotezelor, dar legătura cu intervalele de încredere merită anunțată acum, pentru că simplifică mult lectura articolelor. Pentru o comparație (diferență de medii, diferență de proporții, OR, RR), există o "valoare nulă" care înseamnă "niciun efect": 0 pentru diferențe, 1 pentru rapoarte. Regula este:

- dacă IC 95% nu conține valoarea nulă, testul bilateral corespunzător dă $p < 0,05$;
- dacă IC 95% conține valoarea nulă, $p \geq 0,05$;
- dacă valoarea nulă este exact la o limită a intervalului, $p \approx 0,05$.

În mg2026: diferența TAS bărbați-femei, IC 95% de la -4,9 la 3,8 mmHg, conține 0, iar $p = 0,81$. RR al scăderii ponderale C față de A, IC 95% 1,11-6,17, nu conține 1, iar $p < 0,05$. OR obezitate-diabet, IC 95% 1,86-10,03, nu conține 1, iar $p < 0,001$; faptul că limita inferioară este departe de 1 corespunde unei valori p foarte mici. Regula funcționează exact pentru metodele care folosesc aceeași eroare standard pentru interval și pentru test; pentru alte combinații (de exemplu, un interval Wilson și un test exact) pot apărea mici discordanțe la limită.

Intervalul conține însă mai multă informație decât valoarea p. Valoarea p ne spune doar dacă datele sunt compatibile cu "niciun efect". Intervalul ne spune ce mărimi ale efectului sunt compatibile cu datele: dacă intervalul este îngust și aproape de zero, efectul este mic, chiar dacă semnificativ; dacă este larg, rezultatul este imprecis, oricare ar fi p [70,74].

Studiu de caz: ce lipsește când un raport al șanselor nu are interval de încredere

Cioca și colaboratorii au comparat evoluția COVID-19 la 486 de pacienți (200 fără diabet, 62 cu diabet de tip 1 și 224 cu diabet de tip 2) și au raportat, pentru diabetul de tip 2, un OR de 2,24 pentru internarea în terapie intensivă și un OR de 2,46 pentru ventilația mecanică [8]. În rezumat, aceste valori nu sunt însoțite de intervale de încredere sau de valori p. Ce nu putem ști fără ele? Dacă OR 2,24 provine dintr-un interval de 1,9-2,6 sau de 0,8-6,3; în primul caz, asocierea ar fi sigură și estimată precis, în al doilea, compatibilă și cu absența oricărei asocieri. Cu 62 de pacienți în grupul cu diabet de tip 1 și cu un număr probabil mic de internări în terapie intensivă, intervalul este, foarte probabil, larg. Cititorul critic nu respinge rezultatul, dar îl tratează ca pe o estimare cu precizie necunoscută și caută intervalele în textul complet. Situația inversă, un interval suspect de îngust, apare în studiul lui Albai și colaboratorii asupra scorului ISARIC-4C la 159 de pacienți cu diabet de tip 2 și COVID-19, cu o mortalitate de 25,15% [75]. Sensibilitatea de 87,80% este raportată cu IC 95% 87,66-87,93%, adică o lățime de sub 0,3 puncte procentuale. Pentru o proporție estimată pe aproximativ 40 de pacienți decedați (25,15% din 159), eroarea standard ar fi în jur de $\sqrt{\frac{0,878 \times 0,122}{40}} = 0,052$, deci un interval Wilson de aproximativ ±10 puncte procentuale (în jur de 74-95%). Intervalele atât de înguste provin, probabil, dintr-o eroare de calcul sau de raportare și ilustrează de ce merită să verificați, măcar aproximativ, dacă lățimea unui interval se potrivește cu numărul de subiecți.

4.13 Principii de calcul al dimensiunii eșantionului

4.13.1 De ce se calculează dinainte

"Câți pacienți îmi trebuie?" este, probabil, întrebarea pe care un statistician o aude cel mai des de la clinicieni. Răspunsul nu poate fi "cât mai mulți" și nici "câți apucăm să strângem până la susținere". Un studiu prea mic produce intervale atât de largi încât nu răspunde la întrebare și expune pacienții la proceduri fără beneficiu științific. Un studiu inutil de mare consumă resurse și, uneori, expune mai mulți pacienți decât este necesar unui tratament inferior. Calculul dimensiunii eșantionului se face înainte de colectarea datelor, se scrie în protocol și este cerut de comisiile de etică, de finanțatori și de reviste [18,53].

Toate formulele de mai jos au aceeași logică: cerem o anumită precizie (o marjă de eroare acceptabilă) sau capacitatea de a detecta o anumită diferență, și pornim de la o estimare a variabilității, luată din studii anterioare sau dintr-un studiu-pilot. Dacă variabilitatea este subestimată, eșantionul calculat va fi prea mic.

4.13.2 Pentru estimarea unei medii

Dacă vrem ca marja de eroare a IC 95% (jumătatea lățimii) să fie cel mult E, pornim de la $E = \frac{z \times \sigma}{\sqrt{n}}$ și o rezolvăm în raport cu n:

$$n = \left(\frac{z \times \sigma}{E}\right)^2$$

Exemplu. Vrem să estimăm media TAS într-o populație de pacienți cu risc cardiometabolic cu o marjă de ± 3 mmHg(IC 95%). Din studii anterioare, abaterea standard este în jur de 14 mmHg.

$n = \left(\frac{1{,}96 \times 14}{3}\right)^2 = (9{,}147)^2 = 83{,}7$, rotunjit în sus la 84 de pacienți.

Rotunjirea se face întotdeauna în sus, altfel nu mai obținem precizia cerută. Observați sensibilitatea formulei: pentru o marjă de ±2 mmHg, $n = \left(\frac{1{,}96 \times 14}{2}\right)^2 = 188{,}2$, deci 189 de pacienți; pentru ±5 mmHg, doar 31. Înjumătățirea marjei cvadruplează eșantionul. Pentru glicemia a jeun, cu $\sigma \approx 15{,}76$ mg/dL și o marjă de ±3 mg/dL, ar fi nevoie de 107 pacienți.

4.13.3 Pentru estimarea unei proporții

Formula analogă pentru o proporție este:

$$n = \frac{z^2 \times p_i \times (1 - p_i)}{E^2}$$

unde p_i este proporția anticipată și E marja dorită (în fracție, nu în procente).

Exemplu. Un medic vrea să estimeze prevalența diabetului nediagnosticat la pacienții cu risc din lista lui. Anticipează o prevalență în jur de 20% și dorește o marjă de ±5 puncte procentuale.

$n = \frac{1{,}96^2 \times 0{,}20 \times 0{,}80}{0{,}05^2} = \frac{3{,}8416 \times 0{,}16}{0{,}0025} = 245{,}9$, deci 246 de pacienți.

Dacă nu avem nicio idee despre prevalență, folosim $p_i = 0{,}5$, care dă valoarea maximă a produsului p_i x (1 - p_i) și deci eșantionul cel mai mare, "sigur": $n = \frac{3{,}8416 \times 0{,}25}{0{,}0025} = 384{,}2$, deci 385. Pentru $p_i = 0{,}18$ și o marjă de ±3 puncte procentuale, n = 631. Formula se bazează pe aproximarea Wald; pentru proporții foarte mici (sub 5%) se folosesc variante mai exacte, disponibile în programe.

4.13.4 Pentru compararea a două medii: puterea și mărimea efectului

Când scopul este să comparăm două grupuri (de exemplu, un tratament nou față de cel standard), calculul cere patru ingrediente:

- pragul de semnificație α, de obicei 0,05 bilateral (z = 1,96);
- puterea dorită, 1 - β, de obicei 80% (z = 0,84) sau 90% (z = 1,28): probabilitatea ca studiul să detecteze diferența, dacă ea există cu adevărat;
- diferența minimă relevantă clinic Δ, pe care nu vrem să o ratăm;
- abaterea standard σ a variabilei.

Ultimele două se combină în mărimea standardizată a efectului, d Cohen = Δ / σ [76]. Convențional, d = 0,2 este un efect mic, 0,5 mediu, 0,8 mare. Formula aproximativă pentru numărul de pacienți în fiecare grup este:

n pe grup = 2 x (z(1-α/2) + z(1-β))² / d²

Pentru d = 0,5, α = 0,05 și putere 80%: $n = \frac{2\times(1,96+0,84)^2}{0,25} = \frac{2\times7,85}{0,25} = 62,8$. Programele, care folosesc distribuția t în locul aproximării normale, dau 64 de pacienți pe grup. Pentru d = 0,3, același calcul dă 174,4, iar programele 176 pe grup; pentru d = 0,8, 26 pe grup; pentru d = 0,2, 394 pe grup. Efectele mici cer eșantioane mari: pentru a detecta un efect de două ori mai mic, ne trebuie de patru ori mai mulți pacienți. Dacă cerem putere de 90% în loc de 80%, pentru d = 0,5 numărul crește de la 64 la 86 pe grup.

Un exemplu clinic. Vrem să știm dacă o intervenție scade TAS cu cel puțin 5 mmHg față de grupul de control, cu σ = 14 mmHg. $d = \frac{5}{14} = 0,36$, iar programele indică 125 de pacienți pe grup pentru putere 80%. Să ne întoarcem acum la comparația TAS între sexe din mg2026, cu 74 de bărbați și 86 de femei: puterea acestui eșantion de a detecta o diferență reală de 5 mmHg este de doar aproximativ 61%. Faptul că diferența observată nu a fost semnificativă (p = 0,81) nu exclude, deci, o diferență de 5 mmHg; intervalul de încredere (-4,9 la 3,8 mmHg) spune același lucru, mai direct.

Pentru două proporții, formula este mai lungă, dar logica rămâne. Pentru a detecta o creștere a proporției pacienților cu scădere ponderală ≥ 5% de la 11,3% (brațul A) la 29,6% (brațul C), cu α = 0,05 și putere 80%, ar fi nevoie de aproximativ 76 de pacienți pe braț. Brațele reale din mg2026 au câte 53-54 de pacienți, iar puterea lor pentru această diferență este de aproximativ 67%. Studiul a găsit totuși o diferență semnificativă (IC 95% pentru diferență 3,4-33,2 puncte procentuale), dar cu un interval larg, exact ce ne așteptăm de la un studiu cu putere moderată.

4.13.5 Ajustări practice: pierderile din studiu și populațiile mici

Pierderile din studiu (pacienți care abandonează, care nu mai pot fi contactați, care au date lipsă) reduc eșantionul analizat. Dacă anticipăm o rată de pierdere r, recrutăm n / (1 - r) pacienți. Pentru 64 de pacienți pe grup și pierderi estimate la 20%: 64 / 0,8 = 80 pe grup, adică 160 în total. Greșeala frecventă este să se înmulțească cu 1,2 (64 x 1,2 = 76,8, deci 77), ceea ce subestimează necesarul: din 77 recrutați, cu 20% pierderi rămân doar 61,6.

Corecția pentru populație finită se aplică atunci când eșantionul reprezintă o fracțiune importantă (peste 5-10%) dintr-o populație mică și cunoscută, de exemplu toți pacienții cu risc de pe lista unui medic de familie:

$$n \text{ corectat } = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

Pentru n0 = 246 și o listă de N = 1000 de pacienți cu risc, $n = \frac{246}{1+\frac{245}{1000}} = \frac{246}{1,245} = 197,6$, deci 198. Pentru N = 500, 166 de pacienți. Pentru populații mari (zeci de mii), corecția este neglijabilă.

4.13.6 Programe pentru calculul dimensiunii eșantionului

În practică, nimeni nu face aceste calcule de mână, dar este important să înțelegeți ce introduceți în program. G*Power este un program gratuit, folosit pe scară largă, care acoperă testele t, ANOVA, chi-pătrat, corelația și regresia [77]. Pentru două medii se alege familia "t tests", testul "Means: Difference between two independent means (two groups)", tipul de analiză "A priori" și se introduc mărimea efectului d, α, puterea și raportul dintre grupuri; pentru d = 0,5 rezultatul este 64 pe grup. MedCalc are un meniu dedicat (Sampling) pentru o medie, o proporție, compararea a două medii sau a două proporții, curbe ROC și altele [50]. Epi Info oferă, în modulul StatCalc, calculatoarele "Population survey" (pentru prevalențe, cu corecția pentru populație finită inclusă), "Cohort or cross-sectional" și "Unmatched case-control" [51]. Calculatoarele se folosesc în lucrarea practică 4.

4.13.7 Ce trebuie scris în protocol

Un paragraf corect despre dimensiunea eșantionului permite oricui să refacă exact calculul. El trebuie să conțină: variabila principală pe baza căreia s-a făcut calculul; diferența minimă relevantă clinic sau marja de precizie dorită, cu justificare; estimarea variabilității (abaterea standard sau proporția anticipată) și sursa ei; α și dacă testul este bilateral; puterea; rata de pierderi anticipată; numărul final de pacienți de recrutat; programul folosit. De exemplu: "Pentru a detecta o diferență de 5 mmHg a TAS între grupuri (abatere standard estimată 14 mmHg, după studiul X), cu un test t bilateral, α = 0,05 și putere 80%, sunt necesari 125 de pacienți pe grup (calcul în G*Power). Anticipând 15% pierderi, vom recruta 148 de pacienți pe grup."

Calculul ulterior al puterii (puterea "post-hoc", calculată din efectul observat) are o valoare limitată: este o funcție directă a valorii p și nu aduce informație nouă. Pentru un studiu deja încheiat, intervalul de încredere spune mai clar ce efecte sunt sau nu compatibile cu datele.

4.13.8 Consecințele eșantioanelor prea mici

Un studiu cu putere mică are trei probleme. Prima, evidentă: poate rata un efect real (eroare de tip II, capitolul 5). A doua, mai puțin evidentă: dacă totuși găsește un efect semnificativ, estimarea tinde să fie exagerată, pentru că doar estimările mari, întâmplătoare, trec pragul de semnificație. A treia: intervalele de încredere sunt atât de largi încât rezultatul, semnificativ sau nu, este greu de folosit clinic.

> **Studiu de caz: un studiu cu putere de 29,3%**
>
> Popovici și colaboratorii au comparat, într-o cohortă prospectivă de 100 de pacienți supuși ERCP, sedarea ghidată prin monitorizarea entropiei cu sedarea convențională [78]. Episoadele hipertensive au fost mai rare în grupul ghidat prin entropie (6% față de 26%, p = 0,007), iar scorurile Aldrete de recuperare au fost mai mari la 5 minute (p = 0,003) și la 15 minute (p < 0,001). Autorii semnalează însă onest o putere statistică limitată, de 29,3%. O putere de 29,3% înseamnă că, dacă diferența căutată ar exista cu adevărat, un studiu de această mărime ar avea doar aproximativ 3 șanse din 10 să o detecteze. Un rezultat nesemnificativ obținut în aceste condiții nu poate fi citit ca "nu există diferență"; este, pur și simplu, neconcludent. Designul observațional, nerandomizat, adaugă o a doua limită: grupurile pot diferi și prin alte caracteristici decât metoda de monitorizare. Concluzia corectă, pe care o formulează și titlul ("is associated with"), este că rezultatele generează ipoteze care trebuie confirmate într-un studiu randomizat, cu un eșantion calculat dinainte pentru variabila principală.

4.14 Probleme rezolvate

Problema rezolvată 4.1. IC 95% și IC 99% pentru media glicemiei

Enunț. În mg2026, glicemia a jeun inițială are n = 160, $m = 105{,}44$ mg/dL și $s = 15{,}76$ mg/dL. Calculați de mână IC 95% și IC 99% pentru media populației și interpretați.

Rezolvare. Distribuția glicemiei are o ușoară asimetrie (0,37), dar pentru n = 160 media este aproape normal distribuită (teorema limitei centrale), deci intervalul t este potrivit.

ES $= \frac{15{,}76}{\sqrt{160}} = \frac{15{,}76}{12{,}649} = 1{,}246$ mg/dL.

IC 95%: t(159; 0,975) = 1,975; marja = 1,975 x 1,246 = 2,461; intervalul 105,44 ± 2,46, adică 102,98-107,90 mg/dL.

IC 99%: t(159; 0,995) = 2,607; marja = 2,607 x 1,246 = 3,248; intervalul 105,44 ± 3,25, adică 102,19-108,69 mg/dL.

Verificare în Google Sheets: =CONFIDENCE.T(0,05; STDEV(S2:S161); 160) dă 2,461, iar =CONFIDENCE.T(0,01; STDEV(S2:S161); 160) dă 3,248.

Interpretare statistică. Cu o încredere de 95%, media glicemiei a jeun în populația de adulți cu risc din care provine eșantionul se află între 102,98 și 107,90 mg/dL; intervalul de 99% este mai larg, pentru că cerem o siguranță mai mare. Interpretare clinică. Întregul interval se află în zona glicemiei bazale modificate (100-125 mg/dL), ceea ce confirmă că populația studiată are, în medie, un metabolism glucidic alterat.

Problema rezolvată 4.2. Valori sau medie? Vârsta pacienților

Enunț. Vârsta pacienților din mg2026 are m = 56,29 ani, s = 8,77 ani, n = 160, distribuție normală (Shapiro-Wilk p = 0,687). (a) Estimați, cu o încredere de 95,4%, intervalul în care se află vârstele pacienților. (b) Estimați, cu o încredere de 95,4%, media vârstei în populație. (c) Comentați diferența.

Rezolvare. (a) Pentru valori individuale folosim abaterea standard: $m \pm 2s = 56{,}29 \pm 17{,}54$, adică 38,75-73,83 ani. În eșantion, 94,4% dintre pacienți au vârsta în acest interval.

(b) Pentru medie folosim eroarea standard: ES $= \frac{8{,}77}{\sqrt{160}} = 0{,}693$ ani; $m \pm 2\text{ES} = 56{,}29 \pm 1{,}39$, adică 54,90-57,68 ani.

(c) Primul interval are 35 de ani lățime și descrie diversitatea pacienților: de la adulți tineri la vârstnici. Al doilea are sub 3 ani și descrie cât de precis cunoaștem vârsta medie a populației. Raportul dintre lățimi este √160 ≈ 12,6. La examen, cuvintele-cheie din enunț sunt "valorile" (AS) față de "media" (ES).

Problema rezolvată 4.3. Un eșantion mic

Enunț. La 8 pacienți cu diabet de tip 2, HbA1c după trei luni de tratament are valorile (date fictive): 6,1; 5,8; 6,4; 7,0; 6,2; 5,9; 6,6; 6,0%. Calculați IC 95% pentru media HbA1c.

Rezolvare. Cu doar 8 valori nu putem verifica serios normalitatea; HbA1c are de obicei o distribuție aproximativ simetrică în grupuri omogene, iar valorile nu arată o valoare extremă evidentă, așa că folosim intervalul t, cu prudență.

Media: suma = 50,0, deci $m = \frac{50{,}0}{8} = 6{,}25\%$.

Abaterile față de medie: -0,15; -0,45; 0,15; 0,75; -0,05; -0,35; 0,35; -0,25. Pătratele: 0,0225; 0,2025; 0,0225; 0,5625; 0,0025; 0,1225; 0,1225; 0,0625; suma = 1,12.

$s^2 = \frac{1{,}12}{7} = 0{,}16$; s = 0,40%.

$\text{ES} = \frac{0{,}40}{\sqrt{8}} = \frac{0{,}40}{2{,}828} = 0{,}141\%$.

gl = 7; t(7; 0,975) = 2,365; marja = 2,365 x 0,141 = 0,334.

IC 95%: 6,25 ± 0,33, adică 5,92-6,58%.

Interpretare. Media populației de pacienți tratați similar se află, cu o încredere de 95%, între 5,92% și 6,58%. Dacă am fi folosit greșit z = 1,96, marja ar fi fost 0,28, iar intervalul 5,97-6,53%, cu aproximativ 17% mai îngust decât cel corect. Cu 8 pacienți, diferența dintre t și z nu este o subtilitate.

Problema rezolvată 4.4. IC pentru o proporție: Wald, Wilson și regula lui trei

Enunț. (a) Într-un audit, 12 din 80 de pacienți hipertensivi au TA necontrolată sub tratament. Calculați IC 95% pentru proporție prin metoda Wald și prin metoda Wilson. (b) Într-un alt cabinet, la 40 de pacienți tratați cu un nou antihipertensiv nu s-a observat niciun caz de angioedem. Ce puteți spune despre riscul real?

Rezolvare. $(a)\ p = \frac{12}{80} = 0{,}15$. $\text{ES} = \sqrt{\frac{0{,}15 \times 0{,}85}{80}} = \sqrt{0{,}001594} = 0{,}0399$.

Wald: $0{,}15 \pm 1{,}96 \times 0{,}0399 = 0{,}15 \pm 0{,}078$, adică 7,2%-22,8%.

Wilson: $\frac{z^2}{n} = \frac{3{,}8416}{80} = 0{,}0480$; $\frac{z^2}{2n} = 0{,}0240$. Centrul = (0,15 + 0,024) / 1,048 = 0,1660. Sub radical: $0{,}001594 + \frac{3{,}8416}{4 \times 6400} = 0{,}001594 + 0{,}000150 = 0{,}001744$; radicalul = 0,0418. Marja = $\left(\frac{1{,}96}{1{,}048}\right) \times 0{,}0418 = 1{,}870 \times 0{,}0418 = 0{,}0781$. IC 95% Wilson: 0,166 ± 0,078, adică 8,8%-24,4%.

Intervalul Wilson este deplasat spre 50% și asimetric față de 15% (6,2 puncte sub, 9,4 peste), cum se cuvine pentru o proporție mică într-un eșantion moderat.

(b) Zero evenimente din 40: regula lui trei dă limita superioară a IC 95% aproximativ 3 / 40 = 7,5%. Datele sunt compatibile cu un risc real de angioedem de până la aproximativ 1 caz la 13 pacienți. Absența cazurilor la 40 de pacienți nu dovedește siguranța medicamentului în privința unei reacții rare.

Problema rezolvată 4.5. Diferența a două proporții și raportul prevalențelor

Enunț. În mg2026, la includere, diabetul a fost prezent la 18 din 54 de pacienți obezi (33,3%) și la 11 din 106 pacienți neobezi (10,4%). Calculați IC 95% pentru diferența prevalențelor și pentru raportul prevalențelor.

Rezolvare. Diferența: 0,333 - 0,104 = 0,230 (23,0 puncte procentuale).

$\text{ES} = \sqrt{\frac{0{,}333 \times 0{,}667}{54} + \frac{0{,}104 \times 0{,}896}{106}} = \sqrt{0{,}004115 + 0{,}000878} = \sqrt{0{,}004993} = 0{,}0707$.

IC 95%: $0{,}230 \pm 1{,}96 \times 0{,}0707 = 0{,}230 \pm 0{,}139$, adică 9,1-36,8 puncte procentuale.

Raportul prevalențelor: $\text{RP} = \frac{0{,}333}{0{,}104} = 3{,}21$. Pe scala logaritmică: ln(3,21) = 1,167; $\text{ES}(\ln \text{RP}) = \sqrt{\frac{1}{18} - \frac{1}{54} + \frac{1}{11} - \frac{1}{106}} = \sqrt{0{,}0556 - 0{,}0185 + 0{,}0909 - 0{,}0094} = \sqrt{0{,}1185} = 0{,}344$. $\ln(\text{RP}) \pm 1{,}96 \times 0{,}344 = 1{,}167 \pm 0{,}675$, adică 0,492 până la 1,842; retransformat: $e^{0{,}492} = 1{,}64$ și $e^{1{,}842} = 6{,}31$.

IC 95% pentru RP: 1,64-6,31.

Interpretare. Prevalența diabetului este de aproximativ 3 ori mai mare la obezi, cu un interval compatibil cu un raport de la 1,6 la 6,3; în termeni absoluți, excesul este de 23 de puncte

procentuale (9-37). Ambele intervale exclud valoarea nulă (0, respectiv 1), deci asocierea este semnificativă statistic. Fiind o analiză transversală, ea nu stabilește cauzalitatea (capitolul 10).

Problema rezolvată 4.6. Dimensiunea eșantionului pentru un studiu în cabinet

Enunț. (a) Un medic de familie vrea să estimeze prevalența hipertensiunii necontrolate printre cei 1500 de pacienți hipertensivi de pe listă, anticipând 30%, cu o marjă de ±5 puncte procentuale. Estimează că 10% dintre pacienții invitați nu vor răspunde. (b) Un al doilea medic vrea să compare scăderea glicemiei între o intervenție nouă și consilierea standard: diferența minimă relevantă este 8 mg/dL, iar abaterea standard a scăderii, din mg2026, este în jur de 11 mg/dL; α = 0,05, putere 80%, pierderi estimate 15%.

Rezolvare. (a) Pas 1: $n_0 = \frac{1{,}96^2 \times 0{,}30 \times 0{,}70}{0{,}05^2} = \frac{3{,}8416 \times 0{,}21}{0{,}0025} = 322{,}7$, deci 323. Pas 2, corecția pentru populație finită: $n = \frac{323}{1+\frac{322}{1500}} = \frac{323}{1{,}215} = 265{,}8$, deci 266. Pas 3, nonrăspunsul: 266 / 0,90 = 295,6, deci 296 de pacienți invitați.

(b) $d = \frac{8}{11} = 0{,}73$. n pe grup = $\frac{2 \times (1{,}96+0{,}84)^2}{0{,}73^2} = \frac{15{,}7}{0{,}529} = 29{,}7$, deci 30 (G*Power, cu distribuția t, dă 31). Cu 15% pierderi: 31 / 0,85 = 36,5, deci 37 de pacienți pe grup, 74 în total.

Interpretare. Corecția pentru populație finită a redus eșantionul cu aproape 60 de pacienți, pentru că lista medicului este mică. Pentru al doilea studiu, efectul anticipat este mare (d = 0,73), deci eșantionul este modest; dacă diferența relevantă ar fi doar 4 mg/dL$(d = 0{,}36)$, ar fi nevoie de aproximativ 120 de pacienți pe grup, de patru ori mai mulți.

4.15 Exerciții propuse

1. Colesterolul total din mg2026 are n = 160, $m = 209{,}39$ mg/dL, $s = 36{,}91$ mg/dL. Calculați eroarea standard și IC 95% pentru media populației (t = 1,975).
2. HDL din mg2026 are $m = 49{,}52$ mg/dL, $s = 11{,}78$ mg/dL, n = 160. Calculați intervalele de 68,3%, 95,4% și 99,7% pentru valorile individuale și pentru medie și prezentați-le într-un tabel ca tabelul 4.1.
3. Într-un eșantion de 25 de pacienți cu diabet, HbA1c are media 7,2% și abaterea standard 1,5%. Calculați IC 95% pentru medie. Ce valoare critică ați folosit și de ce?
4. În mg2026, 41 din 160 de pacienți au TAS ≥ 140 mmHg. Calculați IC 95% pentru proporție prin metoda Wald și prin metoda Wilson.
5. Un vaccin nou a fost administrat la 120 de voluntari, fără nicio reacție anafilactică. Care este limita superioară aproximativă a IC 95% pentru riscul de anafilaxie?
6. Două grupuri independente: n1 = 40, m1 = 142 mmHg, s1 = 15 mmHg; n2 = 45, m2 = 134 mmHg, s2 = 14 mmHg. Calculați diferența mediilor, eroarea ei standard și IC 95% aproximativ (cu z = 1,96). Este diferența semnificativă?
7. Într-un studiu caz-control, 20 din 50 de cazuri și 10 din 50 de controale au fost expuse. Calculați OR și IC 95% pe scala logaritmică.
8. Câți pacienți sunt necesari pentru a estima: (a) o medie cu σ = 10 și marja ±2 unități (IC 95%); (b) o proporție complet necunoscută cu marja ±4 puncte procentuale?
9. Pentru compararea a două medii cu d = 0,4, α = 0,05 bilateral și putere 80%, calculați numărul de pacienți pe grup cu formula aproximativă. Câți ar trebui recrutați dacă se anticipează 20% pierderi?

10. Un articol raportează "TAS $131 \pm 1{,}1$ mmHg$(n = 160)$". Ce reprezintă, cel mai probabil, valoarea 1,1? Calculați abaterea standard corespunzătoare. Un alt articol raportează OR = 1,75 (IC 95% 0,95-3,10): este asocierea semnificativă la pragul 0,05?
11. Un IC 95% pentru o medie, calculat pe 40 de pacienți, are lățimea de 10 mmHg. Aproximativ câți pacienți ar fi necesari pentru o lățime de 5 mmHg, presupunând aceeași abatere standard?

Capcane frecvente

- Confundarea abaterii standard cu eroarea standard: AS descrie pacienții, ES descrie precizia mediei. "Media ± ES" face datele să pară mai puțin variabile.
- Folosirea intervalului pentru medie când întrebarea este despre valorile individuale (și invers), mai ales la problemele de tip "estimați valorile/media cu o încredere de 95,4%".
- Interpretarea IC 95% ca "95% dintre pacienți se află în interval" sau ca "probabilitatea de 95% ca media să fie în acest interval anume".
- Folosirea lui z = 1,96 în locul valorii t pentru eșantioane mici.
- Raportarea intervalului Wald pentru proporții mici sau eșantioane mici, cu limite negative sau acoperire insuficientă.
- Concluzia "nu există efect" pe baza unui interval larg care include valoarea nulă; corect este "datele nu permit o concluzie".
- Calcularea intervalului pentru OR sau RR direct pe scala originală, simetric, în loc de scala logaritmică.
- Rotunjirea dimensiunii eșantionului în jos sau ajustarea pentru pierderi prin înmulțire cu (1 + r) în loc de împărțire la (1 - r).
- Calcularea puterii "post-hoc" din efectul observat pentru a "explica" un rezultat nesemnificativ.
- Ignorarea faptului că intervalul de încredere măsoară doar eroarea aleatoare; un eșantion deplasat dă intervale înguste în jurul unei valori greșite.

De reținut

- Eroarea standard a mediei este ES $= \frac{s}{\sqrt{n}}$; pentru TAS din mg2026, 1,097 mmHg, față de AS 13,88 mmHg.
- Intervalul pentru valori individuale este $m \pm k \times s$, intervalul pentru medie este m ± k x ES (k = 1, 2, 3 pentru 68,3%, 95,4%, 99,7%).
- IC pentru medie: $m \pm \frac{t \times s}{\sqrt{n}}$, cu gl = n - 1; pentru TAS, IC 95% 128,96-133,30 mmHg și IC 99% 128,27-133,99 mmHg.
- Un IC 95% înseamnă că procedura de calcul prinde parametrul real în 95% din studiile repetate; lățimea lui scade cu $\sqrt{n}$.
- Pentru proporții se preferă intervalul Wilson (diabet în mg2026: 12,9-24,8%, față de Wald 12,2-24,1%); pentru zero evenimente din n, limita superioară este aproximativ 3 / n.
- Intervalele pentru OR, RR și media geometrică se calculează pe scala logaritmică și sunt asimetrice; estimarea punctuală este media geometrică a limitelor.
- Dacă IC 95% al unei diferențe exclude 0 (sau al unui raport exclude 1), p < 0,05; intervalul arată în plus mărimea și precizia efectului.
- Dimensiunea eșantionului: $n = \left(\frac{z \times \sigma}{E}\right)^2$ pentru o medie (84 pentru ± 3 mmHg, σ 14), $n = \frac{z^2 \times p_i \times (1-p_i)}{E^2}$ pentru o proporție (246 pentru 20% ± 5%), aproximativ 2 x (1,96 + 0,84)² / d² pe grup pentru două medii (64 pentru d = 0,5; 176 pentru d = 0,3).

Întrebări de autoevaluare

1. (CS) În mg2026, TAS are s = 13,88 mmHg și n = 160. Eroarea standard a mediei este aproximativ:
    a) 0,087 mmHg
    b) 1,097 mmHg
    c) 3,72 mmHg
    d) 13,88 mmHg
    e) 27,2 mmHg
2. (CS) Pentru a înjumătăți lățimea unui interval de încredere pentru medie, păstrând aceeași variabilitate și același nivel de încredere, eșantionul trebuie:
    a) dublat
    b) triplat
    c) mărit de patru ori
    d) mărit de opt ori
    e) redus la jumătate
3. (CS) Un studiu raportează OR = 1,40 (IC 95% 0,90-2,18). Afirmația corectă este:
    a) asocierea este semnificativă statistic la pragul 0,05
    b) intervalul include valoarea 1, deci $p \geq 0,05$
    c) intervalul dovedește absența oricărei asocieri
    d) intervalul ar trebui să fie simetric în jurul lui 1,40
    e) valoarea p este sigur mai mică de 0,01
4. (CS) Într-un studiu pe 60 de pacienți nu s-a observat niciun eveniment advers grav. Limita superioară aproximativă a IC 95% pentru riscul real este:
    a) 0%
    b) 1,7%
    c) 3%
    d) 5%
    e) 10%
5. (CS) Câți pacienți sunt necesari pentru a estima o medie cu o marjă de ± 3 mmHg(IC 95%), dacă σ = 14 mmHg?
    a) 19
    b) 47
    c) 84
    d) 176
    e) 246
6. (CM) Despre interpretarea unui IC 95% pentru media TAS de 128,96-133,30 mmHg sunt corecte afirmațiile:
    a) dacă studiul s-ar repeta de multe ori, aproximativ 95% dintre intervalele calculate astfel ar conține media reală
    b) 95% dintre pacienți au TAS între 128,96 și 133,30 mmHg
    c) intervalul exprimă precizia estimării mediei
    d) un IC 99% calculat pe aceleași date ar fi mai larg
    e) intervalul ar deveni mai îngust dacă eșantionul ar fi mai mare
7. (CM) Despre abaterea standard (AS) și eroarea standard (ES) sunt corecte afirmațiile:

a) AS descrie variabilitatea valorilor individuale
b) ES scade când crește volumul eșantionului
c) $\text{ES} = \frac{\text{AS}}{\sqrt{n}}$
d) intervalul de referință pentru valori individuale se construiește cu ES
e) raportarea "media ± ES" în locul "media ± AS" face datele să pară mai puțin variabile

8. (CM) Despre intervalul de încredere pentru o proporție sunt corecte afirmațiile:
   a) metoda Wald poate da limite sub 0% pentru proporții mici
   b) metoda Wilson dă intervale asimetrice pentru proporții apropiate de 0
   c) metoda Wald are acoperire insuficientă în eșantioane mici
   d) intervalul Wilson pentru diabet în mg2026 este 12,9-24,8%
   e) intervalul Wald este întotdeauna mai larg decât cel Wilson
9. (CM) Care dintre următorii factori măresc dimensiunea necesară a eșantionului pentru compararea a două medii?
   a) o diferență minimă relevantă mai mică
   b) o abatere standard mai mare
   c) o putere dorită mai mare (90% în loc de 80%)
   d) un prag α mai mare (0,10 în loc de 0,05)
   e) o rată anticipată de pierderi mai mare
10. (CM) Despre intervalele de încredere pentru rapoarte (OR, RR) sunt corecte afirmațiile:

a) se calculează pe scala logaritmică și se retransformă
b) sunt de obicei asimetrice în jurul estimării punctuale
c) estimarea punctuală este aproximativ media geometrică a limitelor
d) un interval care exclude 1 corespunde unui rezultat semnificativ la pragul de 5%
e) un interval foarte larg indică un efect foarte mare

Capitolul 5. Bazele testării statistice: ipoteze, prag de semnificație, erori de tip I și II, decizia și interpretarea

Obiective

La finalul acestui capitol, studentul va putea:

- să transforme o întrebare clinică într-o ipoteză de cercetare și apoi într-o pereche de ipoteze statistice H0 și H1, scrise în cuvinte și matematic;
- să aleagă între un test bilateral și unul unilateral și să justifice alegerea înainte de a vedea datele;
- să definească corect valoarea p și să recunoască cele mai frecvente interpretări greșite ale ei;
- să aplice regula de decizie pentru un prag α dat și să formuleze concluzia statistică fără expresia "se acceptă H0";
- să descrie erorile de tip I și II, să explice legătura dintre α, β, puterea 1 - β și nivelul de încredere 1 - α și să enumere factorii care influențează puterea;
- să calculeze și să interpreteze o mărime a efectului (d Cohen, dz, r, eta², OR) și să deosebească semnificația statistică de importanța clinică;
- să rezolve de mână un test z și un test t pentru o singură medie, folosind trei căi echivalente: valoarea critică, valoarea p și intervalul de încredere;
- să explice inflația erorii de tip I la comparații multiple și să aplice corecțiile Bonferroni, Holm și Benjamini-Hochberg;
- să raporteze un test statistic după standardele actuale (valoarea p exactă, mărimea efectului, intervalul de încredere) și să recunoască practicile de tip p-hacking și HARKing.

5.1 De la întrebarea clinică la ipoteza testabilă

Capitolele anterioare au descris datele (statistica descriptivă), au arătat cum se comportă o variabilă normală și cum se construiește un interval de încredere pentru o medie sau o proporție. De aici începe partea a doua a cursului, statistica inferențială în sens strict: luăm decizii despre o populație pe baza unui singur eșantion. Toate testele din capitolele 6-9 (testul t, ANOVA, Mann-Whitney, chi-pătrat, testul corelației) folosesc aceeași logică, pe care o construim acum o singură dată, cu grijă. Dacă logica este înțeleasă, fiecare test nou înseamnă doar o formulă nouă pentru "statistica testului"; restul raționamentului rămâne neschimbat [19,63].

5.1.1 Întrebarea de cercetare și ipoteza de cercetare

Orice studiu pornește de la o întrebare clinică. Medicul de familie care a urmărit timp de șase luni pacienții din studiul mg2026 se întreabă dacă aplicația mobilă, adăugată la consiliere, scade glicemia a jeun mai mult decât consilierea singură. Întrebarea este bună, dar încă vagă: ce înseamnă "mai mult", la cine, după cât timp, măsurat cum? Prima muncă a cercetătorului este să precizeze populația, intervenția, comparatorul și rezultatul (schema PICO, discutată în volumul de informatică medicală [1]). Rezultă o ipoteză de cercetare, o afirmație despre lumea reală pe care studiul încearcă să o susțină: "la adulții cu risc cardiometabolic din cabinetele de medicină de familie, adăugarea aplicației de automonitorizare la consiliere produce, în șase luni, o scădere a glicemiei a jeun mai mare decât consilierea singură".

Ipoteza de cercetare este formulată în limbajul clinicianului și exprimă, de obicei, ceea ce cercetătorul se așteaptă să găsească: o diferență, o asociere, un efect. Ea nu poate fi testată direct, pentru că nu avem acces la toți pacienții cu risc cardiometabolic din țară. Avem 160 de pacienți.

Pentru a trece de la acești 160 la populație, ipoteza de cercetare trebuie tradusă într-o formă pe care un calcul o poate verifica: o pereche de ipoteze statistice despre parametrii populației (medii μ, proporții π, coeficienți de corelație ρ).

5.1.2 Ipotezele statistice: H0 și H1

Ipoteza nulă (H0, ipoteza de zero) afirmă că în populație nu există diferența, asocierea sau efectul căutat. Diferența observată în eșantion ar fi atunci doar rezultatul variabilității de eșantionare, adică al întâmplării care decide ce pacienți ajung în studiu. Ipoteza alternativă (H1, uneori notată Ha) neagă ipoteza nulă: în populație diferența există. Cele două ipoteze se exclud reciproc și acoperă împreună toate posibilitățile.

De ce pornim de la H0, adică de la ipoteza pe care, de cele mai multe ori, sperăm să o infirmăm? Pentru că H0 este singura care fixează o valoare precisă a parametrului (de exemplu, diferența dintre medii egală cu zero). Numai pornind de la o valoare precisă putem calcula cât de probabile ar fi datele noastre. Ipoteza alternativă "diferența nu este zero" cuprinde o infinitate de valori posibile și nu ne permite un calcul unic. Raționamentul seamănă cu o demonstrație prin reducere la absurd: presupunem că H0 este adevărată, calculăm ce ar trebui să vedem în acest caz și, dacă datele reale sunt foarte greu de împăcat cu presupunerea, renunțăm la ea.

Pentru exemplul cu aplicația, ipotezele se scriu astfel:

H0: scăderea medie a glicemiei la 6 luni este aceeași în populația tratată cu consiliere + aplicație și în populația tratată doar cu consiliere; matematic, μB = μA.

H1: scăderea medie a glicemiei diferă între cele două populații; matematic, μB ≠ μA.

Observați că ipotezele se referă la populație (litere grecești μ, π, ρ), nu la eșantion. Mediile din eșantion le cunoaștem deja (3,0 mg/dL în brațul A și 11,1 mg/dL în brațul B); nu are rost să facem ipoteze despre ele. La examen, scrierea ipotezelor cu m1 și m2 (mediile eșantioanelor) în loc de μ1 și μ2 este o greșeală frecventă și se depunctează.

Tabelul 5.1. Exemple de traducere a întrebărilor clinice în ipoteze statistice (setul mg2026 și exemple clinice).

Întrebarea clinică	H0 (în cuvinte)	H0 (matematic)	H1 (matematic)
Media TAS a pacienților diferă de 125 mmHg?	media populației este 125 mmHg	$\mu = 125$	μ ≠ 125
TAS diferă între femei și bărbați?	mediile celor două sexe sunt egale	$\mu_F = \mu_M$	μF ≠ μM
Glicemia se modifică între includere și 6 luni (aceiași pacienți)?	media diferențelor individuale este zero	$\mu_D = 0$	μD ≠ 0
Scăderea glicemiei diferă între cele trei brațe?	cele trei medii sunt egale	$\mu_A = \mu_B = \mu_C$	cel puțin o medie diferă
Proporția celor cu scădere ≥5% în greutate diferă între brațele C și A?	proporțiile sunt egale	$\pi_C = \pi_A$	πC ≠ πA
IMC și TAS sunt corelate?	coeficientul de corelație din populație este zero	$\rho = 0$	ρ ≠ 0

5.1.3 Teste bilaterale și unilaterale

Ipoteza alternativă din exemplele de mai sus este bilaterală (engl. two-sided): ea admite o diferență în oricare sens, μB > μA sau μB < μA. Un test unilateral (engl. one-sided) privește un singur sens: H1: μB > μA. Pentru același set de date, valoarea p unilaterală este jumătate din valoarea p bilaterală, dacă diferența observată merge în sensul anunțat. Tocmai aici stă tentația: un

rezultat cu p bilateral 0,08 devine "semnificativ" (p = 0,04) dacă decidem, după ce am văzut datele, că testul "ar fi trebuit" să fie unilateral.

Un test unilateral se justifică rar și numai dacă sunt îndeplinite simultan trei condiții. Prima: sensul este fixat înainte de colectarea datelor, în protocol. A doua: o diferență în sensul opus ar fi imposibilă biologic sau, dacă ar apărea, ar avea aceeași consecință practică precum absența diferenței. A treia: cititorul este informat explicit. Exemplul clasic este studiul de non-inferioritate, în care întrebarea este doar dacă noul tratament nu este mai slab cu mai mult decât o marjă acceptată (secțiunea 5.12). În studiile obișnuite din medicina generală, un medicament nou poate să și înrăutățească evoluția pacientului; de aceea testul implicit, cerut de majoritatea revistelor și folosit în acest curs, este cel bilateral [42,53]. Programele (JASP, MedCalc, Google Sheets) oferă implicit varianta bilaterală; dacă alegeți altceva, trebuie să puteți spune de ce.

5.2 Logica testului: statistica testului și distribuția ei sub H0

Să presupunem că vrem să știm dacă media TAS în populația din care provin cei 160 de pacienți diferă de 125 mmHg, o valoare de referință aleasă pentru exemplu. Media observată este 131,13 mmHg. Diferența de 6,13 mmHg ar putea fi reală, dar ar putea fi și rodul întâmplării: dacă media populației ar fi chiar 125, un eșantion de 160 de pacienți nu ar avea exact media 125, ci o valoare apropiată, mai mare sau mai mică. Întrebarea corectă devine: cât de neobișnuită este o abatere de 6,13 mmHg, dacă H0 ar fi adevărată?

Ca să răspundem, comparăm diferența observată ("semnalul") cu variabilitatea pe care am aștepta-o doar din întâmplare ("zgomotul"). Măsura zgomotului pentru o medie este eroarea standard a mediei, ES $= \frac{s}{\sqrt{n}}$, pe care ați întâlnit-o în capitolul 4. Raportul dintre semnal și zgomot se numește statistica testului:

$$t = \frac{m - \mu_0}{\text{ES}} = \frac{m - \mu_0}{\frac{s}{\sqrt{n}}}$$

Pentru TAS: ES $= \frac{13{,}88}{\sqrt{160}} = \frac{13{,}88}{12{,}649} = 1{,}097$ mmHg, deci $t = \frac{131{,}13 - 125}{1{,}097} = \frac{6{,}13}{1{,}097} = 5{,}59$. Media observată se află la aproape 5,6 erori standard de valoarea afirmată de H0.

Statistica testului are o calitate decisivă pentru raționament: știm cum se distribuie atunci când H0 este adevărată. Dacă am repeta studiul de foarte multe ori, cu eșantioane de 160 de pacienți dintr-o populație cu media exact 125, valorile t obținute ar urma o distribuție Student cu n - 1 = 159 grade de libertate, foarte asemănătoare cu distribuția normală standard: centrată în zero, simetrică, cu aproximativ 95% din valori între -1,975 și +1,975. O valoare t = 5,59 se află mult în afara acestei zone. Sub H0, ea ar fi extraordinar de rară.

Fiecare test are statistica lui (z, t, F, χ^2, U, H), dar ideea este aceeași: un număr care crește când datele se îndepărtează de H0 și a cărui distribuție sub H0 este cunoscută, din teorie sau din tabele. Valorile statisticii testului atât de extreme încât, sub H0, apar cu o probabilitate mai mică decât un prag ales (de obicei 5%) formează regiunea critică sau regiunea de respingere. Granița ei se numește valoare critică.

5.3 Valoarea p

5.3.1 Definiția corectă

Valoarea p este probabilitatea de a obține, prin simplul joc al întâmplării, o statistică a testului cel puțin la fel de extremă ca aceea observată, presupunând că ipoteza nulă este adevărată și că toate condițiile de aplicare ale testului sunt respectate. Fiecare cuvânt din definiție contează. "Cel puțin

la fel de extremă" înseamnă că adunăm probabilitatea valorii observate și a tuturor valorilor mai depărtate de H0; în testul bilateral, adunăm ambele cozi ale distribuției. "Presupunând că H0 este adevărată" înseamnă că p se calculează într-o lume ipotetică în care H0 este adevărată; valoarea p nu ne spune nimic despre probabilitatea ca acea lume să fie cea reală.

Pentru TAS, valoarea p bilaterală asociată cu t = 5,59 și 159 grade de libertate este aproximativ 0,0000001 (în Google Sheets, =T.DIST.2T(5,59;159)). Cu alte cuvinte, dacă media populației ar fi 125 mmHg, doar aproximativ unul din zece milioane de eșantioane de 160 de pacienți ar avea o medie depărtată de 125 cel puțin cât a noastră. O astfel de valoare se raportează "p < 0,001", fiindcă mai multe zecimale nu mai aduc informație utilă.

5.3.2 Ce nu este valoarea p

Valoarea p este, probabil, cel mai des interpretată greșit număr din literatura medicală. Asociația Americană de Statistică a publicat în 2016 o declarație oficială, cu șase principii, tocmai pentru a opri aceste confuzii [74]. Un articol de referință al lui Greenland și colaboratorii enumeră peste douăzeci de interpretări eronate întâlnite curent în articole publicate [70]. Le reținem pe cele care apar și la examenul de licență și rezidențiat.

Valoarea p nu este probabilitatea ca H0 să fie adevărată. Un p = 0,03 nu înseamnă "3% șanse ca tratamentul să nu aibă efect". Probabilitatea ipotezei, dată fiind evidența, cere un alt tip de calcul (raționament bayesian, pe care îl veți întâlni la testele diagnostice în capitolul 11), care ține cont și de cât de plauzibilă era ipoteza înainte de studiu.

Valoarea p nu este probabilitatea ca rezultatul să fie "datorat întâmplării". Formularea din multe cursuri vechi ("p este probabilitatea ca diferențele observate să fi apărut din întâmplare") este o prescurtare comodă, dar înșelătoare. Corect este: p este probabilitatea unor date atât de extreme dacă singurul mecanism în joc ar fi întâmplarea.

Valoarea p nu măsoară mărimea efectului sau importanța lui clinică. Un p = 0,0001 poate însoți o diferență de 0,04 dioptrii, fără nicio relevanță pentru pacient, dacă eșantionul este mare; un p = 0,20 poate însoți o reducere a mortalității cu o treime, dacă eșantionul este mic. Vom reveni cu exemple în secțiunea 5.6.

Un p > 0,05 nu dovedește că H0 este adevărată. Lipsa unei diferențe semnificative statistic nu înseamnă că diferența lipsește; înseamnă că datele nu au fost suficiente pentru a o pune în evidență. Secțiunea 5.4.2 este dedicată acestei idei.

Pragul de 0,05 nu este o graniță naturală. Diferența dintre p = 0,049 și p = 0,051 este neglijabilă ca evidență, deși convenția le pune de o parte și de alta a deciziei. Un rezultat nu devine "adevărat" când p trece sub 0,05 și "fals" când trece peste.

Valoarea p nu spune nimic despre calitatea designului. Un studiu cu selecție părtinitoare a pacienților, cu măsurători greșite sau cu factori de confuzie necontrolați poate produce valori p foarte mici pentru efecte care nu există. Testul statistic verifică doar dacă datele, așa cum sunt, se împacă cu H0; nu verifică dacă datele sunt bune.

5.3.3 Valoarea p exactă și raportarea ei

În trecut, când valoarea p se căuta în tabele tipărite, autorii scriau doar "p < 0,05" sau "NS" (nesemnificativ). Astăzi programele calculează p exact, iar ghidurile de raportare cer valoarea exactă, cu două sau trei zecimale: p = 0,034, p = 0,81, p = 0,0008. Pentru valori foarte mici se scrie p < 0,001. Nu se scrie niciodată p = 0,000, așa cum afișează uneori programele când rotunjesc: probabilitatea nu este zero, ci doar mai mică decât 0,0005. Scrierea exactă permite cititorului să judece singur cât de aproape de prag a fost rezultatul și face posibilă includerea studiului într-o meta-analiză.

5.4 Pragul de semnificație α și regula de decizie

5.4.1 Pragul α și decizia

Pragul de semnificație α este probabilitatea maximă de eroare de tip I (respingerea unei H0 adevărate) pe care cercetătorul o acceptă. El se fixează înainte de analiză, în protocol. Valorile convenționale sunt 0,05, 0,01 și 0,001; în medicina clinică, α = 0,05 este standardul, iar pragurile mai stricte apar când consecințele unui fals pozitiv sunt grave (de exemplu, aprobarea unui medicament) sau când se fac foarte multe teste (studiile genetice folosesc praguri de ordinul 0,00000005).

Regula de decizie este simplă:

- dacă p < α, respingem H0; spunem că rezultatul este semnificativ statistic la pragul α;
- dacă p ≥ α, nu respingem H0; rezultatul nu este semnificativ statistic la pragul α.

Pentru TAS față de 125 mmHg, p < 0,001 < 0,05, deci respingem H0: media TAS în populația de pacienți cu risc cardiometabolic din care provine eșantionul diferă semnificativ de 125 mmHg (este mai mare). Dacă repetăm testul cu valoarea de referință 130 mmHg, obținem $t = \frac{131{,}13-130}{1{,}097} = 1{,}03$ și p = 0,304. Acum p > 0,05: nu respingem H0, iar datele sunt compatibile cu o medie a populației de 130 mmHg. Același eșantion, aceeași medie, două concluzii diferite, pentru că am pus două întrebări diferite.

5.4.2 De ce "nu respingem H0" și nu "acceptăm H0"

Formularea "acceptăm H0", încă prezentă în unele materiale didactice, sugerează că am demonstrat absența diferenței. Nu am demonstrat nimic de acest fel. Testul statistic poate aduce evidență împotriva lui H0, dar nu poate aduce evidență în favoarea ei. Altman și Bland au rezumat ideea într-un titlu devenit proverbial: absența dovezii nu este dovada absenței [79].

Un exemplu din setul mg2026 face lucrul vizibil. Să luăm doar 10 pacienți la întâmplare și să testăm din nou dacă media TAS diferă de 125 mmHg. Cei 10 pacienți au valorile 150, 154, 131, 130, 137, 102, 110, 103, 147 și 136 mmHg, cu media 130,0 și abaterea standard 19,04. Eroarea standard este 19,04 / √10 = 6,02, iar $t = \frac{130{,}0-125}{6{,}02} = 0{,}83$, cu 9 grade de libertate; p = 0,43. Nu respingem H0. Ar fi absurd să "acceptăm" că media este 125 mmHg, de vreme ce pe toți cei 160 de pacienți am găsit o diferență extrem de semnificativă. Intervalul de încredere de 95% pentru media celor 10 pacienți, 116,38-143,62 mmHg, arată de ce: el include 125, dar include la fel de bine 140. Eșantionul mic este compatibil cu aproape orice valoare plauzibilă. Concluzia corectă este: datele nu permit respingerea lui H0; diferența, dacă există, nu a putut fi pusă în evidență cu acest eșantion.

În cuvintele unui clinician, un rezultat nesemnificativ poate însemna două lucruri foarte diferite: efect absent sau efect nedeterminat. Le deosebim privind intervalul de încredere. Dacă intervalul este îngust și cuprinde doar diferențe neimportante clinic, putem spune că un efect relevant este puțin probabil. Dacă intervalul este larg și cuprinde atât zero, cât și diferențe importante, rezultatul este nedeterminat, iar studiul a fost, probabil, prea mic.

5.4.3 Scala verbală a cursului și limitele ei

În tradiția disciplinei, decizia se însoțește de o scală verbală care descrie cât de mică este valoarea p [14,19]:

Tabelul 5.2. Scala verbală folosită în curs pentru α = 0,05.

Valoarea p	Decizia	Formularea
$p \geq 0,05$	nu se respinge H0	diferență nesemnificativă statistic
$0,01 \leq p < 0,05$	se respinge H0	diferență semnificativă
$0,001 \leq p < 0,01$	se respinge H0	diferență foarte semnificativă
$p < 0,001$	se respinge H0	diferență extrem de semnificativă

Scala este utilă la temele și la examenul practic, pentru că obligă studentul să citească valoarea p cu atenție. Are însă două limite pe care trebuie să le cunoașteți. Prima: "extrem de semnificativ" descrie cât de puternică este evidența împotriva lui H0, nu cât de mare sau de important este efectul. Înălțimea diferă între femeile și bărbații din mg2026 cu 13,4 cm ($p < 0,001$), iar TAS diferă între cei 160 de pacienți și valoarea de 125 mmHg tot cu $p < 0,001$, deși în al doilea caz diferența este de 6 mmHg; scala le pune pe aceeași treaptă. A doua: scala nu are trepte sub "nesemnificativ", deci nu face diferența între un rezultat nedeterminat și unul care exclude un efect relevant. De aceea, în lucrările științifice, eticheta verbală nu înlocuiește niciodată valoarea p exactă, mărimea efectului și intervalul de încredere.

Două studii ale colegilor din disciplină arată cum arată în practică aceste trepte. Într-un sondaj printre 121 de studenți la medicină despre competențele în știința datelor, corelația dintre cunoștințele măsurate și cele autoevaluate a fost $R = 0,107$, cu $p = 0,246$, deci nesemnificativă, în timp ce corelația dintre aprecierile a două cursuri a fost $R = 0,869$, $p < 0,001$, extrem de semnificativă [16]. Rezultatul nesemnificativ nu dovedește că studenții nu își pot evalua cunoștințele; el spune că, pe 121 de respondenți, o asociere slabă nu a putut fi deosebită de zero. Intervalul de încredere al lui R, dacă ar fi fost raportat, ar fi arătat cât de largă este zona de incertitudine. În studiul multicentric asupra pacienților hemodializați din 10 centre românești, comparând cohorta din 2015 (n = 1401) cu cea din 2019 (n = 1698), mortalitatea la un an a fost 9,9% față de 10,7% ($p = 0,46$, nesemnificativ), iar prevalența infecției cu virusul hepatitei C a scăzut de la 16,9% la 10,5% ($p < 0,0001$, extrem de semnificativ) [39]. Pe un eșantion de peste 3000 de pacienți, o diferență nesemnificativă de 0,8 puncte procentuale în mortalitate are o interpretare mult mai solidă decât aceeași diferență pe 100 de pacienți: intervalul de încredere este îngust, deci o modificare importantă a mortalității este puțin probabilă.

5.5 Erorile de tip I și II, puterea testului

5.5.1 Cele patru situații posibile

Decizia statistică se ia pe baza unui eșantion, deci poate fi greșită. Adevărul despre populație (H0 adevărată sau falsă) nu îl cunoaștem; decizia noastră (respingem sau nu respingem H0) o cunoaștem. Combinând cele două, apar patru situații.

Tabelul 5.3. Cele patru rezultate posibile ale unui test statistic.

Decizia luată	În realitate, H0 este adevărată (efectul nu există)	În realitate, H0 este falsă (efectul există)
Respingem H0	eroare de tip I (fals pozitiv); probabilitatea α	decizie corectă; probabilitatea 1 - β (puterea)
Nu respingem H0	decizie corectă; probabilitatea 1 - α (nivelul de încredere)	eroare de tip II (fals negativ); probabilitatea β

Eroarea de tip I înseamnă să declarăm un efect care nu există: un medicament ineficient pare eficient, un factor fără legătură cu boala pare factor de risc. Probabilitatea ei, atunci când H0 este adevărată, este exact α, pragul pe care l-am ales. Eroarea de tip II înseamnă să ratăm un efect care există: un tratament util este abandonat pentru că studiul "nu a arătat diferențe". Probabilitatea ei

se notează β. Complementul lui β, 1 - β, este puterea testului: probabilitatea de a detecta efectul, atunci când el există cu adevărat și are o anumită mărime. Complementul lui α, 1 - α, este nivelul de încredere: probabilitatea de a nu respinge H0 când ea este adevărată; aceeași cantitate apare în denumirea intervalului de încredere de 95%.

5.5.2 Două analogii: verdictul în instanță și testul diagnostic

Cea mai veche analogie este procesul penal. Acuzatul este prezumat nevinovat (H0) până când probele îl contrazic dincolo de orice îndoială rezonabilă. Condamnarea unui nevinovat este eroarea de tip I; achitarea unui vinovat este eroarea de tip II. Sistemul juridic tolerează mai multe achitări greșite decât condamnări greșite, așa cum statistica medicală fixează strict α (0,05) și acceptă un β mai mare (de obicei 0,20). Mai observați un detaliu: verdictul este "nevinovat" sau, mai exact, "nu s-a dovedit vinovăția", nu "s-a dovedit nevinovăția". Exact aceeași nuanță separă "nu respingem H0" de "acceptăm H0".

A doua analogie vă va fi familiară din capitolul despre testele diagnostice. Testul statistic se comportă ca un test de laborator aplicat ipotezei: "pozitiv" înseamnă respingerea lui H0, "boala" înseamnă existența efectului. Eroarea de tip I este un rezultat fals pozitiv, iar α corespunde lui 1 - specificitate; eroarea de tip II este un fals negativ, iar puterea 1 - β corespunde sensibilității. La fel ca la un test diagnostic, cele două erori sunt legate: dacă vrem mai puține rezultate fals pozitive (α mai mic), vom avea, la același eșantion, mai multe rezultate fals negative (β mai mare). Analogia merge și mai departe. Valoarea predictivă a unui test pozitiv depinde de prevalența bolii; tot așa, probabilitatea ca un rezultat "semnificativ" să corespundă unui efect real depinde de cât de plauzibile sunt ipotezele testate într-un domeniu. Într-un domeniu în care se testează multe ipoteze puțin plauzibile, o parte importantă din rezultatele semnificative sunt fals pozitive.

5.5.3 Reprezentarea grafică a lui α, β și a puterii

Figura 5.1 așază cele două lumi posibile pe aceeași axă. Curba din stânga este distribuția statisticii testului când H0 este adevărată; curba din dreapta este distribuția ei când există un efect de o anumită mărime. Linia punctată verticală este valoarea critică (1,96 pentru un test z bilateral cu α = 0,05). Aria de sub curba H0, la dreapta valorii critice, este α/2 (cealaltă jumătate se află în coada din stânga, nedesenată). Aria de sub curba H1, la stânga valorii critice, este β: în aceste situații efectul există, dar statistica nu depășește pragul. Aria de sub curba H1, la dreapta valorii critice, este puterea.

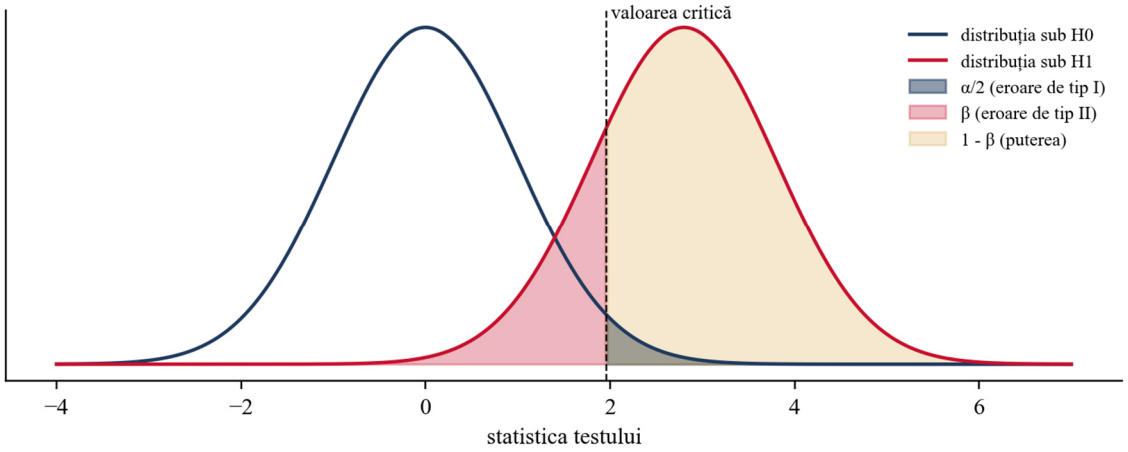


Figura 5.1. Distribuția statisticii testului sub ipoteza nulă (stânga) și sub ipoteza alternativă (dreapta). Valoarea critică (linia punctată) împarte axa: aria H0 din dreapta ei este α/2 (eroarea de tip I), aria H1 din stânga ei este β (eroarea de tip II), iar aria H1 din dreapta ei este puterea 1 - β.

Figura explică imediat de ce α și β variază în sens opus. Dacă mutăm valoarea critică spre dreapta (α mai mic, de exemplu 0,01), aria α scade, dar aria β crește. Singurele căi de a reduce ambele erori în același timp sunt cele care depărtează curbele una de alta sau le îngustează: un efect mai mare sau o eroare standard mai mică, adică mai mulți pacienți ori măsurători mai precise.

5.5.4 Factorii care influențează puterea

Puterea unui test depinde de patru elemente, care trebuie fixate în etapa de planificare a studiului.

Mărimea efectului. Un efect mare (de exemplu, o diferență de 20 mmHg între două tratamente antihipertensive) se detectează ușor, chiar cu puțini pacienți. Un efect mic cere eșantioane mari.

Variabilitatea. Cu cât abaterea standard a variabilei este mai mare, cu atât zgomotul acoperă mai ușor semnalul. Măsurătorile standardizate, aparatele calibrate și criteriile de includere care omogenizează lotul cresc puterea. Designul pereche (aceiași pacienți măsurați de două ori) crește puterea tocmai pentru că elimină variabilitatea dintre pacienți (capitolul 6).

Volumul eșantionului. Eroarea standard scade cu $\sqrt{n}$, deci curbele din figura 5.1 se îngustează când n crește. Este factorul pe care cercetătorul îl controlează cel mai direct.

Pragul α. Un α mai strict (0,01 în loc de 0,05) reduce puterea, iar un test unilateral (justificat) o crește, pentru că toată aria α se pune într-o singură coadă.

Figura 5.2 arată cum crește puterea testului t pentru două grupuri independente odată cu numărul de pacienți din fiecare grup, pentru trei mărimi ale efectului exprimate ca d Cohen (secțiunea 5.6). Pentru un efect mediu, d = 0,5, sunt necesari 64 de pacienți în fiecare grup pentru o putere de 80% la α = 0,05; pentru d = 0,3 sunt necesari 176 pe grup, pentru d = 0,2 aproape 394, iar pentru un efect mare, d = 0,8, sunt suficienți 26 [76]. Dacă dorim o putere de 90% în loc de 80% pentru d = 0,5, numărul urcă la 86 pe grup; dacă păstrăm puterea de 80%, dar coborâm α la 0,01, sunt necesari 96 pe grup. Calculele de acest tip se fac cu programe dedicate, precum G*Power [77], sau cu modulele de calcul al dimensiunii eșantionului din MedCalc și JASP.

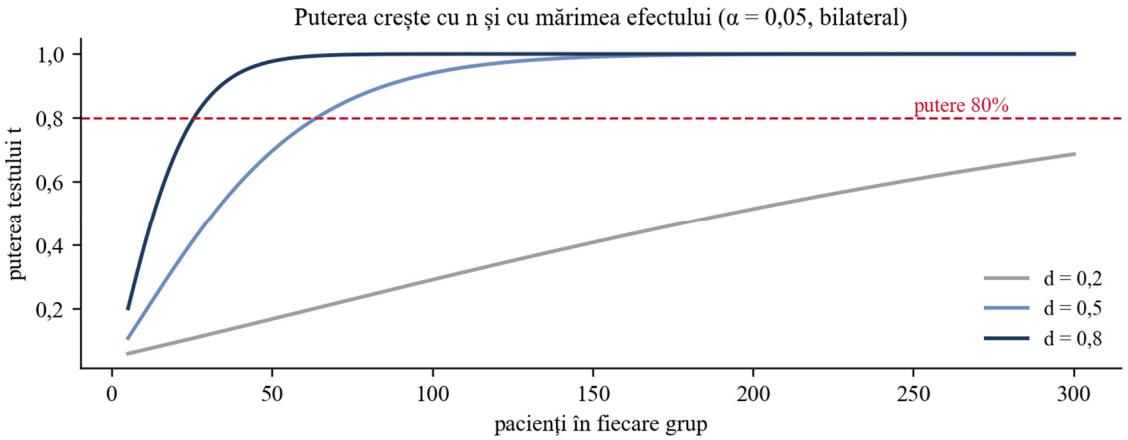


Figura 5.2. Puterea testului t pentru două grupuri independente în funcție de numărul de pacienți din fiecare grup, pentru trei mărimi ale efectului (d = 0,2; 0,5; 0,8), la α = 0,05 bilateral. Linia orizontală marchează puterea de 80%, valoarea minimă acceptată de obicei la planificarea unui studiu.

Convenția uzuală, β = 0,20 (putere 80%), înseamnă că acceptăm să ratăm un efect real într-unul din cinci studii. Asimetria față de α = 0,05 reflectă ideea că declararea unui efect inexistent este considerată mai gravă decât ratarea unuia real. În unele situații (o boală gravă fără alt tratament), raportul ar putea fi discutat, dar convenția rămâne cea folosită de comitetele de etică și de reviste.

Merită observată şi o consecinţă mai puţin intuitivă a puterii mici. Când un studiu subdimensionat obţine totuşi un rezultat semnificativ, efectul estimat este, în medie, exagerat: doar eşantioanele în care întâmplarea a "umflat" diferenţa reuşesc să treacă pragul. De aceea rezultatele spectaculoase din studii mici trebuie confirmate în studii mai mari, iar raportarea puterii (sau, mai bine, a intervalului de încredere) ajută cititorul să judece cât de mult poate avea încredere în estimare.

> **Studiu de caz: puterea statistică limitată într-o cohortă observaţională**
>
> Un studiu prospectiv, observaţional, realizat pe 100 de pacienţi la care s-a efectuat colangiopancreatografie endoscopică retrogradă (ERCP), a comparat două strategii de monitorizare a sedării: sedarea ghidată de entropie (un indice derivat din electroencefalogramă) faţă de monitorizarea standard [78]. Episoadele hipertensive au apărut la 6% faţă de 26% dintre pacienţi ($p = 0,007$), iar scorurile Aldrete de recuperare au fost mai mari la 5 minute ($p = 0,003$) şi la 15 minute ($p < 0,001$). Autorii raportează explicit o putere statistică limitată, de 29,3%. Ce înseamnă acest lucru pentru cititor? O putere de 29,3% înseamnă că, dacă efectul presupus în calcul ar exista cu adevărat, un studiu de această dimensiune l-ar detecta doar în aproximativ trei cazuri din zece. Pentru analiza la care se referă calculul, β ar fi deci de aproximativ 70%, iar un rezultat nesemnificativ ar spune foarte puţin despre absenţa efectului. Pentru rezultatele semnificative, puterea mică are o altă consecinţă: estimările efectului pot fi exagerate, iar confirmarea într-un studiu mai mare este necesară. Al doilea punct de lectură critică este designul. Grupurile nu au fost randomizate, deci diferenţele pot reflecta şi alte caracteristici ale pacienţilor sau ale echipelor. Autorii prezintă rezultatele cu prudenţă, ca asociere ("is associated with"), nu ca efect cauzal, ceea ce este exact formularea corectă pentru un studiu observaţional cu putere limitată. Pentru temele voastre, reţineţi ordinea: întâi designul, apoi puterea, apoi valoarea p.

5.6 Mărimea efectului şi semnificaţia clinică

5.6.1 Măsuri ale mărimii efectului

Valoarea p răspunde la întrebarea "există un efect?", dar nu la întrebarea care îl interesează pe medic: "cât de mare este efectul?". Pentru a doua întrebare folosim mărimea efectului (engl. effect size), o măsură care nu depinde de volumul eşantionului. Mărimile efectului se pot exprima în unităţile originale (o diferenţă de 10,19 mg/dL, o diferenţă de risc de 18,3 puncte procentuale) sau într-o formă standardizată, care permite compararea între studii şi variabile diferite.

Diferenţa medie standardizată d Cohen împarte diferenţa dintre două medii la abaterea standard comună:

$$d = \frac{m_1 - m_2}{s_{\text{comună}}}$$

Un d = 0,5 înseamnă că mediile se află la jumătate de abatere standard una de alta. Pentru datele pereche se foloseşte dz = media diferenţelor / abaterea standard a diferenţelor. Coeficientul de corelaţie r este el însuşi o mărime a efectului, iar r^2 arată proporţia din variabilitate explicată. În ANOVA (capitolul 6), η^2 = SS între grupuri / SS total arată ce proporţie din variabilitatea totală se datorează apartenenţei la grup. Pentru variabilele binare, raportul şanselor (OR) şi riscul relativ (RR) sunt mărimi ale efectului (capitolul 10); pentru ele nu există praguri universale, iar interpretarea depinde de frecvenţa bolii şi de context.

Tabelul 5.4. Praguri orientative pentru interpretarea mărimii efectului (după Cohen [76]); se folosesc doar în lipsa unui criteriu clinic.

Măsura	Efect mic	Efect mediu	Efect mare
d Cohen, dz	0,2	0,5	0,8
r (corelație)	0,1	0,3	0,5
η^2 (ANOVA)	0,01	0,06	0,14

Pragurile lui Cohen sunt convenții generale, propuse inițial pentru științele comportamentale, și nu țin locul judecății clinice. O reducere a TAS cu 0,2 abateri standard (aproximativ 3 mmHg) este "mică" după Cohen, dar aplicată la nivelul unei populații întregi are efecte reale asupra numărului de accidente vasculare.

Câteva exemple din mg2026 fixează ordinul de mărime. Diferența de înălțime dintre bărbați și femei are d = 2,19, un efect enorm, vizibil cu ochiul liber. Scăderea glicemiei între includere și 6 luni are dz = 0,88, un efect mare. Media TAS față de 125 mmHg corespunde unui $d = \frac{131{,}13-125}{13{,}88} = 0{,}44$, efect mic spre mediu. Diferența de TAS dintre sexe are d = -0,04, practic zero. În ANOVA pentru scăderea glicemiei pe brațe, $\eta^2 = 0{,}226$, deci aproape 23% din variabilitatea scăderii glicemiei este explicată de brațul de tratament, un efect mare.

5.6.2 Semnificație statistică și semnificație clinică

Un rezultat este semnificativ statistic când datele sunt greu de împăcat cu H0. Este semnificativ clinic (relevant clinic) când efectul este suficient de mare pentru a schimba decizia medicului sau viața pacientului. Cele două nu coincid automat. Pragul de relevanță clinică se numește diferența minimă importantă clinic (engl. minimal clinically important difference, MCID): cea mai mică modificare pe care pacienții o percep ca benefică sau care justifică schimbarea tratamentului. Ea se stabilește din considerente clinice, din studii cu pacienți sau din ghiduri, nu din statistică, și ar trebui fixată înainte de analiza datelor.

Combinând cele două criterii obținem patru situații pe care merită să le recunoașteți rapid:

- semnificativ statistic și relevant clinic: efect sigur și important; exemplul ideal;
- semnificativ statistic, dar nerelevant clinic: tipic pentru eșantioane foarte mari; efectul există, dar este prea mic pentru a conta;
- nesemnificativ statistic, cu interval de încredere care include efecte importante: rezultat nedeterminat, studiu probabil prea mic;
- nesemnificativ statistic, cu interval de încredere îngust în jurul lui zero: efect relevant puțin probabil.

Primul exemplu din mg2026 este diferența de TAS dintre sexe. Femeile au media 131,37 mmHg (AS 14,47; n = 86), bărbații 130,85 mmHg (AS 13,25; n = 74). Testul t Welch dă t = -0,24, p = 0,81, deci rezultatul este nesemnificativ. Diferența estimată bărbați minus femei este -0,5 mmHg, cu IC 95% de la -4,9 la 3,8 mmHg, iar d = -0,04. Este acesta un efect nedeterminat sau un efect absent? Intervalul de încredere exclude diferențe mai mari de aproximativ 5 mmHg în oricare sens. Dacă acceptăm că o diferență medie de TAS sub 5 mmHg între sexe nu are consecințe practice în acest grup, putem spune că o diferență relevantă clinic este puțin probabilă. Puterea testului cu 86 și 74 de pacienți era de 88% pentru un efect mediu (d = 0,5, adică aproximativ 7 mmHg), dar doar de 24% pentru un efect mic (d = 0,2, aproximativ 3 mmHg). Prin urmare, studiul putea detecta o diferență medie de 7 mmHg, dar nu ne poate spune nimic sigur despre o diferență de 2-3 mmHg.

Al doilea exemplu este scăderea glicemiei a jeun între includere și 6 luni la aceiași 160 de pacienți: de la 105,4 la 95,2 mg/dL, cu diferența medie 10,19 mg/dL (AS 11,57), IC 95% 8,38-

11,99 mg/dL, t = 11,14, 159 grade de libertate, p < 0,001. Este un rezultat extrem de semnificativ. Este și important clinic? Aici trebuie să aducem medicina în discuție. La un pacient cu glicemia a jeun în zona prediabetului (100-125 mg/dL, conform criteriilor ADA [80]), o scădere de 10 mg/dL poate muta valoarea sub pragul de 100 mg/dL; ca medie de grup, efectul are deci relevanță. Dacă, prin absurd, am fi stabilit dinainte că doar o scădere de cel puțin 15 mg/dL contează clinic, intervalul de încredere (8,38-11,99) ar fi arătat că efectul mediu este sigur sub acest prag, deși este extrem de semnificativ statistic. Aceeași cifră, două concluzii clinice diferite, în funcție de pragul de relevanță ales. Mai trebuie observat că un studiu fără grup de control nu poate atribui scăderea intervenției: glicemia ar putea scădea și prin regresie spre medie sau prin efectul participării la studiu. În mg2026 avem însă trei brațe randomizate, iar comparația corectă a eficacității este cea dintre brațe (capitolul 6).

Literatura colegilor oferă exemple pentru ambele capcane. În analiza pe patru ani a îngrijirii cancerului de colon la un spital universitar român, pe 867 de pacienți, autorii au raportat o corelație inversă între vârsta pacientului și durata intervenției chirurgicale, r = -0,088, p = 0,01 [55]. Rezultatul este semnificativ statistic, dar $r^2 = 0{,}0077$, adică vârsta explică mai puțin de 1% din variabilitatea duratei operației. Într-un eșantion atât de mare, aproape orice asociere diferită de zero devine semnificativă. În studiul pe hemodializați, vârsta medie a diferit între cohorte (59,7 ± 12,92 față de 61 ± 12,93 ani, p = 0,005) [39]; o diferență de 1,3 ani între două cohorte de aproape 1500 de pacienți fiecare este foarte semnificativă statistic, dar nu are importanță clinică în sine. Un studiu de concordanță între trei aparate oftalmologice, pe 170 de ochi, a găsit pentru keratometria K1 valori de 43,45 ± 1,64 D cu un aparat și 43,41 ± 1,70 D cu altul, diferență semnificativă (p = 0,019) [81]; o diferență de 0,04 dioptrii este însă mult sub precizia necesară pentru calculul unui implant de cristalin. În sens opus, studiul asupra aparatelor auditive inteligente a raportat un efect uriaș: pacienții cu hipoacuzie neurosenzorială care purtau proteze auditive aveau scoruri de stimă de sine de 35,41 ± 5,32 față de 22,99 ± 4,53 la cei fără proteză (p < 0,001, d Cohen = 2,515), iar ANOVA pe trei grupuri a dat F = 299,00 cu $\eta^2 = 0{,}712$ [82]. Aici semnificația statistică și cea practică merg împreună, deși, fiind un studiu transversal, rezultatul descrie o asociere, nu demonstrează că proteza a crescut stima de sine.

Studiu de caz: o corelație foarte slabă, dar semnificativă

Analiza retrospectivă a 867 de pacienți cu cancer de colon, tratați înainte și în timpul pandemiei COVID-19 într-un spital universitar din România, a comparat perioadele printr-o serie de teste și a constatat o scădere marcată a intervențiilor elective (p < 0,001), o creștere a intervențiilor de urgență (p < 0,001) și o creștere a cazurilor în stadiul IV în al doilea an de pandemie (p = 0,045) [55]. Ca rezultat secundar, rezumatul raportează o corelație inversă între vârstă și durata operației, r = -0,088, p = 0,01. Exercițiul de lectură critică are trei pași. Întâi, semnificația: cu n = 867, eroarea standard a lui r este aproximativ $\frac{1}{\sqrt{n}}$, adică 0,034, iar orice |r| peste aproximativ 0,067 devine semnificativ la α = 0,05; pragul de semnificație este depășit ușor. Apoi, mărimea: $r^2 = 0{,}0077$, sub 1% din variabilitate, o asociere neglijabilă după orice scară (Colton, folosită în curs, consideră "fără corelație" orice |r| sub 0,25). În sfârșit, interpretarea: rezultatul nu permite prezicerea duratei operației pornind de la vârstă și nu justifică nicio decizie clinică. Rezultatele principale ale studiului (modificarea structurii intervențiilor) sunt diferențe de proporții cu relevanță reală pentru organizarea serviciilor. Ce ar fi adăugat valoare raportării? Intervalul de încredere al lui r și, pentru comparațiile de proporții, diferențele absolute cu intervalele lor de încredere. Lecția pentru student este că "semnificativ" nu înseamnă "important", mai ales când n trece de câteva sute.

5.7 Trei căi spre aceeași decizie: valoarea critică, valoarea p și intervalul de încredere

Decizia unui test bilateral la pragul α se poate lua pe trei căi, care duc întotdeauna la aceeași concluzie. Le arătăm pe exemplul TAS, cu H0: μ = 125 mmHg, n = 160, m = 131,13 mmHg, s = 13,88 mmHg, ES = 1,097 mmHg.

Calea valorii critice. Calculăm statistica testului, t = 5,59, și o comparăm cu valoarea critică a distribuției Student pentru 159 grade de libertate și α = 0,05 bilateral, t critic = 1,975 (în Google Sheets, =T.INV.2T(0,05;159)). Regula: respingem H0 dacă |t| > t critic. Deoarece 5,59 > 1,975, respingem H0. Aceasta era calea folosită când calculele se făceau cu tabele tipărite.

Calea valorii p. Calculăm probabilitatea unei valori |t| cel puțin egale cu 5,59 sub H0: p = T.DIST.2T(5,59;159), aproximativ 0,0000001. Regula: respingem H0 dacă p < α. Deoarece p < 0,05, respingem H0. Este calea folosită de toate programele moderne și are avantajul că arată cât de puternică este evidența, nu doar dacă pragul a fost depășit.

Calea intervalului de încredere. Construim intervalul de încredere de 1 - α = 95% pentru media populației: $m \pm t$ critic × ES = $131{,}13 \pm 1{,}975 \times 1{,}097 = 131{,}13 \pm 2{,}17$, adică 128,96-133,30 mmHg (aceleași valori ca în capitolul 4). Regula: respingem H0 dacă valoarea afirmată de H0 (aici 125) se află în afara intervalului. Valoarea 125 este în afara intervalului, deci respingem H0. Această cale are cel mai mare conținut informațional: arată simultan decizia, direcția și mărimea efectului, precum și precizia estimării.

Echivalența nu este o coincidență. Valoarea 125 iese din interval exact atunci când |m - 125| > t critic × ES, adică exact atunci când |t| > t critic, adică exact atunci când p < 0,05. Verificați același lucru pentru valoarea de referință 130: t = 1,03 < 1,975; p = 0,304 > 0,05; 130 se află în intervalul 128,96-133,30. Toate trei căile spun: nu respingem H0.

Tabelul 5.5. Cele trei căi de decizie pentru media TAS (n = 160, m = 131,13 mmHg, α = 0,05).

Valoarea din H0	t calculat	t critic	Valoarea p	IC 95% conține μ0?	Decizia
$\mu_0 = 125$ mmHg	5,59	1,975	< 0,001	nu	se respinge H0
$\mu_0 = 130$ mmHg	1,03	1,975	0,304	da	nu se respinge H0
$\mu_0 = 140$ mmHg	-8,08	1,975	< 0,001	nu	se respinge H0

Corespondența funcționează pentru un IC de 1 - α și un test bilateral la pragul α: IC 95% cu α = 0,05, IC 99% cu α = 0,01. Intervalul de 99% pentru TAS, 128,27-133,99 mmHg, exclude tot 125, deci rezultatul rămâne semnificativ și la α = 0,01.

5.8 Algoritmul unui test statistic

Indiferent de test, rezolvarea urmează aceiași pași. Îi numerotăm aici și îi vom folosi în toate capitolele următoare, precum și în formatul a-g al temelor disciplinei.

1. Formulați ipoteza de cercetare în limbaj clinic (formatul temelor: punctul a).
2. Formulați ipotezele statistice H0 și H1, în cuvinte și matematic, cu parametrii populației (punctul b). Precizați dacă testul este bilateral sau unilateral.
3. Fixați pragul de semnificație α (implicit 0,05) înainte de a privi rezultatele.

4. Identificați tipul variabilelor și designul (o medie comparată cu o valoare, două grupuri independente, două măsurători pe aceiași pacienți, mai mult de două grupuri, variabile calitative) și alegeți testul (punctul c). Algoritmul complet de alegere apare în capitolul 8.
5. Verificați condițiile de aplicare ale testului (normalitatea, egalitatea variațiilor, independența observațiilor, frecvențele așteptate); dacă nu sunt îndeplinite, alegeți alternativa potrivită, de obicei un test neparametric.
6. Calculați statistica testului, de mână sau cu programul (punctul d), și obțineți valoarea p (și/sau valoarea critică).
7. Calculați mărimea efectului și intervalul de încredere pentru efect (punctul e, împreună cu graficul).
8. Luați decizia statistică: $p < \alpha$, se respinge H0; $p \geq \alpha$, nu se respinge H0; adăugați eticheta verbală a cursului (punctul f).
9. Formulați concluzia de cercetare în cuvinte, cu cifrele principale și cu interpretarea clinică, inclusiv limitele (punctul g).

Pasul 3 și pasul 4 se fac înainte de pasul 6. Schimbarea testului sau a pragului după ce ați văzut valoarea p transformă analiza într-o căutare de rezultate convenabile (secțiunea 5.11).

5.9 Testele z și t pentru o singură medie

Cel mai simplu test compară media unui eșantion cu o valoare dată dinainte: o valoare de referință din literatură, o țintă terapeutică, media națională. Cursul disciplinei distinge două variante, testul z și testul t, după ce știm despre variabilitatea populației.

5.9.1 Testul z

Testul z se folosește când abaterea standard a populației, σ, este cunoscută (de exemplu, dintr-un registru național mare sau din specificațiile unui aparat de măsură) și eșantionul este suficient de mare ($n \geq 30$) pentru ca media să urmeze distribuția normală, chiar dacă variabila însăși nu este perfect normală (teorema limită centrală, capitolul 4). Statistica testului este:

$$z = \frac{m - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Sub H0, z urmează distribuția normală standard. Valorile critice bilaterale sunt 1,96 pentru $\alpha = 0{,}05$, 2,576 pentru $\alpha = 0{,}01$ și 3,291 pentru $\alpha = 0{,}001$; pentru un test unilateral cu $\alpha = 0{,}05$, valoarea critică este 1,645.

Să presupunem, pentru exemplu, că un studiu populațional mare a stabilit, pentru adulții de aceeași vârstă din regiune, colesterolul total mediu de 200 mg/dL, cu $\sigma = 40$ mg/dL. În mg2026, media colesterolului este 209,39 mg/dL pe n = 160 de pacienți. Eroarea standard calculată cu σ cunoscut este 40 / √160 = 40 / 12,649 = 3,162 mg/dL, iar $z = \frac{209{,}39 - 200}{3{,}162} = \frac{9{,}39}{3{,}162} = 2{,}97$. Deoarece 2,97 > 1,96 (și chiar > 2,576), respingem H0; valoarea p bilaterală este $2 \times P(Z > 2{,}97) = 0{,}003$, deci diferența este foarte semnificativă. Pacienții cu risc cardiometabolic din studiu au un colesterol mediu mai mare decât populația generală de referință.

5.9.2 Testul t pentru o singură medie

În practică, σ este aproape întotdeauna necunoscut și îl estimăm prin abaterea standard a eșantionului, s. Înlocuirea lui σ cu s adaugă incertitudine, mai ales la eșantioane mici, iar statistica

$$t = \frac{m - \mu_0}{\frac{s}{\sqrt{n}}}$$

nu mai urmează distribuția normală, ci distribuția t a lui Student, descrisă în 1908 de William Gosset, chimist la o fabrică de bere din Dublin, care a publicat sub pseudonimul "Student" [83]. Distribuția t seamănă cu distribuția normală, dar are cozi mai groase; forma ei depinde de numărul gradelor de libertate, gl = n - 1. Cu cât n este mai mare, cu atât distribuția t se apropie de cea normală, iar pentru gl peste 120 diferența devine neglijabilă.

Tabelul 5.6. Valori critice ale distribuției t (bilateral α = 0,05 și 0,01; unilateral α = 0,05).

gl	bilateral 0,05	bilateral 0,01	unilateral 0,05
4	2,776	4,604	2,132
9	2,262	3,250	1,833
15	2,131	2,947	1,753
19	2,093	2,861	1,729
29	2,045	2,756	1,699
40	2,021	2,704	1,684
60	2,000	2,660	1,671
120	1,980	2,617	1,658
159	1,975	2,607	1,654
infinit (z)	1,960	2,576	1,645

Pentru colesterol, cu $s = 36{,}91$ mg/dL în loc de σ = 40, ES $= \frac{36{,}91}{12{,}649} = 2{,}918$ mg/dL și $t = \frac{9{,}39}{2{,}918} = 3{,}22$, cu 159 grade de libertate; p = 0,0016. Concluzia nu se schimbă. La n = 160, alegerea dintre z și t contează puțin; la n = 10 ar fi contat mult (valoarea critică ar fi 2,262 în loc de 1,96).

Regula cursului (testul z pentru n ≥ 30 și σ cunoscut, testul t pentru n < 30 sau σ necunoscut) se poate simplifica pentru practica de azi: când σ nu este cunoscut, folosim testul t la orice volum al eșantionului. Programele statistice fac exact acest lucru; JASP, de exemplu, oferă în meniul T-Tests > One Sample T-Test testul t de Student și, ca opțiune separată, testul z, pentru care cere introducerea valorii σ.

5.9.3 Condițiile de aplicare

Testul t pentru o medie presupune că observațiile sunt independente (fiecare pacient contribuie o singură dată) și că variabila are, în populație, o distribuție aproximativ normală. A doua condiție contează mai ales la eșantioane mici; la n mare, teorema limită centrală face testul robust la abateri moderate de la normalitate. Normalitatea se verifică cu testul Shapiro-Wilk [66] și cu graficul Q-Q (capitolul 3). Dacă variabila este puternic asimetrică și eșantionul este mic, alternativa neparametrică este testul Wilcoxon al rangurilor cu semn pentru o singură mediană (capitolul 7) sau o transformare a datelor (de exemplu, logaritmarea trigliceridelor).

Pentru glicemia a jeun din mg2026, testul Shapiro-Wilk dă p = 0,037, deci o abatere ușoară de la normalitate (asimetrie 0,37). Cu n = 160, testul t rămâne utilizabil, iar testul Wilcoxon duce la aceeași concluzie (problema rezolvată 5.2). Pentru IMC, asimetria este mare (1,42) și Shapiro-Wilk p < 0,001; aici raportarea medianei și a unui test neparametric este mai potrivită, chiar dacă n este mare.

5.10 Comparațiile multiple și inflația erorii de tip I

5.10.1 De ce crește riscul de fals pozitiv

Pragul α = 0,05 controlează eroarea de tip I pentru un singur test. Dacă facem mai multe teste independente, fiecare la α = 0,05, probabilitatea de a obține cel puțin un rezultat fals pozitiv crește rapid:

$$P(\text{cel puțin un fals pozitiv}) = 1 - (1-\alpha)^k = 1 - 0{,}95^k$$

Pentru k = 3 teste, probabilitatea este 0,143; pentru k = 5, 0,226; pentru k = 10, 0,401; pentru k = 20, 0,642. Cu alte cuvinte, un student care compară femeile și bărbații pe 20 de variabile la care, în realitate, nu există nicio diferență are aproape două șanse din trei să "descopere" cel puțin o diferență semnificativă.

Figura 5.3 ilustrează fenomenul în două feluri. În panoul A, am simulat 1000 de studii în care două grupuri de câte 20 de pacienți provin din aceeași populație (TAS cu media 130 și AS 14 mmHg), deci H0 este adevărată de fiecare dată. Valorile p sunt distribuite uniform între 0 și 1, iar 4,9% dintre ele cad sub 0,05: exact proporția de rezultate fals pozitive promisă de α. În panoul B este curba $1 - 0{,}95^k$ și efectul corecției Bonferroni, care menține riscul global aproape de 5%. Aceeași simulare o veți repeta în Google Sheets la lucrarea practică 5.

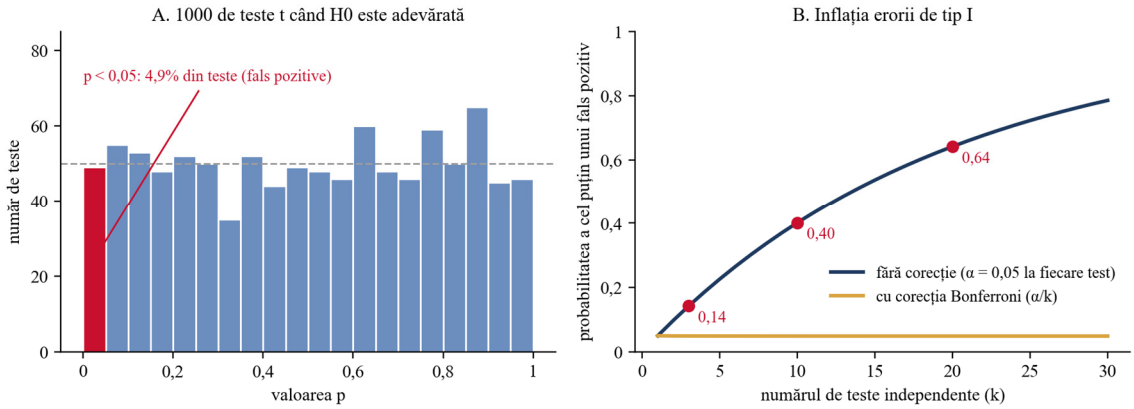


Figura 5.3. A. Distribuția valorilor p în 1000 de teste t simulate când ipoteza nulă este adevărată (două grupuri de câte 20 de valori din aceeași populație normală): valorile p sunt uniform distribuite, iar aproximativ 5% sunt sub 0,05. B. Probabilitatea de a obține cel puțin un rezultat fals pozitiv în funcție de numărul de teste independente, fără corecție și cu corecția Bonferroni. Date simulate.

5.10.2 Corecțiile Bonferroni, Holm și Benjamini-Hochberg

Corecția Bonferroni împarte pragul la numărul de teste: fiecare test se declară semnificativ doar dacă p < α/k. Echivalent, programele raportează valori p ajustate, p × k (plafonate la 1), care se compară cu α obișnuit. Metoda este simplă și sigură, dar conservatoare: la multe teste pierde mult din putere.

Metoda Holm (Bonferroni secvențial) păstrează aceeași garanție, dar este mai puțin conservatoare. Ordonăm valorile p crescător, p(1) ≤ p(2) ≤... ≤ p(k). Comparăm p(1) cu α/k, p(2) cu α/(k - 1), p(3) cu α/(k - 2) și așa mai departe. Ne oprim la primul test care nu trece pragul; acesta și toate cele următoare sunt declarate nesemnificative.

Metoda Benjamini-Hochberg (BH) controlează altă cantitate: nu probabilitatea de a avea vreun fals pozitiv, ci proporția așteptată de falși pozitivi printre rezultatele declarate semnificative (rata falselor descoperiri, engl. false discovery rate, FDR). Ordonăm valorile p crescător și comparăm p(i) cu (i/k) × α. Găsim cel mai mare i pentru care p(i) ≤ (i/k) × α și declarăm semnificative toate testele de la 1 la acel i. Metoda este potrivită când facem multe teste

exploratorii și acceptăm ca o mică parte din "descoperiri" să fie false, cu condiția să știm cât de mare este acea parte.

Pe setul mg2026 putem face exact exercițiul tentant descris mai sus: comparăm femeile și bărbații pe 14 variabile numerice măsurate la includere, cu testul t Welch. Rezultatele, ordonate după valoarea p, sunt în tabelul 5.7.

Tabelul 5.7. Comparația femei-bărbați pe 14 variabile din mg2026, cu valorile p necorectate și ajustate (testul t Welch).

Variabila	Media F	Media M	p necorectat	p Bonferroni	p Holm	p BH
înălțime (cm)	162,60	175,97	< 0,001	< 0,001	< 0,001	< 0,001
greutate (kg)	76,53	90,64	< 0,001	< 0,001	< 0,001	< 0,001
HDL(mg/dL)	53,92	44,41	< 0,001	< 0,001	< 0,001	< 0,001
talie (cm)	86,34	91,09	0,013	0,18	0,14	0,046
trigliceride (mg/dL)	171,21	149,51	0,068	0,95	0,68	0,19
vârstă (ani)	55,42	57,31	0,17	1	1	0,40
FINDRISC	10,02	10,77	0,27	1	1	0,55
glicemie$_0$(mg/d	104,81	106,16	0,59	1	1	0,79
TAD (mmHg)	78,88	79,62	0,59	1	1	0,79
colesterol (mg/dL)	208,08	210,91	0,63	1	1	0,79
HbA1c (%)	5,56	5,59	0,68	1	1	0,79
IMC $\left(\frac{\text{kg}}{m^2}\right)$	28,92	29,18	0,68	1	1	0,79
TAS (mmHg)	131,37	130,85	0,81	1	1	0,88
FC (bătăi/min)	74,69	74,73	0,98	1	1	0,98

Înălțimea, greutatea și HDL diferă între sexe indiferent de corecție; sunt diferențe biologice bine cunoscute. Circumferința taliei are p necorectat 0,013: semnificativă fără corecție, nesemnificativă după Bonferroni (pragul individual devine 0,05/14 = 0,0036) și după Holm, dar încă semnificativă după Benjamini-Hochberg (p ajustat 0,046). Diferența dintre metode nu este o contradicție, ci reflectă întrebări diferite. Dacă nu vrem niciun fals pozitiv în tot tabelul, folosim Bonferroni sau Holm; dacă acceptăm că aproximativ 5% din rezultatele declarate pozitive pot fi false, BH este suficient.

Când este necesară o corecție? Când testele aparțin aceleiași întrebări și concluzia se trage din "cel puțin unul semnificativ": comparații multiple între grupuri după o ANOVA (capitolul 6), multe variabile de rezultat într-un studiu, analize pe multe subgrupuri, controale intermediare repetate ale datelor. Nu este necesară pentru un rezultat principal unic, stabilit dinainte în protocol. Cea mai bună protecție rămâne alegerea unui singur rezultat principal înainte de colectarea datelor; rezultatele secundare se raportează ca exploratorii.

Un exemplu din literatura colegilor arată cât de ușor se adună testele. Studiul transversal asupra asocierii dintre severitatea infecției HIV și tulburările psihiatrice, pe 101 pacienți, raportează șapte rapoarte ale șanselor, cu valori p între 0,038 și sub 0,001 (de exemplu, OR = 5,81, p < 0,001; OR = 2,81, p = 0,038) [84]. Fiecare rezultat luat separat este semnificativ. La șapte teste, pragul Bonferroni $\frac{\text{ar fi}\quad 0{,}05}{7} = 0{,}0071$; rezultatele cu p = 0,016, 0,024 sau 0,038 nu l-ar trece. Ele pot fi reale, dar trebuie privite ca ipoteze de verificat, nu ca rezultate stabilite. Raportarea intervalelor de încredere pentru fiecare OR ar fi ajutat cititorul să judece și precizia estimărilor.

Studiu de caz: corecția Benjamini-Hochberg într-un studiu din departamentul de urgență

Un studiu retrospectiv dintr-un departament de urgență a analizat diferențele dintre femei și bărbați la 213 pacienți (100 de femei, 113 bărbați) tratați pentru suspiciune de infecție bacteriană [9]. Femeile erau mai în vârstă (mediana 77,0 față de 67,0 ani, $p < 0,001$), iar sursa urinară a infecției era mai frecventă la femei (51,0% față de 34,5%, $p = 0,015$). Pentru această asociere, rezumatul raportează OR 2,14 (IC 95% 1,20-3,81; $p = 0,010$) și, după corecția Benjamini-Hochberg pentru testele multiple, p ajustat = 0,039. Autorii au comparat numeroase caracteristici între sexe; corecția pentru testarea multiplă era deci justificată, iar rezultatul a rămas semnificativ și după ea, ceea ce îi întărește credibilitatea. Pentru rezultatele clinice, imaginea este diferită: mortalitatea în spital a fost 7,0% la femei (7 din 100) și 8,0% la bărbați (9 din 113), cu OR 0,87 (IC 95% 0,31-2,43), iar un alt rezultat a apărut la 14 din 100 de femei (14,0%) și la 21 din 113 bărbați (18,6%), OR 0,71 (IC 95% 0,34-1,49). Rezumatul raportează și un OR ajustat de 0,80 (IC 95% 0,27-2,35; $p = 0,690$). Intervalele de încredere sunt foarte largi, de la o reducere de aproape trei ori a șanselor până la o creștere de peste două ori. Cu 16 decese în total, studiul nu poate exclude nici un efect protector, nici un efect nefavorabil important; rezultatul este nedeterminat, nu "absent". Lecțiile sunt trei: corecția pentru comparații multiple se aplică atunci când se testează multe ipoteze în același studiu; un rezultat care rezistă corecției este mai convingător; la evenimente rare, intervalul de încredere spune mai mult decât valoarea p. Comparația dintre OR brut și OR ajustat (aici, ambele nesemnificative) va fi reluată în capitolele despre analiza riscului și regresia logistică.

5.11 p-hacking, HARKing și preînregistrarea

Valoarea p își păstrează sensul doar dacă analiza a fost decisă înainte de a vedea datele. Când cercetătorul încearcă mai multe analize și o raportează doar pe cea "care a ieșit", pragul nominal de 5% nu mai este real. Practica se numește p-hacking (vânătoarea de valori p) și ia forme aparent nevinovate: se mai adaugă câțiva pacienți până când p coboară sub 0,05; se exclud "valorile aberante" doar după ce s-a văzut că rezultatul devine semnificativ; se încearcă testul t, apoi Mann-Whitney, apoi transformarea logaritmică; se împart pacienții în subgrupuri (pe sexe, pe vârste, pe medii de proveniență) până apare un subgrup cu efect; se schimbă rezultatul principal cu unul secundar. Fiecare pas pare rezonabil luat separat, dar împreună transformă cinci încercări într-o singură raportare "semnificativă", cu un risc real de fals pozitiv mult peste 5%.

HARKing (engl. hypothesizing after the results are known) înseamnă formularea ipotezei după ce s-au văzut rezultatele și prezentarea ei ca și cum ar fi fost stabilită dinainte. O asociere observată întâmplător între fumat și o variabilă oarecare devine, în introducerea articolului, "ipoteza noastră principală". Cititorul nu mai poate distinge confirmarea de explorare.

Antidotul este preînregistrarea: protocolul studiului, cu ipoteza principală, rezultatul principal, dimensiunea eșantionului și planul de analiză statistică, se publică într-un registru public înainte de includerea primului pacient. Pentru studiile clinice randomizate, înregistrarea este obligatorie, iar ghidul CONSORT cere raportarea oricărei abateri de la protocol [56]. Pentru studiile observaționale, ghidul STROBE cere descrierea tuturor analizelor efectuate, nu doar a celor semnificative [25]. La nivelul vostru, regula practică este aceasta: scrieți ipotezele și testul ales (punctele a-c din formatul temei) înainte să deschideți fișierul de date în JASP.

Nu orice analiză neplanificată este greșită. Explorarea datelor generează ipoteze noi; problema apare doar când explorarea este prezentată drept confirmare. Formularea onestă este "în analize exploratorii, neplanificate, am observat..." și ea anunță cititorul că rezultatul trebuie verificat într-un studiu nou. Analizele de sensibilitate sunt un exemplu de bună practică: autorii repetă analiza în condiții diferite pentru a vedea dacă rezultatul rezistă. În studiul transversal

asupra tulburărilor interacțiunii intestin-creier, pe 136 de participanți analizați (52 de cazuri și 84 de martori), IMC a fost mai mare la cazuri (28,15 ± 6,49 față de $24{,}47 \pm \frac{4{,}60\ \mathrm{kg}}{m^2}$, $p = 0{,}001$), iar un factor alimentar a avut un OR protector de 0,36 (IC 95% 0,18-0,74; p = 0,009); autorii au repetat analiza după excluderea participanților cu proteina C reactivă peste $\frac{10\mathrm{mg}}{L}$ $(n = 98)$ [71]. Un rezultat care se menține în analiza de sensibilitate este mai credibil decât unul care apare doar într-o singură variantă de analiză.

> **Studiu de caz: comparația în interiorul grupului față de comparația între grupuri**
>
> Un studiu clinic controlat cu placebo, raportat după ghidul CONSORT, a evaluat timp de 3 luni efectul extractului de Aloe ferox (2 capsule/zi, 460 mg) asociat dietei, la 20 de persoane cu obezitate, comparativ cu 20 de martori potriviți, tratați cu dietă și placebo [85]. Rezumatul raportează două tipuri de rezultate: comparații între grupuri la 3 luni și comparații înainte-după în interiorul grupurilor, cu valori p precum 0,03 și 0,002 pentru IMC, 0,01 și 0,039 pentru LDL-colesterol, 0,018 și sub 0,001 pentru glicemia a jeun, 0,032 pentru colesterolul total și 0,006 pentru tensiunea diastolică. Exercițiul de lectură critică nu privește acest studiu în mod special, ci o eroare clasică din literatura clinică, pe care orice cititor trebuie să o poată evita. Să presupunem un studiu în care glicemia scade semnificativ în grupul tratat (p = 0,01 înainte-după) și nesemnificativ în grupul placebo (p = 0,08). Concluzia "tratamentul este eficient, iar placebo nu" este greșită: două rezultate de o parte și de alta a pragului nu demonstrează că efectele diferă între grupuri. Scăderea din grupul placebo poate fi aproape la fel de mare, doar mai imprecis estimată. Singurul test care răspunde la întrebarea despre eficacitate este comparația directă între grupuri: fie compararea scăderilor (diferențele individuale) între grupuri cu testul t pentru eșantioane independente, fie compararea valorilor finale ajustate pentru valoarea inițială (ANCOVA, capitolul 6). Comparațiile în interiorul grupurilor descriu doar evoluția fiecărui grup și sunt influențate de regresia spre medie și de efectul placebo. Când citiți un studiu controlat, căutați întâi comparația între grupuri; în studiul despre Aloe ferox, rezumatul o raportează, iar cititorul trebuie doar să știe pe care dintre valorile p să își sprijine concluzia despre eficacitate. Cu 20 de pacienți pe grup, intervalele de încredere ale diferențelor ar fi completat util valorile p.

5.12 Echivalență și non-inferioritate, pe scurt

Testul clasic poate respinge ipoteza "nicio diferență", dar nu o poate confirma. Uneori însă exact asta vrem să arătăm: că un medicament generic are același efect ca originalul, că o metodă de măsurare ieftină dă aceleași rezultate ca una scumpă sau că un tratament mai comod nu este mai slab decât cel standard. Pentru astfel de întrebări, ipotezele se inversează.

Într-un studiu de echivalență, cercetătorul fixează dinainte o marjă de echivalență Δ (de exemplu, 5 mmHg pentru TAS), adică cea mai mare diferență considerată neimportantă clinic. H0 devine "diferența este cel puțin Δ în modul" (tratamentele nu sunt echivalente), iar H1 "diferența este mai mică decât Δ" (tratamentele sunt echivalente). Practic, echivalența este demonstrată dacă întregul interval de încredere pentru diferență se află între -Δ și +Δ; procedura se numește TOST (două teste unilaterale). Într-un studiu de non-inferioritate interesează un singur sens: noul tratament nu este mai slab decât cel standard cu mai mult de Δ; este cazul rar în care testul unilateral este justificat.

Exemplul TAS pe sexe este instructiv. IC 95% pentru diferență este -4,9 la 3,8 mmHg. Dacă marja de echivalență fixată dinainte ar fi fost 5 mmHg, intervalul s-ar afla în întregime în zona -5 la +5, deci am putea afirma echivalența celor două medii. Dacă marja ar fi fost 3 mmHg, nu am putea afirma nici diferența, nici echivalența. Pentru calculul riguros se folosește de obicei un IC

de 90%, corespunzător celor două teste unilaterale la α = 0,05; ideea de bază rămâne însă aceeași: echivalența se demonstrează cu intervalul de încredere, niciodată cu un p > 0,05.

5.13 Cum se raportează un test statistic

Raportarea unui test într-un articol, o lucrare de licență sau o temă trebuie să permită cititorului să refacă raționamentul. Standardele actuale, cuprinse în recomandările ICMJE și în ghidurile CONSORT și STROBE [25,36,56], cer câteva elemente obligatorii: statisticile descriptive pentru fiecare grup (media și AS, sau mediana și IQR, cu n); numele testului, varianta lui (de exemplu, t Welch, bilateral) și pragul α; valoarea statisticii testului, cu gradele de libertate; valoarea p exactă; mărimea efectului, cu intervalul de încredere de 95%; interpretarea clinică, inclusiv pragul de relevanță, dacă există.

Un exemplu de raportare completă pentru mg2026: "Glicemia a jeun a scăzut de la 105,4 mg/dL la includere la 95,2 mg/dL la 6 luni (diferența medie 10,19 mg/dL, IC 95% 8,38-11,99; testul t pentru eșantioane pereche, t(159) = 11,14, p < 0,001; dz = 0,88). Distribuția diferențelor individuale nu s-a abătut semnificativ de la normalitate (Shapiro-Wilk p = 0,725)."

Exemplu de raportare pentru un rezultat nesemnificativ: "TAS nu a diferit semnificativ între femei ($131{,}37 \pm 14{,}47$ mmHg, $n = 86$) și bărbați ($130{,}85 \pm 13{,}25$ mmHg, $n = 74$); diferența bărbați minus femei a fost -0,5 mmHg (IC 95% -4,9 la 3,8; testul t Welch, $t(157{,}4) = -0{,}24$, $p = 0{,}81$; $d = -0{,}04$)." Observați că fraza nu spune "TAS este egală la femei și bărbați"; intervalul de încredere lasă cititorul să judece ce diferențe sunt excluse.

Câteva formulări de evitat: "p = 0,000" (se scrie p < 0,001); "NS" fără nicio cifră; "tendință spre semnificație" pentru p = 0,07 (fie interpretăm intervalul de încredere, fie spunem că rezultatul este nesemnificativ); "diferență extrem de semnificativă" fără mărimea diferenței; "rezultatele demonstrează că..." într-un studiu observațional.

Studiul retrospectiv asupra evoluției COVID-19 la pacienții cu diabet zaharat de tip 1 și tip 2, pe 486 de pacienți (200 fără diabet, 62 cu diabet de tip 1, 224 cu diabet de tip 2), este un bun exercițiu de lectură [8]. Rezumatul raportează proporții clare (forme severe 33,5% la tipul 2 față de 25,8% la tipul 1; mortalitate 11,6% față de 8,1%) și medii cu AS (durata spitalizării 11,6 ± 7,0 față de 9,1 ± 5,8 zile; glicemia la internare 178,3 ± 34,7 față de 164,8 ± 39,6 mg/dL), dar rapoartele șanselor pentru internarea în terapie intensivă (OR 2,24) și pentru ventilația mecanică (OR 2,46) apar fără interval de încredere și fără valoare p. Fără interval, cititorul nu poate judeca dacă un OR de 2,24 este estimat precis (de exemplu, 1,8-2,8) sau foarte imprecis (de exemplu, 0,9-5,6). În plus, abaterile standard mari în raport cu media duratei de spitalizare sugerează o distribuție asimetrică, pentru care mediana și IQR ar fi fost mai informative. Observația nu scade valoarea clinică a studiului; arată doar cât de mult contează, pentru cititor, câteva cifre în plus în rezumat.

5.14 Probleme rezolvate

Problema rezolvată 5.1. De la întrebare la ipoteze

Enunț. Pentru fiecare dintre întrebările de mai jos, formulați ipoteza de cercetare, ipotezele statistice în cuvinte și matematic și precizați dacă testul este bilateral sau unilateral. (a) În mg2026, frecvența cardiacă de repaus diferă între fumători și nefumători? (b) Proporția pacienților cu diabet confirmat prin TTGO este diferită de 10%, o prevalență ipotetică a populației generale adulte? (c) Un producător susține că o nouă bandă de măsurare a glicemiei nu subestimează valorile cu mai mult de 3 mg/dL față de laboratorul de referință. (d) TAS se modifică între includere și 6 luni la aceiași pacienți?

Rezolvare. (a) Ipoteza de cercetare: frecvența cardiacă de repaus a fumătorilor diferă de cea a nefumătorilor. H0: în populație, media FC este aceeași la fumători și la nefumători, μfumători = μnefumători. H1: mediile diferă, μfumători ≠ μnefumători. Test bilateral: o diferență în oricare sens ar fi interesantă. (b) Ipoteza de cercetare: prevalența diabetului în populația cu risc cardiometabolic diferă de cea din populația generală. H0: π = 0,10; H1: π ≠ 0,10. Test bilateral. (c) Este o întrebare de non-inferioritate: H0: banda subestimează cu cel puțin 3 mg/dL (diferența medie bandă minus laborator ≤ -3); H1: subestimarea medie este mai mică de 3 mg/dL (diferența > -3). Test unilateral, justificat prin natura întrebării, cu marja fixată dinainte. (d) Ipoteza de cercetare: în cele 6 luni de intervenție, TAS se modifică. H0: media diferențelor individuale (TAS0 - TAS6) este zero în populație, μD = 0; H1: μD ≠ 0. Chiar dacă ne așteptăm la o scădere, testul rămâne bilateral: o creștere a TAS ar fi un rezultat important (de exemplu, un efect advers), care nu trebuie ignorat. În mg2026, testul pereche arată o scădere medie de 4,78 mmHg (IC 95% 3,36-6,19; t = 6,68; p < 0,001).

Interpretare. În toate cele patru cazuri, H0 conține semnul egal (sau, la non-inferioritate, "egal ori mai rău"), adică afirmația precisă care permite calculul distribuției statisticii testului.

Problema rezolvată 5.2. Glicemia a jeun față de 100 mg/dL

Enunț. În mg2026, glicemia a jeun la includere are media 105,44 mg/dL, abaterea standard 15,76 mg/dL, n = 160. Pragul ADA pentru glicemia a jeun modificată este 100 mg/dL [80]. Diferă media populației de pacienți cu risc cardiometabolic de 100 mg/dL? Folosiți α = 0,05 și toate cele trei căi de decizie.

Rezolvare. Ipoteza de cercetare: media glicemiei a jeun în populația cu risc cardiometabolic diferă de pragul de 100 mg/dL. H0: $\mu = 100$ mg/dL; H1: μ ≠ 100 mg/dL; bilateral, α = 0,05. Testul: o medie comparată cu o valoare dată, σ necunoscut, deci testul t pentru o singură medie. Condiții: observații independente; Shapiro-Wilk p = 0,037 arată o abatere ușoară de la normalitate (asimetrie 0,37), dar cu n = 160 media are o distribuție practic normală, iar testul t este robust; vom verifica totuși cu testul Wilcoxon.

Calcul. ES $= \frac{15,76}{\sqrt{160}} = \frac{15,76}{12,649} = 1,246$ mg/dL. $t = \frac{105,44-100}{1,246} = \frac{5,44}{1,246} = 4,36$, cu gl = 159.

Calea valorii critice: t critic (159; 0,05 bilateral) = 1,975; 4,36 > 1,975, deci respingem H0. Calea valorii p: p = T.DIST.2T(4,36;159) = 0,00002, adică p < 0,001. Calea IC: IC $95\% = 105,44 \pm 1,975 \times 1,246 = 105,44 \pm 2,46$, adică 102,98-107,90 mg/dL; valoarea 100 nu se află în interval, deci respingem H0. Mărimea efectului: $d = \frac{5,44}{15,76} = 0,35$ (efect mic spre mediu). Testul Wilcoxon pentru mediană față de 100 dă p < 0,001, aceeași concluzie.

Concluzia statistică: p < 0,001, se respinge H0; diferența este extrem de semnificativă. Concluzia de cercetare: media glicemiei a jeun în populația de adulți cu risc cardiometabolic din care provine eșantionul este mai mare decât pragul de 100 mg/dL, cu aproximativ 3-8 mg/dL (IC 95% pentru medie 102,98-107,90 mg/dL). Interpretarea clinică: pacientul "mediu" din studiu se află în zona glicemiei a jeun modificate, ceea ce corespunde criteriului de includere (risc cardiometabolic). Dacă am testa valoarea 105 mg/dL, am obține t = 0,35 și p = 0,73: datele sunt compatibile cu o medie de 105 mg/dL, ceea ce nu înseamnă că media este exact 105.

Problema rezolvată 5.3. Testul z și testul t pentru colesterol

Enunț. Un registru național indică pentru adulții de 35-79 de ani o medie a colesterolului total de 200 mg/dL, cu $\sigma = 40$ mg/dL (valori ipotetice). În mg2026, n = 160, $m = 209,39$ mg/dL, $s = 36,91$ mg/dL. (a) Aplicați testul z. (b) Aplicați testul t. (c) Comentați diferența.

Rezolvare. H0: $\mu = 200$ mg/dL; H1: μ ≠ 200 mg/dL; α = 0,05, bilateral. (a) Cu σ cunoscut: ES $= \frac{40}{\sqrt{160}} = 3{,}162$; $z = \frac{9{,}39}{3{,}162} = 2{,}97$. Valoarea critică 1,96 este depășită, iar 2,97 > 2,576 depășește și pragul pentru α = 0,01. p = 2 × (1 - Φ(2,97)) = 0,003. Se respinge H0; diferența este foarte semnificativă. (b) Cu σ estimat prin s: ES $= \frac{36{,}91}{12{,}649} = 2{,}918$; $t = \frac{9{,}39}{2{,}918} = 3{,}22$, gl = 159; p = 0,0016, tot foarte semnificativ. (c) Cele două teste diferă doar prin estimarea zgomotului. Aici s (36,91) este mai mic decât σ din registru (40), de aceea t este puțin mai mare decât z. Dacă σ din registru descrie corect și populația noastră, testul z este cel potrivit; dacă pacienții noștri sunt mai omogeni decât populația generală, s reflectă mai bine variabilitatea lor și testul t este preferabil. Concluzia de cercetare: pacienții cu risc cardiometabolic au un colesterol mediu mai mare decât populația de referință, cu aproximativ 9 mg/dL (IC 95% pentru medie, calculat cu t: 203,6-215,2 mg/dL).

Problema rezolvată 5.4. Un rezultat nesemnificativ pe un eșantion mic

Enunț. Un student a avut acces doar la 10 pacienți din mg2026, cu TAS: 150, 154, 131, 130, 137, 102, 110, 103, 147, 136 mmHg. Testați dacă media diferă de 125 mmHg și interpretați rezultatul în contextul analizei pe toți cei 160 de pacienți.

Rezolvare. H0: μ = 125; H1: μ ≠ 125; α = 0,05. Suma valorilor este 1300, deci m = 130,0 mmHg. Abaterile față de medie sunt 20, 24, 1, 0, 7, -28, -20, -27, 17, 6; pătratele lor sunt 400, 576, 1, 0, 49, 784, 400, 729, 289, 36, cu suma 3264. Varianța $s^2 = \frac{3264}{9} = 362{,}67$, deci s = 19,04 mmHg. ES $= \frac{19{,}04}{\sqrt{10}} = \frac{19{,}04}{3{,}162} = 6{,}02$ mmHg. $t = \frac{130{,}0-125}{6{,}02} = 0{,}83$, gl = 9. Valoarea critică pentru gl = 9 este 2,262; 0,83 < 2,262, deci nu respingem H0. Valoarea p = T.DIST.2T(0,83;9) = 0,43. IC $95\% = 130{,}0 \pm 2{,}262 \times 6{,}02 = 130{,}0 \pm 13{,}62$, adică 116,38-143,62 mmHg, interval care conține 125.

Interpretare. Concluzia statistică: p = 0,43 > 0,05, nu se respinge H0; diferența este nesemnificativă statistic. Concluzia de cercetare nu poate fi "media TAS este 125 mmHg": intervalul de încredere cuprinde valori de la 116 la 144 mmHg, adică de la tensiune optimă până la hipertensiune de gradul 1. Rezultatul este nedeterminat, nu negativ. Pe toți cei 160 de pacienți, cu ES = 1,097 (de 5,5 ori mai mic), aceeași întrebare primește un răspuns clar: t = 5,59, p < 0,001. Diferența dintre cele două analize nu vine din date contradictorii, ci din puterea testului. Pentru un efect de mărimea celui observat pe tot lotul (d ≈ 0,44), puterea unui test t pe 10 pacienți este foarte mică, de ordinul 20-25%; eșecul de a respinge H0 era previzibil.

Problema rezolvată 5.5. Bonferroni, Holm și Benjamini-Hochberg

Enunț. Un studiu exploratoriu compară două grupuri pe șase rezultate și obține valorile p: 0,012; 0,001; 0,20; 0,041; 0,008; 0,030. Câte rezultate sunt semnificative la α = 0,05 (a) fără corecție, (b) cu Bonferroni, (c) cu Holm, (d) cu Benjamini-Hochberg?

Rezolvare. Ordonăm crescător: p(1) = 0,001; p(2) = 0,008; p(3) = 0,012; p(4) = 0,030; p(5) = 0,041; p(6) = 0,20; k = 6.

(a) Fără corecție, sunt semnificative toate valorile sub 0,05: cinci rezultate. Probabilitatea de a avea cel puțin un fals pozitiv, dacă toate H0 ar fi adevărate, ar fi $1 - 0{,}95^6 = 0{,}265$.

(b) Bonferroni: pragul este 0,05 / 6 = 0,0083. Trec doar 0,001 și 0,008: două rezultate. Valorile p ajustate (p × 6) sunt 0,006; 0,048; 0,072; 0,18; 0,246; 1.

(c) Holm: comparăm p(1) = 0,001 cu 0,05/6 = 0,0083 (trece); p(2) = 0,008 cu 0,05/5 = 0,010 (trece); p(3) = 0,012 cu 0,05/4 = 0,0125 (trece); p(4) = 0,030 cu 0,05/3 = 0,0167 (nu trece). Ne oprim: trei rezultate semnificative.

(d) Benjamini-Hochberg: pragurile (i/6) × 0,05 sunt 0,0083; 0,0167; 0,025; 0,0333; 0,0417; 0,05. Comparăm: 0,001 ≤ 0,0083; 0,008 ≤ 0,0167; 0,012 ≤ 0,025; 0,030 ≤ 0,0333; 0,041 ≤ 0,0417; 0,20 > 0,05. Cel mai mare i pentru care p(i) ≤ prag este i = 5, deci cinci rezultate sunt semnificative.

Interpretare. Metodele răspund la întrebări diferite. Bonferroni și Holm garantează că probabilitatea de a avea vreun fals pozitiv în tot setul nu depășește 5%; Holm este întotdeauna cel puțin la fel de puternic ca Bonferroni și ar trebui preferat. BH acceptă că, în medie, cel mult 5% dintre rezultatele declarate semnificative sunt false; este potrivit pentru analize exploratorii cu multe teste. Alegerea metodei se face înainte de analiză și se raportează în secțiunea de metode.

Problema rezolvată 5.6. Puterea și dimensiunea eșantionului

Enunț. Un medic de familie vrea să compare două programe de educație pentru hipertensiune, printr-un studiu randomizat cu două grupuri. Consideră relevantă clinic o diferență a TAS medii de 5 mmHg; din mg2026 știe că abaterea standard a TAS este aproximativ 14 mmHg. (a) Ce mărime a efectului corespunde? (b) Câți pacienți sunt necesari în fiecare grup pentru o putere de 80% la α = 0,05 bilateral? (c) Ce putere ar avea un studiu cu 53 de pacienți pe grup, ca brațele din mg2026?

Rezolvare. $(a)\ d = \frac{\Delta}{\sigma} = \frac{5}{14} = 0{,}36$, un efect mic spre mediu. (b) Formula aproximativă pentru compararea a două medii este n pe grup = 2 × (z1-α/2 + z1-β)² × σ² / Δ². Pentru α = 0,05 bilateral, $z_1 - \frac{\alpha}{2} = 1{,}96$; pentru putere 80%, z1-β = 0,84. Rezultă $n = \frac{2\times(1{,}96+0{,}84)^2\times 14^2}{5^2} = \frac{2\times 7{,}84\times 196}{25} = 122{,}9$, adică 123 de pacienți pe grup; calculul exact cu distribuția t (G*Power, JASP) dă 125. La aceștia se adaugă de obicei 10-20% pentru pierderile din urmărire. (c) Cu 53 de pacienți pe grup, puterea pentru d = 0,36 este doar de aproximativ 45%: studiul ar rata un efect real de 5 mmHg mai des decât l-ar detecta.

Interpretare. Calculul dimensiunii eșantionului se face înainte de studiu și pornește de la diferența relevantă clinic, nu de la diferența pe care sperăm să o găsim. Un studiu de 53 de pacienți pe braț este suficient pentru efecte mari (scăderea glicemiei din mg2026 diferă între brațe cu efecte standardizate între aproximativ 0,5 și 1,4), dar prea mic pentru o diferență de 5 mmHg la TAS.

Problema rezolvată 5.7. Valoarea p din statistica t

Enunț. Fără program statistic, doar cu tabelul 5.6, decideți la α = 0,05: (a) t = 2,10, gl = 20, bilateral; (b) t = 1,80, gl = 25, bilateral și apoi unilateral (în sensul anunțat dinainte); (c) t = 3,10, gl = 9, bilateral. Verificați apoi valorile p exacte cu Google Sheets.

Rezolvare. (a) Valoarea critică pentru gl = 20 este 2,086; 2,10 > 2,086, deci p este puțin sub 0,05; se respinge H0, la limită. În Sheets, =T.DIST.2T(2,1;20) dă 0,049. (b) Valoarea critică bilaterală pentru gl = 25 este 2,060; 1,80 < 2,060, deci p > 0,05 și nu se respinge H0 (p bilateral = T.DIST.2T(1,8;25) = 0,084). Unilateral, valoarea critică este aproximativ 1,708 (=T.INV(0,95;25)); 1,80 > 1,708, deci p unilateral = 0,042 < 0,05. Același rezultat este "semnificativ" unilateral și "nesemnificativ" bilateral; tocmai de aceea alegerea trebuie făcută și justificată înainte de analiză. (c) Pentru gl = 9, valorile critice sunt 2,262 (α = 0,05) și 3,250 (α = 0,01); 3,10 se află între ele, deci 0,01 < p < 0,05; în Sheets, p = 0,013. Rezultatul este semnificativ, dar nu foarte semnificativ.

5.15 Exerciții propuse

1. Pentru HbA1c la includere, mg2026 are m = 5,58%, s = 0,45%, n = 160. Testați dacă media populației diferă de 5,7% (pragul inferior al intervalului de prediabet după ADA). Calculați t, decideți la α = 0,05 și formulați concluzia de cercetare.

2. Scorul FINDRISC în mg2026 are m = 10,37, s = 4,27, n = 160. Diferă media de 12 (prag ales pentru exercițiu)? Calculați t, IC 95% pentru medie și d.
3. Frecvența cardiacă de repaus la includere: m = 74,71 bătăi/min, s = 9,30, n = 160. Testați H0: μ = 72 bătăi/min. Calculați t, mărimea efectului și comentați relevanța clinică a diferenței.
4. Într-un grup de 36 de copii dintr-o comunitate rurală, hemoglobina medie este 11,2 g/dL. Din studii mari se știe că, la această vârstă, $\sigma = 1{,}2$ g/dL, iar media populației generale este 11,8 g/dL. Aplicați testul z (α = 0,05, bilateral) și calculați valoarea p.
5. Clasificați fiecare situație drept eroare de tip I, eroare de tip II sau decizie corectă: (a) un studiu conclude că un supliment scade colesterolul, dar suplimentul nu are niciun efect; (b) un studiu cu 15 pacienți pe grup nu găsește diferențe semnificative între un antibiotic nou, în realitate superior, și cel standard; (c) un studiu mare nu găsește diferențe între două diete care au, în realitate, același efect.
6. Un student face 8 teste independente la α = 0,05, iar toate ipotezele nule sunt adevărate. (a) Care este probabilitatea de a obține cel puțin un rezultat fals pozitiv? (b) Ce prag individual ar impune corecția Bonferroni? (c) Dar pentru 4 teste?
7. Cinci comparații au dat valorile p: 0,002; 0,009; 0,018; 0,035; 0,25. Câte sunt semnificative la α = 0,05 fără corecție, cu Bonferroni, cu Holm și cu Benjamini-Hochberg?
8. Folosind tabelul 5.6 și apoi Google Sheets, aflați valoarea p bilaterală pentru: (a) t = 2,5 cu gl = 12; (b) t = 1,70 cu gl = 40 (dați și valoarea unilaterală); (c) t = -2,9 cu gl = 159.
9. Un studiu raportează pentru diferența TAS dintre două tratamente: diferența medie 4 mmHg, IC 95% -1 la 9 mmHg. (a) Este p mai mic sau mai mare decât 0,05? (b) Diferența minimă importantă clinic este 5 mmHg. Este rezultatul "efect absent" sau "efect nedeterminat"? (c) Ce ați recomanda pentru un studiu viitor?
10. Pentru IMC la includere, $m = \frac{29{,}04\ \text{kg}}{m^2}$, s = 3,91, n = 160, asimetrie 1,42, Shapiro-Wilk p < 0,001. (a) Calculați t pentru H0: $\mu = \frac{30\ \text{kg}}{m^2}$. (b) Ce test ați prefera și de ce? (c) Cum ați raporta rezultatul?
11. Câți pacienți pe grup sunt necesari pentru a detecta, cu putere de 80% și α = 0,05 bilateral, o diferență de 10 mmHg a TAS între două grupuri, dacă σ = 14 mmHg? Folosiți formula aproximativă din problema rezolvată 5.6.

Capcane frecvente

- Ipotezele statistice se scriu cu parametrii populației (μ, π, ρ), nu cu mediile eșantionului (m1, m2).
- "Se acceptă H0" este o formulare greșită; un rezultat nesemnificativ înseamnă că datele nu permit respingerea lui H0.
- Valoarea p nu este probabilitatea ca H0 să fie adevărată și nici probabilitatea ca rezultatul să fie întâmplător.
- "Extrem de semnificativ" (p < 0,001) descrie tăria evidenței împotriva lui H0, nu mărimea sau importanța clinică a efectului.
- Alegerea testului unilateral după ce s-a văzut sensul diferenței înjumătățește artificial valoarea p.
- Un efect semnificativ într-un grup și nesemnificativ în altul nu dovedește că efectele diferă între grupuri; se compară direct grupurile.
- La multe teste, câteva rezultate "semnificative" apar doar din întâmplare; fără corecție sau fără un rezultat principal stabilit dinainte, ele rămân ipoteze.

- Se scrie p < 0,001, niciodată p = 0,000.

De reținut

- Testul statistic pornește de la H0 (fără efect), calculează cât de neobișnuite ar fi datele dacă H0 ar fi adevărată și respinge H0 când p < α.
- α = probabilitatea erorii de tip I (fals pozitiv); β = probabilitatea erorii de tip II (fals negativ); 1 - β = puterea; 1 - α = nivelul de încredere.
- Puterea crește cu mărimea efectului, cu volumul eșantionului și cu α și scade cu variabilitatea; la planificare se urmărește de obicei o putere de cel puțin 80%.
- Valoarea critică, valoarea p și intervalul de încredere duc la aceeași decizie; intervalul de încredere aduce cea mai multă informație.
- Testul t pentru o medie: $t = \frac{m-\mu_0}{\frac{s}{\sqrt{n}}}$, cu n - 1 grade de libertate; testul z folosește σ cunoscut.
- Un rezultat complet raportat conține statisticile descriptive, testul, statistica cu gradele de libertate, p exact, mărimea efectului cu IC 95% și interpretarea clinică.
- $1 - 0{,}95^k$ dă riscul de cel puțin un fals pozitiv la k teste; Bonferroni (α/k), Holm și Benjamini-Hochberg sunt corecțiile uzuale.

Întrebări de autoevaluare

1. (CS) Valoarea p a unui test statistic reprezintă:
   a) probabilitatea ca ipoteza nulă să fie adevărată
   b) probabilitatea ca ipoteza alternativă să fie adevărată
   c) probabilitatea de a obține un rezultat cel puțin la fel de extrem ca cel observat, dacă ipoteza nulă este adevărată
   d) probabilitatea de a face o eroare de tip II
   e) mărimea efectului standardizată
2. (CS) Un studiu cu 12 pacienți pe grup nu găsește o diferență semnificativă între un medicament nou și placebo (p = 0,24), deși medicamentul este, în realitate, eficient. Eroarea comisă este:
   a) eroare de tip I
   b) eroare de tip II
   c) eroare de eșantionare sistematică
   d) eroare de măsurare
   e) nu este o eroare, deoarece p > 0,05
3. (CS) Pentru media TAS din mg2026 (m = 131,13 mmHg, IC 95% 128,96-133,30 mmHg), testul bilateral la α = 0,05 al ipotezei H0: μ = 132 mmHg va avea ca rezultat:
   a) respingerea H0, deoarece 132 ≠ 131,13
   b) nerespingerea H0, deoarece 132 se află în IC 95%
   c) acceptarea H0, deoarece media populației este sigur 132 mmHg
   d) respingerea H0, deoarece n = 160 este un eșantion mare
   e) decizia nu poate fi luată fără valoarea p
4. (CS) Puterea unui test statistic este:
   a) α
   b) 1 - α
   c) β
   d) 1 - β

e) $\alpha + \beta$

5. (CS) Un cercetător efectuează 10 teste independente la $\alpha = 0,05$, iar toate ipotezele nule sunt adevărate. Probabilitatea de a obține cel puțin un rezultat semnificativ este aproximativ:
   a) 0,05
   b) 0,10
   c) 0,23
   d) 0,40
   e) 0,50
6. (CM) Despre ipotezele statistice, sunt corecte afirmațiile:
   a) ipoteza nulă afirmă de obicei absența diferenței sau a asocierii
   b) ipotezele se formulează despre parametrii populației
   c) ipoteza alternativă bilaterală pentru două medii se scrie $\mu1 \neq \mu2$
   d) ipotezele se formulează după examinarea datelor, pentru a alege sensul testului
   e) H0 și H1 se exclud reciproc
7. (CM) Puterea unui test t pentru două grupuri independente crește dacă:
   a) crește numărul de pacienți din fiecare grup
   b) crește mărimea reală a efectului
   c) crește abaterea standard a variabilei
   d) se trece de la $\alpha = 0,01$ la $\alpha = 0,05$
   e) se trece de la un test unilateral justificat la unul bilateral
8. (CM) În mg2026, TAS nu diferă semnificativ între sexe: diferența -0,5 mmHg, IC 95% -4,9 la 3,8, $p = 0,81$. Sunt corecte afirmațiile:
   a) se acceptă H0, deci TAS medie este identică la femei și bărbați
   b) nu se respinge H0
   c) intervalul de încredere exclude diferențe medii mai mari de aproximativ 5 mmHg
   d) rezultatul demonstrează că sexul nu influențează TAS în nicio populație
   e) mărimea efectului ($d = -0,04$) este practic nulă
9. (CM) Sunt exemple de practici care cresc riscul real de rezultate fals pozitive:
   a) adăugarea de pacienți până când p scade sub 0,05
   b) alegerea testului unilateral după ce s-a văzut sensul diferenței
   c) preînregistrarea protocolului cu rezultatul principal
   d) raportarea doar a subgrupurilor în care efectul a fost semnificativ
   e) aplicarea corecției Holm pentru comparațiile multiple
10. (CM) Despre semnificația statistică și cea clinică, sunt corecte afirmațiile:

a) într-un eșantion foarte mare, o diferență neimportantă clinic poate fi extrem de semnificativă statistic

b) o corelație $r = -0,088$ cu $p = 0,01$ pe 867 de pacienți explică mai puțin de 1% din variabilitate

c) un $p < 0,001$ garantează că efectul este important pentru pacient

d) diferența minimă importantă clinic se stabilește pe criterii clinice, înainte de analiză

e) mărimea efectului și intervalul de încredere completează valoarea p în raportare

Capitolul 6. Teste statistice parametrice uzuale: testul t pentru eșantioane pereche și independente, analiza variației (ANOVA)

Obiective

La finalul acestui capitol, studentul va putea:

- să recunoască designul unui studiu (o medie față de o valoare, două măsurători pe aceiași pacienți, două grupuri independente, trei sau mai multe grupuri) și să aleagă testul parametric potrivit;
- să aplice și să calculeze de mână testul t pentru eșantioane pereche, inclusiv intervalul de încredere al diferenței medii și mărimea efectului dz;
- să aplice testul t pentru eșantioane independente în varianta Student (variață comună) și în varianta Welch, să interpreteze testul Levene și să calculeze d Cohen;
- să explice logica analizei variației (variația dintre grupuri față de variația din interiorul grupurilor) și să construiască de mână tabelul ANOVA (SS, gl, MS, F);
- să verifice condițiile de aplicare (independență, normalitate, omogenitatea variațiilor) și să aleagă alternativa când acestea nu sunt îndeplinite (Welch, Games-Howell, teste neparametrice);
- să aleagă și să interpreteze testele post hoc (Tukey, Bonferroni, Games-Howell, Dunnett) și să explice de ce nu se fac teste t multiple;
- să descrie ANOVA cu măsurători repetate (sfericitate, corecția Greenhouse-Geisser), ideea ANOVA cu doi factori și rolul ANCOVA în ajustarea pentru valoarea inițială;
- să citească critic rezultatele testelor t și ANOVA din articole publicate și să raporteze corect propriile rezultate.

6.1 Familia testelor parametrice

Testele din acest capitol se numesc parametrice pentru că fac presupuneri despre forma distribuției din populație (de obicei, distribuția normală) și compară parametrii acesteia, mediile. În schimbul acestor presupuneri, ele folosesc toată informația din date și au, când condițiile sunt îndeplinite, cea mai mare putere. Testele neparametrice din capitolul 7 (Mann-Whitney, Wilcoxon, Kruskal-Wallis, Friedman) fac mai puține presupuneri și lucrează cu rangurile observațiilor; sunt alternativa naturală când distribuția este puternic asimetrică, eșantionul este mic sau variabila este ordinală [19,63].

Alegerea testului parametric depinde, în primul rând, de designul studiului. Tabelul 6.1 sintetizează situațiile pe care le vom studia; algoritmul complet de alegere a testului, inclusiv pentru variabilele calitative, apare în capitolul 8.

Tabelul 6.1. Teste parametrice pentru compararea mediilor și alternativele lor neparametrice.

Situația	Exemplu în mg2026	Test parametric	Alternativă neparametrică
o medie față de o valoare dată	media TAS față de 125 mmHg	testul z sau t pentru o singură medie	Wilcoxon pentru o mediană
două măsurători pe aceiași pacienți	glicemia la 0 și la 6 luni	testul t pentru eșantioane pereche	Wilcoxon al rangurilor cu semn
două grupuri independente	TAS la femei și la bărbați	testul t pentru eșantioane independente (Student sau Welch)	Mann-Whitney U

Situația	Exemplu în mg2026	Test parametric	Alternativă neparametrică
trei sau mai multe grupuri independente	scăderea glicemiei pe brațele A, B, C	ANOVA cu un factor	Kruskal-Wallis
trei sau mai multe măsurători pe aceiași pacienți	glicemia la 0, 3 și 6 luni	ANOVA cu măsurători repetate	Friedman

Toate testele din tabel au aceeași structură, cea din capitolul 5: o ipoteză nulă de egalitate a mediilor, o statistică a testului care raportează semnalul la zgomot, o distribuție cunoscută a statisticii sub H0 (t sau F), o valoare p și o decizie. Diferă doar modul în care se măsoară semnalul și zgomotul.

6.2 Compararea unei medii cu o valoare dată: testul z și testul t

Prima situație din tabel a fost tratată pe larg în capitolul 5. O reluăm aici pe scurt, cu regula din cursul disciplinei. Testul z se folosește dacă volumul eșantionului este n ≥ 30 și dispersia populației este cunoscută; statistica este $z = \frac{m-\mu_0}{\frac{\sigma}{\sqrt{n}}}$, iar valorile critice provin din distribuția normală standard (1,96 pentru α = 0,05 bilateral). Testul t se folosește dacă n < 30 sau dispersia populației este necunoscută; statistica este $t = \frac{m-\mu_0}{\frac{s}{\sqrt{n}}}$, cu n - 1 grade de libertate. Ambele teste presupun o distribuție aproximativ normală a variabilei în populație, condiție care devine mai puțin importantă pe măsură ce n crește.

În practica de azi, σ este rareori cunoscut, iar programele folosesc testul t indiferent de n. Pentru media TAS din mg2026 față de 125 mmHg am obținut t = 5,59 cu 159 grade de libertate, p < 0,001; pentru glicemia a jeun față de 100 mg/dL, t = 4,36, p < 0,001 (capitolul 5). Ce ne interesează acum este că testul t pentru eșantioane pereche, primul test nou al capitolului, se reduce exact la acest test pentru o singură medie.

6.3 Testul t pentru eșantioane pereche

6.3.1 Când datele sunt pereche

Două serii de valori sunt pereche (dependente) când fiecare valoare dintr-o serie are o corespondentă naturală, unică, în cealaltă. Situațiile clinice tipice sunt patru. Același pacient măsurat înainte și după un tratament (glicemia la includere și la 6 luni). Același pacient măsurat cu două metode sau două aparate (tensiunea arterială măsurată cu aparatul automat și cu cel manual). Două organe pereche ale aceluiași pacient (ochiul drept și ochiul stâng). Perechi potrivite (engl. matched pairs) de pacienți diferiți, aleși să semene ca vârstă, sex și severitate, dintre care unul primește tratamentul și celălalt nu, sau gemeni.

Ce au în comun aceste situații? Valorile din aceeași pereche sunt corelate: un pacient cu glicemia mare la includere va avea, de regulă, glicemia relativ mare și la 6 luni. Testul pereche folosește această corelație în avantajul nostru, eliminând din comparație diferențele dintre pacienți și păstrând doar modificarea din interiorul fiecărui pacient.

6.3.2 Logica: testul lucrează pe diferențe

Ideea este simplă și elegantă. Pentru fiecare pacient calculăm diferența dintre cele două măsurători, di = xi,înainte - xi,după. Cele două serii devin o singură serie de n diferențe. Dacă tratamentul nu are efect, media diferențelor în populație este zero. Testul pereche este, prin urmare, testul t pentru o singură medie aplicat diferențelor, cu valoarea de referință μ0 = 0:

H0: μD = 0 (media diferențelor în populație este zero; între cele două momente nu există modificare)

H1: μD ≠ 0 (bilateral)

Statistica testului:

$$t = \frac{m_D}{\text{ESD}} = \frac{m_D}{\frac{s_D}{\sqrt{n}}}, \text{ cu gl = n - 1}$$

unde mD este media diferențelor, sD abaterea standard a diferențelor, iar n numărul de perechi (nu numărul total de măsurători). Intervalul de încredere de 95% pentru diferența medie este $m_D \pm t$ critic × ESD, iar mărimea efectului pentru date pereche este dz = $\frac{m_D}{s_D}$.

Condițiile de aplicare sunt: perechile sunt independente între ele (pacienții diferiți nu se influențează), iar diferențele au o distribuție aproximativ normală. Atenție la acest detaliu, care apare des la examen: normalitatea se verifică pe diferențe, nu pe fiecare serie separat. Glicemia la includere din mg2026 nu este perfect normală (Shapiro-Wilk p = 0,037), dar diferențele glicemie0 - glicemie6 sunt (p = 0,725), iar testul pereche este pe deplin justificat. Dacă diferențele sunt puternic asimetrice sau au valori extreme, mai ales la n mic, se folosește testul Wilcoxon al rangurilor cu semn (capitolul 7).

6.3.3 Un exemplu calculat de mână

Zece pacienți hipertensivi din cabinetul de medicină de familie au TAS măsurată înainte și după patru săptămâni de tratament cu un inhibitor al enzimei de conversie. Datele și calculul sunt în tabelul 6.2; rezolvarea completă, cu toți pașii formatului a-g, este problema rezolvată 6.1.

Tabelul 6.2. TAS (mmHg) înainte și după 4 săptămâni de tratament la 10 pacienți (date fictive).

Pacient	Înainte	După	Diferența d	d - mD	$(d - m_D)^2$
1	152	140	12	4,2	17,64
2	148	142	6	-1,8	3,24
3	160	148	12	4,2	17,64
4	145	140	5	-2,8	7,84
5	158	150	8	0,2	0,04
6	150	147	3	-4,8	23,04
7	165	150	15	7,2	51,84
8	142	138	4	-3,8	14,44
9	155	146	9	1,2	1,44
10	149	145	4	-3,8	14,44
Total / medie	152,4	144,6	78 / 7,8	0	151,60

Media diferențelor este $m_D = \frac{78}{10} = 7{,}8$ mmHg. Abaterea standard a diferențelor: $s_D = \sqrt{\frac{151{,}60}{9}} = \sqrt{16{,}84} = 4{,}10$ mmHg. Eroarea standard: ESD = $\frac{4{,}10}{\sqrt{10}} = \frac{4{,}10}{3{,}162} = 1{,}30$ mmHg. Statistica testului: $t = \frac{7{,}8}{1{,}30} = 6{,}01$, cu 9 grade de libertate. Valoarea critică pentru gl = 9 și α = 0,05 bilateral este 2,262 (tabelul 5.6), iar pentru α = 0,01 este 3,250; t = 6,01 le depășește pe amândouă. Valoarea p exactă este 0,0002. IC 95% pentru scăderea medie: 7,8 ± 2,262 × 1,30 = 7,8 ± 2,94, adică 4,86-10,74 mmHg. Mărimea efectului dz = $\frac{7{,}8}{4{,}10} = 1{,}90$, un efect foarte mare.

Ce s-ar fi întâmplat dacă am fi ignorat împerecherea și am fi comparat cele două coloane ca și cum ar proveni de la pacienți diferiți? Testul t pentru eșantioane independente ar fi dat t = 2,96 și p = 0,008: tot semnificativ, dar cu o statistică de două ori mai mică. Motivul este că, în analiza independentă, zgomotul include diferențele mari dintre pacienți (TAS inițială variază între 142 și 165 mmHg, AS 7,11), pe când în analiza pereche zgomotul este doar variabilitatea răspunsului la tratament (AS a diferențelor 4,10). Pe date reale, cu efecte mai mici, ignorarea împerecherii transformă des un rezultat semnificativ într-unul nesemnificativ. Greșeala inversă, tratarea unor grupuri independente ca pereche, este și mai gravă, pentru că produce rezultate fără sens: împerecherea trebuie să existe în design, nu să fie inventată la analiză.

6.3.4 Testul pereche pe setul mg2026

În mg2026, glicemia a jeun a fost măsurată la includere și la 6 luni la toți cei 160 de pacienți. Figura 6.1 arată, în panoul A, traiectoriile individuale pentru 25 de pacienți (fiecare linie unește valorile aceluiași pacient) și traiectoria mediei; în panoul B, histograma celor 160 de diferențe individuale.

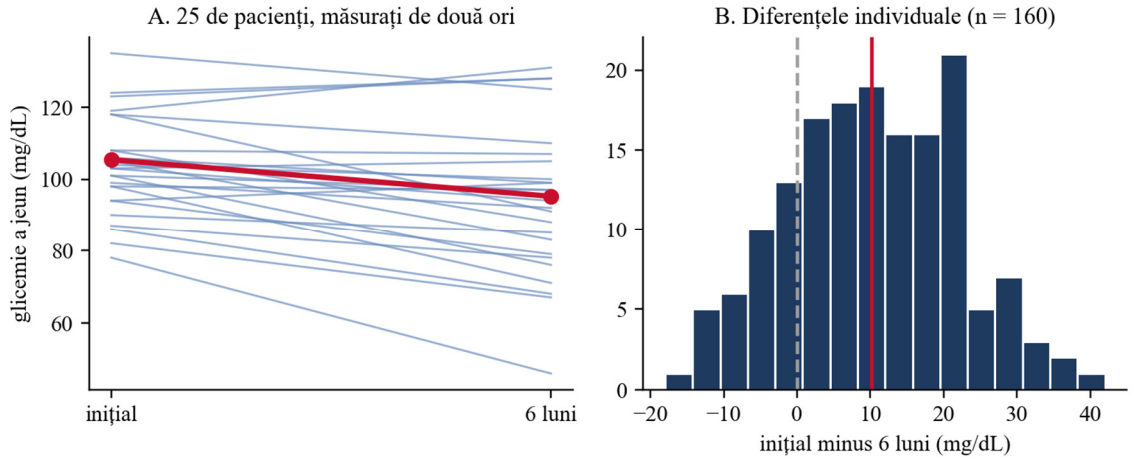


Figura 6.1. Glicemia a jeun la includere și la 6 luni. A. Traiectoriile a 25 de pacienți (linii subțiri) și media grupului (linia groasă). B. Histograma diferențelor individuale (inițial minus 6 luni) la cei 160 de pacienți; linia punctată marchează diferența zero, iar linia continuă media diferențelor (10,19 mg/dL). Date sintetice (setul mg2026).

Graficul spune deja povestea înainte de orice calcul. Majoritatea liniilor coboară, dar nu toate: 125 de pacienți au glicemia mai mică la 6 luni, 34 o au mai mare, iar unul are aceeași valoare. Histograma diferențelor este aproximativ simetrică, centrată în jurul a 10 mg/dL, cu valori între -18 și +42 mg/dL.

Rezultatele numerice: media la includere 105,4 mg/dL, la 6 luni 95,2 mg/dL; diferența medie 10,19 mg/dL, cu abaterea standard a diferențelor 11,57 mg/dL. ESD $= \frac{11{,}57}{\sqrt{160}} = \frac{11{,}57}{12{,}649} =$ 0,915 mg/dL. $t = \frac{10{,}19}{0{,}915} = 11{,}14$, cu 159 grade de libertate, p < 0,001. IC 95% pentru diferența medie: 10,19 ± 1,975 × 0,915 = 10,19 ± 1,81, adică 8,38-11,99 mg/dL. Mărimea efectului dz $= \frac{10{,}19}{11{,}57} = 0{,}88$. Testul Shapiro-Wilk pe diferențe dă p = 0,725, deci condiția de normalitate este îndeplinită; testul Wilcoxon, pentru comparație, dă tot p < 0,001.

Interpretarea are două niveluri. Statistic, scăderea glicemiei este extrem de semnificativă și mare ca efect. Clinic, trebuie să ne amintim că analiza pereche pe toți pacienții nu are grup de control: ea descrie evoluția lotului, nu efectul unei intervenții anume. Întrebarea despre eficacitatea aplicației și a teleconsultației primește răspuns doar prin compararea brațelor (secțiunea 6.5). Aceeași logică se aplică pe perioade mai scurte: între includere și 3 luni, glicemia

scade în medie cu 6,08 mg/dL($\text{IC}\,95\%\,4{,}61 - 7{,}54;\ t = 8{,}22;\ p < 0{,}001;\ \text{dz} = 0{,}65$), iar între 3 și 6 luni cu încă 4,11 mg/dL($\text{IC}\,95\%\,2{,}10 - 6{,}13;\ t = 4{,}03;\ p < 0{,}001;\ \text{dz} = 0{,}32$).

Un exemplu util pentru discuția despre condițiile de aplicare apare dacă privim doar brațul A (consiliere standard, n = 53) și greutatea corporală. Scăderea medie în greutate între includere și 6 luni este 0,71 kg (AS 2,48 kg). Testul t pereche dă t = 2,07, gl = 52, p = 0,043, deci un rezultat semnificativ la limită; testul Wilcoxon pe aceleași date dă p = 0,052, la limită nesemnificativ. Diferențele au o distribuție compatibilă cu normalitatea (Shapiro-Wilk p = 0,696), deci testul t este alegerea justificată. Exemplul arată cât de fragil este un rezultat aflat lângă prag: două teste rezonabile dau decizii opuse, iar scăderea de 0,7 kg în 6 luni este, oricum, fără importanță clinică. Alegerea testului se face după condițiile de aplicare, verificate înainte, nu după valoarea p care convine.

Studiu de caz: datele pereche în farmacologie și în nutriție

Un studiu înainte-după asupra pacienților hipertensivi cu disfuncție endotelială a urmărit nivelul seric al trombospondinei-1 la includere (T0) și după un an de tratament cu perindopril (T1): $8061{,}9 \pm 3684{,}80$ ng/mL la T0 față de 15380 ± 5887 ng/mL la T1, p < 0,001, pe lângă un grup martor de persoane nehipertensive ($9221{,}03 \pm 6510{,}21$ ng/mL) [86]. Al doilea marker raportat a scăzut de la 0,91 ± 0,51 la $0{,}50 \pm 0{,}24$ ng/mL($p < 0{,}001$). Designul este clasic pentru testul pereche (sau pentru Wilcoxon, dacă distribuțiile sunt asimetrice; abaterile standard mari în raport cu mediile sugerează această posibilitate). Exercițiul de lectură critică este următorul: din media și AS la fiecare moment nu se poate reconstitui testul pereche, pentru că testul folosește abaterea standard a diferențelor individuale, care depinde de corelația dintre cele două măsurători. O raportare completă ar include diferența medie (sau mediana diferențelor) cu intervalul ei de încredere. Un exemplu de raportare de acest tip vine din studiul prospectiv pe 921 de studenți urmăriți în primul an de facultate: modificarea medie a IMC a fost -0,45 ± 1,07 kg/m² (IC 95% -0,52 la -0,38; p < 0,001), iar 48,2% dintre participanți au avut o scădere a IMC, 38,7% au rămas stabili și 13,1% au crescut [5]. Putem verifica intervalul: $\text{ESD} = \frac{1{,}07}{\sqrt{921}} = 0{,}035$, iar -0,45 ± 1,96 × 0,035 dă -0,52 la -0,38, exact valorile raportate. Observați și relevanța clinică: o scădere medie de 0,45 kg/m² este extrem de semnificativă la n = 921, dar mică pentru un individ; proporțiile de studenți care cresc sau scad în greutate spun mai mult decât media. Un al treilea exemplu, studiul retrospectiv asupra inhibitorilor SGLT-2 la 200 de pacienți cu diabet de tip 2, raportează modificările înainte-după ca mediane cu IQR (de exemplu, glicemia de la 180,00 la 130,00 mg%, p < 0,001), ceea ce sugerează folosirea testului Wilcoxon, discutat în capitolul următor [59].

6.4 Testul t pentru eșantioane independente

6.4.1 Designul și ipotezele

Două eșantioane sunt independente când provin de la pacienți diferiți și nu există nicio corespondență între valorile unui grup și cele ale celuilalt: femei și bărbați, fumători și nefumători, pacienții tratați cu medicamentul A și cei tratați cu medicamentul B într-un studiu randomizat cu grupuri paralele. Grupurile pot avea dimensiuni diferite. Întrebarea este dacă mediile din cele două populații diferă:

H0: $\mu 1 = \mu 2$ (sau $\mu 1 - \mu 2 = 0$)

H1: $\mu 1 \neq \mu 2$ (bilateral)

6.4.2 Varianta Student, cu varianță comună

Semnalul este diferența dintre mediile eșantioanelor, m1 - m2. Zgomotul este eroarea standard a acestei diferențe. În varianta clasică, Student presupune că cele două populații au aceeași varianță și o estimează printr-o varianță comună (engl. pooled variance), o medie ponderată a celor două varianțe:

$$s^2_{\text{comună}} = \frac{(n_1 - 1) \times {s_1}^2 + (n_2 - 1) \times {s_2}^2}{n_1 + n_2 - 2}$$

$$\text{ES}(m_1 - m_2) = \sqrt{\left[s^2_{\text{comună}} \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]}$$

$t = \frac{m_1 - m_2}{\text{ES}(m_1 - m_2)}$, cu gl = n1 + n2 - 2

Intervalul de încredere de 95% pentru diferența mediilor este $(m_1 - m_2) \pm t$ critic × ES(m1 - m2), iar mărimea efectului este d Cohen = (m1 - m2) / scomună.

6.4.3 Varianta Welch și testul Levene

Dacă varianțele celor două populații diferă, mai ales când și volumele grupurilor sunt inegale, varianta Student poate da valori p greșite. Varianta Welch nu mai presupune varianțe egale: fiecare grup își păstrează propria varianță în calculul erorii standard, iar gradele de libertate se ajustează (formula Welch-Satterthwaite dă de obicei un număr cu zecimale):

$$\text{ES}(m_1 - m_2) = \sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}$$

$$\text{gl} = \frac{\left(\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}\right)^2}{\frac{\left(\frac{{s_1}^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{{s_2}^2}{n_2}\right)^2}{n_2 - 1}}$$

Pentru a decide între cele două variante, programele raportează testul Levene pentru egalitatea varianțelor, cu H0: $\sigma 1^2 = \sigma 2^2$. Dacă p Levene ≥ 0,05, nu avem motive să respingem egalitatea varianțelor și se poate folosi varianta Student; dacă p Levene < 0,05, se folosește varianta Welch. O regulă practică rapidă: dacă abaterea standard mai mare depășește de peste două ori abaterea standard mai mică, varianța comună nu mai este o presupunere rezonabilă.

Tot mai mulți statisticieni recomandă folosirea implicită a testului Welch, fără testul Levene prealabil: când varianțele sunt egale, Welch dă practic același rezultat ca Student, iar când nu sunt egale, îl protejează pe cercetător de o eroare. JASP oferă ambele variante prin simpla bifare a opțiunilor Student și Welch; în Google Sheets, funcția T.TEST are tipul 2 pentru varianța comună și tipul 3 pentru varianțe inegale. În acest curs, raportați testul Levene și alegeți varianta în funcție de el, dar rețineți că alegerea Welch nu este niciodată o greșeală gravă.

6.4.4 Grupuri de dimensiuni inegale

Testul t nu cere grupuri egale; în mg2026 avem 86 de femei și 74 de bărbați, iar în studiile observaționale dezechilibrele pot fi mult mai mari (de exemplu, 224 de pacienți cu diabet de tip 2 față de 62 cu diabet de tip 1 [8]). Trei lucruri trebuie totuși reținute. Primul: la același număr total de pacienți, puterea este maximă când grupurile sunt egale; un studiu cu 20 + 140 de pacienți are mult mai puțină putere decât unul cu 80 + 80. Al doilea: combinația dintre grupuri inegale și varianțe inegale este situația în care varianta Student greșește cel mai mult; dacă grupul mic are

varianța mai mare, testul Student devine prea permisiv (prea multe rezultate fals pozitive), iar dacă grupul mic are varianța mai mică, devine prea conservator. Exercițiul 3 din acest capitol arată o astfel de situație, în care Student și Welch duc la decizii opuse. Al treilea: diferențele mari de dimensiune între grupuri ridică întrebări despre selecție, pe care niciun test nu le rezolvă.

6.4.5 Testul independent pe setul mg2026

Rezultatul nesemnificativ folosit în capitolul 5 provine dintr-un test independent: TAS la femei $131{,}37 \pm 14{,}47$ mmHg$(n = 86)$ și la bărbați $130{,}85 \pm 13{,}25$ mmHg$(n = 74)$. Calculul de mână cu varianța comună: $s^2_{\text{comună}} = \frac{85\times14{,}47^2+73\times13{,}25^2}{158} = \frac{85\times209{,}38+73\times175{,}56}{158} = \frac{17797{,}4+12816{,}0}{158} = 193{,}76$, deci scomună = 13,92 mmHg. ES $= \sqrt{\left[193{,}76 \times \left(\frac{1}{86}+\frac{1}{74}\right)\right]} = \sqrt{193{,}76 \times 0{,}02514} = \sqrt{4{,}871} = 2{,}207$ mmHg. $t = \frac{130{,}85-131{,}37}{2{,}207} = -\frac{0{,}52}{2{,}207} = -0{,}24$, gl = 158, p = 0,81. Testul Levene dă p = 0,529, deci variațele pot fi considerate egale; varianta Welch dă practic același rezultat (t = -0,24, gl = 157,4, p = 0,81). Diferența bărbați minus femei este -0,5 mmHg (IC 95% -4,9 la 3,8), iar d = -0,04. Testul Mann-Whitney, pentru comparație, dă p = 0,80. Nu se respinge H0; după cum am discutat în capitolul 5, intervalul de încredere exclude diferențe medii mai mari de aproximativ 5 mmHg.

Glicemia a jeun pe sexe dă un rezultat asemănător: bărbații au media $106{,}2$ mg/dL$(\text{AS } 15{,}19;\ n = 74)$, femeile $104{,}8$ mg/dL$(\text{AS } 16{,}30;\ n = 86)$. Testul Levene p = 0,426; testul t Welch t = 0,54, gl = 157,0, p = 0,589; diferența bărbați minus femei 1,35 mg/dL (IC 95% -3,57 la 6,27), d = 0,09. Nesemnificativ. Ca să vedeți și un rezultat semnificativ pe același tip de comparație, priviți înălțimea: femeile $162{,}6 \pm 6{,}06$ cm, bărbații $175{,}97 \pm 6{,}18$ cm; Levene p = 0,890; testul t Student t = 13,79, gl = 158, p < 0,001; diferența 13,37 cm (IC 95% 11,45-15,28), d = 2,19. Aici diferența este evidentă și fără statistică, dar testul o cuantifică.

Literatura colegilor conține numeroase comparații de două medii. În studiul asupra pacienților cu insuficiență cardiacă ischemică cu fracție de ejecție redusă, comparați cu 60 de martori sănătoși de aceeași vârstă, osteoprotegerina a fost mult mai mare la pacienți (4,7 ± 0,25 față de $1{,}3 \pm 0{,}67$ ng/mL, $p < 0{,}001$), iar vitamina D3 mult mai mică (20,49 ± 7,31 față de 37,09 ± 4,59 ng/mL, p < 0,001) [87]. În studiul de diagnostic prin ecografie cu contrast, parametrul CEUS-PAT a fost 0,83 ± 0,09 la cei 37 de pacienți cu ciroză și 0,49 ± 0,11 la cei 27 de voluntari sănătoși (p < 0,0001) [2]; dacă calculăm d din valorile publicate, obținem aproximativ 3,4, o separare aproape completă a celor două grupuri. Asemenea diferențe uriașe sunt tipice designurilor caz-martor cu martori perfect sănătoși, care exagerează capacitatea de discriminare a unui test față de situația reală, în care medicul are de deosebit ciroza de alte hepatopatii (efectul de spectru, capitolul 11). În studiul transversal asupra tulburărilor interacțiunii intestin-creier, IMC a fost 28,15 ± 6,49 kg/m² la cele 52 de cazuri și 24,47 ± 4,60 kg/m² la cei 84 de martori (p = 0,001) [71]; aici grupurile sunt inegale, iar abaterile standard diferă (raport 1,4), situație în care varianta Welch este cea prudentă.

> **Studiu de caz: când abaterea standard se apropie de medie**
>
> Un studiu de diagnostic realizat în departamentul de urgență pe 85 de pacienți cu alertă de accident vascular cerebral (69 cu AVC, 16 fără AVC) a comparat doi biomarkeri serici [61]. Proteina acidă fibrilară glială (GFAP) a fost $142{,}91 \pm 102{,}19$ pg/mL la pacienții cu AVC și $37{,}76 \pm 19{,}92$ pg/mL la cei fără AVC (p < 0,001); UCH-L1 a fost 1307,68 ± 967,54 față de $189{,}81 \pm 92{,}69$ pg/mL$(p < 0{,}001)$. Diferențele sunt mari și clinic interesante, iar pentru deosebirea AVC hemoragic de cel ischemic GFAP a atins o acuratețe de 94,2% la un prag de 77,15 pg/mL. Pentru un student de anul I, rezumatul este un exercițiu excelent despre condițiile testului t. Primul semnal: abaterea standard este aproape cât media (102,19 față de 142,91;

> 967,54 față de 1307,68). O variabilă care nu poate fi negativă și are AS apropiată de medie este aproape sigur asimetrică spre dreapta; media ±2ASar include valori negative, imposibile biologic. Al doilea semnal: variațele diferă enorm între grupuri (raportul abaterilor standard este de aproximativ 5 pentru GFAP și 10 pentru UCH-L1), iar grupurile sunt foarte inegale (69 față de 16). În aceste condiții, testul t Student cu varianță comună nu este potrivit; variantele corecte sunt testul Welch pe date logaritmate sau testul Mann-Whitney, iar descrierea prin mediană și IQR ar reda mai fidel distribuțiile. Concluzia clinică a studiului (biomarkerii sunt mult mai mari în AVC) nu se schimbă, pentru că diferențele sunt foarte mari; la diferențe mai mici, alegerea testului ar putea schimba decizia. Rezumatul nu numește testul folosit, așa că exercițiul nu este o critică, ci o listă de lucruri pe care le verificați înainte de a alege testul pentru propriile date.

6.5 Analiza variației (ANOVA) cu un factor

6.5.1 De ce nu facem mai multe teste t

În mg2026 avem trei brațe de tratament. Întrebarea "diferă scăderea glicemiei între brațe?" ar putea fi abordată cu trei teste t: A față de B, A față de C, B față de C. Problema este cea din capitolul 5: fiecare test are riscul de 5% de fals pozitiv, iar riscul ca cel puțin unul dintre cele trei să fie fals pozitiv, dacă nu există nicio diferență reală, crește la aproximativ $1 - 0{,}95^3 = 0{,}143$ (testele nu sunt perfect independente, dar ordinul de mărime este corect). Pentru 4 grupuri sunt 6 comparații, pentru 5 grupuri 10, pentru k grupuri $k(k - 1)/2$, iar riscul global crește rapid. Analiza variației (ANOVA, engl. analysis of variance), dezvoltată de R. A. Fisher în anii 1920, răspunde la întrebarea globală printr-un singur test, cu riscul de eroare de tip I menținut la α [42,53].

H0: $\mu 1 = \mu 2 = \ldots = \mu k$ (toate mediile populațiilor sunt egale)

H1: cel puțin o medie diferă de celelalte

Atenție la formularea lui H1: ANOVA nu spune care medie diferă. Pentru asta sunt necesare testele post hoc (secțiunea 6.6).

6.5.2 Logica: variația dintre grupuri față de variația din interiorul grupurilor

Numele "analiza variației" pare ciudat pentru un test care compară medii. El vine din modul în care se face comparația. Variabilitatea totală a datelor (cât diferă fiecare valoare de media generală) se descompune în două părți. Variația dintre grupuri arată cât de mult diferă mediile grupurilor de media generală; dacă H0 este adevărată, ea este mică, produsă doar de întâmplare. Variația din interiorul grupurilor arată cât de mult diferă fiecare valoare de media propriului grup; ea reflectă variabilitatea biologică și de măsurare și nu depinde de H0. Raportul dintre cele două, după împărțirea fiecăreia la gradele ei de libertate, este statistica F:

F = MS între grupuri / MS în interiorul grupurilor = (variația explicată de grup) / (variația neexplicată)

Dacă H0 este adevărată, ambele MS estimează aceeași varianță, iar F este în jur de 1. Dacă mediile diferă, MS între grupuri crește, iar F devine mult mai mare decât 1. Distribuția lui F sub H0 este distribuția F a lui Fisher, cu două numere de grade de libertate: k - 1 la numărător și N - k la numitor (N fiind numărul total de observații).

Figura 6.2 arată logica pe un exemplu mic: scorul durerii (0-10) la 15 pacienți tratați cu trei analgezice, câte 5 pacienți pe grup. Săgețile arată distanța dintre fiecare medie de grup și media generală (variația dintre grupuri); liniile subțiri verticale arată distanța dintre fiecare pacient și media grupului său (variația din interiorul grupurilor). Mediile sunt bine separate, iar valorile din fiecare grup sunt strânse în jurul mediei: F va fi mare.

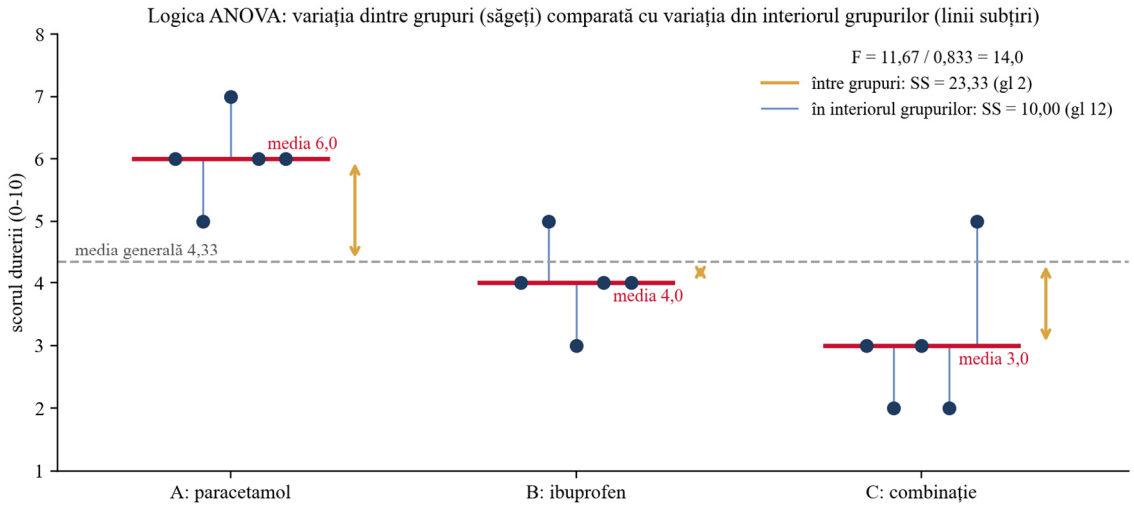


Figura 6.2. Logica ANOVA pe un exemplu cu trei analgezice (A paracetamol, B ibuprofen, C combinație; câte 5 pacienți, scorul durerii 0-10). Liniile roșii orizontale sunt mediile grupurilor, linia punctată este media generală (4,33). Săgețile reprezintă variația dintre grupuri, liniile subțiri reprezintă variația din interiorul grupurilor. Date fictive.

6.5.3 Sumele pătratelor, gradele de libertate și tabelul ANOVA

Descompunerea se face cu sume de pătrate (SS, engl. sum of squares). Notăm cu xij valoarea pacientului j din grupul i, cu mi media grupului i, cu ni numărul de pacienți din grupul i, cu M media generală și cu N numărul total de pacienți.

SS total $= \sum(x_{ij} - M)^2$, cu gl = N - 1

SS între grupuri $= \sum n_i \times (m_i - M)^2$, cu gl = k - 1

SS în interiorul grupurilor (reziduală) = $\sum(x_{ij} - m_i)^2 = \sum(n_i - 1) \times {s_i}^2$, cu gl = N - k

Proprietatea fundamentală este aditivitatea: SS total = SS între + SS în interior, iar gl total = gl între + gl în interior. Mediile pătratelor (MS, engl. mean squares) se obțin împărțind fiecare SS la gradele ei de libertate: MS între = SS între / (k - 1), MS în interior = SS în interior / (N - k). Apoi F = MS între / MS în interior, iar valoarea p este aria de sub distribuția F la dreapta valorii calculate. Toate aceste cantități se așază într-un tabel standard, pe care îl produc toate programele.

Tabelul 6.3. Structura tabelului ANOVA cu un factor.

Sursa variației	SS	gl	MS	F	p
între grupuri	$\sum n_i(m_i - M)^2$	k - 1	SS între / (k - 1)	MS între / MS în interior	din distribuția F
în interiorul grupurilor	$\sum(n_i - 1)\, {s_i}^2$	N - k	SS în interior / (N - k)		
total	$\sum(x_{ij} - M)^2$	N - 1			

Formula SS în interior = $\sum(n_i - 1) \times {s_i}^2$ este foarte utilă: ea permite construirea tabelului ANOVA doar din n, media și abaterea standard a fiecărui grup, adică din cifrele raportate în orice articol. O vom folosi pe datele mg2026 în problema rezolvată 6.6.

O măsură a mărimii efectului în ANOVA este eta pătrat: η^2 = SS între / SS total, proporția din variabilitatea totală explicată de apartenența la grup. Pragurile orientative ale lui Cohen sunt 0,01 (mic), 0,06 (mediu) și 0,14 (mare) [76].

6.5.4 Tabelul ANOVA pentru exemplul cu analgezicele

Pe datele din figura 6.2 (A: 6, 5, 7, 6, 6; B: 4, 5, 3, 4, 4; C: 3, 2, 3, 2, 5), mediile grupurilor sunt 6, 4 și 3, iar media generală este 65/15 = 4,33. SS între = 5 × [(6 - 4,33)² + (4 - 4,33)² + (3 - 4,33)²] = 5 × (2,778 + 0,111 + 1,778) = 5 × 4,667 = 23,33. SS în interior se obține adunând, în fiecare grup, pătratele abaterilor față de media grupului: în grupul A, 0 + 1 + 1 + 0 + 0 = 2; în grupul B, 0 + 1 + 1 + 0 + 0 = 2; în grupul C, 0 + 1 + 0 + 1 + 4 = 6; total 10,00. SS total = 33,33 (verificare: 23,33 + 10,00).

Tabelul 6.4. Tabelul ANOVA pentru scorul durerii la trei analgezice (date fictive).

Sursa variației	SS	gl	MS	F	p
între grupuri	23,33	2	11,67	14,00	0,0007
în interiorul grupurilor	10,00	12	0,833		
total	33,33	14			

Valoarea critică F(2; 12) pentru α = 0,05 este 3,89, iar pentru α = 0,01 este 6,93; F = 14,00 le depășește pe amândouă, iar p = 0,0007. Respingem H0: cel puțin un analgezic diferă de celelalte în privința scorului mediu al durerii. $\eta^2 = \frac{23,33}{33,33} = 0,70$: 70% din variabilitatea scorului este explicată de tipul de analgezic, un efect foarte mare (firesc, pentru un exemplu construit ca să fie clar). Care analgezice diferă între ele aflăm cu testul Tukey, în problema rezolvată 6.3.

6.5.5 Condițiile de aplicare

ANOVA cu un factor are trei condiții, asemănătoare cu cele ale testului t independent.

Independența observațiilor. Fiecare pacient apare o singură dată, într-un singur grup, iar pacienții nu se influențează între ei. Este condiția cea mai importantă și se asigură prin design; niciun test statistic nu o poate verifica sau repara. Măsurătorile repetate pe aceiași pacienți cer ANOVA cu măsurători repetate (secțiunea 6.7).

Normalitatea în fiecare grup (sau, echivalent, a reziduurilor, adică a abaterilor față de media grupului). Se verifică cu Shapiro-Wilk pe fiecare grup sau cu graficul Q-Q al reziduurilor. La grupuri mari și de dimensiuni apropiate, ANOVA este robustă la abateri moderate.

Omogenitatea varianțelor (homoscedasticitatea). Se verifică cu testul Levene (H0: varianțele sunt egale) sau cu testul Bartlett, mai sensibil la abaterea de la normalitate. Regula practică: ANOVA clasică rămâne acceptabilă dacă raportul dintre cea mai mare și cea mai mică abatere standard este sub 2 și grupurile au dimensiuni apropiate.

Când condițiile nu sunt îndeplinite, avem alternative. Pentru variațe inegale: ANOVA Welch (F cu grade de libertate ajustate) și testul post hoc Games-Howell. Pentru distribuții puternic asimetrice sau date ordinale: testul Kruskal-Wallis (capitolul 7) ori o transformare logaritmică.

6.5.6 ANOVA pe setul mg2026: scăderea glicemiei pe brațe

Ajungem la întrebarea principală a studiului mg2026. Variabila analizată este scăderea glicemiei a jeun la 6 luni, calculată pentru fiecare pacient ca glicemie0 - glicemie6 (valori pozitive înseamnă scădere). Figura 6.3 arată distribuția ei pe cele trei brațe.

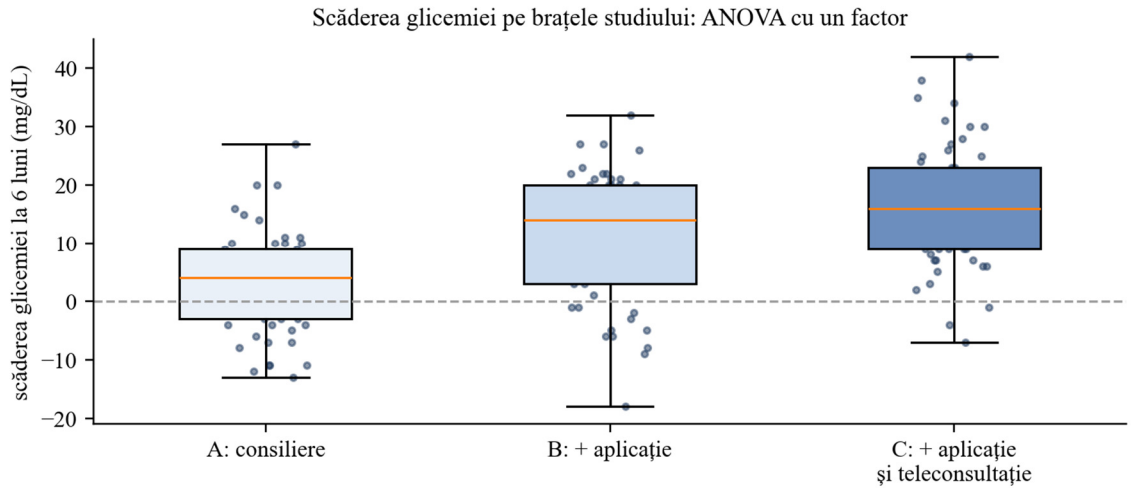


Figura 6.3. Scăderea glicemiei a jeun la 6 luni (glicemie0 - glicemie6) pe brațele studiului: A consiliere standard, B consiliere + aplicație, C consiliere + aplicație + teleconsultație. Box-plot cu valorile individuale suprapuse; linia punctată marchează scăderea zero. Date sintetice (setul mg2026).

Graficul sugerează o creștere a scăderii glicemiei de la brațul A la brațul C, cu dispersii asemănătoare. Statisticile descriptive: A, media 3,0 mg/dL(AS 8,8; $n = 53$); B, 11,1 mg/dL(AS 11,2; $n = 53$); C, 16,3 mg/dL(AS 10,6; $n = 54$); media generală 10,19 mg/dL.

Verificarea condițiilor. Pacienții sunt independenți și apar fiecare într-un singur braț (randomizare individuală). Testul Shapiro-Wilk pe fiecare braț: A p = 0,575, B p = 0,052, C p = 0,903; normalitatea este acceptabilă, brațul B fiind la limită, iar grupurile sunt mari și egale. Raportul dintre cea mai mare și cea mai mică AS este 11,2 / 8,8 = 1,27, sub 2. Testul Levene dă p = 0,163 în varianta centrată pe mediană (Brown-Forsythe) și p = 0,089 în varianta centrată pe medie; testul Bartlett, p = 0,192. Nu avem motive să respingem omogenitatea variațelor.

Tabelul 6.5. Tabelul ANOVA pentru scăderea glicemiei la 6 luni pe brațe (mg2026).

Sursa variației	SS	gl	MS	F	p
între brațe	4806,9	2	2403,4	22,93	< 0,001
în interiorul brațelor	16459,5	157	104,8		
total	21266,4	159			

F(2; 157) = 22,93, cu p < 0,001 (valoarea critică pentru α = 0,05 este 3,05). $\eta^2 = \frac{4806,9}{21266,4} = 0,226$: aproape 23% din variabilitatea scăderii glicemiei este explicată de brațul de tratament, un efect mare. ANOVA Welch, care nu presupune variațe egale, confirmă: F(2; 103,5) = 26,05, p < 0,001. Testul neparametric Kruskal-Wallis dă H = 36,63, p < 0,001.

Concluzia statistică: se respinge H0; scăderea medie a glicemiei diferă extrem de semnificativ între cel puțin două brațe. Concluzia de cercetare cere testele post hoc, pentru a vedea care brațe diferă.

6.5.7 Al doilea rezultat: scăderea în greutate

Aceeași analiză pentru scăderea în greutate la 6 luni (greutate0 - greutate6, kg) dă: A 0,71 ± 2,48 kg, B 2,84 ± 2,36 kg, C 3,37 ± 2,39 kg. Abaterile standard sunt aproape identice (Levene p = 0,79), distribuțiile sunt compatibile cu normalitatea în toate brațele (Shapiro-Wilk p între 0,17 și 0,70). ANOVA: SS între = 211,9 (gl 2), SS în interior = 911,3 (gl 157), MS între = 105,9, MS în interior = 5,80, F(2; 157) = 18,25, p < 0,001, $\eta^2 = 0,189$. Brațele cu aplicație au slăbit mai mult decât brațul cu consiliere standard; dacă și teleconsultația aduce un beneficiu suplimentar față de aplicație vom vedea tot din testele post hoc.

6.6 Testele post hoc

6.6.1 Ce problemă rezolvă

Un F semnificativ spune doar că mediile nu sunt toate egale. Pentru a afla care perechi diferă, se fac comparații ulterioare, numite teste post hoc (lat. "după aceasta"). Ele compară perechile de grupuri, dar ajustează pragul sau valorile p astfel încât riscul global de eroare de tip I să rămână la α, exact ideea corecțiilor pentru comparații multiple din capitolul 5. Există multe teste post hoc; în medicina clinică se folosesc patru.

Testul Tukey HSD (engl. honestly significant difference) compară toate perechile posibile și este alegerea standard când variațele sunt omogene și grupurile au dimensiuni apropiate. Folosește distribuția "intervalului studentizat" q și calculează diferența minimă semnificativă:

$$HSD = q(k; N - k; \alpha) \times \sqrt{(MS \text{ în interior} / n)}$$

unde n este numărul de pacienți dintr-un grup (pentru grupuri inegale, programele folosesc varianta Tukey-Kramer). Două medii diferă semnificativ dacă diferența lor depășește HSD.

Corecția Bonferroni aplicată testelor t perechi înmulțește fiecare valoare p cu numărul de comparații. Este simplă, dar conservatoare când comparațiile sunt multe; varianta Holm este preferabilă.

Testul Games-Howell nu presupune variațe egale și nici grupuri egale. Este alegerea potrivită când testul Levene este semnificativ sau când abaterile standard diferă mult, de obicei împreună cu ANOVA Welch.

Testul Dunnett compară fiecare grup de tratament doar cu un grup de control, nu toate perechile între ele. Fiind mai puține comparații, are mai multă putere. Este potrivit când întrebarea este "care tratamente diferă de control?" și nu interesează comparațiile dintre tratamente.

6.6.2 Testul Tukey pe setul mg2026

Pentru scăderea glicemiei pe brațe, testul Tukey dă rezultatele din tabelul 6.6.

Tabelul 6.6. Comparațiile post hoc pentru scăderea glicemiei la 6 luni (mg2026).

Comparația	Diferența medie (mg/dL)	IC 95% Tukey	p Tukey	p Games-Howell	p Bonferroni
B - A	8,1	3,4-12,8	0,0002	0,0002	0,0002
C - A	13,3	8,6-18,0	< 0,001	< 0,001	< 0,001
C - B	5,2	0,6-9,9	0,024	0,038	0,044

Toate cele trei perechi diferă semnificativ. Aplicația (brațul B) aduce o scădere suplimentară de 8,1 mg/dL față de consilierea standard, iar teleconsultația lunară (brațul C) încă 5,2 mg/dL față de aplicație. Diferența C - B este cea mai mică și cea mai aproape de prag: rămâne semnificativă cu toate cele trei metode, dar intervalul ei de încredere începe la 0,6 mg/dL, deci beneficiul suplimentar al teleconsultației ar putea fi și foarte mic. Observați cum metodele mai conservatoare (Games-Howell, Bonferroni) dau valori p puțin mai mari pentru această pereche; dacă p Tukey ar fi fost 0,045, alegerea metodei ar fi putut schimba decizia, de aceea ea se stabilește înainte de analiză.

Pentru scăderea în greutate, Tukey arată că brațul B (diferența față de A 2,13 kg, $p < 0,001$) și brațul C (2,66 kg, $p < 0,001$) diferă de A, dar C nu diferă semnificativ de B (0,53 kg, $p = 0,489$). Combinând cele două rezultate, aplicația pare să aducă cea mai mare parte a beneficiului asupra greutății, iar teleconsultația adaugă un beneficiu suplimentar mai ales asupra glicemiei. Formulăm prudent, pentru că lipsa unei diferențe semnificative C - B la greutate nu dovedește egalitatea celor două brațe.

6.6.3 Alte variante de analiză și ce nu se face

Dacă întrebarea ar fi fost doar "aplicația și teleconsultația sunt mai bune decât consilierea standard?", testul Dunnett, cu brațul A drept control, ar fi fost alegerea firească. Ce nu se face: nu se aplică teste t simple, necorectate, între toate perechile; nu se fac teste post hoc după un F nesemnificativ, în căutarea unei perechi "care iese" (cu excepția comparațiilor planificate dinainte, justificate în protocol); nu se interpretează doar perechile semnificative, ignorând mărimea diferențelor și intervalele de încredere.

Un exemplu publicat de analiză pe mai multe grupuri ordonate vine din ecografia hepatică. Parametrul UGAP a crescut monoton cu gradul de steatoză stabilit prin metoda de referință: S_0 198,3 $\pm \frac{25,7 d_B}{m}$, S_1 216,86 $\pm \frac{26,3 d_B}{m}$, S_2 237,79 $\pm \frac{26,3 d_B}{m}$, S_3 270,8 $\pm \frac{31,62 d_B}{m}$ ($p < 0,001$), cu grupuri de 48, 56, 14 și 59 de subiecți [58]. Abaterile standard sunt foarte apropiate (raport 1,2), deci ANOVA clasică este potrivită; grupul S2 este însă mic, iar comparațiile care îl implică au mai puțină putere. Când grupurile sunt ordonate natural (stadii, grade), pe lângă ANOVA se poate testa și tendința liniară, care răspunde direct la întrebarea "crește parametrul odată cu gradul?". Alte două studii ale colegilor compară trei grupuri fără să numească testul în rezumat: un studiu prospectiv pe 150 de gravide împărțite în trei grupuri egale de câte 50 a găsit diferențe semnificative ale rigidității arteriale ($p < 0,001$), ale frecvenței cardiace ($p = 0,006$), ale presiunii sistolice aortice ($p = 0,02$), ale indicelui de augmentare brahial ($p = 0,01$) și ale presiunii pulsului ($p = 0,049$) [88], iar un studiu cu trei grupuri paralele de pacienți după AVC (22, 18 și 30 de pacienți) a raportat rezultate mai bune în grupurile cu terapie combinată ($p < 0,05$) și ameliorări față de momentul inițial la 12 luni ($p < 0,05$) [89]. În ambele cazuri, cititorul ar avea nevoie de numele testului (ANOVA sau Kruskal-Wallis), de testele post hoc folosite și de mediile cu intervalele lor de încredere pentru a judeca mărimea diferențelor; când aceleași paciente sau aceiași pacienți sunt evaluați la mai multe momente, este necesară și o analiză pentru măsurători repetate.

Studiu de caz: ANOVA, Tukey și eta² în două studii ale colegilor

Studiul transversal asupra protezelor auditive inteligente a comparat stima de sine (scala SES) la 245 de persoane împărțite în trei grupuri: 73 cu auz normal, 86 de purtători de proteze auditive și 86 de pacienți cu hipoacuzie neurosenzorială fără proteză [82]. ANOVA cu un factor a dat $F = 299,00$, $p < 0,001$, cu $\eta^2 = 0,712$, iar testul post hoc Tukey HSD a arătat că purtătorii de proteze au scoruri mai mari decât cei fără proteză (35,41 ± 5,32 față de $22,99 \pm 4,53$; $p < 0,001$; d Cohen = 2,515). Un η^2 de 0,712 înseamnă că 71% din variabilitatea stimei de sine este asociată cu grupul, un efect excepțional de mare în științele comportamentale. Câteva observații de lectură critică: abaterile standard sunt apropiate (5,32 și 4,53), deci condiția de omogenitate pare îndeplinită și Tukey este testul potrivit; grupurile au dimensiuni apropiate; designul este transversal, deci rezultatul arată o asociere puternică, dar nu stabilește că proteza a produs creșterea stimei de sine (persoanele care aleg să poarte proteză pot diferi și prin alte caracteristici). Al doilea studiu a comparat deformarea longitudinală globală a ventriculului stâng (GLS) la 148 de pacienți din trei grupuri: 62 cu insuficiență cardiacă cu fracție de ejecție păstrată, 46 hipertensivi și 40 de martori de aceeași vârstă [90]. Diferența globală a fost semnificativă ($p < 0,001$), iar comparațiile pe perechi au arătat că martorii (-20,2 ± 1,4%) diferă de hipertensivi ($-18,4 \pm 3,0\%$, $p = 0,031$) și de pacienții cu insuficiență cardiacă ($-17,6 \pm 2,3\%$, $p < 0,001$). Aici abaterea standard a hipertensivilor este de peste două ori mai mare decât a martorilor; într-o reanaliză, ANOVA Welch cu testul Games-Howell ar fi o verificare utilă. Observați și că diferența martori-hipertensivi, deși semnificativă, are un p mult mai aproape de prag decât diferența martori-insuficiență cardiacă, ceea ce corespunde unei diferențe medii mai mici (1,8 față de 2,6 puncte procentuale).

6.7 ANOVA cu măsurători repetate

6.7.1 Când aceiaşi pacienţi sunt măsuraţi de mai multe ori

În mg2026, glicemia a fost măsurată la includere, la 3 luni şi la 6 luni la aceiaşi pacienți. Cele trei serii nu sunt independente, deci ANOVA cu un factor nu este potrivită; nici trei teste t pereche necorectate nu sunt (inflația erorii de tip I). Testul potrivit este ANOVA cu măsurători repetate, generalizarea testului t pereche la trei sau mai multe momente. Ca și testul pereche, el elimină din zgomot diferențele constante dintre pacienți și compară doar modificările din interiorul fiecărui pacient.

H0: μ0 luni = μ3 luni = μ6 luni (media glicemiei nu se modifică în timp)

H1: media diferă la cel puțin un moment

6.7.2 Sfericitatea şi corecţia Greenhouse-Geisser

Pe lângă normalitate, ANOVA cu măsurători repetate are o condiție specifică, sfericitatea: variațele diferențelor dintre toate perechile de momente (0-3, 0-6, 3-6) trebuie să fie egale. Se verifică cu testul Mauchly (H0: sfericitatea este respectată). Dacă testul Mauchly este semnificativ, gradele de libertate ale testului F se înmulțesc cu un factor de corecție ε (epsilon) subunitar, cel mai des estimat prin metoda Greenhouse-Geisser (mai conservatoare) sau Huynh-Feldt. Corecția nu schimbă valoarea F, ci gradele de libertate, deci valoarea p crește puțin. Pentru două momente de timp, sfericitatea este automat îndeplinită, iar ANOVA cu măsurători repetate dă exact același rezultat ca testul t pereche ($F = t^2$).

6.7.3 Rezultatul pe mg2026

Mediile glicemiei sunt 105,4 / 99,4 / 95,2 mg/dL la 0 / 3 / 6 luni. ANOVA cu măsurători repetate: SS timp = 8405,5 (gl 2), SS eroare = 20570,5 (gl 318), MS timp = 4202,8, MS eroare = 64,69, F(2; 318) = 64,97, p < 0,001. Mărimea efectului, eta² parțial = 8405,5 / (8405,5 + 20570,5) = 0,29. Testul Mauchly este semnificativ (W = 0,873, p < 0,001), deci sfericitatea nu este respectată; corecția Greenhouse-Geisser are ε = 0,887, gradele de libertate devin 1,77 și 282,2, iar p rămâne < 0,001. Comparațiile post hoc cu corecție Bonferroni arată că toate cele trei perechi de momente diferă (0-3 luni, 0-6 luni și 3-6 luni, toate p < 0,001). Testul neparametric echivalent, Friedman, dă χ^2 = 74,67, p < 0,001 (capitolul 7).

Concluzia: glicemia medie scade semnificativ și progresiv pe parcursul celor 6 luni, cu o scădere mai mare în primele 3 luni (6,1 mg/dL) decât în următoarele 3 (4,1 mg/dL). Ca și la testul pereche, analiza pe tot lotul descrie evoluția în timp, nu efectul unei intervenții anume; pentru întrebarea "diferă evoluția în timp între brațe?" este necesar un model mixt, cu factorul timp (în interiorul pacienților) și factorul braț (între pacienți), al cărui termen de interacțiune timp × braț răspunde exact la această întrebare. JASP permite acest model din același meniu (Repeated Measures ANOVA, cu brațul adăugat la "Between Subject Factors").

> **Studiu de caz: măsurători repetate într-un studiu pilot pediatric fără grup de control**
>
> Un studiu pilot prospectiv, cu un singur braț, a urmărit 21 de copii cu dermatită atopică și disbioză intestinală confirmată în laborator, tratați cu o terapie personalizată țintită pe microbiota intestinală, cu evaluări la includere, la 30 și la 90 de zile [91]. Scorul POEM a scăzut de la 14,67 ± 5,22 la includere la 7,10 ± 3,82 la 90 de zile, iar scorul SCORAD de la 41,66 ± 15,28 la 17,93 ± 11,68 (ambele p < 0,001); aproximativ 76% dintre copii au obținut o reducere de cel puțin 50% a scorului POEM și 71% a scorului SCORAD. Analiza statistică folosită, ANOVA cu măsurători repetate, este alegerea corectă pentru trei evaluări ale aceloraşi copii; raportarea ei completă ar include testul de sfericitate și, dacă este cazul, corecția Greenhouse-Geisser, precum și comparațiile post hoc între momente. Lecția principală privește

> însă designul. Într-un studiu cu un singur braț, scăderea scorurilor nu poate fi atribuită cu certitudine tratamentului: dermatita atopică are o evoluție fluctuantă, copiii sunt adesea incluși într-un moment de exacerbare (deci regresia spre medie este probabilă), iar participarea la studiu schimbă îngrijirea. Autorii descriu corect studiul ca pilot; rezultatele foarte semnificative statistic justifică un studiu randomizat controlat, nu o recomandare terapeutică. Este exact diferența dintre testul pereche pe tot lotul mg2026 (evoluție în timp) și comparația dintre brațele randomizate (efectul intervenției).

6.8 ANOVA cu doi factori și ANCOVA, pe scurt

6.8.1 ANOVA cu doi factori

Uneori ne interesează simultan efectul a doi factori asupra unei variabile numerice, de exemplu brațul de tratament și sexul pacientului asupra scăderii glicemiei. ANOVA cu doi factori (engl. two-way ANOVA) testează trei ipoteze: efectul principal al primului factor (brațul), efectul principal al celui de-al doilea factor (sexul) și interacțiunea dintre ei (efectul brațului diferă între femei și bărbați?). Interacțiunea este adesea cea mai interesantă întrebare clinică: un tratament care funcționează doar la femei sau doar la vârstnici.

În mg2026, ANOVA brat × sex pentru scăderea glicemiei dă: efectul brațului $F(2; 154) = 23{,}03$, $p < 0{,}001$; efectul sexului $F(1; 154) = 0{,}66$, $p = 0{,}416$; interacțiunea $F(2; 154) = 0{,}53$, $p = 0{,}591$. Mediile pe celule sunt, la femei și la bărbați: A 2,6 și 3,6 mg/dL; B 12,7 și 9,6 mg/dL; C 17,2 și 15,4 mg/dL. Concluzia: efectul intervenției nu diferă semnificativ între sexe, deci putem raporta efectul brațului pentru tot lotul. Atenție, ca întotdeauna: interacțiunile au putere mică, iar lipsa unei interacțiuni semnificative nu dovedește că efectul este identic la femei și la bărbați.

6.8.2 ANCOVA: ajustarea pentru valoarea inițială

Analiza covarianței (ANCOVA) combină ANOVA cu regresia (capitolul 9): compară mediile grupurilor după ajustarea pentru o variabilă numerică, numită covariabilă. În studiile randomizate, cea mai frecventă aplicație este ajustarea pentru valoarea inițială a rezultatului. Chiar și cu randomizare, brațele pot porni de la valori puțin diferite: în mg2026, glicemia medie la includere a fost 107,8 mg/dL în brațul A, 106,3 în B și 102,3 în C. ANCOVA compară glicemiile la 6 luni ca și cum toți pacienții ar fi pornit de la aceeași glicemie inițială.

Rezultatul pe mg2026: după ajustarea pentru glicemia inițială, glicemia la 6 luni diferă între brațe, $F(2; 156) = 22{,}26$, $p < 0{,}001$; față de brațul A, glicemia ajustată este mai mică cu 8,07 mg/dL în brațul B (IC 95% 4,12-12,01) și cu 13,29 mg/dL în brațul C (IC 95% 9,32-17,25). Estimările sunt aproape identice cu diferențele dintre scăderile medii din ANOVA (8,1 și 13,3 mg/dL), ceea ce arată că micile dezechilibre de la includere nu au influențat concluzia. ANCOVA este analiza recomandată pentru studiile randomizate cu rezultat numeric măsurat înainte și după, pentru că are, în general, mai multă putere decât compararea simplă a valorilor finale sau a diferențelor. Ghidul CONSORT recomandă ca ajustările să fie stabilite în protocol, nu alese după vizualizarea datelor, și descurajează testarea statistică a diferențelor de la includere dintre brațele randomizate [56]; dacă randomizarea a fost corectă, orice diferență inițială este, prin definiție, întâmplătoare.

6.9 Raportarea testelor t și ANOVA

O raportare completă a unui test t sau a unei ANOVA cuprinde: statisticile descriptive pe grupuri (n, media, AS) sau pe momente; verificarea condițiilor (normalitatea, omogenitatea variațelor, sfericitatea, cu testul folosit); varianta testului (Student sau Welch; ANOVA clasică sau Welch); statistica testului cu gradele de libertate ($t(159) = 11{,}14$; $F(2; 157) = 22{,}93$); valoarea p exactă;

mărimea efectului (d, dz, eta²) și intervalul de încredere pentru diferențele dintre medii; testul post hoc și rezultatele lui; interpretarea clinică [48,53].

Exemplu pentru mg2026: "Scăderea glicemiei a jeun la 6 luni a diferit între brațe (A: 3,0 ± 8,8; B: 11,1 ± 11,2; C: 16,3 ± 10,6 mg/dL; ANOVA cu un factor, $F(2;\ 157) = 22{,}93$, $p < 0{,}001$, $\eta^2 = 0{,}23$; testul Levene p = 0,163). Comparațiile post hoc Tukey au arătat scăderi mai mari în brațul B față de A (diferența 8,1 mg/dL, IC 95% 3,4-12,8; p < 0,001), în brațul C față de A (13,3 mg/dL, IC 95% 8,6-18,0; p < 0,001) și în brațul C față de B (5,2 mg/dL, IC 95% 0,6 − 9,9; $p = 0{,}024$)."

Pentru grafice, box-plotul cu punctele individuale (figura 6.3) sau graficul mediilor cu IC 95% sunt preferabile graficelor cu bare și "antene" de AS, care ascund distribuția datelor. Pentru date pereche, graficul cu linii individuale (figura 6.1 A) arată ce nu poate arăta niciun tabel: câți pacienți se îmbunătățesc, câți se înrăutățesc și cât de mult.

6.10 Probleme rezolvate

Problema rezolvată 6.1. Testul t pereche calculat de mână

Enunț. La 10 pacienți hipertensivi s-a măsurat TAS înainte și după 4 săptămâni de tratament cu un inhibitor al enzimei de conversie (tabelul 6.2). A scăzut semnificativ TAS? Rezolvați în formatul a-g, cu α = 0,05.

Rezolvare. (a) Ipoteza de cercetare: tratamentul de 4 săptămâni modifică TAS a pacienților hipertensivi. (b) H0: media diferențelor TAS înainte - TAS după este zero în populație, μD = 0; H1: μD ≠ 0 (bilateral). (c) Metoda: aceiași pacienți măsurați de două ori, deci date pereche, variabilă numerică; testul t pentru eșantioane pereche, cu condiția ca diferențele să fie aproximativ normale. (d) Calcul: diferențele sunt 12, 6, 12, 5, 8, 3, 15, 4, 9, 4; suma 78, media mD = 7,8 mmHg. Suma pătratelor abaterilor față de medie este 151,60 (tabelul 6.2), deci $s_D{}^2 = \frac{151{,}60}{9} = 16{,}84$ și sD = 4,10 mmHg. ESD = $\frac{4{,}10}{\sqrt{10}} = 1{,}30$ mmHg. $t = \frac{7{,}8}{1{,}30} = 6{,}01$, gl = 9. Testul Shapiro-Wilk pe diferențe: p = 0,347, normalitatea este acceptabilă. (e) Rezultatul: t critic (9; 0,05) = 2,262 < 6,01; p = 0,0002. IC 95% pentru scăderea medie: 7,8 ± 2,262 × 1,30 = 4,86-10,74 mmHg. dz = $\frac{7{,}8}{4{,}10} = 1{,}90$. Graficul potrivit: linii individuale înainte-după, ca în figura 6.1 A. (f) Concluzia statistică: p < 0,001, se respinge H0; diferența este extrem de semnificativă. (g) Concluzia de cercetare: după 4 săptămâni de tratament, TAS a scăzut în medie cu 7,8 mmHg (IC 95% 4,9-10,7 mmHg); toți cei 10 pacienți au avut o scădere, între 3 și 15 mmHg. Limita principală: lipsa unui grup de control; o parte din scădere poate reflecta regresia spre medie (pacienții au fost incluși cu valori mari) sau obișnuirea cu măsurarea.

Verificare cu programul: în Google Sheets, =T.TEST(B2:B11;C2:C11;2;1) dă 0,0002 (tipul 1 = pereche). Dacă s-ar folosi greșit tipul 2 (eșantioane independente), rezultatul ar fi p = 0,008; analiza ar ignora faptul că valorile provin de la aceiași pacienți.

Problema rezolvată 6.2. Testul t pentru eșantioane independente calculat de mână

Enunț. Glicemia a jeun (mg/dL) după 3 luni de tratament a fost măsurată la 8 pacienți tratați cu metformin și dietă și la 10 pacienți tratați doar cu dietă (date fictive). Grupul 1 (metformin + dietă): 112, 118, 105, 120, 110, 115, 108, 124. Grupul 2 (dietă): 125, 130, 118, 135, 128, 122, 140, 126, 132, 119. Diferă glicemia medie între grupuri?

Rezolvare. (a) Ipoteza de cercetare: glicemia a jeun la 3 luni diferă între pacienții tratați cu metformin și dietă și cei tratați doar cu dietă. (b) H0: μ1 = μ2; H1: μ1 ≠ μ2. (c) Două grupuri

independente, variabilă numerică: testul t pentru eșantioane independente; verificăm normalitatea în fiecare grup (Shapiro-Wilk p = 0,965 și p = 0,932) și egalitatea variațelor.

(d) Calcul. Grupul 1: suma 912, m1 = 114,0; abaterile față de medie sunt -2, 4, -9, 6, -4, 1, -6, 10, iar suma pătratelor lor este 4 + 16 + 81 + 36 + 16 + 1 + 36 + 100 = 290; $s_1{}^2 = \frac{290}{7} = 41{,}43$; s1 = 6,44. Grupul 2: suma 1275, m2 = 127,5; abaterile sunt -2,5; 2,5; -9,5; 7,5; 0,5; -5,5; 12,5; -1,5; 4,5; -8,5, cu suma pătratelor 440,5; $s_2{}^2 = \frac{440{,}5}{9} = 48{,}94$; s2 = 7,00. Raportul variațelor 48,94 / 41,43 = 1,18, iar testul Levene dă p = 0,887: variațele pot fi considerate egale.

Varianța comună: $s^2_{\text{comună}} = \frac{290+440{,}5}{8+10-2} = \frac{730{,}5}{16} = 45{,}66$; scomună = 6,76. ES = $\sqrt{\left[45{,}66 \times \left(\frac{1}{8} + \frac{1}{10}\right)\right]} = \sqrt{45{,}66 \times 0{,}225} = \sqrt{10{,}27} = 3{,}205$. $t = \frac{127{,}5-114{,}0}{3{,}205} = \frac{13{,}5}{3{,}205} = 4{,}21$, gl = 16.

(e) t critic (16; 0,05) = 2,120 < 4,21; p = 0,0007. IC 95% pentru diferență: 13,5 ± 2,120 × 3,205 = 13,5 ± 6,79, adică 6,71-20,29 mg/dL. d Cohen = 13,5 / 6,76 = 2,00. Varianta Welch, pentru comparație: ES = $\sqrt{\frac{41{,}43}{8} + \frac{48{,}94}{10}} = \sqrt{5{,}179 + 4{,}894} = \sqrt{10{,}07} = 3{,}174$; t = 4,25; gl = 15,6; p = 0,0006, practic același rezultat. (f) p < 0,001, se respinge H0; diferența este extrem de semnificativă. (g) La 3 luni, pacienții tratați cu metformin și dietă au glicemia a jeun în medie cu 13,5 mg/dL mai mică decât cei tratați doar cu dietă (IC 95% 6,7-20,3 mg/dL). Într-un studiu real ar trebui verificat dacă grupurile erau comparabile la includere (ideal, prin randomizare); altfel, diferența poate reflecta selecția pacienților.

Problema rezolvată 6.3. Tabelul ANOVA și testul Tukey de mână

Enunț. Scorul durerii (0-10) la 2 ore după administrarea a trei analgezice, la câte 5 pacienți (date fictive): A (paracetamol) 6, 5, 7, 6, 6; B (ibuprofen) 4, 5, 3, 4, 4; C (combinație) 3, 2, 3, 2, 5. Diferă scorul mediu al durerii între analgezice? Dacă da, care diferă între ele?

Rezolvare. (a) Ipoteza de cercetare: eficacitatea analgezică (scorul durerii la 2 ore) diferă între cele trei tratamente. (b) H0: μA = μB = μC; H1: cel puțin o medie diferă. (c) Trei grupuri independente, variabilă numerică: ANOVA cu un factor, urmată de testul Tukey dacă F este semnificativ. Condiții: independența prin design; normalitatea greu de evaluat la 5 valori pe grup (în practică, la grupuri atât de mici, s-ar raporta și testul Kruskal-Wallis); omogenitatea variațelor: $s^2A = \frac{2}{4} = 0{,}5$, $s^2B = \frac{2}{4} = 0{,}5$, $s^2C = \frac{6}{4} = 1{,}5$; testul Levene p = 0,557.

(d) Calcul. Mediile: $m_A = \frac{30}{5} = 6$, $m_B = \frac{20}{5} = 4$, $m_C = \frac{15}{5} = 3$; media generală $M = \frac{65}{15} = 4{,}33$. SS între = 5 × [(6 - 4,33)² + (4 - 4,33)² + (3 - 4,33)²] = 23,33, gl = 2, MS între = 11,67. SS în interior = 2 + 2 + 6 = 10,00 (echivalent: 4 × 0,5 + 4 × 0,5 + 4 × 1,5), gl = 12, MS în interior = 0,833. $F = \frac{11{,}67}{0{,}833} = 14{,}0$. Tabelul complet este tabelul 6.4.

(e) F critic (2; 12; 0,05) = 3,89 < 14,0; p = 0,0007; $\eta^2 = 0{,}70$. Testul Tukey: valoarea q pentru k = 3 grupuri și 12 grade de libertate reziduale, la α = 0,05, este 3,77 (tabelele intervalului studentizat). HSD = $3{,}77 \times \sqrt{\frac{0{,}833}{5}} = 3{,}77 \times 0{,}408 = 1{,}54$. Diferențele: A - B = 2 > 1,54, semnificativă (p Tukey = 0,012); A - C = 3 > 1,54, semnificativă (p = 0,0006); B - C = 1 < 1,54, nesemnificativă (p = 0,23). (f) Se respinge H0 global; diferențele sunt extrem de semnificative. (g) Ibuprofenul și combinația reduc durerea mai mult decât paracetamolul (cu 2, respectiv 3 puncte pe scala 0-10), iar diferența de 1 punct dintre combinație și ibuprofen nu a putut fi pusă în evidență cu 5 pacienți pe grup. Clinic, o diferență de 2 puncte pe o scală a durerii de 0-10 este considerată, de obicei, perceptibilă de pacient.

Problema rezolvată 6.4. Testul pereche din statisticile rezumative (mg2026)

Enunț. Între includere și 3 luni, diferențele individuale ale glicemiei a jeun (glicemie0 - glicemie3) la cei 160 de pacienți din mg2026 au media 6,08 mg/dL și abaterea standard 9,35 mg/dL. Testați dacă glicemia s-a modificat și calculați IC 95% și dz.

Rezolvare. H0: μD = 0; H1: μD ≠ 0. ESD $= \frac{9,35}{\sqrt{160}} = \frac{9,35}{12,649} = 0,739$ mg/dL. $t = \frac{6,08}{0,739} = 8,22$, gl = 159; t critic 1,975; p < 0,001. IC 95%: 6,08 ± 1,975 × 0,739 = 6,08 ± 1,46, adică 4,62-7,54 mg/dL. dz $= \frac{6,08}{9,35} = 0,65$ (efect mediu spre mare). Normalitatea diferențelor: Shapiro-Wilk p = 0,133. Concluzia: glicemia a scăzut extrem de semnificativ în primele 3 luni, în medie cu aproximativ 6 mg/dL (IC 95% 4,6-7,5). Comparați cu scăderea la 6 luni (10,19 mg/dL, dz = 0,88): aproximativ trei cincimi din scăderea totală s-au produs în primele trei luni. Observați că, pentru calcul, este nevoie de abaterea standard a diferențelor; abaterile standard ale glicemiei la fiecare moment (15,76 la includere) nu sunt suficiente.

Problema rezolvată 6.5. Testul independent din statisticile rezumative (mg2026)

Enunț. În mg2026, vârsta femeilor este 55,42 ± 8,95 ani (n = 86), iar a bărbaților 57,31 ± 8,50 ani (n = 74). Diferă vârsta medie între sexe? De ce ar conta acest lucru pentru alte comparații între sexe?

Rezolvare. H0: μF = μM; H1: μF ≠ μM. Condiții: Shapiro-Wilk p = 0,804 (F) și p = 0,311 (M); Levene p = 0,617. Varianța comună: $s^2_{\text{comună}} = \frac{85\times8,95^2+73\times8,50^2}{158} = \frac{85\times80,10+73\times72,25}{158} = \frac{6808,5+5274,3}{158} = 76,47$; scomună = 8,74 ani. ES $= 8,74 \times \sqrt{\frac{1}{86}+\frac{1}{74}} = 8,74 \times 0,1586 = 1,387$ ani. $t = \frac{57,31-55,42}{1,387} = \frac{1,89}{1,387} = 1,36$, gl = 158; t critic 1,975; p = 0,17 (varianta Welch: t = 1,37, gl = 156,4, p = 0,173). IC 95% pentru diferență: -0,84 la 4,62 ani. $d = \frac{1,89}{8,74} = 0,22$.

Concluzia: nu se respinge H0; vârsta medie nu diferă semnificativ între sexe. Intervalul de încredere admite însă o diferență de până la aproximativ 4,6 ani în favoarea bărbaților. De ce contează? Pentru că TAS crește cu vârsta (în mg2026, aproximativ 0,4 mmHg pe an, capitolul 9). O diferență de vârstă între sexe ar putea masca sau crea o diferență de TAS între sexe. Aici diferența de vârstă este mică și nesemnificativă, dar într-un studiu observațional comparabilitatea grupurilor pentru factorii de confuzie trebuie verificată și, la nevoie, controlată prin regresie multiplă sau ANCOVA.

Problema rezolvată 6.6. Tabelul ANOVA reconstruit din n, medii și AS (mg2026)

Enunț. Un articol raportează scăderea glicemiei la 6 luni pe trei brațe: A 3,0 ± 8,8 mg/dL($n = 53$), B 11,1 ± 11,2 mg/dL($n = 53$), C 16,3 ± 10,6 mg/dL($n = 54$). Reconstruiți tabelul ANOVA.

Rezolvare. N = 160, k = 3. Media generală este media ponderată: $M = \frac{53\times3,0+53\times11,1+54\times16,3}{160} = \frac{159,0+588,3+880,2}{160} = \frac{1627,5}{160} = 10,17$ mg/dL (cu mediile nerotunjite, 10,19).

SS între = 53 × (3,0 - 10,17)² + 53 × (11,1 - 10,17)² + 54 × (16,3 - 10,17)² = 53 × 51,41 + 53 × 0,86 + 54 × 37,58 = 2724,7 + 45,8 + 2029,2 = 4799,7; gl = 2; MS între = 2399,9.

SS în interior = 52 × 8,8² + 52 × 11,2² + 53 × 10,6² = 52 × 77,44 + 52 × 125,44 + 53 × 112,36 = 4026,9 + 6522,9 + 5955,1 = 16504,8; gl = 157; MS în interior = 105,1.

$F = \frac{2399,9}{105,1} = 22,8$; p < 0,001. $\eta^2 = \frac{4799,7}{4799,7+16504,8} = 0,225$.

Interpretare. Rezultatul reconstruit ($F = 22{,}8$, $\eta^2 = 0{,}225$) este foarte apropiat de cel calculat de program pe datele individuale ($F = 22{,}93$, $\eta^2 = 0{,}226$, tabelul 6.5); micile diferențe provin din rotunjirea mediilor și a abaterilor standard la o zecimală. Tehnica este utilă la lectura critică a articolelor: dacă un rezumat raportează medii, AS și n pe grupuri, puteți verifica ordinul de mărime al lui F și al valorii p raportate.

Problema rezolvată 6.7. TAS după statusul de fumător: citirea unui rezultat ANOVA

Enunț. În mg2026, TAS la includere pe categorii de fumat este: foști fumători $127{,}55 \pm 14{,}39$ mmHg($n = 33$), fumători $137{,}04 \pm 14{,}63$ mmHg($n = 53$), nefumători $128{,}50 \pm 11{,}76$ mmHg($n = 74$). Programul raportează: SS între = 2785,6, SS în interior = 27846,6, Levene p = 0,114, Shapiro-Wilk p între 0,23 și 0,86 pe grupuri; Tukey: fumători - foști fumători 9,49 mmHg (p = 0,0045), fumători - nefumători 8,54 mmHg (p = 0,0014), nefumători - foști fumători 0,95 mmHg (p = 0,94). Completați tabelul ANOVA și interpretați.

Rezolvare. gl între = 3 - 1 = 2; gl în interior = 160 - 3 = 157. MS între = 2785,6 / 2 = 1392,8; MS în interior = 27846,6 / 157 = 177,4. $F = \frac{1392{,}8}{177{,}4} = 7{,}85$; p = 0,0006 (F critic 3,05). $\eta^2 = \frac{2785{,}6}{2785{,}6+27846{,}6} = \frac{2785{,}6}{30632{,}2} = 0{,}091$, un efect mediu spre mare. Condițiile sunt îndeplinite (Levene p > 0,05; normalitate acceptabilă; raportul $\frac{\text{AS } 14{,}63}{11{,}76} = 1{,}24$).

Concluzia statistică: se respinge H0, TAS medie diferă foarte semnificativ între categoriile de fumat. Testul Tukey arată că fumătorii actuali au TAS mai mare cu aproximativ 9 mmHg decât foștii fumători și decât nefumătorii, iar foștii fumători nu diferă de nefumători. Concluzia de cercetare trebuie formulată prudent: studiul este, pentru această întrebare, o analiză transversală; fumătorii pot diferi și prin alți factori (vârstă, IMC, consum de alcool). În capitolul 9 veți vedea că, după ajustarea pentru IMC, vârstă și sex în regresia multiplă, diferența fumători - foști fumători rămâne semnificativă (6,74 mmHg, p = 0,012), dar mai mică. Rezultatul este compatibil cu efectul presor cunoscut al fumatului activ și cu reversibilitatea lui după renunțare, fără să o demonstreze.

6.11 Exerciții propuse

1. La 8 pacienți cu diabet zaharat de tip 2, HbA1c (%) a fost măsurată înainte și după 3 luni de educație terapeutică: înainte 7,8; 8,2; 7,5; 9,0; 8,4; 7,9; 8,8; 8,1; după 7,2; 7,9; 7,4; 8,1; 7,8; 7,6; 8,0; 7,7. Aplicați testul t pereche (α = 0,05): calculați diferențele, mD, sD, t, IC 95% și dz.
2. Un studiu compară TAS la două grupuri independente: grupul 1, n = 25, m = 142 mmHg, s = 12; grupul 2, n = 30, m = 135 mmHg, s = 14. Calculați testul t cu varianță comună, valoarea p, IC 95% și d. Comentați rezultatul în raport cu pragul de 0,05.
3. Două grupuri: grupul 1, n = 12, m = 4,8, s = 1,1; grupul 2, n = 40, m = 6,1, s = 3,2 (o concentrație serică, în unități arbitrare). Calculați testul t în varianta Student și în varianta Welch. Explicați de ce diferă concluziile.
4. Trei grupuri de câte 10 pacienți au TAS medie 120, 128 și 131 mmHg, cu abaterile standard 10, 12 și 11 mmHg. Construiți tabelul ANOVA, calculați F, eta² și decideți la α = 0,05 (F critic (2; 27) = 3,35).
5. Completați tabelul ANOVA: 4 grupuri, N = 40 de pacienți, SS între = 360, SS în interior = 1440. Calculați gl, MS, F, eta² și decideți la α = 0,05 (F critic (3; 36) = 2,87).
6. În studiul asupra tulburărilor interacțiunii intestin-creier [71], IMC a fost 28,15 ± 6,49 kg/m² la 52 de cazuri și 24,47 ± 4,60 kg/m² la 84 de martori. Folosind doar aceste valori,

calculați testul t Welch și d Cohen. Este rezultatul compatibil cu valoarea p raportată (p = 0,001)?

7. În studiul asupra COVID-19 la pacienții cu diabet [8], durata spitalizării a fost 11,6 ± 7,0 zile la 224 de pacienți cu diabet de tip 2 și 9,1 ± 5,8 zile la 62 de pacienți cu diabet de tip 1. Calculați, ca exercițiu, testul t Welch din aceste valori. Ce vă spune raportul dintre AS și medie despre forma distribuției și ce test ați prefera pe datele individuale?
8. În mg2026, diferențele individuale TAS0 - TAS6 au media 4,78 mmHg și abaterea standard 9,05 mmHg (n = 160). Calculați t, IC 95% pentru diferența medie și dz. Este scăderea relevantă clinic?
9. Pentru fiecare situație, numiți testul parametric potrivit și alternativa lui neparametrică: (a) frecvența cardiacă a pacienților cu BPOC înainte și după un tratament bronhodilatator; (b) înălțimea studenților pe sexe; (c) glicemia la 0, 3 și 6 luni la aceiași pacienți; (d) glicemia la 6 luni la pacienții asigurați și neasigurați; (e) indicele de permeabilitate bronșică la pacienții tratați cu trei medicamente diferite; (f) media IMC a unui grup față de valoarea 25 kg/m².
10. Pentru datele din exercițiul 4, calculați diferența minimă semnificativă Tukey (q(3; 27; 0,05) = 3,51) și verificați dacă vreo pereche de grupuri diferă. Este rezultatul consecvent cu ANOVA?
11. Un rezumat raportează pentru un marker măsurat la aceiași pacienți: $8061{,}9 \pm 3684{,}80$ ng/mL înainte și 15380 ± 5887 ng/mL după tratament, p < 0,001 [86]. Explicați de ce nu se poate recalcula testul t pereche din aceste cifre și ce informație suplimentară ar fi necesară.

Capcane frecvente

- Datele pereche analizate cu testul pentru eșantioane independente pierd putere; grupurile independente analizate ca pereche produc rezultate fără sens.
- Normalitatea pentru testul pereche se verifică pe diferențe, nu pe fiecare serie separat.
- Numărul de grade de libertate în testul pereche este n - 1, unde n este numărul de perechi, nu numărul total de măsurători.
- La grupuri inegale cu variațe inegale, testul Student poate da valori p greșite; se folosește testul Welch.
- Trei sau mai multe grupuri nu se compară prin teste t multiple necorectate; se face ANOVA și apoi un test post hoc.
- Un F semnificativ nu spune care grupuri diferă; testele post hoc nu se fac după un F nesemnificativ, în căutarea unei perechi semnificative.
- Măsurătorile repetate pe aceiași pacienți cer ANOVA cu măsurători repetate (cu verificarea sfericității), nu ANOVA cu un factor.
- O AS apropiată de medie la o variabilă pozitivă semnalează asimetrie; media ± AS și testul t pot fi nepotrivite.

De reținut

- Testul t pereche: $t = \frac{m_D}{\frac{s_D}{\sqrt{n}}}$, gl = n - 1; este testul t pentru o medie aplicat diferențelor; dz = $\frac{m_D}{s_D}$.
- Testul t independent Student: varianța comună $s^2_{\text{comună}} = \frac{(n_1-1)\,{s_1}^2+(n_2-1)\,{s_2}^2}{n_1+n_2-2}$, gl = n1 + n2 - 2; varianta Welch nu presupune variațe egale; testul Levene verifică egalitatea variațelor.
- ANOVA compară variația dintre grupuri cu variația din interiorul grupurilor: F = MS între / MS în interior, cu gl k - 1 și N - k; SS total = SS între + SS în interior.

- η^2 = SS între / SS total este mărimea efectului în ANOVA; d Cohen este mărimea efectului pentru două medii.
- Condițiile ANOVA: independență, normalitate în grupuri, omogenitatea varianțelor; alternative: ANOVA Welch cu Games-Howell, Kruskal-Wallis.
- Post hoc: Tukey (toate perechile, variațe egale), Bonferroni sau Holm (puține comparații), Games-Howell (variațe inegale), Dunnett (fiecare grup față de control).
- În mg2026: t pereche glicemie 0-6 luni t(159) = 11,14, p < 0,001; TAS pe sexe p = 0,81; ANOVA scăderea glicemiei pe brațe F(2; 157) = 22,93, p < 0,001, $\eta^2 = 0{,}226$; ANOVA cu măsurători repetate F = 64,97, p < 0,001.

Întrebări de autoevaluare

1. (CS) Glicemia a jeun a fost măsurată la includere și la 6 luni la aceiași 160 de pacienți. Testul parametric potrivit pentru a compara cele două momente este:
    a) testul t pentru eșantioane independente
    b) testul t pentru eșantioane pereche
    c) ANOVA cu un factor
    d) testul chi-pătrat
    e) testul z pentru o proporție
2. (CS) Într-un test t pereche cu 25 de pacienți, numărul gradelor de libertate este:
    a) 48
    b) 50
    c) 24
    d) 25
    e) 23
3. (CS) Într-o ANOVA cu un factor, 4 grupuri cu un total de 60 de pacienți, gradele de libertate ale numărătorului și numitorului statisticii F sunt:
    a) 4 și 60
    b) 3 și 56
    c) 3 și 59
    d) 4 și 56
    e) 59 și 3
4. (CS) Testul Levene are p = 0,003 la compararea a două grupuri independente de 15 și 60 de pacienți. Varianta potrivită a testului t este:
    a) testul t pereche
    b) testul t Student cu varianță comună
    c) testul t Welch
    d) testul z
    e) testul t pentru o singură medie
5. (CS) Într-o ANOVA, SS între grupuri = 300 și SS total = 1200. Eta² este:
    a) 0,20
    b) 0,25
    c) 0,33
    d) 4,00
    e) nu se poate calcula fără gradele de libertate

6. (CM) Despre testul t pentru eșantioane pereche, sunt corecte afirmațiile:
    a) lucrează pe diferențele individuale dintre cele două măsurători
    b) normalitatea se verifică pe diferențe
    c) are, în general, mai multă putere decât testul independent aplicat aceloraşi date pereche
    d) cere ca cele două serii să aibă variante egale
    e) alternativa neparametrică este testul Wilcoxon al rangurilor cu semn
7. (CM) Condițiile de aplicare ale ANOVA cu un factor sunt:
    a) independența observațiilor
    b) distribuția aproximativ normală în fiecare grup
    c) omogenitatea variantelor
    d) grupuri obligatoriu de dimensiuni egale
    e) cel puțin 30 de pacienți în fiecare grup
8. (CM) În mg2026, ANOVA pentru scăderea glicemiei pe brațe dă $F(2; 157) = 22{,}93$, $p < 0{,}001$, $\eta^2 = 0{,}226$, iar testul Tukey arată diferențe semnificative pentru toate cele trei perechi. Sunt corecte afirmațiile:
    a) se respinge ipoteza că cele trei medii sunt egale
    b) aproximativ 23% din variabilitatea scăderii glicemiei este explicată de braț
    c) ANOVA singură arată că brațul C diferă de brațul B
    d) testul Tukey menține riscul global de eroare de tip I la nivelul α
    e) rezultatul se datorează exclusiv numărului mare de pacienți
9. (CM) Sunt alegeri potrivite când variantele grupurilor diferă mult:
    a) testul t Welch pentru două grupuri
    b) ANOVA Welch pentru trei sau mai multe grupuri
    c) testul post hoc Games-Howell
    d) testul t Student cu varianță comună
    e) testul Tukey clasic, fără nicio verificare
10. (CM) Despre ANOVA cu măsurători repetate, sunt corecte afirmațiile:

a) se aplică atunci când aceiași pacienți sunt măsurați la trei sau mai multe momente

b) sfericitatea se verifică cu testul Mauchly

c) corecția Greenhouse-Geisser modifică gradele de libertate ale testului F

d) alternativa neparametrică este testul Kruskal-Wallis

e) pentru două momente de timp, rezultatul este echivalent cu cel al testului t pereche

Capitolul 7. Teste statistice neparametrice uzuale: Mann-Whitney U, Wilcoxon, Kruskal-Wallis, Friedman și testul chi-pătrat

Obiective

La finalul acestui capitol, studentul va putea:

- să recunoască situațiile în care condițiile testelor parametrice nu sunt îndeplinite (date ordinale, asimetrie marcată, eșantioane mici, valori extreme) și să aleagă alternativa neparametrică potrivită;
- să atribuie ranguri unei serii de valori, inclusiv atunci când există valori egale, și să verifice suma rangurilor;
- să calculeze de mână statistica U Mann-Whitney, statistica T Wilcoxon, statistica H Kruskal-Wallis și statistica Friedman pe eșantioane mici și să le compare cu valorile critice;
- să interpreteze corect ipotezele testelor pe ranguri (distribuții, mediane) și să raporteze o mărime a efectului (corelația biserială a rangurilor, epsilon pătrat, W Kendall);
- să construiască un tabel de contingență, să calculeze frecvențele așteptate, statistica chi-pătrat și gradele de libertate, și să verifice condițiile de aplicare;
- să aleagă între testul chi-pătrat, corecția Yates, testul exact Fisher, testul McNemar, testul pentru trend și testul de concordanță;
- să calculeze și să interpreteze coeficientul phi și V Cramér, înțelegând că valoarea p nu măsoară tăria asocierii;
- să citească critic secțiunea de metode și rezultatele unor articole publicate de colegii din disciplină care au folosit aceste teste.

7.1 Când nu mai putem folosi testele parametrice

În capitolul 6 am lucrat cu testul t și cu ANOVA. Ambele compară medii și se sprijină pe câteva presupuneri: variabila este numerică, distribuția în fiecare grup este aproximativ normală (sau eșantionul este suficient de mare încât media să se comporte normal), iar varianțele grupurilor nu diferă prea mult. În practica medicală, aceste presupuneri cad mai des decât ne-am dori. Pacientul nu răspunde la întrebarea "cât de mulțumit sunteți?" cu un număr măsurat în miligrame, ci alege o căsuță de la 1 la 5. Durata de spitalizare are o coadă lungă spre dreapta, pentru că un singur pacient complicat poate sta 40 de zile când ceilalți stau 5. Iar într-un studiu-pilot cu 12 pacienți nu avem cum să verificăm convingător normalitatea.

Testele neparametrice au fost construite exact pentru aceste situații. Ideea lor comună este simplă: în loc să lucrăm cu valorile propriu-zise, lucrăm cu poziția lor într-un clasament, adică cu rangurile. Un pacient care a stat 40 de zile în spital nu mai trage media în sus cu toată greutatea lui; el devine doar "ultimul din clasament". Pierdem o parte din informație, dar câștigăm robustețe.

7.1.1 Patru situații tipice

Prima situație este cea a **datelor ordinale**. Scala Likert de satisfacție, stadiul NYHA al insuficienței cardiace, gradul de fibroză hepatică F0-F4, scorul mMRC al dispneei sau nivelul de educație au categorii ordonate, dar distanța dintre categorii nu este neapărat egală. Trecerea de la "nemulțumit" la "neutru" nu este obligatoriu echivalentă cu trecerea de la "mulțumit" la "foarte mulțumit". Media unor astfel de coduri are un sens discutabil, iar testul t presupune tocmai că

diferențele dintre valori au sens aritmetic. Pentru date ordinale, testele pe ranguri sunt alegerea firească.

A doua situație este **asimetria marcată** a unei variabile numerice. În setul mg2026, trigliceridele au asimetria 0,95 și testul Shapiro-Wilk dă $p < 0,001$, iar indicele de masă corporală (IMC) are asimetria 1,42 și excesul 3,68. Când grupurile sunt mari, testul t rezistă destul de bine la asimetrie, datorită teoremei limită centrale. Când grupurile sunt mici sau asimetria este puternică, media devine un rezumat nereprezentativ, iar testul pe ranguri este mai sigur. O alternativă discutată în capitolul 3 este transformarea logaritmică, care pentru trigliceride aduce asimetria la -0,28 și SW $p = 0,200$.

A treia situație este **eșantionul mic**. Cu 6-8 pacienți pe grup, testul de normalitate are o putere foarte mică: un rezultat "$p > 0,05$" la Shapiro-Wilk nu dovedește normalitatea, doar arată că nu avem destule date ca să o infirmăm. În astfel de studii, mulți autori aleg din start un test neparametric, iar alegerea este rezonabilă dacă a fost prevăzută în protocol.

A patra situație este prezența **valorilor extreme** (engl. outliers) care nu sunt erori de măsurare. Un pacient cu trigliceride 442 mg/dL este real și nu are voie să fie șters doar pentru că strică media. Testele pe ranguri îl păstrează, dar îi limitează influența: indiferent dacă valoarea lui este 442 sau 4420, rangul rămâne același.

7.1.2 Ce înseamnă "neparametric" și ce nu înseamnă

Termenul "neparametric" vine de la faptul că aceste teste nu presupun o anumită formă a distribuției (de exemplu, normala cu parametrii μ și σ). Asta nu înseamnă că ele nu au nicio condiție. Toate presupun observații independente (sau perechi independente între ele), iar unele presupun că distribuțiile comparate au aceeași formă, diferind doar prin poziție, dacă vrem să interpretăm rezultatul ca diferență de mediane. Testul chi-pătrat, la rândul lui, cere frecvențe așteptate suficient de mari.

O confuzie frecventă la examen este ideea că testul neparametric "compară medianele" în orice condiții. Mai exact, testul Mann-Whitney verifică dacă valorile dintr-un grup tind să fie mai mari decât cele din celălalt grup. Dacă cele două distribuții au aceeași formă și aceeași împrăștiere, acest lucru este echivalent cu o diferență între mediane. Dacă formele diferă mult, testul poate ieși semnificativ chiar la mediane egale. În raportare spunem de aceea "distribuțiile diferă" sau "valorile au fost mai mari în grupul C", și dăm medianele și intervalele intercuartile ca descriere. Tabelul 7.1 pune față în față testele parametrice din capitolul 6 și echivalentele lor neparametrice.

Tabelul 7.1. Corespondența dintre testele parametrice și cele neparametrice.

Situația	Test parametric	Test neparametric	Ce se compară în varianta neparametrică
2 grupuri independente	testul t nepereche (Welch)	Mann-Whitney U (Wilcoxon rank-sum)	rangurile celor două grupuri
2 măsurători pe aceiași subiecți	testul t pereche	Wilcoxon (semne și ranguri); testul semnului	rangurile diferențelor, cu semn
≥ 3 grupuri independente	ANOVA unifactorială	Kruskal-Wallis, apoi Dunn	rangurile medii ale grupurilor
≥ 3 măsurători pe aceiași subiecți	ANOVA cu măsurători repetate	Friedman, apoi Wilcoxon cu corecție	rangurile în interiorul fiecărui subiect
asocierea a două variabile	corelația Pearson	corelația Spearman	rangurile ambelor variabile
două variabile calitative	(nu există variantă parametrică)	chi-pătrat, Fisher, McNemar	frecvențele observate față de cele așteptate

7.1.3 Eficiența relativă a testelor neparametrice

Prețul plătit pentru robustețe este o pierdere de putere atunci când datele chiar sunt normale. Această pierdere se măsoară prin eficiența relativă asimptotică: raportul dintre numărul de subiecți de care are nevoie testul parametric și numărul de care are nevoie testul neparametric pentru aceeași putere. Pentru Mann-Whitney comparat cu testul t, la date perfect normale, eficiența este $3/\pi$, adică aproximativ 0,955 (95,5%) [53,63]. Concret, ce face testul t cu 95 de pacienți face Mann-Whitney cu aproximativ 100. Pierderea este mică. Pentru distribuții cu cozi groase (multe valori extreme), eficiența testului pe ranguri poate depăși 1, adică testul neparametric devine mai puternic decât testul t, iar pentru nicio distribuție continuă nu coboară sub aproximativ 0,86. Testul Wilcoxon pentru date pereche are aceeași eficiență de 95,5% față de testul t pereche.

Testul semnului, care păstrează doar direcția diferenței (plus sau minus) și ignoră mărimea ei, este mult mai risipitor: eficiența lui la date normale este $2/\pi$, aproximativ 0,64. Vom vedea în secțiunea 7.4 un exemplu din mg2026 în care aceeași întrebare primește p = 0,043 la testul t pereche, p = 0,052 la Wilcoxon și p = 0,092 la testul semnului, exact în ordinea eficienței.

Concluzia practică este liniștitoare. Dacă ai un motiv real să alegi un test neparametric (date ordinale, asimetrie puternică la n mic, valori extreme), nu pierzi aproape nimic. Dacă datele sunt numerice și aproximativ normale, testul parametric rămâne prima alegere, pentru că estimează direct diferența de medii cu interval de încredere, ceea ce este mai ușor de interpretat clinic.

7.2 Rangurile: materia primă a testelor neparametrice

7.2.1 Atribuirea rangurilor, cu valori egale

Rangul unei valori este poziția ei în seria ordonată crescător: cea mai mică valoare primește rangul 1, următoarea rangul 2 și așa mai departe. Problema apare la valorile egale (engl. ties), foarte frecvente la scorurile ordinale. Regula este ca valorile egale să primească **media rangurilor** pe care le-ar fi ocupat dacă ar fi fost ușor diferite.

Să luăm opt scoruri de durere pe scala 0-10, notate de pacienți după o extracție dentară: 3, 5, 5, 2, 7, 5, 9, 3. Le ordonăm: 2, 3, 3, 5, 5, 5, 7, 9. Valoarea 2 ocupă poziția 1, deci rangul 1. Cele două valori 3 ocupă pozițiile 2 și 3, deci fiecare primește (2 + 3)/2 = 2,5. Cele trei valori 5 ocupă pozițiile 4, 5 și 6, deci fiecare primește (4 + 5 + 6)/3 = 5. Valoarea 7 primește rangul 7, iar valoarea 9 rangul 8.

Tabelul 7.2. Atribuirea rangurilor cu valori egale (scoruri de durere, n = 8).

Valoarea originală, în ordinea culegerii	**3**	**5**	**5**	**2**	**7**	**5**	**9**	**3**
Rangul	2,5	5	5	1	7	5	8	2,5

7.2.2 Verificarea: suma rangurilor

Înainte de orice calcul, verifică întotdeauna suma rangurilor. Pentru N observații, suma rangurilor 1, 2,..., N este:

$$\frac{N(N+1)}{2}$$

Pentru N = 8, suma trebuie să fie $\frac{8\times 9}{2} = 36$. Adunăm: 2,5 + 5 + 5 + 1 + 7 + 5 + 8 + 2,5 = 36. Corect. Regula mediei pentru valorile egale păstrează suma, de aceea verificarea funcționează și

când există egalități. La examen, o sumă greșită a rangurilor este primul semn că s-a sărit o valoare sau s-a calculat greșit un rang mediu, iar verificarea durează zece secunde.

Valorile egale au și un efect asupra variabilității rangurilor: o reduc puțin. De aceea, formulele testelor au o "corecție pentru egalități", pe care programele o aplică automat. Pe eșantioanele mici calculate de mână o vom arăta o dată, ca să știi de unde vin micile diferențe dintre rezultatul tău și cel din JASP.

7.3 Testul Mann-Whitney U

Testul a fost propus în forma cu statistica U de Mann și Whitney, în 1947 [92], pornind de la testul sumei rangurilor descris de Wilcoxon cu doi ani mai devreme [93]. De aceea, în unele programe apare sub numele "Wilcoxon rank-sum test"; este același test. Îl folosim pentru a compara două grupuri independente atunci când variabila este ordinală sau numerică fără distribuție normală.

7.3.1 Ipotezele

În cuvinte, ipoteza nulă spune că valorile din cele două populații provin din aceeași distribuție: dacă alegem la întâmplare un pacient din grupul 1 și unul din grupul 2, probabilitatea ca primul să aibă valoarea mai mare este 0,5. Ipoteza alternativă bilaterală spune că această probabilitate diferă de 0,5, adică valorile dintr-un grup tind să fie sistematic mai mari.

H0: P(X1 > X2) = 0,5 (distribuțiile sunt identice) H1: P(X1 > X2) ≠ 0,5

Dacă acceptăm presupunerea că distribuțiile au aceeași formă, ipotezele se pot scrie prin mediane: H0: Me1 = Me2, H1: Me1 ≠ Me2. Această formulare este cea folosită în temele disciplinei și este acceptabilă, cu condiția să știi presupunerea din spate.

7.3.2 Calculul lui U de mână

Pașii sunt aceiași pentru orice eșantion mic:

1. se pun toate valorile din cele două grupuri într-o singură serie și se atribuie ranguri, cu media rangurilor pentru valorile egale;
2. se calculează suma rangurilor din fiecare grup, R1 și R2, și se verifică $R_1 + R_2 = \frac{N(N+1)}{2}$;
3. se calculează $U_1 = R_1 - \frac{n_1(n_1+1)}{2}$ și $U_2 = R_2 - \frac{n_2(n_2+1)}{2}$; verificare: U1 + U2 = n1 x n2;
4. statistica testului este U = minimul dintre U1 și U2;
5. se compară U cu valoarea critică din tabel pentru n1, n2 și α; **H0 se respinge dacă U este mai mic sau egal cu valoarea critică** (atenție, invers decât la t și chi-pătrat).

Un exemplu: într-un studiu-pilot, 5 pacienți operați au primit o aplicație de telefon cu exerciții de respirație și reamintiri pentru analgezice, iar 6 pacienți au primit îngrijirea obișnuită. Scorul durerii la 24 de ore (0-10) a fost: aplicație 2, 3, 3, 5, 6; control 4, 6, 7, 7, 8, 9. Figura 7.1 arată valorile și rangurile lor în eșantionul comun.

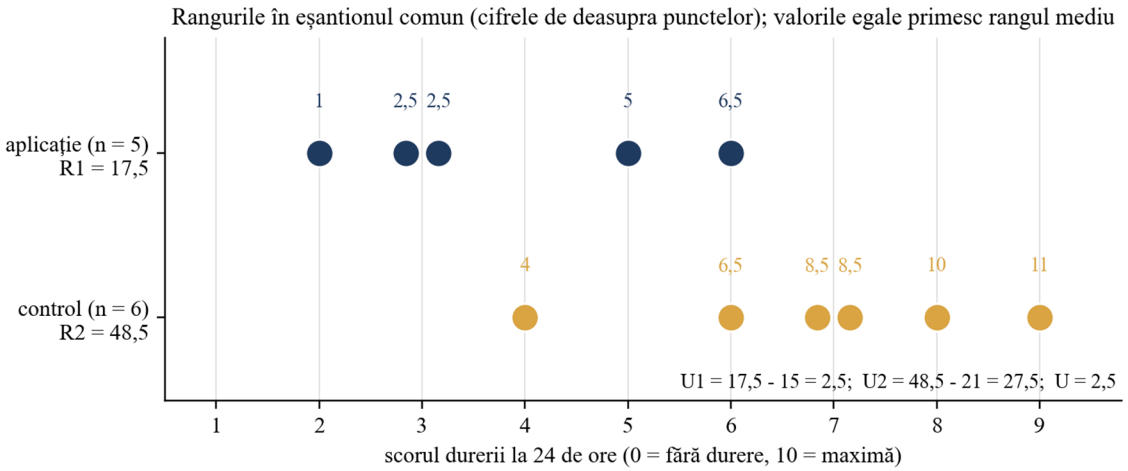


Figura 7.1. Calculul statisticii U Mann-Whitney pe un exemplu mic: scorul durerii la 24 de ore la 5 pacienți cu aplicație și 6 pacienți de control. Cifrele de deasupra punctelor sunt rangurile în eșantionul comun de 11 valori; valorile egale (3 și 3, 6 și 6, 7 și 7) primesc rangul mediu. Date fictive, construite pentru calcul manual.

Seria comună ordonată este 2, 3, 3, 4, 5, 6, 6, 7, 7, 8, 9 (N = 11). Rangurile: 2 → 1; cele două valori 3 → 2,5; 4 → 4; 5 → 5; cele două valori 6 (una din fiecare grup) → 6,5; cele două valori 7 → 8,5; 8 → 10; 9 → 11.

$$R_1(\text{ aplicație }) = 1 + 2{,}5 + 2{,}5 + 5 + 6{,}5 = 17{,}5 R_2(\text{ control }) = 4 + 6{,}5 + 8{,}5 + 8{,}5 + 10 + 11 = 48{,}5$$

Verificare: $17{,}5 + 48{,}5 = 66 = \frac{11 \times 12}{2}$. Corect.

$$U_1 = 17{,}5 - \frac{5 \times 6}{2} = 17{,}5 - 15 = 2{,}5 U_2 = 48{,}5 - \frac{6 \times 7}{2} = 48{,}5 - 21 = 27{,}5$$

Verificare: 2,5 + 27,5 = 30 = 5 x 6. Statistica testului este U = 2,5.

Ce înseamnă U, intuitiv? Dacă formăm toate cele 5 x 6 = 30 de perechi (un pacient cu aplicație, un pacient de control), U1 numără de câte ori pacientul cu aplicație are scorul mai mare (egalitățile contează o jumătate). Doar 2,5 din 30 de perechi "favorizează" grupul cu aplicație ca având mai multă durere. Cu alte cuvinte, în 27,5 din 30 de comparații pacientul cu aplicație are mai puțină durere.

Valoarea critică pentru n1 = 5, n2 = 6, test bilateral, α = 0,05, este 3 (tabelul 7.15, la începutul secțiunii 7.11). Deoarece U = 2,5 ≤ 3, respingem H0. Valoarea p exactă dată de program este 0,030. Concluzia statistică: diferența dintre distribuțiile scorului durerii este semnificativă la pragul de 0,05. Concluzia de cercetare: pacienții care au folosit aplicația au raportat durere mai mică la 24 de ore (mediana 3 față de 7). Fiind un studiu-pilot cu 11 pacienți, rezultatul trebuie confirmat pe un eșantion mai mare.

7.3.3 Aproximarea normală

Pentru grupuri mai mari (în practică, ambele grupuri peste 10-20 de subiecți), tabelele nu mai sunt necesare. Sub H0, U are aproximativ o distribuție normală cu media și abaterea standard:

$$\mu_U = \frac{n_1 \times n_2}{2} \sigma_U = \sqrt{\left[\frac{n_1 \times n_2 \times (N+1)}{12}\right]}$$

iar statistica

$$z = \frac{U - \mu_U}{\sigma_U}$$

se compară cu 1,96 pentru α = 0,05 bilateral. Dacă există multe valori egale, σU se corectează în jos, iar programele fac asta automat. Pe exemplul de mai sus, μU = 15, σU = √(30 x 12/12) = 5,48 (5,44 cu corecția pentru egalități), deci $z = \frac{2{,}5-15}{5{,}44} = -2{,}30$, cu p bilateral aproximativ 0,02. La grupuri atât de mici aproximarea este doar orientativă; valoarea exactă (0,030) este cea corectă.

7.3.4 Mărimea efectului: corelația biserială a rangurilor

Valoarea p ne spune cât de surprinzătoare ar fi diferența dacă H0 ar fi adevărată, dar nu cât de mare este. Pentru Mann-Whitney, cea mai simplă măsură a efectului este corelația biserială a rangurilor (engl. rank-biserial correlation):

$$r_{\text{rb}} = 1 - \frac{2U}{n_1 \times n_2}$$

unde U este minimul dintre U1 și U2. Valoarea ei merge de la 0 (nicio diferență) la 1 (toate valorile dintr-un grup sunt mai mari decât toate valorile din celălalt). Pe exemplul cu aplicația, $r_{\text{rb}} = 1 - \frac{2\times 2{,}5}{30} = 0{,}83$, un efect foarte mare. O altă variantă, folosită de unii autori, este $r = \frac{z}{\sqrt{N}}$. Orientativ, se consideră 0,1 efect mic, 0,3 mediu și 0,5 mare [76], dar pragurile sunt convenții, nu legi.

7.3.5 Aplicație pe setul mg2026: satisfacția în brațul C față de brațul A

La 6 luni, pacienții din studiul mg2026 și-au notat satisfacția față de programul primit pe o scală de la 1 (deloc mulțumit) la 5 (foarte mulțumit). Frecvențele pe brațe sunt în tabelul 7.3, iar distribuția procentuală este în figura 7.2.

Tabelul 7.3. Satisfacția la 6 luni pe brațe (număr de pacienți), setul mg2026.

Brațul	1	2	3	4	5	Total	Mediana (IQR)
A: consiliere	1	13	28	9	2	53	3 (2-3)
B: + aplicație	3	3	14	24	9	53	4 (3-4)
C: + aplicație și teleconsultație	0	3	12	22	17	54	4 (3-5)
Total	4	19	54	55	28	160	4 (3-4)

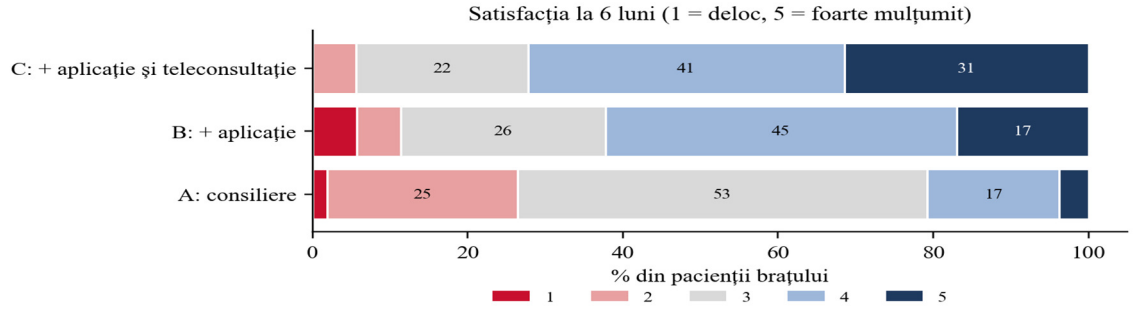


Figura 7.2. Distribuția satisfacției la 6 luni (1 = deloc mulțumit, 5 = foarte mulțumit), în procente din pacienții fiecărui braț. În brațul C predomină notele 4 și 5, în brațul A nota 3. Date sintetice (setul mg2026).

Întrebarea de cercetare: diferă satisfacția pacienților din brațul C de cea a pacienților din brațul A? Variabila este ordinală, deci folosim Mann-Whitney. Nu are rost să testăm normalitatea unei variabile cu cinci valori posibile.

Rezultatul din program: medianele sunt 4 (IQR 3-5) în brațul C și 3 (IQR 2-3) în brațul A, U = 2265,5, p < 0,001. Să vedem de unde vine cifra. Sumele rangurilor în eșantionul comun de 107 pacienți sunt RC = 3750,5 și RA = 2027,5. Atunci $UC = 3750{,}5 - \frac{54\times55}{2} = 3750{,}5 - 1485 = 2265{,}5$, iar UA = 54 x 53 - 2265,5 = 596,5. Programele raportează uneori U-ul unui grup, alteori pe cel al celuilalt; ambele duc la aceeași valoare p. Aproximarea normală: $\mu_U = \frac{54\times53}{2} = 1431$, σU = 153,5 (cu corecția pentru numeroasele egalități), deci $z = \frac{2265{,}5-1431}{153{,}5} = 5{,}44$, mult peste 1,96.

Mărimea efectului: $r_{\text{rb}} = 1 - \frac{2\times596{,}5}{2862} = 0{,}58$, un efect mare. Interpretarea directă: dacă alegem la întâmplare un pacient din brațul C și unul din brațul A, probabilitatea ca primul să fie mai mulțumit (cu egalitățile împărțite pe din două) este (1 + 0,58)/2 = 0,79.

Concluzia statistică: respingem H0; satisfacția diferă semnificativ între brațele C și A (p < 0,001). Concluzia clinică: pacienții care au primit aplicație și teleconsultație lunară au fost mai mulțumiți decât cei care au primit doar consiliere, iar diferența este semnificativă statistic și, în plus, mare. Atenție însă: satisfacția nu este același lucru cu eficacitatea; ea poate reflecta și atenția suplimentară primită în brațul C.

> **Studiu de caz: aplicația mobilă pentru astm, trei teste neparametrice într-un singur articol**
>
> Munteanu și colaboratorii, între care și dr. Frandeș din disciplina noastră, au evaluat într-un studiu-pilot controlat o aplicație de telefon menită să îmbunătățească aderența la tratament și autogestionarea bolii la 93 de pacienți cu astm, urmăriți un an [94]. Designul combină grupuri independente (tratament standard față de tratament plus aplicație) cu măsurători repetate pe aceiași pacienți. Autorii au ales consecvent teste neparametrice: Mann-Whitney U pentru compararea grupurilor (p < 0,001), testul Friedman pentru măsurătorile repetate, cu corecție Bonferroni pentru comparațiile două câte două (p < 0,001), și testul chi-pătrat Pearson pentru un tabel 2 x 2 în care proporțiile comparate au fost 10,30% și 46,30%, cu $\chi^2(1) = 13{,}707$, p < 0,001. Lectura acestui rezumat este un exercițiu bun de potrivire a testului cu întrebarea: comparație între grupuri independente, comparație în timp pe aceiași pacienți, comparație de proporții. Ce ar fi adăugat valoare: mărimea efectului pentru Mann-Whitney (de exemplu corelația biserială a rangurilor) și, pentru tabelul 2 x 2, riscul relativ cu interval de încredere, care arată clinicianului cât de mare este diferența, pe lângă faptul că există. Fiind un studiu-pilot, rezultatele generează ipoteze pentru un studiu randomizat mai mare.

7.4 Testul Wilcoxon pentru date pereche și testul semnului

Testul Wilcoxon al semnelor și rangurilor (engl. Wilcoxon signed-rank test) [93] este alternativa neparametrică a testului t pereche. Îl folosim când avem două măsurători pe aceiași subiecți (înainte și după tratament, ochiul drept și ochiul stâng, două metode aplicate aceleiași probe) și diferențele nu au o distribuție normală sau variabila este ordinală.

7.4.1 Logica testului

Pentru fiecare subiect calculăm diferența d = înainte - după. Dacă tratamentul nu are efect, diferențele pozitive și negative ar trebui să fie la fel de frecvente și la fel de mari. Testul ordonează valorile absolute ale diferențelor, le atribuie ranguri și apoi readaugă semnul. Sub H0, suma rangurilor pozitive (W+) ar trebui să fie apropiată de suma rangurilor negative (W-). Diferențele egale cu zero se elimină (varianta clasică), iar n devine numărul de diferențe nenule.

H0: distribuția diferențelor este simetrică în jurul lui 0 (mediana diferențelor = 0) H1: mediana diferențelor ≠ 0

7.4.2 Calculul de mână

Zece pacienți supraponderali au urmat un program de trei luni de consiliere nutrițională. Greutatea (kg) înainte și după este dată în tabelul 7.4.

Tabelul 7.4. Calculul testului Wilcoxon pe 10 pacienți (date fictive).

Pacient	Înainte	După	d	abs(d)	Rangul lui abs(d)	Rang cu semn
1	92,0	89,5	2,5	2,5	6	+6
2	85,5	85,5	0,0	0	eliminat	-
3	101,2	97,0	4,2	4,2	8	+8
4	78,4	79,6	-1,2	1,2	2	-2
5	88,0	85,8	2,2	2,2	4,5	+4,5
6	95,3	93,1	2,2	2,2	4,5	+4,5
7	83,1	83,9	-0,8	0,8	1	-1
8	110,4	104,2	6,2	6,2	9	+9
9	90,2	87,0	3,2	3,2	7	+7
10	87,6	86,1	1,5	1,5	3	+3

Pacientul 2 are diferența zero și se elimină, deci n = 9. Valorile absolute ordonate: 0,8 (rang 1), 1,2 (rang 2), 1,5 (rang 3), 2,2 și 2,2 (rangurile 4 și 5, deci 4,5 fiecare), 2,5 (rang 6), 3,2 (rang 7), 4,2 (rang 8), 6,2 (rang 9).

$$W^+ = 6 + 8 + 4{,}5 + 4{,}5 + 9 + 7 + 3 = 42\, W^- = 2 + 1 = 3$$

Verificare: $W^+ + W^- = 45 = \frac{9\times10}{2}$. Corect. Statistica testului este T = minimul dintre W+ și W-, adică T = 3.

Valoarea critică pentru n = 9, bilateral, α = 0,05, este 5 (tabelul 7.16). Ca la Mann-Whitney, **H0 se respinge dacă T este mai mic sau egal cu valoarea critică**. Cum 3 ≤ 5, respingem H0; p exact = 0,020. Pentru n mai mare se folosește aproximarea normală:

$$\mu_T = \frac{n(n+1)}{4} = 22{,}5 \sigma_T = \sqrt{\left[\frac{n(n+1)\,(2n+1)}{24}\right]} = \sqrt{\frac{9\times10\times19}{24}} = 8{,}44 z = \frac{3-22{,}5}{8{,}44} = -2{,}31,\ p \approx 0{,}021$$

Concluzia: greutatea a scăzut semnificativ după program (mediana diferențelor 2,2 kg; 7 din 9 pacienți cu diferență nenulă au slăbit).

7.4.3 Testul semnului

Testul semnului este cel mai simplu test pentru date pereche. El numără doar câte diferențe sunt pozitive și câte negative, ignorând mărimea lor. Sub H0, fiecare diferență nenulă are probabilitatea 0,5 să fie pozitivă, deci numărul de diferențe pozitive urmează o distribuție binomială cu n și p = 0,5.

În exemplul de mai sus avem 7 diferențe pozitive din 9. Probabilitatea bilaterală de a obține 7 sau mai multe (sau, simetric, 2 sau mai puține) la o binomială cu n = 9 și p = 0,5 este 0,18. Testul semnului nu găsește o diferență semnificativă acolo unde Wilcoxon găsește p = 0,020. Explicația: cei doi pacienți care s-au îngrășat au luat foarte puțin în greutate (1,2 și 0,8 kg), iar cei care au slăbit au slăbit mult mai mult. Wilcoxon folosește această informație, testul semnului o aruncă. Testul semnului rămâne util atunci când avem doar direcția schimbării ("mai bine", "la fel", "mai rău"), de exemplu în evaluarea globală a pacientului.

7.4.4 Aplicație pe mg2026: greutatea în brațul A, un rezultat la limită

În brațul A (doar consiliere, n = 53), mediana diferenței de greutate dintre includere și 6 luni este 1,1 kg. 32 de pacienți au slăbit, 19 au luat în greutate și 2 au rămas la fel. Testul Wilcoxon dă W- = 456 și W+ = 870 pentru n = 51 de diferențe nenule, cu p = 0,052.

Aici se vede clar de ce nu are sens să tratăm pragul de 0,05 ca pe un zid. Cu p = 0,052, nu respingem H0, dar asta nu înseamnă că programul din brațul A nu are niciun efect: înseamnă că datele nu sunt suficient de convingătoare la pragul ales. Vom formula concluzia prudent: "la 6 luni, în brațul cu consiliere standard, scăderea în greutate (mediana 1,1 kg) nu a atins pragul de semnificație statistică (Wilcoxon, p = 0,052)". Absența dovezii nu este dovada absenței [79].

Același exemplu ilustrează eficiența relativă discutată în secțiunea 7.1.3. Diferențele de greutate din brațul A au, de fapt, o distribuție compatibilă cu normala (Shapiro-Wilk pe diferențe p = 0,696), deci testul potrivit era testul t pereche. Acesta dă diferența medie 0,71 kg (AS 2,48), t = 2,07, gl 52, p = 0,043. Testul semnului (32 de plusuri din 51) dă p = 0,092. Trei teste, aceeași întrebare, trei valori p ordonate exact după cantitatea de informație folosită: media și varianța (t), rangurile (Wilcoxon), doar semnele (testul semnului). Lecția nu este "alege testul care iese semnificativ", ci "alege testul potrivit înainte de a vedea rezultatul"; aici, testul t pereche era justificat de distribuția diferențelor. În celelalte două brațe nu există nicio ambiguitate: Wilcoxon dă p < 0,001 atât în brațul B (mediana 2,3 kg), cât și în brațul C (mediana 3,55 kg).

Un exemplu din cercetarea colegilor arată cum arată datele pereche raportate prin mediane. Gherbon și colaboratorii, cu dr. Frandeș, au urmărit retrospectiv 200 de pacienți români cu diabet zaharat de tip 2 tratați cu inhibitori SGLT-2, comparând valorile de la începutul tratamentului cu cele de după un an [59]. Datele sunt pereche (aceiași pacienți, două momente), iar rezultatele sunt raportate ca mediană și interval intercuartil: glicemia a scăzut de la 180,00 mg% (IQR 154,50-207,00) la 130,00 mg% (IQR 117,50-150,00), p < 0,001; HbA1c de la 8,40% (IQR 7,98-9,15%) la 7,30% (IQR 6,90-7,95%), p < 0,001; greutatea de la 90,50 kg (82,00-106,50) la 89,00 kg (77,50-100,00), p = 0,018; tensiunea sistolică de la 140,00 mmHg (130,00-160,00) la 120,00 mmHg (115,50-130,00), p < 0,001. Circumferința abdominală apare în schimb ca medie ± AS, de la 107,05 (± 16,39) la 102,50 (± 15,11), p = 0,042. Rezumatul nu numește testul, dar alegerea medianelor sugerează un test pereche neparametric (Wilcoxon) pentru variabilele asimetrice și testul t pereche pentru circumferința abdominală. Două observații de lectură critică: rezumatul numește aceste mediane "medii", o confuzie de terminologie pe care un student atent o sesizează imediat, iar pentru datele pereche cea mai informativă cifră ar fi fost mediana diferențelor individuale (cu IQR sau interval de încredere), alături de medianele celor două momente. Diferența dintre mediane (de exemplu 90,50 - 89,00 = 1,50 kg) nu este obligatoriu egală cu mediana diferențelor.

7.5 Testul Kruskal-Wallis

Când avem trei sau mai multe grupuri independente și variabila este ordinală sau numerică fără distribuție normală, folosim testul Kruskal-Wallis [95], alternativa neparametrică a ANOVA unifactoriale. Se poate privi ca o generalizare a testului Mann-Whitney la k grupuri.

7.5.1 Ipotezele și statistica H

H0: cele k populații au aceeași distribuție (în cazul formelor similare: Me1 = Me2 =... = Mek) H1: cel puțin o populație diferă de celelalte (cel puțin o mediană diferă)

Toate cele N valori se ordonează împreună și primesc ranguri. Pentru fiecare grup i se calculează suma rangurilor Ri. Dacă H0 este adevărată, rangul mediu din fiecare grup ar trebui să fie apropiat de rangul mediu general, (N + 1)/2. Statistica testului măsoară cât de mult se abat rangurile medii ale grupurilor de această valoare:

$$H = \left[\frac{12}{N(N+1)}\right] \times \sum\left(\frac{{R_i}^2}{n_i}\right) - 3(N+1)$$

Pentru grupuri nu foarte mici (de regulă, cel puțin 5 subiecți în fiecare grup), H urmează aproximativ o distribuție chi-pătrat cu k - 1 grade de libertate. H0 se respinge dacă H depășește valoarea critică: 5,99 pentru 2 grade de libertate, 7,81 pentru 3, 9,49 pentru 4 (α = 0,05). Când există valori egale, H se împarte la factorul de corecție

$$C = 1 - \frac{\sum(t^3 - t)}{N^3 - N}$$

unde t este numărul de valori dintr-un grup de egalități. Corecția mărește puțin H.

7.5.2 Calculul de mână

Cincisprezece pacienți operați au primit una din trei scheme de analgezie: schema 1 (paracetamol), schema 2 (paracetamol + antiinflamator), schema 3 (analgezie multimodală). Scorul durerii la 24 de ore (0-10) a fost:

- schema 1: 6, 7, 5, 8, 7
- schema 2: 4, 5, 3, 5, 6
- schema 3: 2, 3, 4, 1, 3

Ordonăm cele 15 valori: 1, 2, 3, 3, 3, 4, 4, 5, 5, 5, 6, 6, 7, 7, 8. Rangurile: 1 → 1; 2 → 2; cele trei valori 3 ocupă pozițiile 3, 4, 5 → rangul 4; cele două valori 4 ocupă pozițiile 6, 7 → 6,5; cele trei valori 5 ocupă pozițiile 8, 9, 10 → 9; cele două valori 6 → 11,5; cele două valori 7 → 13,5; 8 → 15.

Tabelul 7.5. Rangurile pentru testul Kruskal-Wallis (date fictive).

Schema 1	Rang	Schema 2	Rang	Schema 3	Rang
6	11,5	4	6,5	2	2
7	13,5	5	9	3	4
5	9	3	4	4	6,5
8	15	5	9	1	1
7	13,5	6	11,5	3	4
R1	62,5	R2	40	R3	17,5

Verificare: $62{,}5 + 40 + 17{,}5 = 120 = \frac{15\times16}{2}$. Corect.

$$H = \left[\frac{12}{15 \times 16}\right] \times \left(\frac{62{,}5^2}{5} + \frac{40^2}{5} + \frac{17{,}5^2}{5}\right) - 3 \times 16 = 0{,}05 \times (781{,}25 + 320 + 61{,}25) - 48$$
$$= 0{,}05 \times 1162{,}5 - 48 = 58{,}125 - 48 = 10{,}125$$

Corecția pentru egalități: avem două grupuri de câte trei valori egale (valorile 3 și 5) și trei grupuri de câte două (valorile 4, 6 și 7). $\sum(t^3 - t) = 2 \times 24 + 3 \times 6 = 66$, iar $C = 1 - \frac{66}{3375-15} = 0{,}980$. H corectat $= \frac{10{,}125}{0{,}980} = 10{,}33$.

Cu 2 grade de libertate, valoarea critică este 5,99. Cum 10,33 > 5,99, respingem H0 (p = 0,006). Medianele sunt 7, 5 și 3. Concluzie: scorul durerii diferă între cele trei scheme. Testul nu spune însă care scheme diferă între ele; pentru asta avem nevoie de comparații post hoc.

7.5.3 Comparațiile post hoc: testul Dunn

După un Kruskal-Wallis semnificativ, procedura cea mai folosită este testul Dunn. Pentru fiecare pereche de grupuri, el compară rangurile medii (calculate în ierarhia comună a tuturor celor N subiecți) printr-o statistică z:

z = (rang mediu i - rang mediu j)/√{[N(N + 1)/12] x (1/ni + 1/nj)}

Valorile p obținute se corectează pentru numărul de comparații. Cea mai simplă corecție este Bonferroni: fiecare p se înmulțește cu numărul de comparații (3 pentru trei grupuri), sau, echivalent, pragul 0,05 se împarte la 3. Corecția Holm, aplicată în trepte, este mai puțin conservatoare. O greșeală frecventă este să faci trei teste Mann-Whitney separate, fiecare la pragul 0,05, fără nicio corecție: probabilitatea de a obține cel puțin un rezultat fals pozitiv crește atunci de la 5% la aproape 14%.

7.5.4 Aplicație pe mg2026: scăderea glicemiei și satisfacția, pe trei brațe

În capitolul 6 am comparat scăderea glicemiei a jeun la 6 luni între cele trei brațe cu ANOVA: F(2;157) = 22,93, p < 0,001. Testul Kruskal-Wallis pe aceleași date dă H = 36,63, gl 2, p < 0,001. Calculul urmează exact pașii de mai sus, doar că pe 160 de pacienți. Sumele rangurilor sunt 2696 (A), 4554,5 (B) și 5629,5 (C), adică ranguri medii de 50,9, 85,9 și 104,3, față de rangul mediu general (160 + 1)/2 = 80,5.

$$H = \left[\frac{12}{160\times161}\right] \times \left(\frac{2696^2}{53} + \frac{4554{,}5^2}{53} + \frac{5629{,}5^2}{54}\right) - 3 \times 161 = 0{,}000466 \times (137\,139{,}9 + 391\,386{,}2 + 586\,875{,}4) - 483 = 36{,}60$$

, iar cu corecția pentru egalități 36,63.

Medianele scăderii glicemiei sunt 4 mg/dL (IQR -3 la 9) în brațul A, 14 mg/dL (3 la 20) în brațul B și 16 mg/dL (9 la 23) în brațul C. Testul Dunn cu corecție Bonferroni arată diferențe semnificative între A și B (p = 0,0003) și între A și C (p < 0,001), dar nu între B și C (p ajustat = 0,12; neajustat 0,041). Este interesant de comparat cu testul Tukey din capitolul 6, care găsise și diferența C-B semnificativă (p = 0,024). Nu este o contradicție: testele pe ranguri sunt puțin mai conservatoare când datele sunt aproximativ normale, iar corecția Bonferroni este mai strictă decât Tukey. Pentru această variabilă numerică, cu distribuție rezonabilă în fiecare braț și Levene p = 0,163, analiza principală rămâne ANOVA cu Tukey; Kruskal-Wallis servește ca analiză de sensibilitate care confirmă direcția rezultatului.

Satisfacția, fiind ordinală, se analizează direct cu Kruskal-Wallis: H = 32,16, gl 2, p < 0,001. Rangurile medii sunt 53,5 (A), 86,7 (B) și 100,9 (C). Comparațiile Dunn cu Bonferroni: A față de B p = 0,0003, A față de C p < 0,001, B față de C p = 0,29. Așadar, ambele brațe cu aplicație au produs pacienți mai mulțumiți decât consilierea simplă, iar adăugarea teleconsultației nu a schimbat semnificativ satisfacția față de aplicația singură. Ca verificare, testul Mann-Whitney direct între B și C dă p = 0,079.

7.5.5 Mărimea efectului pentru Kruskal-Wallis

Analogul lui eta pătrat din ANOVA este epsilon pătrat:

$$\varepsilon^2 = \frac{H}{N-1}$$

Pentru scăderea glicemiei, $\varepsilon^2 = \frac{36{,}63}{159} = 0{,}23$; pentru satisfacție, $\varepsilon^2 = \frac{32{,}16}{159} = 0{,}20$. Ambele sunt efecte mari, comparabile cu $\eta^2 = 0{,}226$ obținut prin ANOVA pentru glicemie. Raportarea acestei mărimi este o practică bună și apare din ce în ce mai des în articole, ca în studiul HIV prezentat mai jos.

> **Studiu de caz: Kruskal-Wallis în trei articole ale colegilor**
>
> Gherbon și colaboratorii, cu dr. Frandeș, au analizat transversal 453 de adulți cu diabet de tip 2, la care distresul sever legat de diabet a fost identificat la 19 participanți (4,2%) [69]. Pentru a compara scorurile între trei categorii ordonate au folosit testul Kruskal-Wallis, cu p = 0,003, apoi o regresie logistică multivariabilă în care vârsta a avut aOR = 1,163 pe an (IC 95% 1,055-1,282; p = 0,002). Logica este clară: scorurile de distres sunt ordinale și asimetrice (puțini pacienți cu valori mari), deci un test pe ranguri este potrivit. Hasan și colaboratorii, cu dr. Tudor, au studiat 90 de pacienți la o lună după chirurgie toracică și au comparat anxietatea

și intensitatea durerii între trei grupuri definite de stilul de coping, cu Kruskal-Wallis: p = 0,028, p = 0,022, p = 0,042 și p = 0,007 pentru patru variabile [96]. Scorurile chestionarelor validate sunt variabile ordinale, iar testul este bine ales; întrebarea pe care o pune un cititor atent este câte comparații s-au făcut în total și dacă s-a ținut cont de ele, pentru că trei din cele patru valori p sunt între 0,02 și 0,05. Marin și colaboratorii, cu dr. R. Moleriu și dr. Muntean, au studiat 462 de adulți cu HIV aflați sub tratament antiretroviral [68]. Răspunsurile discordante imuno-virologice au apărut la 30,7% dintre pacienți (14,5% și 16,2% în cele două tipuri de discordanță), iar comparațiile între cele trei grupuri de răspuns au folosit Kruskal-Wallis cu mărimea efectului: H = 153,62, p < 0,001, ε^2 = 0,33 și H = 115,10, p < 0,001, ε^2 = 0,25. Acesta este modelul de raportare complet: statistica, valoarea p și mărimea efectului. Un ε^2 de 0,33 înseamnă că aproximativ o treime din variabilitatea rangurilor este explicată de apartenența la grup.

7.6 Testul Friedman

Testul Friedman [97] este alternativa neparametrică a ANOVA cu măsurători repetate. Îl folosim când aceiași subiecți sunt evaluați în trei sau mai multe momente (sau condiții, sau cu trei sau mai multe metode), iar variabila este ordinală sau numerică fără distribuție normală.

7.6.1 Logica testului

Diferența față de Kruskal-Wallis este că rangurile se atribuie **în interiorul fiecărui subiect**, nu în eșantionul comun. Pentru fiecare pacient, cele k măsurători primesc rangurile 1, 2,..., k. Astfel, diferențele dintre pacienți (unul are glicemii în jur de 90, altul în jur de 140) dispar, și rămâne doar ordinea măsurătorilor la același pacient. Dacă nu există niciun efect al timpului, fiecare moment ar trebui să primească, în medie, rangul (k + 1)/2.

H0: distribuția valorilor este aceeași în toate cele k momente H1: cel puțin un moment diferă de celelalte

Cu n subiecți și k momente, iar Rj suma rangurilor momentului j:

$$\chi^2 F = \left[\frac{12}{n \times k \times (k+1)}\right] \times \sum {R_j}^2 - 3n(k+1)$$

Statistica urmează aproximativ o distribuție chi-pătrat cu k - 1 grade de libertate.

7.6.2 Calculul de mână

Șase pacienți cu lombalgie și-au notat durerea (0-10) la începutul unui program de kinetoterapie, la 2 săptămâni și la 6 săptămâni.

Tabelul 7.6. Rangurile în interiorul fiecărui pacient pentru testul Friedman (date fictive).

Pacient	Început	2 săpt.	6 săpt.	Ranguri (început; 2 săpt.; 6 săpt.)
1	7	5	3	3; 2; 1
2	8	6	4	3; 2; 1
3	6	6	5	2,5; 2,5; 1
4	9	7	4	3; 2; 1
5	5	4	4	3; 1,5; 1,5
6	7	4	2	3; 2; 1
Suma rangurilor Rj				17,5; 12; 6,5

Verificare: fiecare rând are suma rangurilor 1 + 2 + 3 = 6, deci totalul este 6 x 6 = 36 = 17,5 + 12 + 6,5. Corect.

$$\chi^2F = \left[\frac{12}{6 \times 3 \times 4}\right] \times (17{,}5^2 + 12^2 + 6{,}5^2) - 3 \times 6 \times 4$$
$$= \left(\frac{12}{72}\right) \times (306{,}25 + 144 + 42{,}25) - 72 = 0{,}1667 \times 492{,}5 - 72 = 82{,}08 - 72$$
$$= 10{,}08$$

Cu 2 grade de libertate, valoarea critică este 5,99. Respingem H0. Programele aplică și aici o corecție pentru cele două egalități (pacienții 3 și 5) și raportează 11,00, cu p = 0,004. Durerea a scăzut semnificativ în cursul programului.

7.6.3 Aplicație pe mg2026: glicemia la 0, 3 și 6 luni

Glicemia a jeun a fost măsurată la includere, la 3 luni și la 6 luni la toți cei 160 de pacienți. Medianele sunt 104,5, 99,0 și 92,5 mg/dL (mediile 105,4, 99,4 și 95,2). Figura 7.3 prezintă mediana și intervalul intercuartil pe brațe.

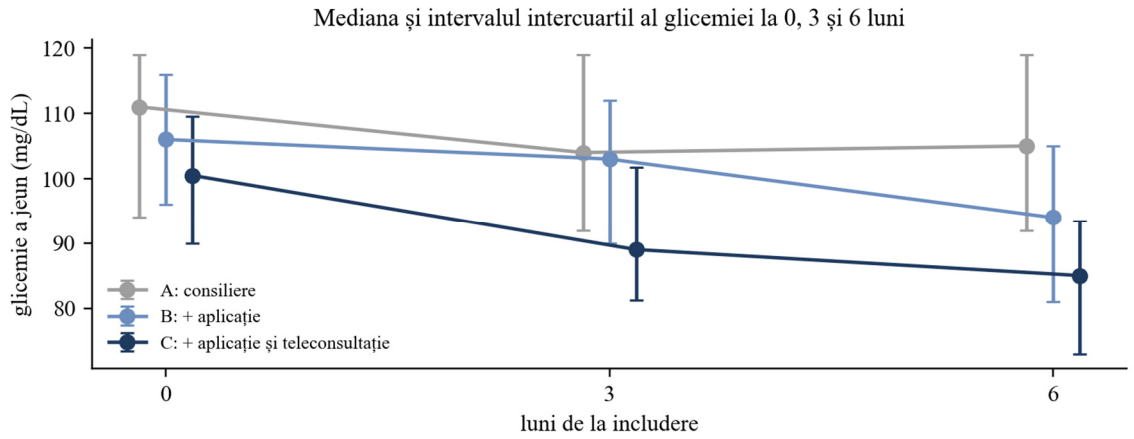


Figura 7.3. Mediana și intervalul intercuartil al glicemiei a jeun la 0, 3 și 6 luni, pe brațe. În brațul C scăderea este cea mai abruptă și continuă până la 6 luni; în brațul A mediana scade la 3 luni și apoi rămâne pe loc. Date sintetice (setul mg2026).

Testul Friedman pe toți pacienții: sumele rangurilor sunt 406 (0 luni), 295 (3 luni) și 259 (6 luni), adică ranguri medii de 2,54, 1,84 și 1,62. Fără corecție, $\chi^2F = \left[\frac{12}{160 \times 3 \times 4}\right] \times (406^2 + 295^2 + 259^2) - 3 \times 160 \times 4 = 73{,}39$; cu corecția pentru egalități, 74,67, gl 2, p < 0,001. Pentru comparație, ANOVA cu măsurători repetate dă F = 64,97, p < 0,001. Ambele teste spun același lucru.

Mărimea efectului pentru Friedman este coeficientul de concordanță W Kendall:

$$W = \frac{\chi^2F}{n(k-1)} = \frac{74{,}67}{160 \times 2} = 0{,}23$$

W merge de la 0 (nicio tendință comună) la 1 (toți pacienții au exact aceeași ordine a măsurătorilor). O valoare de 0,23 indică o tendință comună moderată: mulți pacienți au scăderi, dar nu toți în aceeași ordine.

Comparațiile post hoc se fac cu testul Wilcoxon pentru fiecare pereche de momente, cu corecție Bonferroni (3 comparații, deci fiecare p se înmulțește cu 3 sau pragul devine 0,05/3 = 0,0167):

Tabelul 7.7. Comparațiile post hoc după testul Friedman (glicemie, n = 160).

Comparația	Mediana diferențelor (mg/dL)	p Wilcoxon	p ajustat Bonferroni
0 luni față de 3 luni	7,0	< 0,001	< 0,001
3 luni față de 6 luni	4,0	0,0002	0,0007
0 luni față de 6 luni	9,5	< 0,001	< 0,001

Toate trei comparațiile rămân semnificative după corecție. Analiza pe brațe este mai instructivă: în brațul C, Friedman dă $\chi^2F = 64{,}63$, $p < 0{,}001$ (medianele 100,5, 89,0 și 85,0 mg/dL), în brațul B $\chi^2F = 21{,}07$, $p < 0{,}001$, iar în brațul A doar $\chi^2F = 6{,}04$, $p = 0{,}049$ (medianele 111, 104 și 105 mg/dL). În brațul A rezultatul global trece la limită pragul de 0,05, dar niciuna dintre comparațiile Wilcoxon nu rămâne semnificativă după corecția Bonferroni (cel mai mic p ajustat este 0,075, pentru 0 față de 6 luni). Situația este des întâlnită în practică: testul global "vede" o tendință slabă, dar puterea nu ajunge pentru a localiza diferența după corecție.

Testul Friedman apare și în studiile de comparare a metodelor de măsurare. Al Barri și colaboratorii, cu dr. Tudor, au comparat trei dispozitive de keratometrie și biometrie oculară la pacienți candidați la operația de cataractă: 170 de ochi de la 102 pacienți, vârsta medie 69,12 ± 10,26 ani [81]. Fiecare ochi a fost măsurat cu toate cele trei aparate, deci avem trei măsurători legate, exact situația pentru testul Friedman, urmat de comparații Wilcoxon două câte două. Pentru keratometria K1, mediile au fost 43,45 ± 1,64 D (Argos) și 43,41 ± 1,70 D (Aladdin), cu $p < 0{,}001$ global și $p = 0{,}019$ pentru Argos față de Aladdin. Adâncimea camerei anterioare a fost 3,19 ± 0,42 față de 3,13 ± 0,41mm, $p < 0{,}001$, iar grosimea corneei a fost practic identică între Aladdin și Sirius (540,27 ± 33,44 față de 540,47 ± 33,78 µm, $p = 0{,}169$). Componentele vectoriale ale astigmatismului J0 ($p = 0{,}277$) și J45 ($p = 0{,}084$) nu au diferit. Articolul arată foarte bine diferența dintre semnificația statistică și cea clinică: o diferență de 0,04 D în K1 sau de 0,06 mm în adâncimea camerei anterioare poate fi "semnificativă", pentru că măsurătorile pereche elimină variabilitatea dintre ochi, dar este mult sub pragul care ar schimba calculul cristalinului artificial. A doua observație privește independența: 170 de ochi provin de la 102 pacienți, deci unii pacienți contribuie cu doi ochi, iar ochii aceluiași pacient nu sunt complet independenți. Revenim la această temă în capitolul 8, la unitatea de analiză.

7.7 Testul chi-pătrat pentru tabele de contingență

Până acum am lucrat cu variabile numerice sau ordinale. Când ambele variabile sunt calitative (nominale sau ordinale cu puține categorii) și ne interesează dacă sunt asociate, datele se rezumă într-un tabel de contingență, iar instrumentul clasic este testul chi-pătrat (χ^2) propus de Karl Pearson în 1900 [98]. Este, probabil, testul cel mai des întâlnit în articolele medicale, pentru că multe rezultate clinice sunt de tip "da/nu": vindecat sau nu, internat sau nu, cu complicație sau fără.

7.7.1 Tabelul de contingență și ipotezele

Un tabel de contingență r x c are r rânduri (categoriile unei variabile) și c coloane (categoriile celeilalte). În fiecare celulă se trece numărul de subiecți care au simultan categoria rândului și pe cea a coloanei. Totalurile pe rânduri și pe coloane se numesc totaluri marginale.

Exemplul de lucru vine din analiza transversală a pacienților mg2026 la includere: există o asociere între obezitate (IMC ≥ 30 kg/m²) și diabetul confirmat prin testul oral de toleranță la glucoză (TTGO)?

Tabelul 7.8. Obezitatea și diabetul confirmat prin TTGO la includere, setul mg2026 (frecvențe observate).

	Diabet TTGO da	**Diabet TTGO nu**	**Total**
Obezi (IMC ≥ 30)	18	36	54
Neobezi (IMC < 30)	11	95	106
Total	29	131	160

Prevalența diabetului este 18/54 = 33,3% la obezi și 11/106 = 10,4% la neobezi. Ipotezele:

H0: obezitatea și diabetul sunt independente (prevalența diabetului este aceeași la obezi și la neobezi: p1 = p2) H1: există o asociere între obezitate și diabet (p1 ≠ p2)

7.7.2 Frecvențele așteptate

Ideea testului este să comparăm ce am observat cu ce ne-am aștepta să observăm dacă H0 ar fi adevărată. Dacă obezitatea nu are legătură cu diabetul, atunci proporția generală de diabetici, 29/160 = 18,1%, ar trebui să se regăsească la fel la obezi și la neobezi. Printre cei 54 de obezi ne-am aștepta atunci la $\frac{54\times29}{160} = 9{,}79$ diabetici. Regula generală pentru orice celulă:

E = (totalul rândului x totalul coloanei)/totalul general

Tabelul 7.9. Frecvențele așteptate sub ipoteza de independență.

	Diabet TTGO da	**Diabet TTGO nu**	**Total**
Obezi	$\frac{54 \times 29}{160} = 9{,}79$	$\frac{54 \times 131}{160} = 44{,}21$	54
Neobezi	$\frac{106 \times 29}{160} = 19{,}21$	$\frac{106 \times 131}{160} = 86{,}79$	106
Total	29	131	160

Verificare: frecvențele așteptate au aceleași totaluri marginale ca cele observate (9,79 + 44,21 = 54; 9,79 + 19,21 = 29). Dacă totalurile nu se potrivesc, ai greșit un calcul. Frecvențele așteptate pot fi zecimale; nu se rotunjesc la numere întregi.

7.7.3 Statistica, gradele de libertate, decizia

Statistica chi-pătrat adună, pentru toate celulele, abaterea pătratică relativă dintre observat (O) și așteptat (E):

$$\chi^2 = \frac{\sum(O-E)^2}{E}$$

Tabelul 7.10. Contribuțiile celulelor la χ^2 (obezitate x diabet).

Celula	**O**	**E**	**O - E**	$\frac{(O-E)^2}{E}$
obez, diabet	18	9,79	8,21	6,891
obez, fără diabet	36	44,21	-8,21	1,525
neobez, diabet	11	19,21	-8,21	3,510
neobez, fără diabet	95	86,79	8,21	0,777
Total	160	160	0	$\chi^2 = 12{,}70$

Observați că într-un tabel 2 x 2 toate abaterile O - E au aceeași valoare absolută (aici 8,21). Pentru tabelele 2 x 2 există și o formulă scurtă, utilă la examen:

$$\chi^2 = \frac{N(\mathrm{ad}-\mathrm{bc})^2}{(a+b)\,(c+d)\,(a+c)\,(b+d)}$$

= 160 x (18 x 95 - 36 x 11)²/(54 x 106 x 29 x 131) = 160 x 1314²/21 745 116 = 12,70

Numărul gradelor de libertate pentru un tabel r x c este:

$$\text{gl} = (r - 1)\,(c - 1)$$

Pentru 2 x 2, gl = 1. Raționamentul: odată fixate totalurile marginale, o singură celulă poate fi aleasă liber, celelalte rezultă prin scădere. Valoarea critică pentru gl = 1 și α = 0,05 este 3,84 (pentru gl = 2 este 5,99; pentru gl = 3, 7,81; pentru gl = 4, 9,49). Cum 12,70 > 3,84, respingem H0; programul dă p = 0,0004 (raportăm p < 0,001).

Concluzia statistică: există o asociere semnificativă între obezitate și diabetul confirmat prin TTGO. Concluzia clinică: în această populație cu risc cardiometabolic, diabetul a fost de aproximativ trei ori mai frecvent la obezi (33,3% față de 10,4%; raportul prevalențelor 3,21, IC 95% 1,64-6,31; OR 4,32, IC 95% 1,86-10,03). Fiind un studiu transversal, nu putem spune că obezitatea a "produs" diabetul; spunem doar că cele două apar împreună mai des decât ar fi de așteptat prin întâmplare. Măsurile de risc și designurile de studiu sunt detaliate în capitolul 10.

Chi-pătratul este întotdeauna un test "bilateral" prin construcție: el detectează abateri în orice direcție. Direcția asocierii o citim din tabel (procentele pe rânduri), nu din valoarea χ^2.

7.7.4 Un tabel 3 x 2 pe mg2026: brațul și scăderea în greutate de cel puțin 5%

Obiectivul secundar al studiului mg2026 a fost proporția de pacienți care au slăbit cel puțin 5% din greutatea inițială în 6 luni. Figura 7.4 prezintă proporțiile pe brațe.

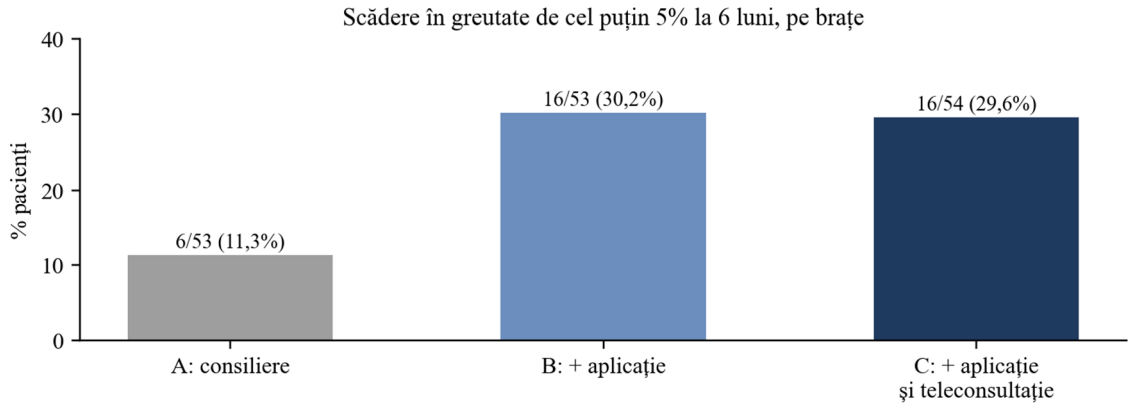


Figura 7.4. Proporția pacienților cu scădere în greutate de cel puțin 5% la 6 luni, pe brațe: 6/53 (11,3%) în brațul A, 16/53 (30,2%) în brațul B și 16/54 (29,6%) în brațul C. Testul chi-pătrat pe tabelul 3 x 2: χ^2 = 6,77, gl 2, p = 0,034. Date sintetice (setul mg2026).

Tabelul 7.11. Brațul de studiu și scăderea în greutate ≥ 5%: frecvențe observate, așteptate (în paranteză) și contribuții la χ^2.

Brațul	Scădere ≥ 5%	Fără scădere ≥ 5%	Total	% cu scădere
A	6 (12,59; contribuție 3,447)	47 (40,41; 1,074)	53	11,3
B	16 (12,59; 0,925)	37 (40,41; 0,288)	53	30,2
C	16 (12,82; 0,786)	38 (41,18; 0,245)	54	29,6
Total	38	122	160	23,8

Frecvențele așteptate se calculează la fel: pentru brațul A și coloana "scădere ≥ 5%", $E = \frac{53 \times 38}{160} = 12{,}59$. Suma contribuțiilor este χ^2 = 3,447 + 1,074 + 0,925 + 0,288 + 0,786 + 0,245 = 6,77. Gradele de libertate: (3 - 1)(2 - 1) = 2. Valoarea critică 5,99 este depășită, p = 0,034. Cea mai mică frecvență așteptată este 12,6, deci condițiile de aplicare sunt îndeplinite.

Tabelul contribuțiilor spune și **unde** este asocierea: aproape jumătate din χ^2 (3,447 din 6,77) vine dintr-o singură celulă, brațul A cu scădere ≥ 5%, unde am observat 6 pacienți în loc de 12,6 așteptați. Brațele B și C sunt foarte apropiate între ele.

Un χ^2 semnificativ într-un tabel mai mare decât 2 x 2 este, ca la ANOVA, un rezultat global. Dacă vrem să comparăm brațele două câte două, facem tabele 2 x 2 și corectăm pentru comparații multiple. Fără corecție: A față de B $\chi^2 = 5{,}74$, $p = 0{,}017$; A față de C $\chi^2 = 5{,}49$, $p = 0{,}019$; B față de C $\chi^2 = 0{,}00$, $p = 0{,}95$. Cu Bonferroni pentru trei comparații, pragul devine 0,0167, iar A față de B și A față de C ajung chiar la limită. Pentru întrebarea principală a studiului, care fusese stabilită în protocol ca fiind comparația C față de A, nu avem nevoie de corecție, iar rezultatul se exprimă cel mai util prin riscul relativ: 29,6% față de 11,3%, RR 2,62 (IC 95% 1,11-6,17), diferența absolută 18,3 puncte procentuale (IC 95% 3,4-33,2), NNT aproximativ 6.

7.7.5 Condițiile de aplicare

Testul chi-pătrat este un test aproximativ: distribuția statisticii se apropie de distribuția teoretică χ^2 doar când frecvențele așteptate sunt suficient de mari. Regulile folosite în practică [42,53]:

- observațiile sunt independente: fiecare subiect apare într-o singură celulă, o singură dată;
- se lucrează cu frecvențe absolute (numere de subiecți), niciodată cu procente sau medii;
- pentru tabelele 2 x 2, toate frecvențele așteptate trebuie să fie cel puțin 5;
- pentru tabelele mai mari, cel mult 20% dintre celule pot avea frecvența așteptată sub 5 (regula de 80%) și nicio celulă nu poate avea frecvența așteptată sub 1.

Atenție: condiția privește frecvențele **așteptate**, nu pe cele observate. O celulă cu 0 observați poate fi perfect acceptabilă dacă frecvența așteptată acolo este 6. Când condiția nu este îndeplinită, avem trei soluții: testul exact Fisher (pentru 2 x 2 și, în multe programe, pentru tabele mai mari), comasarea unor categorii cu sens clinic (de exemplu "fost fumător" cu "fumător", dacă întrebarea o permite) sau creșterea eșantionului.

Mărimea inegală a grupurilor nu este, în sine, o problemă pentru testul chi-pătrat; contează doar frecvențele așteptate. Un exemplu din cercetarea colegilor: Marin și colaboratorii, cu dr. L. Moleriu, au comparat retrospectiv răspunsul virologic la 462 de pacienți cu HIV-1 subtip F tratați cu darunavir potențat fie cu ritonavir (384 de pacienți), fie cu cobicistat (78 de pacienți) [99]. Rezumatul raportează un procent de 84,6% în grupul cu cobicistat, comparat cu 76,8%. Grupurile sunt foarte inegale (384 și 78), dar cu zeci de pacienți în fiecare celulă frecvențele așteptate sunt mari, iar compararea a două proporții prin chi-pătrat (sau Fisher, pentru celulele cu puține evenimente) este perfect validă. Problema reală a unui astfel de design este alta: într-un studiu observațional, medicul a ales schema după caracteristicile pacientului, deci diferența poate reflecta tipul de pacienți tratați cu fiecare schemă (confuzie prin indicație), nu doar efectul medicamentului.

7.7.6 Corecția Yates

Pentru tabelele 2 x 2, Frank Yates a propus o corecție de continuitate: din fiecare abatere absolută se scade 0,5 înainte de ridicarea la pătrat.

$$\chi^2\text{Yates} = \frac{\sum(|O-E|-0{,}5)^2}{E}$$

Corecția compensează faptul că frecvențele observate sunt numere întregi, iar distribuția χ^2 este continuă. Pe exemplul obezitate x diabet, χ^2Yates = 11,20 ($p = 0{,}0008$), față de 12,70 fără corecție. Corecția face testul mai conservator. Mulți statisticieni consideră astăzi că ea corectează prea mult și preferă, pentru tabele mici, direct testul Fisher; JASP o oferă ca opțiune ("continuity correction"), iar Epi Info o afișează automat alături de varianta necorectată. Regula practică pentru examen: la un tabel 2 x 2 cu frecvențe așteptate între 5 și 10, raportează varianta cu corecție Yates sau testul Fisher; când toate frecvențele așteptate sunt mari, diferența dintre variante nu schimbă concluzia.

7.7.7 Testul exact Fisher

Testul exact Fisher nu folosește nicio aproximare. Considerând totalurile marginale fixe, el calculează probabilitatea exactă a tabelului observat și a tuturor tabelelor "cel puțin la fel de extreme", folosind distribuția hipergeometrică. Pentru un tabel cu celulele a, b, c, d și N subiecți:

$$P = \frac{(a+b)!\ (c+d)!\ (a+c)!\ (b+d)!}{N!\ a!\ b!\ c!\ d!}$$

Exemplu: într-un audit al infecțiilor postoperatorii, 8 pacienți au primit o schemă nouă de profilaxie antibiotică (1 infecție) și 10 pacienți schema veche (6 infecții).

Tabelul 7.12. Infecția postoperatorie după schema de profilaxie (date fictive).

	Infecție	**Fără infecție**	**Total**
Schema nouă	1	7	8
Schema veche	6	4	10
Total	7	11	18

Frecvențele așteptate în coloana "infecție" sunt $\frac{8\times7}{18} = 3{,}11$ și $\frac{10\times7}{18} = 3{,}89$, ambele sub 5. Testul chi-pătrat nu este valid aici, deși programul îl calculează fără să protesteze: χ^2 = 4,22, p = 0,040. Testul Fisher calculează probabilitatea fiecărui tabel posibil cu aceleași totaluri, după numărul a de infecții în schema nouă (a poate fi de la 0 la 7): P(a = 0) = 0,0038; P(a = 1) = 0,0528 (tabelul nostru); P(a = 2) = 0,2217; P(a = 3) = 0,3695; P(a = 4) = 0,2640; P(a = 5) = 0,0792; P(a = 6) = 0,0088; P(a = 7) = 0,0003. Valoarea p bilaterală adună probabilitățile tabelelor care sunt cel mult la fel de probabile ca tabelul observat: 0,0038 + 0,0528 + 0,0088 + 0,0003 = 0,066.

Concluzia se schimbă: cu testul potrivit (Fisher), p = 0,066, nu respingem H0. Chi-pătratul folosit greșit ar fi dat un rezultat "semnificativ" (0,040), iar corecția Yates (p = 0,117) merge în aceeași direcție ca Fisher. Cu doar 18 pacienți, nu putem afirma că schema nouă reduce infecțiile, deși diferența observată (12,5% față de 60%) merită un studiu mai mare. Calculul manual al lui Fisher este laborios; la examen se cere de obicei să recunoști când trebuie folosit și să interpretezi valoarea p dată de program.

> **Studiu de caz: tabele mici în urgență, la pacienții cu accident vascular și COVID de lungă durată**
>
> Popa și colaboratorii, cu dr. Tudor, au analizat retrospectiv 132 de pacienți care au ajuns în Unitatea de Primiri Urgențe cu alertă de accident vascular cerebral, dintre care 26 aveau COVID de lungă durată și 106 nu [100]. Grupul de 26 este mic, iar unele simptome sunt rare, așa că mai multe tabele 2 x 2 au frecvențe așteptate mici. Autorii au folosit testul exact Fisher (în rezumat apare formularea "Fisher's Exact Test chi-squared", p = 0,027, care amestecă denumirile a două teste diferite). Cefaleea a apărut la 19,2% față de 5,7% (OR = 3,97, p = 0,040), iar asimetria facială la 42,3% față de 19,8% (OR = 2,97, p = 0,022). Să facem un calcul aproximativ din procente: 19,2% din 26 înseamnă aproximativ 5 pacienți, iar 5,7% din 106 aproximativ 6 pacienți, deci în total circa 11 pacienți cu cefalee. Frecvența așteptată de pacienți cu cefalee în grupul cu COVID de lungă durată ar fi atunci 26 x 11/132, adică aproximativ 2,2, mult sub 5: testul Fisher era exact alegerea corectă. Trombectomia mecanică a fost mai frecventă în grupul cu COVID de lungă durată (30,8% față de 16,0%), iar o altă comparație a dat 26,9% față de 14,2%, p = 0,143, fără semnificație statistică. Cu doar 26 de pacienți într-un grup, o diferență de aproape 13-15 puncte procentuale poate rămâne nesemnificativă din lipsă de putere; "nesemnificativ" nu înseamnă aici "fără diferență". Ce ar fi completat raportarea: intervalele de încredere ale OR, care ar arăta cât de imprecise sunt estimările obținute pe grupuri mici.

7.8 Variante ale testului chi-pătrat

7.8.1 Testul McNemar pentru proporții pereche

Testul chi-pătrat clasic presupune grupuri independente. Când aceeași variabilă dihotomică este măsurată de două ori pe aceiași subiecți (înainte și după, sau cu două teste diagnostice aplicate aceluiași pacient), datele sunt pereche și se folosește testul McNemar. Este echivalentul, pentru variabile binare, al testului t pereche.

Tabelul se construiește diferit: rândurile sunt starea la primul moment, coloanele starea la al doilea moment, iar fiecare pacient apare o singură dată. Pacienții care nu și-au schimbat starea (celulele de pe diagonală) nu aduc informație despre schimbare. Testul folosește doar celulele discordante, b și c:

χ^2McNemar = (b - c)²/(b + c), cu 1 grad de libertate; cu corecție de continuitate: (|b - c| - 1)²/(b + c)

Aplicație pe mg2026: câți pacienți aveau glicemia a jeun ≥ 100 mg/dL (glicemie modificată a jeun sau mai mult) la includere și câți la 6 luni?

Tabelul 7.13. Glicemia a jeun ≥ 100 mg/dL la includere și la 6 luni, aceiași 160 de pacienți (setul mg2026).

	La 6 luni ≥ 100	**La 6 luni < 100**	**Total**
La includere ≥ 100	61	36 (b)	97
La includere < 100	1 (c)	62	63
Total	62	98	160

Proporția cu glicemie ≥ 100 a scăzut de la 97/160 = 60,6% la 62/160 = 38,8%. Celulele discordante: 36 de pacienți au trecut sub 100, un singur pacient a trecut peste 100. $\chi^2 = \frac{(36-1)^2}{36+1} = \frac{1225}{37} = 33{,}11$; cu corecție, (35 - 1)²/37 = 31,24; p < 0,001. A fost greșit să aplicăm aici testul chi-pătrat obișnuit pe un tabel "momentul x starea" cu 320 de observații: același pacient ar apărea de două ori, iar independența ar fi încălcată. În cercetarea colegilor, Iovanescu și colaboratorii, cu dr. Frandeș, au folosit testul McNemar într-un studiu de screening al neoplaziei colorectale în care mai multe investigații (testul imunochimic fecal, sigmoidoscopia, colonoscopia) au fost evaluate la aceleași persoane (McNemar, P < 0,001); 252 de persoane au făcut ulterior colonoscopia, iar la 75 dintre ele (29,76%) s-a găsit neoplazie avansată [101]. Când două teste diagnostice sunt aplicate acelorași persoane, compararea proporțiilor lor de rezultate pozitive cere un test pentru date pereche.

Un exemplu mai puțin spectaculos: tensiunea sistolică ≥ 140 mmHg la includere și la 6 luni. Au trecut sub 140 un număr de 16 pacienți și au trecut peste 140 un număr de 8. χ^2McNemar = (16 - 8)²/24 = 2,67, p = 0,10 (cu corecție 2,04, p = 0,15). Deși proporția celor cu TAS ≥ 140 a scăzut de la 41/160 la 33/160, schimbarea nu este semnificativă statistic, pentru că se bazează pe doar 24 de pacienți discordanți.

7.8.2 Testul chi-pătrat pentru trend

Când una dintre variabile are categorii ordonate (grupe de vârstă, stadii, doze, sau brațe cu intensitate crescătoare a intervenției), testul chi-pătrat obișnuit ignoră ordinea: ar da același rezultat dacă am amesteca rândurile. Testul pentru trend (Cochran-Armitage) întreabă mai precis: crește (sau scade) proporția de "da" odată cu categoria? Fiecărei categorii i se atribuie un scor (de exemplu 0, 1, 2), iar testul are un singur grad de libertate, ceea ce îl face mai puternic decât testul global atunci când există o tendință monotonă.

În mg2026, brațele A, B și C au intensitate crescătoare (consiliere; + aplicație; + aplicație și teleconsultație). Testul pentru trend cu scorurile 0, 1, 2 dă χ^2trend = 4,92, gl 1, p = 0,027. Proporțiile sunt 11,3%, 30,2% și 29,6%: trendul există, dar este dat aproape în întregime de saltul de la A la B; de la B la C proporția nu mai crește. Testul pentru trend se găsește în Epi Info (StatCalc, "Chi square for trend") și în MedCalc.

7.8.3 Testul chi-pătrat de concordanță

Testul de concordanță (engl. goodness of fit) compară distribuția unei singure variabile calitative cu o distribuție teoretică stabilită dinainte. Frecvențele așteptate nu mai vin din totalurile marginale, ci din ipoteză: E = N x proporția teoretică. Gradele de libertate sunt k - 1, unde k este numărul de categorii (dacă nu s-a estimat niciun parametru din date).

Primul exemplu: randomizarea în mg2026 a urmărit alocarea egală în cele trei brațe. Observat 53, 53, 54; așteptat 53,33 în fiecare. χ^2 = 0,0125, gl 2, p = 0,99: alocarea respectă foarte bine raportul 1:1:1. Al doilea: în eșantion sunt 86 de femei și 74 de bărbați. Față de un raport teoretic 50:50 (80 și 80), $\chi^2 = \frac{(86-80)^2}{80} + \frac{(74-80)^2}{80} = 0{,}45 + 0{,}45 = 0{,}90$, gl 1, p = 0,34. Al treilea: să presupunem că, într-un registru regional, distribuția nivelului de educație la adulții de 40-70 de ani ar fi 15% gimnaziu, 35% liceu sau școală profesională, 35% universitar și 15% postuniversitar (proporții ipotetice, folosite doar pentru exercițiu). Observat în mg2026: 20, 57, 64, 19; așteptat 24, 56, 56, 24. χ^2 = 0,67 + 0,02 + 1,14 + 1,04 = 2,87, gl 3, p = 0,41. Eșantionul nu diferă semnificativ de distribuția de referință, ceea ce sprijină, fără să o dovedească, reprezentativitatea lui pentru această caracteristică.

Testul pentru trend apare firesc atunci când grupurile au o ordine naturală. Piroș și colaboratorii, cu dr. Lupușoru, dr. L. Moleriu și dr. Muntean, au studiat transversal 693 de studenți de 18-24 de ani [6]. Prevalența prediabetului a fost 21,1% (diabet 1,4%). Studenții au fost grupați prin analiză de clustere în trei tipare alimentare: orientat spre sănătate (prediabet 15,4%), mixt (20,0%) și orientat spre fast-food (27,3%), iar testul chi-pătrat pe tabelul 3 x 2 a dat p = 0,003. Categoriile sunt ordonate după calitatea alimentației, iar autorii au mers mai departe exact în direcția discutată mai sus: un test pentru trend (p-trend < 0,001) și o regresie logistică în care fiecare categorie de tipar a crescut șansa de prediabet cu OR = 1,78 pe categorie (IC 95% 1,38-2,30; p < 0,001). Lectura corectă: există o asociere graduală între tiparul alimentar și prediabet, compatibilă cu o relație de tip doză-răspuns. Fiind un studiu transversal, asocierea nu demonstrează cauzalitatea, iar clusterele sunt construite din aceleași date, deci rezultatul trebuie confirmat într-o altă populație.

7.9 Dincolo de valoarea p: phi, V Cramér și alte măsuri ale asocierii

Valoarea χ^2 și valoarea p cresc odată cu mărimea eșantionului. Dacă am dubla toate frecvențele dintr-un tabel, procentele ar rămâne aceleași, dar χ^2 s-ar dubla, iar p ar scădea. De aceea, χ^2 și p nu spun cât de **puternică** este asocierea. Pentru asta folosim măsuri care nu depind de N.

Pentru un tabel 2 x 2, coeficientul phi:

$$\varphi = \sqrt{\frac{\chi^2}{N}}$$

Pentru tabele mai mari, coeficientul V Cramér:

$$V = \sqrt{\left[\frac{\chi^2}{N \times (m-1)}\right]}$$, unde m este cel mai mic dintre r și c

Ambele iau valori între 0 (independență) și 1 (asociere perfectă). Orientativ, pentru tabele cu m - 1 = 1, 0,1 înseamnă asociere slabă, 0,3 moderată și 0,5 puternică [76]. Pe mg2026:

Tabelul 7.14. Testul chi-pătrat și mărimea asocierii pe câteva tabele din mg2026.

Tabelul	χ^2	gl	p	φ sau V
obezitate x diabet TTGO (2 x 2)	12,70	1	< 0,001	$\varphi = 0{,}28$
brațul x scădere ≥ 5% (3 x 2)	6,77	2	0,034	$V = 0{,}21$
sexul x statutul de fumător (2 x 3)	0,54	2	0,76	$V = 0{,}06$
vârsta ≥ 50 ani x diabet TTGO (2 x 2)	4,36	1	0,037	$\varphi = 0{,}17$

Ultimul rând merită un comentariu. Diabetul a apărut la 2 din 34 de pacienți sub 50 de ani (5,9%) și la 27 din 126 de pacienți de 50 de ani sau peste (21,4%). Testul fără corecție dă p = 0,037, Fisher p = 0,044, iar corecția Yates p = 0,066, pentru că una dintre frecvențele așteptate este doar 6,2. Rezultatul este la limită, iar asocierea, măsurată prin φ = 0,17, este slabă. Concluzia onestă: un semnal compatibil cu creșterea riscului de diabet cu vârsta, dar fragil în acest eșantion.

Pentru tabelele 2 x 2 din studiile clinice și epidemiologice, măsurile cele mai utile clinic nu sunt φ sau V, ci cele care au unități interpretabile: diferența de proporții (cu IC 95%), riscul relativ sau raportul prevalențelor, odds ratio și numărul necesar de tratat. Chi-pătratul răspunde la întrebarea "există asociere?", iar aceste măsuri răspund la "cât de mare este și în ce direcție?". Un articol bun le raportează pe amândouă. Capitolul 10 este dedicat lor.

Studiu de caz: comportamentele alimentare de comoditate și clasele de IMC la studenți

Moleriu și colaboratorii, cu dr. L. Moleriu, dr. Lupușoru și dr. Muntean, au analizat 921 de studenți de 18-24 de ani [4]. Dintre ei, 24,1% au fost supraponderali sau obezi (19,4% supraponderali, 4,7% obezi). Metodele combină exact instrumentele din acest capitol și din capitolul 10: testul chi-pătrat pentru asocierea dintre comportamentele alimentare și clasa de IMC, V Cramér pentru tăria acestor asocieri, apoi regresie logistică multivariabilă cu odds ratio ajustate (de exemplu aOR 7,20, IC 95% 4,26-12,17 și aOR 1,98, IC 95% 1,26-3,10). Un scor compozit de risc alimentar arată o relație doză-răspuns: proporția celor cu exces ponderal crește de la 19,8% la scorul 0 la 50,0% la scorul 4, cu un OR ajustat de 1,43 pentru fiecare punct (IC 95% 1,23-1,66; p < 0,001). Pe un eșantion de 921 de persoane, chiar și asocieri slabe pot ieși semnificative statistic, iar V Cramér este tocmai măsura care ne ajută să le deosebim pe cele slabe de cele puternice. Designul este transversal: asocierea nu stabilește sensul cauzal, iar unele comportamente pot fi consecința, nu cauza excesului ponderal.

7.10 Raportarea rezultatelor testelor neparametrice și a testului chi-pătrat

Un rezultat bine raportat permite cititorului să refacă raționamentul. Pentru testele pe ranguri, descrierea se face prin mediană și interval intercuartil (sau minim-maxim, la eșantioane foarte mici), nu prin medie și abatere standard. Apoi se dau numele testului, statistica, valoarea p exactă (cu trei zecimale, iar sub 0,001 se scrie "p < 0,001") și, ideal, o mărime a efectului. Pentru chi-pătrat se dau frecvențele absolute și procentele (pe rândurile care corespund grupurilor comparate), χ^2, gradele de libertate, p și o măsură a efectului cu interval de încredere.

Câteva modele, pe datele din acest capitol:

- "Satisfacția la 6 luni a fost mai mare în brațul C decât în brațul A (mediana 4, IQR 3-5, față de 3, IQR 2-3; Mann-Whitney U = 2265,5, p < 0,001; corelația biserială a rangurilor 0,58)."

- "Scăderea glicemiei a jeun a diferit între brațe (Kruskal-Wallis H = 36,63, gl 2, p < 0,001, ε² = 0,23); comparațiile Dunn cu corecție Bonferroni au arătat diferențe între A și B (p < 0,001) și între A și C (p < 0,001), nu și între B și C (p = 0,12)."
- "Glicemia a jeun a scăzut în cursul studiului (mediane 104,5, 99,0 și 92,5 mg/dL la 0, 3 și 6 luni; Friedman χ^2 = 74,67, gl 2, p < 0,001; W Kendall = 0,23)."
- "Proporția pacienților cu scădere în greutate ≥ 5% a diferit între brațe (11,3%, 30,2% și 29,6%; χ^2 = 6,77, gl 2, p = 0,034; V Cramér = 0,21)."

Ce nu facem: nu scriem "p = 0,000" (programul rotunjește, dar probabilitatea nu este zero), nu raportăm doar "p < 0,05" când avem valoarea exactă, nu prezentăm media unor scoruri Likert ca și cum ar fi măsurători fizice și nu scriem "testul a fost nesemnificativ, deci grupurile sunt egale".

Un exemplu de raportare corectă a scorurilor ordinale vine dintr-un studiu care vă privește direct. Lungeanu și colaboratorii, cu dr. Lupușoru, au întrebat 121 de studenți la medicină, printr-un chestionar anonim, cum evaluează cursurile legate de competențele digitale și de știința datelor [16]. Aprecierile, pe o scală de la 1 la 10, sunt variabile ordinale, iar autorii le-au raportat exact cum recomandă acest capitol: mediană și IQR. Un curs a avut mediana 9 (IQR 7-10), cursul de informatică medicală 7 (5-9), iar biostatistica 8 (5-9). Relațiile dintre aprecieri au fost evaluate cu corelația Spearman, care lucrează tot pe ranguri: o corelație foarte puternică, R = 0,869 (p < 0,001), și o corelație practic absentă între cunoștințele măsurate și cele autoevaluate, R = 0,107 (p = 0,246). Al doilea rezultat este probabil cel mai interesant pentru voi: faptul că un student se simte pregătit nu spune aproape nimic despre cât știe de fapt. Un "nesemnificativ" bine interpretat, aici, este o informație valoroasă, nu un eșec al studiului.

7.11 Probleme rezolvate

Pentru calculele de mână pe eșantioane mici aveți nevoie de valorile critice ale testelor Mann-Whitney și Wilcoxon, care nu se găsesc în tabelele clasice ale distribuțiilor din anexă. Tabelele 7.15 și 7.16 le dau pentru testul bilateral la α = 0,05; au fost calculate din distribuțiile exacte ale statisticilor. Regula de decizie este aceeași la ambele: **H0 se respinge dacă statistica calculată este mai mică sau egală cu valoarea din tabel**. O liniuță înseamnă că, pentru acele mărimi ale grupurilor, nicio valoare nu poate atinge semnificația la acest prag.

Tabelul 7.15. Valori critice ale lui U Mann-Whitney (test bilateral, α = 0,05); n1 și n2 sunt mărimile celor două grupuri.

n1 n2	3	4	5	6	7	8	9	10
3	-	-	0	1	1	2	2	3
4	-	0	1	2	3	4	4	5
5	0	1	2	3	5	6	7	8
6	1	2	3	5	6	8	10	11
7	1	3	5	6	8	10	12	14
8	2	4	6	8	10	13	15	17
9	2	4	7	10	12	15	17	20
10	3	5	8	11	14	17	20	23

Tabelul 7.16. Valori critice ale lui T Wilcoxon pentru date pereche (test bilateral, α = 0,05); n este numărul de diferențe nenule.

n	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
T critic	0	2	3	5	8	10	13	17	21	25	29	34	40	46	52

Problema rezolvată 7.1. Durata spitalizării în pneumonie, Mann-Whitney de mână

Într-o secție de medicină internă, 6 pacienți cu pneumonie comunitară au trecut precoce de la antibiotic intravenos la antibiotic oral, iar 7 pacienți au primit schema intravenoasă standard. Durata spitalizării (zile) a fost: trecere precoce 4, 5, 5, 6, 8, 12; schema standard 6, 7, 7, 9, 10, 11, 15. Diferă durata spitalizării între cele două strategii (α = 0,05)?

Ipoteza de cercetare. Trecerea precoce la antibioticul oral scurtează spitalizarea.

Ipotezele statistice. H0: duratele de spitalizare provin din aceeași distribuție în cele două grupuri (Me1 = Me2). H1: distribuțiile diferă (Me1 ≠ Me2), test bilateral.

Alegerea testului. Variabila este numerică, dar durata spitalizării este tipic asimetrică (valoarea 15 zile și valoarea 12 zile trag media), grupurile sunt foarte mici (6 și 7) și independente. Alegem Mann-Whitney U.

Calcul. Seria comună ordonată (N = 13): 4, 5, 5, 6, 6, 7, 7, 8, 9, 10, 11, 12, 15. Rangurile: 4 → 1; 5 și 5 → 2,5; 6 și 6 (câte una din fiecare grup) → 4,5; 7 și 7 → 6,5; 8 → 8; 9 → 9; 10 → 10; 11 → 11; 12 → 12; 15 → 13.

R1 (trecere precoce) = 1 + 2,5 + 2,5 + 4,5 + 8 + 12 = 30,5 R2 (standard) = 4,5 + 6,5 + 6,5 + 9 + 10 + 11 + 13 = 60,5

Verificare: $30{,}5 + 60{,}5 = 91 = \frac{13\times14}{2}$.

$U_1 = 30{,}5 - \frac{6\times7}{2} = 9{,}5$; $U_2 = 60{,}5 - \frac{7\times8}{2} = 32{,}5$; verificare 9,5 + 32,5 = 42 = 6 x 7. U = 9,5.

Valoarea critică pentru n1 = 6, n2 = 7, bilateral, α = 0,05 este 6. Cum 9,5 > 6, nu respingem H0. Valoarea p exactă este 0,138. Mărimea efectului: $r_{\text{rb}} = 1 - \frac{2\times9{,}5}{42} = 0{,}55$.

Concluzia statistică. Nu există o diferență semnificativă statistic între distribuțiile duratei de spitalizare (U = 9,5, p = 0,138).

Concluzia clinică. Medianele sunt 5,5 zile față de 9 zile, iar efectul estimat este mare (0,55), dar studiul este prea mic pentru a-l confirma. Rezultatul nu arată că strategiile sunt echivalente; el justifică un studiu cu mai mulți pacienți, al cărui număr de pacienți se calculează în prealabil (capitolele 4 și 5).

Problema rezolvată 7.2. Durerea înainte și după kinetoterapie, Wilcoxon și testul semnului

Zece pacienți cu gonartroză și-au notat durerea (0-10) înainte și după 4 săptămâni de kinetoterapie: înainte 7, 6, 8, 5, 9, 6, 7, 8, 6, 7; după 4, 5, 5, 4, 6, 6, 3, 5, 4, 7. A scăzut durerea?

Ipotezele. H0: mediana diferențelor (înainte - după) este 0. H1: mediana diferențelor ≠ 0.

Alegerea testului. Aceiași pacienți, două momente, deci date pereche; scorul este ordinal. Alegem Wilcoxon pentru date pereche.

Calcul. Diferențele: 3, 1, 3, 1, 3, 0, 4, 3, 2, 0. Două diferențe sunt zero și se elimină, n = 8. Valorile absolute ordonate: 1, 1 (rangurile 1 și 2, deci 1,5 fiecare); 2 (rang 3); 3, 3, 3, 3 (rangurile

4, 5, 6, 7, deci 5,5 fiecare); 4 (rang 8). Toate diferențele nenule sunt pozitive, deci W+ = 1,5 + 1,5 + 3 + 4 x 5,5 + 8 = 36 și W- = 0. Verificare: $36 + 0 = \frac{8 \times 9}{2}$. T = 0.

Valoarea critică pentru n = 8, bilateral, α = 0,05 este 3. Cum 0 ≤ 3, respingem H0; p exact = 0,008. Testul semnului dă același rezultat aici: 8 plusuri din 8, p = 2 x 0,5 = 0,008.

Concluzii. Durerea a scăzut semnificativ după kinetoterapie (mediana de la 7 la 5; mediana diferențelor 2,5 puncte; Wilcoxon T = 0, p = 0,008). Clinic, o scădere de 2-3 puncte pe scala 0-10 este de obicei percepută de pacient ca relevantă. Lipsa unui grup de control nu permite însă separarea efectului kinetoterapiei de evoluția naturală sau de efectul placebo.

Problema rezolvată 7.3. Vindecarea plăgilor cu trei tipuri de pansament, Kruskal-Wallis

Numărul de zile până la vindecarea completă a unei plăgi superficiale a fost înregistrat la 14 pacienți: pansament clasic (n = 4) 12, 14, 15, 18; pansament hidrocoloid (n = 5) 9, 10, 11, 13, 14; pansament cu argint (n = 5) 7, 8, 8, 10, 12. Diferă timpul de vindecare între pansamente?

Ipotezele. H0: distribuția timpului de vindecare este aceeași pentru cele trei pansamente. H1: cel puțin un pansament diferă.

Alegerea testului. Trei grupuri independente, eșantioane foarte mici, variabilă numerică de tip "durată": Kruskal-Wallis.

Calcul. Seria comună ordonată (N = 14): 7, 8, 8, 9, 10, 10, 11, 12, 12, 13, 14, 14, 15, 18. Rangurile: 7 → 1; 8 și 8 → 2,5; 9 → 4; 10 și 10 → 5,5; 11 → 7; 12 și 12 → 8,5; 13 → 10; 14 și 14 → 11,5; 15 → 13; 18 → 14.

R1 (clasic) = 8,5 + 11,5 + 13 + 14 = 47 R2 (hidrocoloid) = 4 + 5,5 + 7 + 10 + 11,5 = 38 R3 (argint) = 1 + 2,5 + 2,5 + 5,5 + 8,5 = 20

Verificare: $47 + 38 + 20 = 105 = \frac{14 \times 15}{2}$.

$$H = \left[\frac{12}{14 \times 15}\right] \times \left(\frac{47^2}{4} + \frac{38^2}{5} + \frac{20^2}{5}\right) - 3 \times 15 = 0{,}05714 \times (552{,}25 + 288{,}8 + 80) - 45$$
$$= 0{,}05714 \times 921{,}05 - 45 = 7{,}63$$

Cu corecția pentru egalități, H = 7,70. Cu 2 grade de libertate, 7,70 > 5,99, deci respingem H0 (p = 0,021). Atenție: un grup are doar 4 pacienți, deci aproximarea chi-pătrat este la limita acceptabilității; programele care oferă varianta exactă trebuie preferate.

Post hoc (Dunn, Bonferroni). Rangurile medii sunt 11,75 (clasic), 7,6 (hidrocoloid) și 4,0 (argint). Doar diferența clasic față de argint rămâne semnificativă după corecție (z = 2,76, p ajustat = 0,017); clasic față de hidrocoloid p ajustat = 0,42, hidrocoloid față de argint p ajustat = 0,52.

Concluzii. Timpul de vindecare a diferit între pansamente (mediane 14,5, 11 și 8 zile; H = 7,70, p = 0,021), iar diferența se localizează între pansamentul clasic și cel cu argint. Pe 14 pacienți, rezultatul este orientativ.

Problema rezolvată 7.4. Oboseala după COVID-19 la trei momente, Friedman

Șapte pacienți cu simptome persistente după COVID-19 și-au notat oboseala (0-10) la 1, 3 și 6 luni după infecție:

Tabelul 7.17. Scorul oboselii și rangurile în interiorul fiecărui pacient (date fictive).

Pacient	1 lună	3 luni	6 luni	Ranguri
1	8	6	4	3; 2; 1
2	7	7	5	2,5; 2,5; 1
3	9	6	5	3; 2; 1
4	6	5	5	3; 1,5; 1,5
5	8	5	3	3; 2; 1
6	7	6	4	3; 2; 1
7	9	8	6	3; 2; 1
Rj				20,5; 14; 7,5

Ipotezele. H0: distribuția scorului de oboseală este aceeași la cele trei momente. H1: cel puțin un moment diferă.

Alegerea testului. Aceiași pacienți, trei momente, scor ordinal: Friedman.

Calcul. Verificare: 20,5 + 14 + 7,5 = 42 = 7 x 6.

$$\chi^2 F = \left[\frac{12}{7 \times 3 \times 4}\right] \times (20{,}5^2 + 14^2 + 7{,}5^2) - 3 \times 7 \times 4$$
$$= 0{,}1429 \times (420{,}25 + 196 + 56{,}25) - 84 = 0{,}1429 \times 672{,}5 - 84 = 12{,}07$$

Cu corecția pentru egalități, 13,00; gl 2; 13,00 > 5,99, deci respingem H0 (p = 0,0015). Medianele: 8, 6 și 5.

Post hoc. Wilcoxon pentru fiecare pereche: 1 față de 3 luni p = 0,031; 3 față de 6 luni p = 0,031; 1 față de 6 luni p = 0,016. Cu Bonferroni (x 3): 0,094, 0,094 și 0,047. Doar comparația 1 față de 6 luni rămâne semnificativă. Cu 7 pacienți, cea mai mică valoare p pe care o poate da testul Wilcoxon exact bilateral este 0,016, deci comparațiile post hoc au o putere foarte limitată.

Concluzii. Oboseala a scăzut semnificativ în primele 6 luni (Friedman χ^2 = 13,00, p = 0,0015); diferența dintre 1 și 6 luni este cea care rezistă corecției. Fără grup de comparație, nu putem atribui scăderea vreunui tratament.

Problema rezolvată 7.5. Fumatul și BPOC, chi-pătrat 2 x 2 complet

Într-un studiu transversal în cabinetele de medicină de familie au fost evaluate spirometric 300 de persoane de peste 40 de ani: dintre cei 120 de fumători, 30 aveau BPOC; dintre cei 180 de nefumători, 15 aveau BPOC. Este fumatul asociat cu BPOC?

Ipotezele. H0: proporția de BPOC este aceeași la fumători și nefumători (p1 = p2). H1: p1 ≠ p2.

Alegerea testului. Două variabile nominale dihotomice, grupuri independente: chi-pătrat pe tabel 2 x 2 (verificăm condițiile).

Tabelul 7.18. Fumatul și BPOC: observat (așteptat).

	BPOC da	BPOC nu	Total
Fumători	30 (18)	90 (102)	120
Nefumători	15 (27)	165 (153)	180
Total	45	255	300

Frecvențe așteptate. $E_{11} = \frac{120 \times 45}{300} = 18$; $E_{12} = \frac{120 \times 255}{300} = 102$; $E_{21} = \frac{180 \times 45}{300} = 27$; $E_{22} = \frac{180 \times 255}{300} = 153$. Toate ≥ 5, testul este valid.

Statistica. |O - E| = 12 în fiecare celulă.

$$\chi^2 = \frac{144}{18} + \frac{144}{102} + \frac{144}{27} + \frac{144}{153} = 8{,}000 + 1{,}412 + 5{,}333 + 0{,}941 = 15{,}69$$

gl = 1; valoarea critică 3,84; p < 0,001. Cu corecția Yates: (12 - 0,5)² = 132,25, iar χ^2Yates = $132{,}25 \times \left(\frac{1}{18} + \frac{1}{102} + \frac{1}{27} + \frac{1}{153}\right) = 14{,}41$, p < 0,001. Testul Fisher dă tot p < 0,001. Coeficientul φ = √(15,69/300) = 0,23.

Măsuri de efect. Prevalența BPOC: 25,0% la fumători, 8,3% la nefumători. Raportul prevalențelor = $\frac{0{,}250}{0{,}083} = 3{,}0(;\ \text{IC}\,95\%\ 1{,}69 - 5{,}33)$; OR = $\frac{30\times165}{90\times15} = 3{,}67(;\ \text{IC}\,95\%\ 1{,}87 - 7{,}17)$.

Concluzii. Există o asociere semnificativă între fumat și BPOC (χ^2 = 15,69, p < 0,001); prevalența BPOC este de trei ori mai mare la fumători. Asocierea este de intensitate slabă spre moderată ca φ, dar clinic importantă, pentru că BPOC este frecventă. Designul transversal nu stabilește singur cauzalitatea, însă rezultatul este în acord cu dovezile din studiile de cohortă.

Problema rezolvată 7.6. Statutul de fumător și diabetul în mg2026, tabel 3 x 2

În setul mg2026, diabetul confirmat prin TTGO a apărut la 3 din 33 de foști fumători, 10 din 53 de fumători și 16 din 74 de nefumători. Este statutul de fumător asociat cu diabetul la includere?

Ipotezele. H0: proporția de diabet este aceeași în cele trei categorii de fumat. H1: cel puțin o proporție diferă.

Calculul frecvențelor așteptate (29 de diabetici din 160, deci 18,1%): foști fumători $\frac{33\times29}{160} = 5{,}98$; fumători $\frac{53\times29}{160} = 9{,}61$; nefumători $\frac{74\times29}{160} = 13{,}41$; în coloana "fără diabet": 27,02, 43,39 și 60,59. Toate ≥ 5 (cea mai mică 5,98), deci testul este valid.

Contribuțiile: foști fumători (3 - 5,98)²/5,98 = 1,486 și (30 - 27,02)²/27,02 = 0,329; fumători (10 - 9,61)²/9,61 = 0,016 și (43 - 43,39)²/43,39 = 0,004; nefumători (16 - 13,41)²/13,41 = 0,499 și (58 - 60,59)²/60,59 = 0,111. χ^2 = 2,44, gl = (3 - 1)(2 - 1) = 2, p = 0,29. V Cramér $= \sqrt{\frac{2{,}44}{160}} = 0{,}12$.

Concluzii. Nu am găsit o asociere semnificativă între statutul de fumător și diabet (9,1%, 18,9% și 21,6%; χ^2 = 2,44, p = 0,29). Procentul mai mic la foști fumători se bazează pe doar 3 cazuri; cu 33 de persoane în acea categorie, estimarea este foarte imprecisă. Fumatul rămâne un factor de risc cardiovascular dovedit; acest eșantion mic nu are puterea de a evidenția o asociere cu diabetul.

Problema rezolvată 7.7. Două teste rapide pe aceiași pacienți, McNemar

La 100 de pacienți cu faringită s-au aplicat două teste rapide pentru streptococ (A și B). Ambele pozitive: 40; A pozitiv și B negativ: 15; A negativ și B pozitiv: 5; ambele negative: 40. Diferă proporția de rezultate pozitive între cele două teste?

Alegerea testului. Fiecare pacient are două rezultate binare, deci date pereche: McNemar. Testul chi-pătrat obișnuit ar fi greșit, pentru că aceiași pacienți ar apărea în ambele rânduri.

Calcul. Proporțiile de pozitivi: $\frac{A\,55}{100} = 55\%$, $\frac{B\,45}{100} = 45\%$. Celulele discordante: b = 15, c = 5.

$\chi^2 = \frac{(15-5)^2}{15+5} = \frac{100}{20} = 5{,}00$, gl 1, p = 0,025; cu corecție de continuitate: (10 - 1)²/20 = 4,05, p = 0,044; varianta exactă (binomială) p = 0,041

Concluzii. Testul A dă semnificativ mai multe rezultate pozitive decât testul B. Testul McNemar nu spune care test este mai bun; pentru asta ar trebui comparate ambele cu standardul de referință (cultura din exsudatul faringian), adică ar trebui calculate sensibilitatea și specificitatea (capitolele 11 și 12).

Problema rezolvată 7.8. Delirul postoperator într-un studiu mic, testul exact Fisher

Într-un studiu-pilot la vârstnici operați, delirul postoperator a apărut la 1 din 12 pacienți care au primit un sedativ nou și la 6 din 10 pacienți care au primit îngrijirea obișnuită. Diferă frecvența delirului?

Verificarea condițiilor. Totalul cu delir este 7 din 22. Frecvențele așteptate: $\frac{12\times7}{22} = 3{,}82$ și $\frac{10\times7}{22} = 3{,}18$ în coloana "delir"; 8,18 și 6,82 în coloana "fără delir". Două celule din patru au E < 5, deci folosim testul exact Fisher.

Calcul. Cu totalurile fixate (12 și 10 pe rânduri, 7 și 15 pe coloane), numărul a de cazuri de delir în grupul cu sedativ poate fi de la 0 la 7. Probabilitățile hipergeometrice sunt: P(0) = 0,0007; P(1) = 0,0148 (tabelul observat); P(2) = 0,0975; P(3) = 0,2709; P(4) = 0,3483; P(5) = 0,2090; P(6) = 0,0542; P(7) = 0,0046. Adunăm tabelele cel mult la fel de probabile ca cel observat: 0,0007 + 0,0148 + 0,0046 = 0,020.

Concluzii. Fisher p = 0,020: delirul a fost semnificativ mai rar în grupul cu sedativ nou (8,3% față de 60,0%). Testul chi-pătrat aplicat greșit ar fi dat p = 0,010, iar cu corecție Yates p = 0,033; concluzia nu se schimbă aici, dar valoarea corectă este cea exactă. Cu 22 de pacienți, efectul trebuie confirmat într-un studiu randomizat adecvat dimensionat.

7.12 Exerciții propuse

1. Atribuiți ranguri valorilor 12, 15, 15, 9, 20, 15, 11, 9 (în ordinea dată) și verificați suma rangurilor.
2. Scorul de anxietate (0-10) înainte de o endoscopie a fost: grup cu informare video (n = 4) 3, 4, 4, 6; grup fără informare (n = 5) 5, 7, 8, 8, 9. Calculați R1, R2, U1, U2 și U. Valoarea critică pentru n1 = 4, n2 = 5, bilateral, α = 0,05, este 1. Ce decideți?
3. Tensiunea sistolică (mmHg) la 8 pacienți, înainte și după 4 săptămâni de tratament: înainte 140, 152, 138, 160, 145, 150, 148, 155; după 132, 150, 140, 148, 139, 150, 141, 146. Aplicați testul Wilcoxon de mână (valoarea critică pentru n = 7 este 2). Ce observați la pacientul 6?
4. Într-un studiu cu trei grupuri de câte 6 pacienți (N = 18), sumele rangurilor sunt 40, 60 și 71. Verificați suma rangurilor, calculați H și decideți (α = 0,05).
5. Zece pacienți au fost evaluați în trei momente; sumele rangurilor în interiorul pacienților sunt 14, 20 și 26. Calculați statistica Friedman și decideți (α = 0,05). Calculați W Kendall.
6. Într-un sezon gripal, 10 din 200 de persoane vaccinate și 24 din 150 de persoane nevaccinate au avut gripă confirmată. Construiți tabelul, calculați frecvențele așteptate, χ^2 (cu și fără Yates), riscul relativ și φ.
7. Un tabel 3 x 3 are frecvențele observate: rândul 1: 8, 12, 5; rândul 2: 10, 20, 6; rândul 3: 2, 4, 3. Calculați frecvențele așteptate pentru rândul 3. Este aplicabil testul chi-pătrat? Ce soluții propuneți?
8. La 80 de pacienți s-a evaluat prezența hipertensiunii înainte și după un program de scădere în greutate. 12 pacienți au trecut de la "hipertensiv" la "normotensiv", iar 4 de la "normotensiv" la "hipertensiv". Ce test folosiți? Calculați statistica, cu și fără corecție.
9. Într-un eșantion de 200 de donatori, grupele sanguine O, A, B, AB au fost observate de 70, 85, 30 și 15 ori. Testați concordanța cu proporțiile ipotetice 35%, 40%, 17%, 8%.

10. Un efect advers a apărut la 3 din 12 pacienți cu medicamentul X și la 0 din 12 pacienți cu placebo. Calculați frecvențele așteptate, alegeți testul și interpretați valoarea p dată de program (Fisher bilateral p = 0,217).
11. Pe setul mg2026, în JASP, comparați satisfacția la 6 luni între brațele B și A. Raportați medianele, U, p și corelația biserială a rangurilor.

Capcane frecvente

- Aplicarea testului Mann-Whitney sau a testului chi-pătrat obișnuit pe date pereche (aceiași pacienți măsurați de două ori). Pentru date pereche: Wilcoxon, respectiv McNemar.
- Raportarea mediei și abaterii standard pentru scoruri Likert și apoi aplicarea unui test pe ranguri: descrierea trebuie să se potrivească testului (mediană și IQR).
- Verificarea condiției de aplicare a testului chi-pătrat pe frecvențele observate în loc de cele așteptate.
- Introducerea procentelor în locul frecvențelor absolute în tabelul de contingență.
- Concluzia "medianele sunt egale" după un test nesemnificativ; corect: "nu am găsit o diferență semnificativă".
- Mai multe teste Mann-Whitney între trei grupuri, fiecare la pragul 0,05, fără test global și fără corecție.
- Compararea lui U sau T cu valoarea critică în direcția greșită: la Mann-Whitney și Wilcoxon, H0 se respinge când statistica este **mai mică sau egală** cu valoarea critică.
- Interpretarea unui χ^2 mare drept "asociere puternică": tăria se citește din φ, V Cramér sau din RR/OR, nu din p.

De reținut

- Testele neparametrice lucrează pe ranguri; sunt potrivite pentru date ordinale, distribuții asimetrice la eșantioane mici și valori extreme reale. Eficiența lor la date normale este aproximativ 95%.
- Mann-Whitney: 2 grupuri independente; Wilcoxon: 2 măsurători pereche; Kruskal-Wallis: ≥ 3 grupuri independente (post hoc Dunn); Friedman: ≥ 3 măsurători repetate (post hoc Wilcoxon cu corecție).
- La valori egale se atribuie rangul mediu; suma rangurilor este întotdeauna $N(N + 1)/2$.
- Frecvența așteptată: E = totalul rândului x totalul coloanei/N; gl = (r - 1)(c - 1).
- Chi-pătrat valid: E ≥ 5 în toate celulele unui tabel 2 x 2; în tabelele mai mari, cel mult 20% din celule cu E < 5 și niciuna cu E < 1. Altfel: Fisher sau comasarea categoriilor.
- McNemar pentru proporții pereche, testul pentru trend pentru categorii ordonate, testul de concordanță pentru o singură variabilă comparată cu o distribuție teoretică.
- Raportarea completă: descriere (mediană și IQR, respectiv n și %), statistica testului, gl, p exact și o mărime a efectului (r_{rb}, ε^2, W, φ, V, RR sau OR cu IC 95%).

Întrebări de autoevaluare

1. (CS) Un studiu compară scorul de calitate a vieții (scală ordinală 0-100, distribuție puternic asimetrică) între 15 pacienți operați laparoscopic și 15 pacienți operați clasic. Testul potrivit este:
    a) testul t pereche
    b) testul Wilcoxon pentru date pereche
    c) testul Mann-Whitney U
    d) testul Friedman
    e) testul McNemar
2. (CS) Valorile 4, 7, 7, 2, 7, 9 primesc, în ordinea dată, rangurile:
    a) 2; 3; 4; 1; 5; 6
    b) 2; 4; 4; 1; 4; 6

c) 2; 3,5; 3,5; 1; 3,5; 6

d) 2; 4; 4; 1; 5; 6

e) 1; 4; 4; 2; 4; 6

3. (CS) Un tabel de contingență are 4 rânduri și 3 coloane. Numărul gradelor de libertate pentru testul chi-pătrat este:

a) 12

b) 11

c) 7

d) 6

e) 5

4. (CS) Într-un tabel 2 x 2, totalul primului rând este 40, totalul primei coloane este 30, iar totalul general este 200. Frecvența așteptată în celula de la intersecția lor este:

a) 6

b) 7,5

c) 12

d) 15

e) 30

5. (CS) La 60 de pacienți, prezența durerii toracice a fost înregistrată înainte și după tratament. Testul potrivit pentru a compara proporțiile este:

a) chi-pătrat Pearson pe un tabel "moment x durere" cu 120 de observații

b) testul exact Fisher

c) testul McNemar

d) testul Kruskal-Wallis

e) testul pentru trend Cochran-Armitage

6. (CM) Despre testul Kruskal-Wallis sunt adevărate afirmațiile:

a) este alternativa neparametrică a ANOVA unifactoriale

b) rangurile se atribuie în interiorul fiecărui subiect

c) statistica H se compară cu distribuția chi-pătrat cu k - 1 grade de libertate

d) un rezultat semnificativ arată exact care grupuri diferă între ele

e) după un rezultat semnificativ se pot face comparații Dunn cu corecție pentru comparații multiple

7. (CM) Testul chi-pătrat pe un tabel 2 x 2 NU este adecvat (fără alte măsuri) dacă:

a) o frecvență așteptată este 3,2

b) o frecvență observată este 0, iar toate frecvențele așteptate sunt peste 5

c) aceiași pacienți sunt numărați atât în rândul "înainte", cât și în rândul "după"

d) tabelul conține procente în loc de numere de pacienți

e) grupurile au mărimi inegale (300 și 80 de pacienți)

8. (CM) În setul mg2026, Mann-Whitney pentru satisfacție, brațul C față de A, a dat U = 2265,5, p < 0,001, corelația biserială a rangurilor 0,58. Sunt corecte:

a) satisfacția a fost mai mare în brațul C

b) efectul este mare

c) medianele sunt 4 (IQR 3-5) și 3 (IQR 2-3)

d) rezultatul dovedește că brațul C este mai eficient în controlul glicemiei

e) probabilitatea ca un pacient din C să fie mai mulțumit decât unul din A este aproximativ 0,79

9. (CM) Despre testul Friedman sunt adevărate:
   a) se aplică la ≥ 3 măsurători pe aceiași subiecți
   b) rangurile se atribuie în eșantionul comun al tuturor măsurătorilor
   c) coeficientul W Kendall poate fi folosit ca mărime a efectului
   d) comparațiile post hoc se pot face cu testul Wilcoxon și corecție Bonferroni
   e) este alternativa neparametrică a ANOVA cu măsurători repetate
10. (CM) Pentru tabelul obezitate x diabet din mg2026 ($\chi^2 = 12{,}70$, $p < 0{,}001$, $\varphi = 0{,}28$) sunt corecte:

a) asocierea este semnificativă statistic
b) φ arată o asociere de intensitate slabă spre moderată
c) valoarea p mică arată că asocierea este foarte puternică
d) designul transversal nu permite concluzia că obezitatea a cauzat diabetul
e) frecvența așteptată pentru celula "obez cu diabet" este 9,79

Capitolul 8. Selecția și aplicarea testelor statistice: criterii de alegere, rezolvarea de probleme și studii de caz

Obiective

La finalul acestui capitol, studentul va putea:

- să identifice scopul unei analize (diferență, asociere, acord, predicție) pornind de la întrebarea de cercetare;
- să clasifice variabila rezultat și variabila de grupare și să recunoască designul (grupuri independente sau date pereche, numărul de grupuri sau de momente);
- să aplice pas cu pas algoritmul de alegere a testului statistic și să justifice alegerea, inclusiv alternativa neparametrică sau categorială;
- să recunoască unitatea de analiză corectă și situațiile în care observațiile nu sunt independente;
- să explice de ce testul trebuie stabilit înainte de analiza datelor și ce se întâmplă când este ales după valoarea p;
- să recunoască cele mai frecvente greșeli de alegere a testului în descrieri de tip "Metode" și să propună testul corect;
- să redacteze rezolvarea unei probleme în formatul standardizat al disciplinei (a-g), de la ipoteza de cercetare la concluzia clinică;
- să evalueze critic potrivirea dintre design, tipul variabilei și testul raportat în articole publicate de colegii din disciplină.

8.1 Alegerea testului este o decizie de proiectare, nu de calcul

Programele statistice moderne calculează orice test în câteva secunde. JASP, MedCalc sau Epi Info nu vor protesta dacă le cereți un test t pe date care sunt de fapt pereche sau un chi-pătrat pe un tabel cu frecvențe așteptate de 1,5. Rezultatul va arăta impecabil, cu trei zecimale și un grafic colorat, și va fi greșit. Din acest motiv, partea grea a analizei statistice nu este calculul, ci alegerea: ce întrebare punem, ce tip de date avem, ce test corespunde acestei combinații.

În capitolele 5, 6 și 7 am văzut testele unul câte unul. Acest capitol le pune laolaltă. Vom construi un algoritm de decizie, îl vom aplica pe 22 de probleme rezolvate pe setul mg2026, vom exersa recunoașterea testelor greșite și vom citi secțiunile de metode din articolele colegilor din disciplină. La examenul practic, cerința tipică este exact aceasta: "Ce test statistic aplicați pentru a vedea dacă...?", urmată de rezolvarea completă în formatul a-g. Experiența de la laborator arată că studenții care greșesc la examen nu greșesc, de regulă, la citirea valorii p, ci la primul pas: au confundat date pereche cu grupuri independente sau o variabilă ordinală cu una numerică.

Alegerea testului se face, în mod ideal, înainte de culegerea datelor, în protocolul studiului, odată cu definirea variabilei principale și cu calculul dimensiunii eșantionului [18,56]. Testul decurge din design, iar designul decurge din întrebare. Dacă întrebarea este bine formulată ("diferă scăderea glicemiei la 6 luni între cele trei brațe?"), testul aproape că se alege singur.

8.2 Algoritmul de alegere a testului

Figura 8.1 rezumă algoritmul folosit în acest volum. El nu este singurul posibil, dar acoperă marea majoritate a situațiilor din cercetarea clinică de bază.

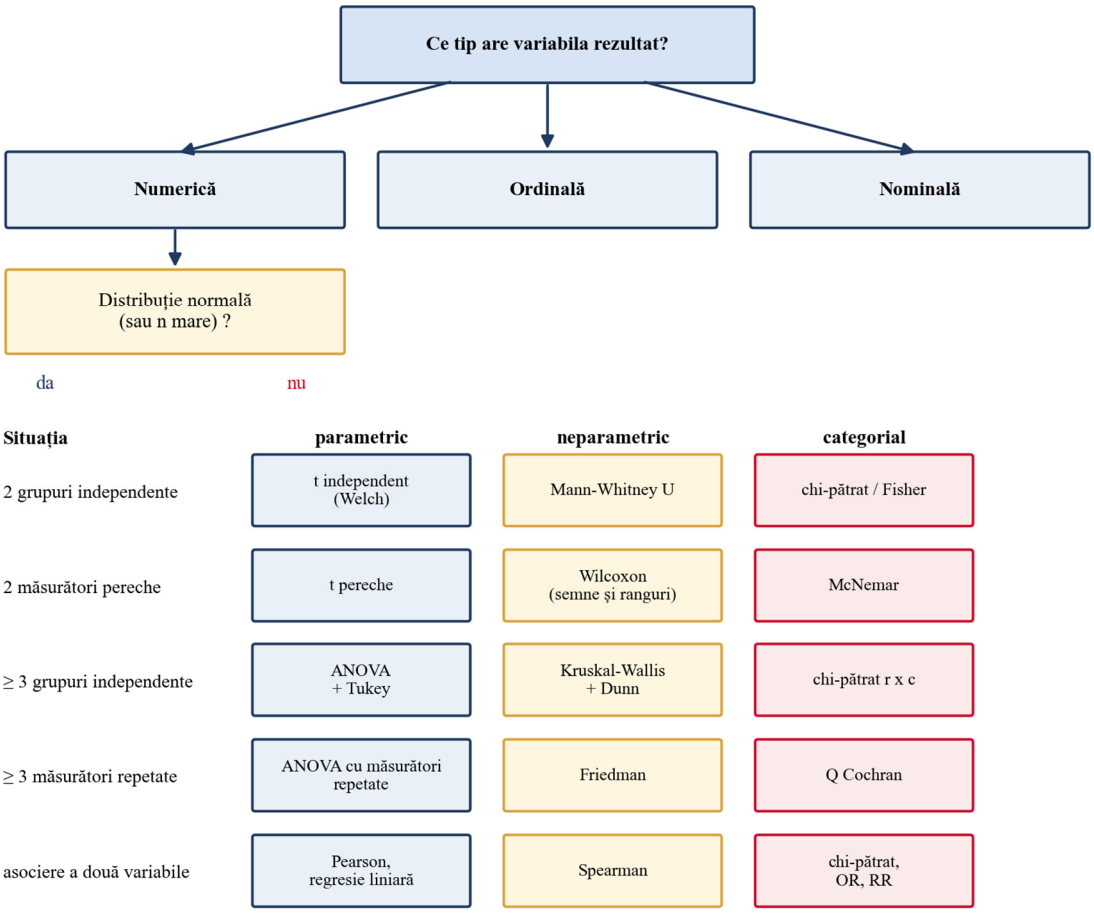


Figura 8.1. Algoritmul de alegere a testului statistic: tipul variabilei rezultat (numerică, ordinală, nominală), apoi situația (2 grupuri independente, 2 măsurători pereche, ≥ 3 grupuri independente, ≥ 3 măsurători repetate, asocierea a două variabile) și, pentru variabilele numerice, distribuția (normală sau n mare față de asimetrie la n mic). Coloanele dau varianta parametrică, neparametrică și categorială.

Algoritmul se parcurge în cinci pași. Ordinea contează: un pas greșit la început nu poate fi reparat mai târziu.

8.2.1 Pasul 1: care este scopul analizei?

Prima întrebare nu privește datele, ci intenția. Există patru scopuri mari, iar fiecare cere o familie diferită de metode:

- **diferența** între grupuri sau între momente: "Este glicemia mai mică în brațul C decât în brațul A?" (teste t, ANOVA, Mann-Whitney, Kruskal-Wallis, Wilcoxon, Friedman, chi-pătrat, McNemar);
- **asocierea** între două variabile măsurate pe aceiași subiecți: "Crește tensiunea sistolică odată cu IMC?" (corelația Pearson sau Spearman, chi-pătrat pentru variabile calitative, OR și RR);
- **acordul** între două metode sau doi evaluatori care măsoară același lucru: "Dau tensiometrul automat și cel manual aceleași valori?" (metoda Bland-Altman [102], coeficientul de corelație intraclasă, kappa Cohen [103]);
- **predicția** unui rezultat din unul sau mai mulți factori: "Cât de bine prezice FINDRISC diabetul?" (regresia liniară sau logistică, curba ROC).

Confuzia cea mai frecventă este între asociere și acord. Două tensiometre pot avea o corelație de 0,95 și totuși unul să dea sistematic cu 10 mmHg mai mult. Corelația măsoară dacă valorile cresc

împreună; acordul măsoară dacă sunt aceleași. Pentru acord nu se folosește nici testul t pereche singur, nici corelația, ci analiza diferențelor (media diferențelor și limitele de acord).

8.2.2 Pasul 2: ce tip are variabila rezultat?

Variabila rezultat (dependentă) este cea pe care vrem să o comparăm sau să o explicăm: glicemia, satisfacția, prezența diabetului. Variabila de grupare (independentă) definește grupurile: brațul, sexul, mediul de rezidență. Tipul variabilei rezultat decide coloana din figura 8.1.

- **Numerică** (continuă sau discretă cu multe valori): glicemie, TAS, greutate, IMC, vârstă, scor FINDRISC 0-26.
- **Ordinală**: satisfacția 1-5, nivelul de educație 1-4, activitatea fizică autoevaluată 1-5, stadiul unei boli.
- **Nominală** (inclusiv dihotomică): diabet da/nu, scădere ≥ 5% da/nu, statutul de fumător, sexul.

Două situații cer atenție. Un scor cu multe valori (FINDRISC 0-26, un scor de calitate a vieții 0-100) este tratat de obicei ca numeric, dacă distribuția o permite. O variabilă numerică dihotomizată (glicemia ≥ 100 mg/dL da/nu) devine nominală, dar pierde informație și putere; dihotomizarea se justifică doar când pragul are sens clinic și a fost stabilit dinainte.

8.2.3 Pasul 3: câte grupuri sau câte momente?

Numărăm grupurile sau momentele comparate: unul (comparația cu o valoare de referință), două sau trei și mai multe. Distincția dintre "două" și "trei sau mai multe" nu este formală. Cu trei grupuri, trei teste t separate ridică riscul global de eroare de tip I de la 5% la 14,3% ($1 - 0{,}95^3$); cu cinci grupuri, cele zece comparații posibile îl duc la 40%. De aceea, pentru trei sau mai multe grupuri folosim un test global (ANOVA, Kruskal-Wallis, chi-pătrat r x c) și abia apoi comparații post hoc cu corecție.

8.2.4 Pasul 4: grupuri independente sau date pereche?

Datele sunt **pereche** (dependente) când fiecare valoare dintr-o serie are un corespondent natural în cealaltă: același pacient înainte și după, ochiul drept și ochiul stâng, două metode pe aceeași probă, perechi potrivite în studiile caz-control. Datele sunt **independente** când grupurile sunt formate din subiecți diferiți, fără legătură între ei. Întrebarea de control este simplă: "pot uni cu o linie fiecare valoare din prima coloană cu o anumită valoare din a doua?" Dacă da, datele sunt pereche.

Diferența de rezultat poate fi mare. În mg2026, glicemia a scăzut de la 105,44 mg/dL la includere la 95,25 mg/dL la 6 luni. Testul t pereche (corect) dă t = 11,14, cu eroarea standard a diferenței de 0,91 mg/dL. Dacă tratăm greșit cele două serii ca pe două grupuri independente (testul Welch), eroarea standard crește la 2,03 mg/dL, iar t scade la 5,03. Motivul este corelația puternică dintre glicemia unui pacient la 0 și la 6 luni (r = 0,82): testul pereche elimină variabilitatea dintre pacienți și rămâne doar cu variabilitatea schimbării. Aici ambele variante ies semnificative, dar pe un eșantion mai mic testul greșit ar fi putut rata un efect real.

8.2.5 Pasul 5: distribuția și mărimea eșantionului

Doar pentru variabilele numerice, ultima întrebare este dacă putem folosi un test parametric. Criteriile, discutate în capitolele 3 și 6:

- forma distribuției în fiecare grup (histogramă, diagrama Q-Q, asimetria) și, ca indicație, testul Shapiro-Wilk [66]; pentru testul t pereche se verifică distribuția **diferențelor**, nu a celor două serii;
- mărimea eșantionului: cu cel puțin 30-40 de subiecți pe grup și asimetrie moderată, testul t și ANOVA rămân robuste; cu grupuri mici, distribuția contează mult mai mult;

- omogenitatea varianțelor (testul Levene); dacă este încălcată, se folosește varianta Welch a testului t sau a ANOVA, nu automat un test neparametric;
- valorile extreme reale și variabilele mărginite (scoruri cu efect de plafon).

Un test Shapiro-Wilk "semnificativ" pe un eșantion mare nu obligă la un test neparametric. În mg2026, glicemia la includere are SW p = 0,037 la n = 160, dar asimetria este doar 0,37, iar histograma arată o distribuție aproape simetrică. Pentru 160 de pacienți, testul t este perfect adecvat. Invers, un SW "nesemnificativ" pe 8 pacienți nu garantează normalitatea, ci arată doar puterea mică a testului de normalitate.

8.2.6 Algoritmul pe scurt

Pentru lucrul rapid la laborator, algoritmul se poate scrie ca o listă de verificare:

1. Scrie întrebarea de cercetare într-o propoziție. Ce vrei să afli: diferență, asociere, acord sau predicție?
2. Identifică variabila rezultat și tipul ei (numerică, ordinală, nominală).
3. Identifică variabila de grupare sau momentele: câte categorii are?
4. Verifică dacă datele sunt independente sau pereche.
5. Pentru variabila numerică: verifică distribuția în fiecare grup (sau a diferențelor) și omogenitatea varianțelor; decide între varianta parametrică, Welch sau neparametrică.
6. Pentru variabilele calitative: calculează frecvențele așteptate și decide între chi-pătrat, Yates, Fisher sau McNemar.
7. Pentru trei sau mai multe grupuri: test global, apoi comparații post hoc cu corecție.
8. Stabilește dinainte mărimea efectului pe care o vei raporta (diferența de medii cu IC 95%, d Cohen, r_rb, RR, OR, φ, V).

8.3 Tabelul-sinteză al testelor

Tabelul 8.1 adună în același loc situațiile, testele și locul în care le găsiți în programele folosite la lucrările practice (JASP 0.98.1 [47], MedCalc 24.0 [50], Epi Info 7.2.7 [51]). Denumirile meniurilor pot diferi ușor între versiuni; în lucrările practice 6, 7 și 8 pașii sunt descriși în detaliu.

Tabelul 8.1. Sinteza testelor statistice uzuale și a meniurilor din programe.

Situația	Variabilă numerică, distribuție normală (parametric)	Variabilă ordinală sau numerică asimetrică (neparametric)	Variabilă nominală (categorial)	JASP	MedCalc	Epi Info
1 eșantion față de o valoare de referință	t pentru un eșantion	Wilcoxon pentru un eșantion	testul binomial; chi-pătrat de concordanță	T-Tests > One Sample T-Test; Frequencies > Binomial Test, Multinomial Test	One sample t-test; Signed rank sum test	StatCalc: Binomial
2 grupuri independente	t nepereche (Student; Welch la varianțe inegale)	Mann-Whitney U	chi-pătrat 2 x 2; Fisher la E < 5	T-Tests > Independent Samples T-Test (Student, Welch, Mann-Whitney); Frequencies >	Independent samples t-test; Mann-Whitney test; Chi-squared test; Fisher's exact test	Classic Analysis: MEANS cu variabilă de grupare; TABLES; StatCalc:

Situația	**Variabilă numerică, distribuție normală (parametric)**	**Variabilă ordinală sau numerică asimetrică (neparametric)**	**Variabilă nominală (categorial)**	**JASP**	**MedCalc**	**Epi Info**
				Contingency Tables		Tables 2 x 2
2 măsurători pereche	t pereche	Wilcoxon (semne și ranguri); testul semnului	McNemar	T-Tests > Paired Samples T-Test (Student, Wilcoxon signed-rank)	Paired samples t-test; Wilcoxon test (paired samples); McNemar test	StatCalc: perechi potrivite (McNemar)
≥ 3 grupuri independente	ANOVA unifactorială + Tukey (Welch la varianțe inegale)	Kruskal-Wallis + Dunn	chi-pătrat r x c	ANOVA > ANOVA (Post Hoc Tests; Nonparametrics: Kruskal-Wallis, Dunn)	One-way analysis of variance; Kruskal-Wallis test	MEANS (ANOVA și Kruskal-Wallis); TABLES
≥ 3 măsurători repetate	ANOVA cu măsurători repetate	Friedman + Wilcoxon cu corecție	Q Cochran	ANOVA > Repeated Measures ANOVA (Nonparametrics: Friedman)	Repeated measures analysis of variance; Friedman test; Cochran's Q test	nu există o procedură dedicată
proporție față de categorii ordonate	-	Mann-Whitney sau Kruskal-Wallis pe variabila ordinală	chi-pătrat pentru trend	Frequencies > Contingency Tables	Chi-squared test for trend	StatCalc: Chi square for trend
asocierea a două variabile numerice sau ordinale	corelația Pearson	corelația Spearman, Kendall	-	Regression > Correlation	Correlation; Rank correlation	Classic Analysis: REGRESS
asocierea a două variabile nominale	-	-	chi-pătrat, Fisher; φ, V Cramér; RR, OR cu IC 95%	Frequencies > Contingency Tables (Phi and Cramér's V; Log odds ratio)	Chi-squared test; Relative risk; Odds ratio	TABLES; StatCalc: Tables 2 x 2
acordul a două metode sau evaluatori	Bland-Altman; coeficientul de corelație intraclasă	kappa ponderat	kappa Cohen	modulul Reliability	Bland-Altman plot; Intraclass correlation; Inter-rater agreement (kappa)	-
predicția unui rezultat numeric	regresie liniară simplă sau multiplă	-	-	Regression > Linear Regression	Multiple regression	Classic Analysis: REGRESS
predicția unui rezultat binar	regresie logistică; curba ROC	-	-	Regression > Logistic Regression	Logistic regression; ROC curve analysis	Classic Analysis: LOGISTIC

Situația	Variabilă numerică, distribuție normală (parametric)	Variabilă ordinală sau numerică asimetrică (neparametric)	Variabilă nominală (categorial)	JASP	MedCalc	Epi Info
timpul până la eveniment	Kaplan-Meier, log-rank, regresie Cox	-	-	modulul Survival	Kaplan-Meier survival analysis; Cox proportional-hazards regression	Classic Analysis: Kaplan-Meier, Cox

Un student care stăpânește primele cinci rânduri ale tabelului poate rezolva aproape toate problemele de la examenul practic. Rândurile pentru acord, predicție și supraviețuire sunt detaliate în capitolele 9-13.

8.4 Unitatea de analiză și independența observațiilor

Toate testele din tabelul 8.1 presupun că observațiile sunt independente (sau, la testele pereche, că perechile sunt independente între ele). Presupunerea pare banală, dar este încălcată des, mai ales când un pacient contribuie cu mai multe valori.

Unitatea de analiză este entitatea pe care s-a făcut randomizarea sau eșantionarea și despre care vrem să tragem concluzia: de regulă pacientul. Problemele apar când analizăm altceva:

- **ochi, dinți, articulații, leziuni**: dacă un pacient contribuie cu doi ochi, valorile celor doi ochi sunt corelate. În studiul colegilor despre biometria oculară, 170 de ochi au provenit de la 102 pacienți [81]. În stomatologie, suprafețele aceluiași dinte și dinții aceluiași pacient se grupează (efect de cluster) [104]; a trata 600 de dinți de la 30 de pacienți ca pe 600 de observații independente umflă artificial n și produce valori p prea mici.
- **vizite repetate**: un pacient cu 5 internări într-un an nu înseamnă 5 pacienți. Dacă analizăm internările, trebuie folosite metode pentru date repetate.
- **clustere**: în mg2026, cei 160 de pacienți provin din 12 cabinete de medicină de familie. Pacienții aceluiași medic pot semăna între ei (același stil de consiliere, aceeași zonă). Dacă randomizarea s-ar fi făcut pe cabinete (toți pacienții unui cabinet în același braț), unitatea de analiză ar fi devenit cabinetul, iar analiza ar fi trebuit să țină cont de această grupare. În scenariul nostru randomizarea a fost individuală, deci testele obișnuite sunt acceptabile, dar o analiză de sensibilitate ar putea include cabinetul ca factor.
- **evenimente în loc de persoane**: 30 de episoade de hipoglicemie pot proveni de la 30 de pacienți sau de la 3 pacienți. Un chi-pătrat pe episoade ar încălca independența.

Regula practică: înainte de orice test, întreabă-te "câte unități independente am, de fapt?". Răspunsul este n-ul real al analizei.

8.5 Testul se stabilește înainte de a vedea datele

Un test statistic își păstrează eroarea de tip I la 5% doar dacă a fost ales independent de rezultat. Dacă încercăm pe aceleași date mai multe teste (t, Welch, Mann-Whitney, cu și fără valori extreme, cu și fără transformare) și îl raportăm pe cel care "iese", probabilitatea reală de rezultat fals pozitiv crește mult peste 5%, chiar dacă fiecare test, luat separat, este corect [70,74]. Fenomenul este cunoscut în literatură drept "grădina cărărilor care se bifurcă": fiecare decizie aparent nevinovată deschide o nouă șansă de a obține $p < 0,05$ din întâmplare.

Setul mg2026 oferă un exemplu concret, prezentat în figura 8.2. Una dintre variabilele secundare este scăderea tensiunii sistolice la 6 luni. Mediile sunt 2,79 mmHg (AS 8,96) în brațul

A, 4,53 (AS 9,63) în brațul B și 6,96 (AS 8,18) în brațul C. Distribuțiile sunt compatibile cu normala în fiecare braț (SW p între 0,07 și 0,43), iar variațele sunt omogene (Levene p = 0,67). Testul potrivit, stabilit în planul de analiză, este ANOVA clasică.

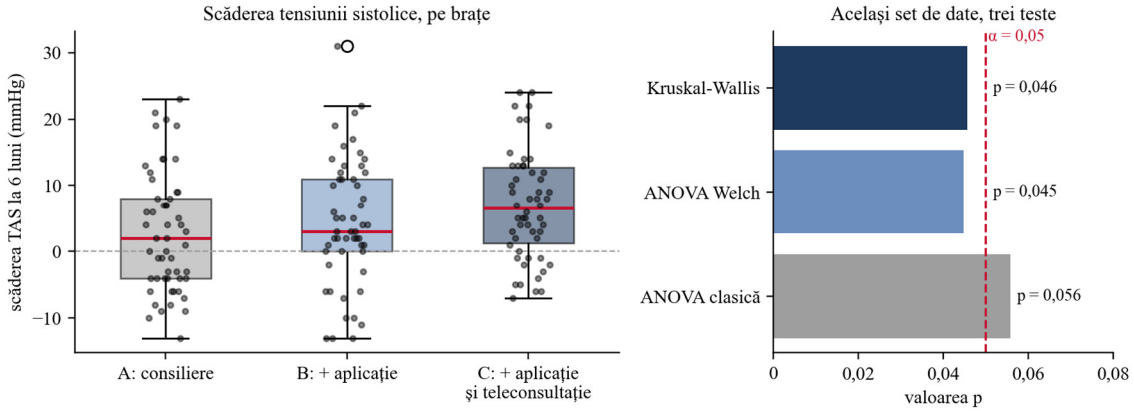


Figura 8.2. Stânga: scăderea tensiunii arteriale sistolice la 6 luni, pe brațe (cutii cu mediana și quartilele, puncte individuale). Dreapta: valorile p ale trei teste aplicate aceloraşi date: ANOVA clasică 0,056, ANOVA Welch 0,045, Kruskal-Wallis 0,046. Linia roșie marchează pragul 0,05. Date sintetice (setul mg2026).

ANOVA clasică dă F(2;157) = 2,94, p = 0,056. Varianta Welch dă p = 0,045, iar Kruskal-Wallis dă p = 0,046. Un cercetător care alege testul după rezultat va raporta "p < 0,05" și va scrie că teleconsultația scade tensiunea. Un cercetător care respectă planul va raporta p = 0,056, η^2 = 0,036, și va spune că datele sugerează o scădere mai mare în brațul C (diferența C-A de 4,17 mmHg, IC 95% Tukey 0,08-8,26), dar că testul global nu a atins pragul prestabilit. A doua formulare este cea corectă. Diferența dintre 0,045 și 0,056 este clinic neglijabilă; problema nu ține de cifre, ci de onestitatea procedurii.

Soluțiile sunt simple și fac parte din buna practică: testul principal, variabila principală și pragul se scriu în protocol înainte de începerea studiului; analizele suplimentare sunt numite explicit "analize de sensibilitate" sau "exploratorii"; iar dacă o condiție de aplicare se dovedește încălcată, regula de rezervă ("dacă Levene p < 0,05, folosim Welch") este scrisă și ea dinainte. Ghidurile de raportare ale studiilor clinice cer ca modificările față de planul inițial să fie declarate [56].

8.6 Greșeli frecvente în alegerea și aplicarea testului

8.6.1 Test pentru grupuri independente aplicat pe date pereche

Am văzut în secțiunea 8.2.4 efectul pe glicemia din mg2026: t = 5,03 în loc de t = 11,14. Aceeași greșeală apare la variabilele binare: un chi-pătrat obișnuit pe un tabel "moment (0 sau 6 luni) x glicemie ≥ 100 (da/nu)" cu 320 de observații, în loc de testul McNemar pe 160 de pacienți. Semnalul de alarmă în citirea unui articol: aceiași pacienți măsurați de două ori și cuvântul "independent" sau testul chi-pătrat Pearson în secțiunea de metode.

8.6.2 Mai multe teste t în loc de ANOVA

Cu trei brațe, tentația este să comparăm A cu B, A cu C și B cu C prin trei teste t. Fiecare comparație are 5% risc de rezultat fals pozitiv, iar riscul global ajunge la 14,3%. Procedura corectă: ANOVA (sau Kruskal-Wallis), apoi comparații post hoc cu corecție (Tukey, Bonferroni, Holm, Dunn). Excepția legitimă este o comparație principală stabilită dinainte (de exemplu C față de A ca obiectiv primar), care se poate testa direct.

8.6.3 Chi-pătrat cu frecvențe așteptate mici

În capitolul 7, pe un tabel cu 18 pacienți și frecvențe așteptate de 3,1 și 3,9, chi-pătratul a dat p = 0,040, iar testul corect (Fisher) p = 0,066. Programele afișează uneori un avertisment discret ("x celule au frecvența așteptată sub 5"), pe care mulți îl ignoră. Verificarea frecvențelor așteptate face parte din rezolvare și se scrie în formatul a-g, la justificarea metodei.

8.6.4 Alegerea testului după valoarea p

Exemplul din secțiunea 8.5 este tipic. Variante frecvente ale aceleiași greșeli: trecerea la un test unilateral după ce testul bilateral a dat p = 0,08; eliminarea "valorilor extreme" doar până când rezultatul devine semnificativ; schimbarea variabilei principale cu una secundară care a ieșit semnificativă; testarea a 20 de subgrupuri și raportarea celui cu p < 0,05.

8.6.5 Media scorurilor ordinale tratată ca măsurătoare

Satisfacția din mg2026 are mediile 2,96 (A), 3,62 (B) și 3,98 (C). ANOVA pe aceste scoruri dă F = 17,32, p < 0,001, iar Kruskal-Wallis H = 32,16, p < 0,001: concluzia nu se schimbă. Problema nu este, aici, valoarea p, ci interpretarea: "satisfacția medie a crescut cu 1,02 puncte" presupune că diferența dintre "neutru" și "mulțumit" este egală cu cea dintre "mulțumit" și "foarte mulțumit", ceea ce scala nu garantează. Raportarea corectă folosește mediane și distribuția pe categorii (figura 7.2 din capitolul 7). Pe eșantioane mici sau la distribuții cu efect de plafon, media scorurilor ordinale poate induce și concluzii statistice greșite.

8.6.6 Alte greșeli pe care le întâlnim des

Testarea normalității pe întregul eșantion în loc de fiecare grup (sau pe diferențe, la datele pereche) este o greșeală tehnică frecventă: două grupuri normale cu medii diferite formează împreună o distribuție bimodală, iar testul de normalitate "eșuează" fără motiv. O altă greșeală, specifică studiilor randomizate, este testarea statistică a diferențelor de la includere între brațe (tabelul 1 cu valori p). Dacă randomizarea a fost corectă, orice diferență la includere este, prin definiție, întâmplătoare; ghidul CONSORT recomandă descrierea, nu testarea lor [56]. În mg2026, nivelul de activitate fizică la includere are p = 0,08 între brațe (Kruskal-Wallis), iar statutul de fumător p = 0,07 (chi-pătrat): ambele sunt fluctuații ale randomizării, care se discută prin mărimea lor clinică, nu prin valoarea p. Ultima greșeală frecventă este ignorarea testărilor multiple: un articol care raportează 7 odds ratio pentru 7 tulburări diferite, fiecare la pragul 0,05, are o probabilitate de aproximativ 30% de a obține cel puțin un rezultat fals pozitiv chiar dacă nicio asociere nu ar fi reală.

8.7 Raportarea unui rezultat în format standardizat (a-g)

La temele și la examenul practic al disciplinei, fiecare problemă se rezolvă în același format, cu șapte componente. Formatul nu este o formalitate birocratică: el obligă la ordinea corectă a gândirii (întâi întrebarea, apoi ipotezele, apoi metoda și abia la urmă calculul) și face ca rezolvarea să poată fi verificată de altcineva.

- **(a) Ipoteza de cercetare**: ce crede cercetătorul, în limbaj clinic. "Pacienții din brațul C au o scădere mai mare a glicemiei decât cei din brațul A."
- **(b) Ipotezele statistice**, în cuvinte și matematic: H0 și H1, cu parametrul populației (μ, Me, p) și cu precizarea dacă testul este bilateral.
- **(c) Metoda aleasă și justificarea**: tipul variabilelor, designul (independent sau pereche, numărul de grupuri), verificarea condițiilor (normalitate pe grupuri, Levene, frecvențe așteptate) și alternativa care ar fi fost folosită dacă o condiție nu era îndeplinită.

- **(d) Modul de rezolvare și programul**: programul și meniul (de exemplu JASP, T-Tests > Independent Samples T-Test, cu opțiunile bifate), sau calculul de mână pe pași.
- **(e) Rezultatul, cu grafic**: statisticile descriptive pe grupuri, statistica testului, gradele de libertate, valoarea p, mărimea efectului cu IC 95% și graficul potrivit (box-plot sau grafic de medii cu IC pentru variabile numerice, grafic cu bare pentru proporții).
- **(f) Concluzia statistică**: "p < 0,05, deci respingem H0" sau "p ≥ 0,05, deci nu respingem H0", cu formularea "există / nu există o diferență semnificativă statistic între...".
- **(g) Concluzia de cercetare**, în cuvinte: ce înseamnă rezultatul pentru întrebarea clinică, cu mărimea efectului și cu limitele designului.

Tabelul 8.2 dă un exemplu complet, pe întrebarea principală a studiului mg2026.

Tabelul 8.2. Model de rezolvare în formatul a-g: scăderea glicemiei la 6 luni pe brațe (setul mg2026).

Componenta	Conținut
(a) Ipoteza de cercetare	Scăderea glicemiei a jeun la 6 luni diferă în funcție de intervenția primită (consiliere, consiliere + aplicație, consiliere + aplicație + teleconsultație).
(b) Ipotezele statistice	H0: scăderea medie a glicemiei este aceeași în cele trei brațe (μA = μB = μC). H1: cel puțin o medie diferă de celelalte.
(c) Metoda și justificarea	Variabila rezultat (scăderea glicemiei, mg/dL) este numerică; variabila de grupare (brațul) este nominală cu 3 categorii; grupurile sunt independente. Distribuțiile sunt aproximativ normale în fiecare braț, variațele omogene (Levene p = 0,163), deci ANOVA unifactorială, urmată de testul Tukey. Alternativa la încălcarea condițiilor: Kruskal-Wallis cu Dunn.
(d) Rezolvarea și programul	JASP 0.98.1, ANOVA > ANOVA; Dependent Variable: scăderea glicemiei; Fixed Factors: brat; Assumption Checks: Homogeneity tests; Post Hoc Tests: Tukey; Estimates of effect size: η².
(e) Rezultatul	A: $3{,}0 \pm 8{,}8 (n\,53)$; B: $11{,}1 \pm 11{,}2 (n\,53)$; C: $16{,}3 \pm 10{,}6 (n\,54)$; F(2;157) = 22,93, p < 0,001, η² = 0,226. Tukey: B-A 8,1 (IC 95% 3,4-12,8; p = 0,0002), C-A 13,3 (8,6-18,0; p < 0,001), C-B 5,2 (0,6-9,9; p = 0,024). Grafic: box-plot al scăderii glicemiei pe brațe, ca în capitolul 6.
(f) Concluzia statistică	p < 0,001 < 0,05, deci respingem H0: există diferențe semnificative statistic între brațe; toate cele trei comparații două câte două sunt semnificative.
(g) Concluzia de cercetare	Adăugarea aplicației mobile și, în plus, a teleconsultației lunare a dus la o scădere mai mare a glicemiei a jeun la 6 luni (în medie cu 8 și, respectiv, 13 mg/dL mai mult decât consilierea simplă). Efectul este mare (η² = 0,226). Datele sunt simulate, iar într-un studiu real ar trebui verificată și relevanța clinică a unei scăderi de această mărime.

8.8 Probleme rezolvate

Problemele de mai jos folosesc setul mg2026 (fișierul mg2026_curat.csv de pe Moodle) și urmează stilul subiectelor de examen și al fișelor "Teste statistice - recapitulare" și "Probleme BPOC" ale disciplinei. Pentru fiecare problemă sunt date întrebarea, tipul variabilelor, designul, testul ales cu alternativa, ipotezele, rezultatul și concluziile. Toate cifrele au fost calculate pe setul

de date; le puteți reproduce în JASP sau MedCalc. Pragul de semnificație este α = 0,05, iar testele sunt bilaterale.

Problema rezolvată 8.1. Glicemia a jeun la pacienții din mediul urban şi rural

Întrebarea. Diferă glicemia a jeun la includere între pacienții din mediul urban și cei din mediul rural?

Variabile și design. Rezultat: glicemia la includere (numerică, mg/dL). Grupare: mediul (nominală, 2 categorii). Două grupuri independente.

Testul ales și alternativa. Shapiro-Wilk pe grupuri: urban p = 0,076, rural p = 0,054; grupuri mari (95 și 65). Levene p = 0,117. Test t nepereche (raportăm varianta Welch, care nu presupune varianțe egale); alternativa: Mann-Whitney U. În JASP: T-Tests > Independent Samples T-Test, cu Student, Welch, Mann-Whitney și Assumption checks bifate.

Ipoteze. H0: μurban = μrural; H1: μurban ≠ μrural.

Rezultat. Urban $105{,}82 \pm 15{,}30$ mg/dL$(n\,95)$, rural $104{,}88 \pm 16{,}52$ mg/dL$(n\,65)$. Welch t = -0,37, gl 130,5, p = 0,715; diferența rural - urban -0,94 mg/dL (IC 95% -6,05 la 4,16); d Cohen -0,06. Mann-Whitney p = 0,81.

Concluzia statistică. p = 0,715 > 0,05, nu respingem H0: nu există o diferență semnificativă statistic.

Concluzia clinică. Glicemia a jeun a fost practic aceeași în cele două medii. Intervalul de încredere arată că o diferență reală mai mare de aproximativ 6 mg/dL în oricare direcție este puțin probabilă, ceea ce este mai informativ decât simplul "nesemnificativ".

Problema rezolvată 8.2. HDL-colesterolul la femei şi la bărbaţi

Întrebarea. Diferă HDL-colesterolul între femei și bărbați?

Variabile și design. Rezultat: HDL (numerică, mg/dL). Grupare: sexul (2 categorii). Grupuri independente.

Testul ales și alternativa. SW: femei p = 0,41, bărbați p = 0,48, distribuții normale. Levene p = 0,034, varianțe inegale, deci testul t Welch. Alternativa: Mann-Whitney.

Ipoteze. H0: μF = μM; H1: μF ≠ μM.

Rezultat. Femei $53{,}92 \pm 11{,}70$ mg/dL$(n\,86)$, bărbați $44{,}41 \pm 9{,}66$ mg/dL$(n\,74)$. Welch t = 5,63, gl 157,7, p < 0,001; diferența F - M 9,51 mg/dL (IC 95% 6,18-12,85); d = 0,88. Mann-Whitney p < 0,001.

Concluzia statistică. p < 0,001, respingem H0: diferența este semnificativă statistic.

Concluzia clinică. Femeile au HDL în medie cu aproximativ 9,5 mg/dL mai mare decât bărbații, un efect mare, în acord cu diferențele fiziologice cunoscute. Consecință practică: pragurile de HDL "scăzut" diferă pe sexe.

Problema rezolvată 8.3. IMC la femei şi la bărbaţi, o variabilă asimetrică

Întrebarea. Diferă IMC-ul la includere între femei și bărbați?

Variabile și design. Rezultat: IMC (numerică, kg/m²), cu asimetrie pozitivă (1,42 în tot eșantionul). Grupare: sexul. Grupuri independente.

Testul ales și alternativa. SW: femei p = 0,0001, bărbați p < 0,001; distribuțiile sunt asimetrice. Alegem Mann-Whitney U, iar testul t Welch îl raportăm ca analiză de sensibilitate (grupurile sunt mari, deci și el este robust).

Ipoteze. H0: distribuțiile IMC sunt identice la femei și la bărbați (MeF = MeM); H1: distribuțiile diferă.

Rezultat. Femei: mediana 28,25 kg/m² (IQR 26,02-31,25), bărbați: 28,25 (IQR 26,32-31,50). Mann-Whitney U = 3020,5, p = 0,58; r_rb = 0,05. Welch p = 0,68.

Concluzia statistică. p = 0,58, nu respingem H0.

Concluzia clinică. IMC-ul nu diferă între sexe în acest eșantion; ambele teste duc la aceeași concluzie. Când grupurile sunt mari și testele concordă, alegerea dintre ele nu schimbă interpretarea; contează ca descrierea (mediană și IQR) să fie în acord cu testul ales.

Problema rezolvată 8.4. Scorul FINDRISC la pacienții cu și fără diabet

Întrebarea. Au pacienții cu diabet confirmat prin TTGO un scor FINDRISC mai mare decât cei fără diabet?

Variabile și design. Rezultat: FINDRISC (scor 0-26, tratat ca numeric). Grupare: diabetul (da/nu). Grupuri independente, foarte inegale (29 și 131).

Testul ales și alternativa. SW pe grupuri: p = 0,17 și 0,33; Levene p = 0,67. Test t nepereche; alternativa: Mann-Whitney.

Ipoteze. H0: μdiabet = μfără diabet; H1: μdiabet ≠ μfără diabet.

Rezultat. Fără diabet 9,40 ± 3,70 (mediana 10), cu diabet 14,76 ± 3,96 (mediana 14). Student t = -6,97, p < 0,001 (Welch t = -6,67, gl 39,5, p < 0,001); diferența 5,36 puncte (IC 95% 3,74-6,99); d = 1,43. Mann-Whitney p < 0,001, r_rb = 0,71.

Concluzia statistică. Respingem H0.

Concluzia clinică. Scorul FINDRISC este în medie cu peste 5 puncte mai mare la pacienții cu diabet, un efect foarte mare. Dacă întrebarea ar fi fost "cât de bine identifică FINDRISC pacienții cu diabet?", scopul ar fi fost predicția, iar metoda potrivită ar fi fost curba ROC (AUC 0,852 în mg2026), nu testul t. Același set de date, două întrebări, două metode diferite.

Problema rezolvată 8.5. Tensiunea sistolică la pacienții cu și fără diabet, când testele nu concordă

Întrebarea. Diferă TAS la includere între pacienții cu și fără diabet?

Variabile și design. Rezultat: TAS (numerică, mmHg). Grupare: diabetul (da/nu). Grupuri independente (29 și 131).

Testul ales și alternativa. SW: p = 0,19 și 0,20; Levene p = 0,53. Testul stabilit dinainte este t nepereche; alternativa, Mann-Whitney.

Ipoteze. H0: μdiabet = μfără diabet; H1: μdiabet ≠ μfără diabet.

Rezultat. Fără diabet 129,96 ± 13,86 mmHg (mediana 131), cu diabet 136,41 ± 12,92 mmHg (mediana 132). Student t = -2,30, p = 0,023 (Welch t = -2,40, p = 0,021); diferența 6,45 mmHg (IC 95% 1,04-11,87); d = 0,47. Mann-Whitney p = 0,065.

Concluzia statistică. Conform testului ales dinainte (t nepereche), p = 0,023 < 0,05, respingem H0.

Concluzia clinică. Pacienții cu diabet au avut TAS în medie cu aproximativ 6,5 mmHg mai mare, un efect mic spre mediu, estimat imprecis (IC de la 1 la 12 mmHg). De ce dă Mann-Whitney p = 0,065? Medianele sunt foarte apropiate (131 și 132), iar diferența de medii vine mai ales din coada superioară a distribuției la diabetici. Testele răspund la întrebări ușor diferite. Nu avem voie să alegem testul după rezultat; dacă protocolul prevedea testul t și condițiile lui sunt îndeplinite, raportăm testul t, iar discordanța o menționăm ca analiză de sensibilitate.

Problema rezolvată 8.6. Greutatea la includere și la 6 luni, pe tot eșantionul

Întrebarea. S-a schimbat greutatea pacienților între includere și 6 luni?

Variabile și design. Rezultat: greutatea (numerică, kg), măsurată de două ori la aceiași 160 de pacienți. Date pereche.

Testul ales și alternativa. Pentru date pereche se verifică distribuția diferențelor: SW p = 0,86, normală. Test t pereche; alternativa: Wilcoxon. În JASP: T-Tests > Paired Samples T-Test, perechea greutate0 - greutate6.

Ipoteze. H0: media diferențelor $\mu d = 0$; H1: $\mu d \neq 0$.

Rezultat. Greutate medie 83,05 kg la includere și 80,74 kg la 6 luni. Diferența medie 2,31 kg (AS 2,66), IC 95% 1,90-2,73; t = 11,00, gl 159, p < 0,001; dz = 0,87. Wilcoxon p < 0,001.

Concluzia statistică. Respingem H0.

Concluzia clinică. Pacienții au slăbit în medie 2,3 kg în 6 luni. Testul pereche pe tot eșantionul arată că s-a produs o schimbare, dar nu ce a produs-o: fără compararea brațelor, nu putem separa efectul intervențiilor de efectul participării la studiu. Întrebarea despre eficacitate se rezolvă comparând scăderea în greutate între brațe (ANOVA, F = 18,25, p < 0,001), adică problema L5 din lucrarea practică 8.

Problema rezolvată 8.7. Tensiunea sistolică la includere și la 6 luni

Întrebarea. A scăzut TAS între includere și 6 luni?

Variabile și design. TAS, numerică, aceiași pacienți la două momente: date pereche.

Testul ales și alternativa. Diferențele au SW p = 0,13. Test t pereche; alternativa: Wilcoxon.

Ipoteze. H0: $\mu d = 0$; H1: $\mu d \neq 0$.

Rezultat. TAS medie 131,13 mmHg la includere și 126,36 mmHg la 6 luni; diferența medie 4,78 mmHg (AS 9,05), IC 95% 3,36-6,19; t = 6,68, gl 159, p < 0,001; dz = 0,53. Wilcoxon p < 0,001.

Concluzia statistică. Respingem H0.

Concluzia clinică. TAS a scăzut în medie cu aproximativ 4,8 mmHg. La nivel populațional, o scădere de această mărime este considerată relevantă pentru riscul cardiovascular, dar și aici lipsa unui grup de comparație netratat limitează interpretarea cauzală (regresia spre medie și efectul de măsurare repetată pot contribui).

Problema rezolvată 8.8. Scăderea glicemiei pe cele trei brațe

Întrebarea. Diferă scăderea glicemiei la 6 luni între brațele A, B și C?

Variabile și design. Rezultat numeric, grupare nominală cu 3 categorii, grupuri independente.

Testul ales și alternativa. ANOVA unifactorială cu Tukey (Levene p = 0,163); alternativa: Kruskal-Wallis cu Dunn.

Ipoteze. H0: $\mu A = \mu B = \mu C$; H1: cel puțin o medie diferă.

Rezultat. Rezolvarea completă este în tabelul 8.2: F(2;157) = 22,93, p < 0,001, $\eta^2 = 0,226$; Tukey semnificativ pentru toate cele trei perechi. Kruskal-Wallis H = 36,63, p < 0,001.

Concluzia statistică. Respingem H0.

Concluzia clinică. Intervențiile cu aplicație au produs scăderi mai mari ale glicemiei, iar efectul crește odată cu intensitatea intervenției.

Problema rezolvată 8.9. Tensiunea sistolică după statutul de fumător

Întrebarea. Diferă TAS la includere între nefumători, foşti fumători şi fumători?

Variabile şi design. Rezultat: TAS (numerică). Grupare: fumatul (nominală, 3 categorii). Grupuri independente.

Testul ales şi alternativa. SW pe grupuri: 0,37, 0,86, 0,23; Levene p = 0,114. ANOVA cu Tukey; alternativa: Kruskal-Wallis.

Ipoteze. H0: μnefumători = μfoşti fumători = μfumători; H1: cel puţin o medie diferă.

Rezultat. Foşti fumători $127{,}55 \pm 14{,}39 (n\,33)$, fumători $137{,}04 \pm 14{,}63 (n\,53)$, nefumători $128{,}50 \pm 11{,}76 (n\,74)$. F(2;157) = 7,85, p < 0,001 (p = 0,0006), η² = 0,091. Tukey: fumători - foşti fumători 9,49 mmHg (IC 95% 2,50-16,48; p = 0,005); fumători - nefumători 8,54 mmHg (IC 95% 2,87-14,21; p = 0,001); nefumători - foşti fumători 0,95 mmHg (p = 0,94). Kruskal-Wallis H = 12,77, p = 0,002.

Concluzia statistică. Respingem H0; fumătorii diferă de ambele celelalte grupuri.

Concluzia clinică. Fumătorii actuali au TAS cu aproximativ 9 mmHg mai mare decât nefumătorii şi foştii fumători, iar foştii fumători nu diferă de nefumători. Rezultatul este în acord cu regresia multiplă din capitolul 9, în care fumatul rămâne asociat cu TAS (+6,74 mmHg faţă de foştii fumători) după ajustarea pentru IMC, vârstă şi sex. Fiind o analiză transversală, ea nu stabileşte singură cauzalitatea.

Problema rezolvată 8.10. Scăderea tensiunii sistolice pe braţe, un rezultat la limită

Întrebarea. Diferă scăderea TAS la 6 luni între cele trei braţe?

Variabile şi design. Rezultat numeric, trei grupuri independente.

Testul ales şi alternativa. SW pe braţe 0,07-0,43; Levene p = 0,67. Planul de analiză prevede ANOVA clasică; alternativa în caz de varianţe inegale: Welch; în caz de asimetrie: Kruskal-Wallis.

Ipoteze. H0: μA = μB = μC; H1: cel puţin o medie diferă.

Rezultat. A 2,79 ± 8,96; B 4,53 ± 9,63; $C\ 6{,}96 \pm 8{,}18$ mmHg. F(2;157) = 2,94, p = 0,056, η² = 0,036 (figura 8.2).

Concluzia statistică. p = 0,056 > 0,05, nu respingem H0.

Concluzia clinică. Datele sugerează o scădere mai mare a TAS în braţul C, dar testul global prestabilit nu atinge pragul de semnificaţie. Faptul că Welch (p = 0,045) şi Kruskal-Wallis (p = 0,046) trec sub 0,05 nu schimbă concluzia, pentru că nu acestea erau testele planificate, iar condiţiile ANOVA clasice sunt îndeplinite. Rezultatul este un semnal care merită investigat ca obiectiv principal într-un studiu viitor. Comparaţiile post hoc nu se interpretează ca dovadă după un test global nesemnificativ.

Problema rezolvată 8.11. Satisfacţia după nivelul de educaţie

Întrebarea. Diferă satisfacţia la 6 luni după nivelul de educaţie?

Variabile şi design. Rezultat: satisfacţia (ordinală 1-5). Grupare: educaţia (ordinală, 4 niveluri). Grupuri independente.

Testul ales şi alternativa. Rezultat ordinal, 4 grupuri: Kruskal-Wallis. Pentru că şi educaţia este ordonată, o întrebare complementară, "creşte satisfacţia odată cu educaţia?", se testează mai eficient cu corelaţia Spearman.

Ipoteze. H0: distribuţia satisfacţiei este aceeaşi în cele 4 niveluri de educaţie; H1: cel puţin un nivel diferă.

Rezultat. Medianele: gimnaziu 3 (n 20), liceu sau școală profesională 4 (n 57), universitar 3 (n 64), postuniversitar 4 (n 19). Kruskal-Wallis H = 4,12, gl 3, p = 0,25. Spearman rho = 0,125, p = 0,115.

Concluzia statistică. Nu respingem H0 (nici pentru diferență, nici pentru trend).

Concluzia clinică. Satisfacția nu diferă semnificativ după nivelul de educație. Absența diferenței este un rezultat util pentru implementare: programul pare să fie acceptat la fel de bine de pacienți cu niveluri diferite de educație, deși grupurile extreme (20 și 19 pacienți) sunt mici.

Problema rezolvată 8.12. Activitatea fizică la includere pe brațe: are sens testul?

Întrebarea. Diferă activitatea fizică autoevaluată (1-5) la includere între brațe?

Variabile și design. Rezultat ordinal, trei grupuri independente, măsurat înainte de intervenție, într-un studiu randomizat.

Testul ales și alternativa. Tehnic, Kruskal-Wallis (alternativa: chi-pătrat pe tabelul 3 x 5, dar o celulă are frecvențe foarte mici, deoarece doar 3 pacienți au nivelul 5).

Ipoteze. H0: distribuția activității este aceeași în cele trei brațe; H1: cel puțin un braț diferă.

Rezultat. Mediana 3 în toate brațele (mediile codurilor 2,98, 2,58 și 2,61); Kruskal-Wallis H = 5,01, gl 2, p = 0,082.

Concluzia statistică. Nu respingem H0.

Concluzia clinică. Mai important decât rezultatul este faptul că întrebarea însăși este discutabilă. Într-un studiu randomizat corect, diferențele de la includere sunt întâmplătoare prin definiție, iar valoarea p nu spune nimic despre calitatea randomizării. Ce contează este dacă o diferență, oricât de mică, este suficient de mare clinic încât să influențeze rezultatul; dacă da, ea se ajustează în analiză (de exemplu prin regresie), indiferent de valoarea p [56].

Problema rezolvată 8.13. Glicemia la 0, 3 și 6 luni în brațul C

Întrebarea. S-a modificat glicemia a jeun în timp la pacienții din brațul C?

Variabile și design. Rezultat numeric, aceiași 54 de pacienți la 3 momente: măsurători repetate.

Testul ales și alternativa. În brațul C, glicemia la includere are o distribuție asimetrică (asimetria 1,05; SW p = 0,003), diferențele 0-3 luni nu sunt normale (SW p = 0,027), iar condiția de sfericitate a ANOVA cu măsurători repetate este încălcată (Mauchly p = 0,002). Alegem testul Friedman cu post hoc Wilcoxon și corecție Bonferroni; alternativa este ANOVA cu măsurători repetate cu corecția Greenhouse-Geisser.

Ipoteze. H0: distribuția glicemiei este aceeași la 0, 3 și 6 luni; H1: cel puțin un moment diferă.

Rezultat. Medianele: 100,5 / 89,0 / 85,0 mg/dL. Friedman $\chi^2 = 64{,}63$, gl 2, $p < 0{,}001$. Post hoc (Bonferroni): 0 față de 3 luni $p < 0{,}001$; 3 față de 6 luni p = 0,011; 0 față de 6 luni $p < 0{,}001$.

Concluzia statistică. Respingem H0; toate cele trei comparații rămân semnificative după corecție.

Concluzia clinică. În brațul C, glicemia a scăzut în primele 3 luni și a continuat să scadă până la 6 luni, cu o scădere totală a medianei de 15,5 mg/dL. Comparația cu brațele A și B arată că scăderea nu se explică doar prin trecerea timpului (vezi capitolul 7, secțiunea 7.6.3).

Problema rezolvată 8.14. Statutul de fumător după sex

Întrebarea. Este statutul de fumător asociat cu sexul?

Variabile și design. Două variabile nominale (sex: 2 categorii; fumat: 3 categorii), subiecți independenți: tabel 2 x 3.

Testul ales și alternativa. Chi-pătrat Pearson; cea mai mică frecvență așteptată este 15,26, condiția este îndeplinită. Alternativa la E mici: Fisher sau comasarea categoriilor.

Ipoteze. H0: fumatul și sexul sunt independente; H1: există o asociere.

Rezultat. Femei: foști fumători 16 (18,6%), fumătoare 30 (34,9%), nefumătoare 40 (46,5%). Bărbați: 17 (23,0%), 23 (31,1%), 34 (45,9%). $\chi^2 = 0,54$, gl 2, $p = 0,76$; V Cramér = 0,06.

Concluzia statistică. Nu respingem H0.

Concluzia clinică. Distribuția statutului de fumător este asemănătoare la femei și la bărbați în acest eșantion. În populația generală din România, prevalența fumatului diferă de obicei între sexe; un eșantion de pacienți cu risc cardiometabolic recrutați în cabinete nu reflectă neapărat populația generală.

Problema rezolvată 8.15. Diabetul după sex

Întrebarea. Diferă prevalența diabetului confirmat prin TTGO între femei și bărbați?

Variabile și design. Două variabile dihotomice, subiecți independenți: tabel 2 x 2.

Testul ales și alternativa. Frecvențele așteptate: cea mai mică 13,4, deci chi-pătrat Pearson; Fisher ca verificare.

Ipoteze. H0: $pF = pM$; H1: $pF \neq pM$.

Rezultat. Femei 15/86 (17,4%), bărbați 14/74 (18,9%). $\chi^2 = 0,06$, gl 1, $p = 0,81$ (Yates $p = 0,97$; Fisher $p = 0,84$).

Concluzia statistică. Nu respingem H0.

Concluzia clinică. Prevalența diabetului nedetectat anterior a fost practic aceeași la femei și la bărbați. În modelul de regresie logistică pentru diabet (obezitate, vârstă și sex ca predictori), sexul rămâne fără efect și după ajustare (OR 0,97, IC 95% 0,40-2,38).

Problema rezolvată 8.16. Vârsta de cel puțin 50 de ani și diabetul: ce variantă raportăm?

Întrebarea. Este vârsta ≥ 50 de ani asociată cu diabetul confirmat prin TTGO?

Variabile și design. Vârsta dihotomizată la 50 de ani (prag stabilit dinainte, cu sens clinic pentru screening) și diabetul (da/nu); tabel 2 x 2, subiecți independenți.

Testul ales și alternativa. Cea mai mică frecvență așteptată este 6,2 (34 x 29/160): condiția $E \geq 5$ este îndeplinită, dar la limită. Raportăm chi-pătrat, iar ca verificare testul Fisher.

Ipoteze. H0: $p \geq 50 = p < 50$; H1: $p \geq 50 \neq p < 50$.

Rezultat. Sub 50 de ani: 2/34 (5,9%); 50 de ani sau peste: 27/126 (21,4%). $\chi^2 = 4,36$, gl 1, $p = 0,037$; Yates $\chi^2 = 3,38$, $p = 0,066$; Fisher $p = 0,044$; $\varphi = 0,17$.

Concluzia statistică. Cu testul Pearson și cu testul Fisher, $p < 0,05$: respingem H0. Corecția Yates, mai conservatoare, dă $p = 0,066$.

Concluzia clinică. Diabetul a fost mai frecvent după 50 de ani, dar rezultatul depinde de doar 2 cazuri în grupul tânăr și este fragil. Vârsta tratată ca variabilă continuă în regresia logistică (OR 1,11 pe an, IC 95% 1,05-1,18) dă o imagine mai stabilă decât dihotomizarea. Lecția: dihotomizarea unei variabile numerice pierde informație.

Problema rezolvată 8.17. Scăderea în greutate ≥ 5% în brațul C față de brațul A

Întrebarea. Crește teleconsultația, adăugată la aplicație, proporția pacienților care slăbesc cel puțin 5%, comparativ cu consilierea simplă?

Variabile și design. Rezultat dihotomic (scădere ≥ 5% da/nu), două brațe randomizate independente (comparația principală stabilită în protocol).

Testul ales și alternativa. Chi-pătrat 2 x 2 (cea mai mică frecvență așteptată peste 10) și măsuri de risc: RR, diferența de riscuri, NNT.

Ipoteze. H0: pC = pA; H1: pC ≠ pA.

Rezultat. C: 16/54 (29,6%); A: 6/53 (11,3%). χ^2 = 5,49, gl 1, p = 0,019 (Yates 4,43, p = 0,035). RR = 2,62 (IC 95% 1,11-6,17); diferența absolută 18,3 puncte procentuale (IC 95% 3,4-33,2); NNT 5,5, rotunjit la 6 (IC 95% 3,0-29,1).

Concluzia statistică. Respingem H0.

Concluzia clinică. Pentru fiecare 6 pacienți tratați cu aplicație și teleconsultație în loc de consiliere simplă, încă un pacient atinge o scădere în greutate de cel puțin 5%. Intervalul de încredere al NNT este larg (de la 3 la 29), deci mărimea beneficiului este estimată imprecis.

Problema rezolvată 8.18. Obezitatea la includere și la 6 luni: McNemar

Întrebarea. S-a schimbat proporția pacienților obezi (IMC ≥ 30) între includere și 6 luni?

Variabile și design. Variabilă dihotomică măsurată de două ori pe aceiași pacienți: date pereche. IMC la 6 luni s-a calculat din greutatea la 6 luni și înălțime.

Testul ales și alternativa. Testul McNemar (cu corecție de continuitate sau varianta exactă, pentru că numărul perechilor discordante este mic).

Ipoteze. H0: proporția de obezi este aceeași la cele două momente; H1: proporția diferă.

Rezultat. Obezi la ambele momente 47; obezi doar la includere 7; obezi doar la 6 luni 4; neobezi la ambele 102. Proporția: 54/160 (33,8%) la includere, 51/160 (31,9%) la 6 luni. χ^2McNemar = (7 - 4)²/11 = 0,82, p = 0,37 (cu corecție 0,36, p = 0,55; exact p = 0,55).

Concluzia statistică. Nu respingem H0.

Concluzia clinică. Deși pacienții au slăbit în medie 2,3 kg (problema 8.6), clasificarea ca obez s-a schimbat la foarte puțini pacienți: o scădere modestă în greutate nu mută ușor un pacient peste pragul de 30 kg/m². Comparația cu problema 8.6 arată din nou cât de multă informație se pierde prin dihotomizare.

Problema rezolvată 8.19. IMC și tensiunea sistolică: asociere

Întrebarea. Este IMC-ul asociat cu TAS la includere?

Variabile și design. Două variabile numerice măsurate pe aceiași pacienți; scopul este asocierea.

Testul ales și alternativa. Corelația Pearson, cu verificarea diagramei de dispersie; alternativa, din cauza asimetriei IMC: Spearman.

Ipoteze. H0: ρ = 0 (nu există corelație liniară în populație); H1: ρ ≠ 0.

Rezultat. r = 0,405, p < 0,001; Spearman rho = 0,372, p < 0,001; $R^2 = 0{,}164$.

Concluzia statistică. Respingem H0.

Concluzia clinică. TAS crește odată cu IMC, cu o corelație moderată. IMC explică aproximativ 16% din variabilitatea TAS; restul de 84% ține de alți factori (vârstă, fumat, genetică, sare, medicație). Corelația nu arată singură cauzalitatea. Detalii în capitolul 9.

Problema rezolvată 8.20. Prezicerea tensiunii sistolice din IMC

Întrebarea. Ce TAS ne așteptăm să aibă un pacient cu IMC 30 kg/m²?

Variabile și design. Scopul este predicția unei variabile numerice dintr-o altă variabilă numerică.

Testul ales și alternativa. Regresie liniară simplă, cu verificarea reziduurilor; pentru mai mulți predictori, regresie multiplă.

Ipoteze. H0: panta $\beta = 0$; H1: $\beta \neq 0$.

Rezultat. TAS = 89,47 + 1,435 x IMC; IC 95% pentru pantă 0,925-1,944; $p < 0,001$; eroarea standard a estimării 12,73 mmHg. Predicție la IMC 30: 132,5 mmHg; la IMC 35: 139,7 mmHg.

Concluzia statistică. Respingem H0: panta este semnificativ diferită de 0.

Concluzia clinică. Fiecare unitate de IMC în plus se asociază, în medie, cu o TAS mai mare cu aproximativ 1,4 mmHg. Pentru un pacient individual, predicția este însă imprecisă: cu o eroare standard a estimării de 12,7 mmHg, TAS reală a unui pacient cu IMC 30 poate fi cu ușurință între 108 și 157 mmHg. Modelul descrie tendința medie, nu înlocuiește măsurarea tensiunii.

Problema rezolvată 8.21. Satisfacţia în braţul C faţă de braţul B

Întrebarea. Aduce teleconsultația, adăugată la aplicație, o satisfacție mai mare?

Variabile și design. Rezultat ordinal (1-5), două grupuri independente.

Testul ales și alternativa. Mann-Whitney U. Nu are rost testul t pe un scor cu 5 valori.

Ipoteze. H0: distribuția satisfacției este aceeași în brațele C și B; H1: distribuțiile diferă.

Rezultat. Mediana 4 (IQR 3-5) în C și 4 (IQR 3-4) în B; U = 1698, p = 0,079; r_rb = 0,19.

Concluzia statistică. Nu respingem H0.

Concluzia clinică. Satisfacția tinde să fie puțin mai mare în brațul C (31% foarte mulțumiți, față de 17% în B), dar diferența nu este semnificativă statistic, iar efectul estimat este mic. Pentru decizia de a oferi teleconsultație, criteriul principal ar trebui să fie rezultatul clinic și costul, nu satisfacția.

Problema rezolvată 8.22. Nivelul de educaţie după mediul de rezidenţă

Întrebarea. Diferă nivelul de educație între pacienții din mediul urban și cei din mediul rural?

Variabile și design. Educația este ordinală (4 niveluri), mediul este nominal (2 categorii); subiecți independenți.

Testul ales și alternativa. Două abordări sunt acceptabile: chi-pătrat pe tabelul 2 x 4 (ignoră ordinea nivelurilor) sau Mann-Whitney pe educație (folosește ordinea și este mai puternic dacă diferența este o deplasare în ansamblu).

Ipoteze. H0: distribuția nivelului de educație este aceeași în urban și rural; H1: distribuțiile diferă.

Rezultat. Rural: 7 (10,8%), 22 (33,8%), 30 (46,2%), 6 (9,2%); urban: 13 (13,7%), 35 (36,8%), 34 (35,8%), 13 (13,7%). Cea mai mică frecvență așteptată 7,7. $\chi^2 = 2,04$, gl 3, $p = 0,56$; V Cramér = 0,11. Mann-Whitney $p = 0,69$.

Concluzia statistică. Nu respingem H0 cu niciuna dintre abordări.

Concluzia clinică. Nivelul de educație este asemănător în cele două medii în acest eșantion. Este un rezultat care ar surprinde într-un sondaj populațional; aici, el reflectă probabil felul în care au fost recrutați pacienții în cabinete (eșantion de conveniență), nu realitatea demografică.

8.9 Recunoașterea testului greșit: zece descrieri de tip "Metode"

Fragmentele de mai jos imită formulări întâlnite în articole și în lucrări de licență. Sunt fictive, construite pentru exercițiu. Pentru fiecare, citiți întâi descrierea și încercați să găsiți singuri problema, apoi comparați cu explicația.

Descrierea 1. "HbA1c a fost măsurată la includere și după 6 luni de tratament la 45 de pacienți. Valorile au fost comparate cu testul t Student pentru eșantioane independente (p = 0,09)." Problema: aceiași pacienți, două momente, deci date pereche. Testul pentru grupuri independente ignoră corelația dintre cele două măsurători și pierde putere. Corect: testul t pereche (dacă diferențele sunt aproximativ normale) sau Wilcoxon pentru date pereche.

Descrierea 2. "Durata spitalizării (mediana 6 zile, IQR 4-15) a fost comparată între cele două secții (n = 18 și n = 22) prin testul t Student." Problema: autorii descriu chiar ei variabila prin mediană și un IQR foarte asimetric (de la 4 la 15 în jurul unei mediane de 6), iar grupurile sunt mici. Descrierea și testul nu se potrivesc. Corect: Mann-Whitney U, cu raportarea medianelor pe grupuri.

Descrierea 3. "Tensiunea sistolică a fost comparată între cele patru clase de IMC (normoponderal, supraponderal, obezitate de gradul I, obezitate de gradul II-III) prin teste t două câte două; o valoare p < 0,05 a fost considerată semnificativă." Problema: patru grupuri înseamnă șase comparații, iar riscul global de eroare de tip I depășește 26%. Corect: ANOVA unifactorială (sau Kruskal-Wallis), urmată de comparații post hoc cu corecție; dacă interesează creșterea monotonă a TAS cu clasa de IMC, un test pentru trend sau o regresie cu IMC continuu.

Descrierea 4. "Complicațiile postoperatorii au apărut la 2 din 15 pacienți operați laparoscopic și la 7 din 14 pacienți operați clasic (chi-pătrat Pearson, p = 0,033)." Problema: frecvențele așteptate în coloana "complicații" sunt 4,66 și 4,34, ambele sub 5. Corect: testul exact Fisher, care dă aici p = 0,050, adică exact la limită și, strict, nesemnificativ. Concluzia "operația laparoscopică reduce semnificativ complicațiile" nu este susținută de date.

Descrierea 5. "Colesterolul total a fost comparat între nefumători (n = 40), foști fumători (n = 25) și fumători (n = 35) cu testul Friedman." Problema: testul Friedman este pentru măsurători repetate pe aceiași subiecți, nu pentru grupuri independente. Corect: ANOVA unifactorială sau, dacă distribuțiile sunt asimetrice, Kruskal-Wallis.

Descrierea 6. "Prevalența hipertensiunii necontrolate la 0, 6 și 12 luni la aceiași 80 de pacienți a fost comparată printr-un test chi-pătrat pe tabelul 3 x 2 (moment x control tensional)." Problema: tabelul conține 240 de "observații" care sunt, de fapt, 80 de pacienți numărați de trei ori. Corect: testul Q Cochran pentru o variabilă binară măsurată repetat (sau McNemar pentru perechi de momente, cu corecție).

Descrierea 7. "Acordul dintre tensiometrul automat și cel manual a fost excelent (corelația Pearson r = 0,93, p < 0,001)." Problema: corelația măsoară asocierea, nu acordul. Un aparat care dă sistematic cu 10 mmHg mai mult poate avea r = 0,99. Corect: metoda Bland-Altman (media diferențelor și limitele de acord de 95%) și, eventual, coeficientul de corelație intraclasă [102].

Descrierea 8. "Am analizat 412 dinți de la 52 de copii, comparând frecvența cariilor la dinții sigilați și nesigilați prin testul chi-pătrat." Problema: unitatea de analiză. Dinții aceluiași copil nu sunt independenți (aceeași igienă, aceeași dietă, aceeași salivă), deci n real este mult mai mic decât 412, iar valoarea p este prea optimistă [104]. Corect: metode care țin cont de grupare (de exemplu modele cu efect al pacientului) sau analiza la nivel de pacient.

Descrierea 9. "Testul Shapiro-Wilk aplicat întregului eșantion (n = 300, ambele grupuri reunite) a dat p = 0,03, de aceea am folosit testul Mann-Whitney." Problema: normalitatea se verifică în fiecare grup, nu pe grupurile reunite (dacă mediile diferă, amestecul lor nu este normal

chiar când fiecare grup este). În plus, la 150 de subiecți pe grup, testul t este robust la abateri moderate. Decizia trebuie luată pe baza formei distribuției în fiecare grup și stabilită dinainte.

Descrierea 10. "Într-o cohortă prospectivă urmărită până la 5 ani, cu durate de urmărire diferite de la un pacient la altul, mortalitatea a fost comparată între cei cu și fără diabet prin chi-pătrat, iar asocierea a fost exprimată ca odds ratio." Problema: când durata de urmărire variază, proporția simplă de decese ignoră cât timp a fost urmărit fiecare pacient și ce s-a întâmplat cu cei pierduți din urmărire. Corect: analiza de supraviețuire (curbe Kaplan-Meier, testul log-rank, regresia Cox cu hazard ratio și IC 95%).

8.10 Lectura critică a secțiunii Metode: articole ale colegilor din disciplină

8.10.1 Cum citim secțiunea Metode

Când citim un articol, aplicăm algoritmul din figura 8.1 "invers": pornim de la designul și variabilele declarate și verificăm dacă testul raportat este cel pe care l-am fi ales noi. Șase întrebări sunt suficiente pentru o primă lectură:

1. Care este designul (transversal, cohortă, caz-control, studiu clinic, înainte-după, comparare de metode)?
2. Care este variabila rezultat principală și de ce tip este?
3. Datele sunt independente sau pereche? Care este unitatea de analiză?
4. Testul numit (sau sugerat de modul de raportare) se potrivește cu tipul variabilei și cu designul?
5. Descrierea (medie ± AS sau mediană și IQR) se potrivește cu testul?
6. Este raportată o mărime a efectului cu interval de încredere, și este măsura de efect compatibilă cu designul (OR, RR, HR)?

Rezumatele au un spațiu limitat, iar multe detalii apar doar în textul complet. Exercițiul de mai jos lucrează exclusiv cu informația din rezumat, așa cum este ea înregistrată în catalogul bibliografiei acestui volum; observațiile critice sunt puncte de discuție, nu verdicte asupra articolelor.

8.10.2 Tabelul de lectură

Tabelul 8.3. Potrivirea dintre design, variabila rezultat și metoda statistică în articole ale colegilor din disciplină (pe baza rezumatelor).

Articolul	Designul și n	Variabila rezultat	Testul sau măsura raportată	Se potrivește?	Comentariu
Padurean și colab. 2026 [82]	transversal, 3 grupuri: 73, 86, 86	scorul stimei de sine (numeric)	ANOVA cu Tukey; d Cohen; η^2; Pearson; regresie multiplă	da	$F = 299{,}00$, $p < 0{,}001$, $\eta^2 = 0{,}712$; $d = 2{,}515$ pentru purtătorii de proteze auditive față de nepurtători; raportare model, cu mărimi ale efectului
Munteanu și colab. 2020 [94]	pilot controlat, măsurători repetate un an, n = 93	scoruri de aderență și control (ordinale); exacerbări (binar)	Mann-Whitney; Friedman cu Bonferroni; chi-pătrat 2 x 2	da	$\chi^2(1) = 13{,}707$, $p < 0{,}001$ pentru 10,30% față de 46,30%; ar fi util RR cu IC 95% și o mărime a

Articolul	Designul și n	Variabila rezultat	Testul sau măsura raportată	Se potrivește?	Comentariu
					efectului pentru Mann-Whitney
Gherbon și colab. 2024 [59]	retrospectiv înainte-după, 1 an, n = 200	glicemie, HbA1c, greutate, TAS (numerice, asimetrice)	mediane (IQR) înainte și după; testul nenumit	probabil da	design pereche: Wilcoxon este testul sugerat; medianele sunt numite "medii" în rezumat; lipsește mediana diferențelor individuale
Buda și colab. 2019 [86]	înainte-după un an de perindopril, cu lot martor; n nedeclarat în rezumat	trombospondina-1 serică (numerică)	comparație T0 față de T1, p < 0,001	parțial	8061,9 ± 3684,80 față de 15380 ± 5887 ng/mL: AS mare față de medie sugerează asimetrie (Wilcoxon sau logaritmare); comparația cu martorii (9221,03 ± 6510,21) cere un test pentru grupuri independente; n trebuie declarat
Gherbon și colab. 2021 [85]	studiu controlat cu placebo, 20 + 20, 3 luni	IMC, lipide, glicemie, TAD (numerice)	comparații între grupuri la 3 luni și în interiorul grupurilor înainte-după	parțial	rezumatul dă valori p din ambele tipuri de comparații; doar comparația între grupuri (ideal pe modificări sau prin ANCOVA) răspunde la întrebarea despre efect; la 20 de pacienți pe grup, verificarea distribuției contează
Danila și colab. 2024 [105]	declarat transversal, potrivire 168 + 168	severitate, evoluție (binar); indici inflamatori	ROC (AUC 0,624 și 0,752); hazard ratio 2,62 (IC 1,29-5,04)	nu în totalitate	HR presupune urmărire în timp (design de cohortă); procentele cu trei zecimale (70,312%) sugerează o precizie falsă
Ion și colab. 2024 [106]	cohortă prospectivă, n = 141	COVID-19 sever (binar)	ROC la două momente; regresie logistică; HR	parțial	regresia logistică produce OR, nu HR (SOFA HR 3,07, IC 2,29-

Articolul	Designul și n	Variabila rezultat	Testul sau măsura raportată	Se potrivește?	Comentariu
					4,12); trebuie precizat dacă s-a folosit regresie Cox cu timp până la eveniment
Cioca și colab. 2024 [8]	cohortă retrospectivă, 3 grupuri: 200, 62, 224	severitate, mortalitate (binar); durata internării (numerică)	proporții, medie ± AS, OR	parțial	OR 2,24 (ATI) și 2,46 (ventilație) fără IC și fără p; durata internării 11,6 ± 7,0 față de 9,1 ± 5,8 zile este asimetrică (mediană, Mann-Whitney); 3 grupuri cer întâi un test global
Stoichescu-Hogea și colab. 2021 [90]	3 grupuri: 62, 46, 40	strain longitudinal global (numeric)	comparația a 3 medii cu comparații două câte două; Spearman; ROC	da, cu o observație	AS diferite (1,4% la martori față de 3,0% la hipertensivi): ANOVA Welch sau post hoc Games-Howell ar fi fost de luat în calcul
Popa și colab. 2025 [100]	cohortă retrospectivă, 26 + 106	simptome, tratamente (binar)	Fisher; OR	da	grup mic și evenimente rare, deci Fisher este corect; lipsesc IC pentru OR (de exemplu OR = 3,97 pentru cefalee)
Piroș și colab. 2026 [6]	transversal, n = 693	prediabet (binar)	chi-pătrat pe 3 clustere (p = 0,003); trend; regresie logistică	da	categorii ordonate, trend raportat (p-trend < 0,001); clusterele sunt construite din aceleași date
Moleriu și colab. 2026 [4]	transversal, n = 921	exces ponderal (binar)	chi-pătrat, V Cramér; regresie logistică (aOR)	da	raportează tăria asocierii (V) alături de p; design transversal, deci asociere, nu cauzalitate
Miulescu și colab. 2026 [91]	pilot cu un singur braț, n = 21, 3 momente	scorurile POEM și SCORAD (numerice)	ANOVA cu măsurători repetate	da, cu limite	POEM 14,67 ± 5,22 la 7,10 ± 3,82; la 21 de copii, Friedman ar fi o analiză de sensibilitate utilă; fără grup de control,

Articolul	Designul și n	Variabila rezultat	Testul sau măsura raportată	Se potrivește?	Comentariu
					efectul nu se poate atribui terapiei
Popa și colab. 2025 [61]	diagnostic, 69 cu AVC și 16 fără	GFAP, UCH-L1 (numerice)	comparația mediilor; prag diagnostic	parțial	GFAP 142,91 ± 102,19 față de 37,76 ± 19,92 pg/mL; AS apropiată de medie arată asimetrie: mediană (IQR) și Mann-Whitney
Christodorescu și colab. 2026 [23]	cohortă prospectivă, 12 luni, n = 278	evenimentul primar în timp	Kaplan-Meier cu log-rank; Cox multivariabil	da	log-rank p = 0,027, dar după ajustare p = 0,240: exemplu de confuzie
Marc și colab. 2023 [39]	transversal repetat, 1401 față de 1698	prevalența VHC, VHB (binar); vârsta (numerică)	comparații de proporții și medii	da	prevalența VHC 16,9% față de 10,5%, p < 0,0001; vârsta 59,7 față de 61 de ani, p = 0,005: semnificativ statistic, dar clinic neglijabil

Câteva concluzii generale ies din tabel. Majoritatea articolelor folosesc teste potrivite cu designul. Problemele care apar cel mai des nu țin de alegerea testului principal, ci de raportare: testul nu este numit, lipsesc intervalele de încredere pentru OR, variabilele asimetrice sunt descrise prin medie ± AS, iar comparațiile în interiorul grupului sunt prezentate ca dovadă a efectului. Acestea sunt exact punctele pe care le verifică un recenzent și pe care le veți întâlni la redactarea propriei lucrări de licență.

> **Studiu de caz: o raportare model pentru trei grupuri**
>
> Pădurean și colaboratorii, cu dr. Frandeș și dr. Muntean, au comparat stima de sine la 245 de persoane: 73 cu auz normal, 86 de purtători de proteze auditive și 86 de persoane cu hipoacuzie care nu poartă proteze [82]. Rezultatul, un scor numeric, este comparat între trei grupuri independente, deci ANOVA unifactorială este alegerea firească, urmată de testul Tukey pentru comparațiile două câte două. Autorii raportează testul global ($F = 299{,}00$, $p < 0{,}001$), mărimea efectului globală ($\eta^2 = 0{,}712$) și mărimea efectului pentru comparația cu interes clinic: purtătorii de proteze au avut scoruri de 35,41 ± 5,32 față de 22,99 ± 4,53 la nepurtători ($p < 0{,}001$, d Cohen = 2,515). Asocierile cu scorurile de handicap auditiv și de tinitus sunt date prin corelația Pearson ($r = -0{,}573$ și $r = -0{,}443$), iar modelul de regresie multiplă explică aproximativ 65% din variabilitate. Este un exemplu complet de aplicare a formatului din acest capitol: test potrivit, test global apoi post hoc, mărimi ale efectului. Limita designului este cea a oricărui studiu transversal: nu știm dacă protezele cresc stima de sine sau dacă persoanele cu stimă de sine mai mare aleg mai ușor să poarte proteze.

8.10.3 Când designul și măsura de efect nu se potrivesc

Două dintre articolele din tabel ridică aceeași întrebare: este măsura de efect compatibilă cu designul declarat? Odds ratio, riscul relativ și hazard ratio nu sunt sinonime. OR se obține din

tabele 2 x 2 sau din regresia logistică și se poate calcula în orice design; RR cere un design în care incidența poate fi estimată (cohortă, studiu clinic); HR cere date de tip "timp până la eveniment", adică urmărirea pacienților în timp, și se obține de regulă prin regresia Cox.

Studiu de caz: hazard ratio într-un studiu transversal și într-o regresie logistică

Danila și colaboratorii, cu dr. Cioca, au evaluat indici inflamatori derivați din hemoleucogramă la 336 de pacienți cu COVID-19 (168 cu diabet, 168 fără diabet, potriviți după sex, IMC și severitate) [105]. Studiul este declarat transversal. Partea de acuratețe diagnostică este coerentă cu acest design: pentru dNLR, la pragul 2,685, sensibilitatea a fost 70,312%, specificitatea 65,978%, AUC = 0,624, iar pentru ALRI AUC = 0,752. Rezumatul raportează însă și un hazard ratio pentru SII la 7 zile, 2,62 (IC 1,29-5,04, p < 0,001). Un hazard ratio presupune pacienți urmăriți în timp și momente ale evenimentului, ceea ce nu se potrivește cu eticheta "transversal"; fie designul este, de fapt, o cohortă retrospectivă, fie măsura raportată este alta (de exemplu un OR). Merită observat și că un IC de la 1,29 la 5,04 este destul de larg pentru o valoare p atât de mică, un detaliu pe care un cititor l-ar verifica în textul complet. Ion și colaboratorii, tot cu dr. Cioca, au urmărit prospectiv 141 de pacienți cu diabet și COVID-19 și au comparat scorurile clinice prin curbe ROC (de exemplu SOFA AUC 0,771 la internare și 0,873 în ziua a cincea) [106]. Metodele menționează regresia logistică, dar rezultatele sunt date ca hazard ratio (SOFA HR 3,07, IC 95% 2,29-4,12; APACHE II HR 2,96, IC 95% 2,21-3,96). Regresia logistică produce odds ratio; hazard ratio vine din regresia Cox. Ambele observații sunt exerciții de lectură atentă și nu pun în discuție valoarea clinică a scorurilor, dar ilustrează o regulă utilă: numele măsurii de efect trebuie să corespundă modelului statistic și designului.

8.10.4 Comparația în interiorul grupului nu înlocuiește comparația între grupuri

O greșeală de interpretare foarte răspândită este să concluzionăm că un tratament funcționează pentru că "s-a obținut o scădere semnificativă față de valoarea inițială în grupul tratat". Testul înainte-după din interiorul unui grup arată doar că s-a produs o schimbare. Aceeași schimbare poate apărea și în grupul de control, prin regresie spre medie, efect placebo, sezonalitate sau simpla participare la studiu. În mg2026, greutatea a scăzut semnificativ chiar și în brațul cu consiliere simplă (testul t pereche, p = 0,043). Doar comparația între brațe (ANOVA, F = 18,25, p < 0,001) arată efectul intervențiilor.

Studiu de caz: Aloe ferox și perindopril, comparații în interiorul grupului și între grupuri

Gherbon și colaboratorii, cu dr. Frandeș, au testat într-un studiu controlat cu placebo, raportat după ghidul CONSORT, efectul a 2 capsule/zi de Aloe ferox (460 mg) la 20 de persoane cu obezitate, comparativ cu 20 de persoane potrivite care au primit dietă și placebo [85]. Metodele descriu două tipuri de comparații: între grupuri la 3 luni și înainte-după în interiorul grupurilor. Rezumatul dă, pentru aceleași variabile, două serii de valori p: de exemplu, pentru IMC P = 0,03 și P = 0,002, pentru LDLc P = 0,01 și P = 0,039, pentru glicemia a jeun P = 0,018 și P < 0,001, la care se adaugă colesterolul total (P = 0,032) și tensiunea diastolică (P = 0,006). Cititorul trebuie să verifice în textul complet care valoare răspunde la care întrebare, pentru că ele nu au aceeași greutate. O scădere față de valoarea inițială în grupul tratat nu dovedește singură efectul suplimentului: o scădere a IMC poate apărea și sub dietă cu placebo. Comparația care răspunde la întrebarea clinică este cea între grupuri, ideal pe modificările de la început la 3 luni sau prin ANCOVA cu valoarea inițială drept covariabilă. La 20 de pacienți pe grup, alegerea dintre testul t și Mann-Whitney trebuie justificată prin forma distribuțiilor. Buda și colaboratorii, cu dr. Tudor, au comparat trombospondina-1 serică la pacienți hipertensivi cu disfuncție endotelială înainte și după un an

de perindopril: 8061,9 ± 3684,80 față de 15380 ± 5887 ng/mL, $p < 0,001$ [86]. Datele sunt pereche, deci testul potrivit este testul t pereche sau Wilcoxon, aplicat diferențelor individuale. O abatere standard de aproape jumătate din medie sugerează o distribuție asimetrică, deci varianta neparametrică sau logaritmarea sunt de luat în calcul. Lotul martor fără hipertensiune ($9221,03 \pm 6510,21$ ng/mL) aduce o a doua întrebare, cu alt test: comparația cu martorii se face cu un test pentru grupuri independente. Rezumatul nu declară numărul de pacienți, informație fără de care nicio valoare p nu poate fi interpretată.

8.11 Exerciții propuse

Pentru fiecare dintre situațiile 1-6, precizați: tipul variabilei rezultat, designul (independent sau pereche, numărul de grupuri), testul ales și alternativa, ipotezele statistice.

1. Un medic de familie vrea să afle dacă numărul de vizite pe an la cabinet (0, 1, 2,..., cu mulți pacienți cu 0-2 vizite și câțiva cu peste 15) diferă între pacienții cu și fără diabet.
2. Saturația de oxigen a fost măsurată la 25 de pacienți cu BPOC în repaus și după testul de mers de 6 minute.
3. Se compară proporția de pacienți vaccinați antigripal în trei județe (450, 380 și 520 de pacienți).
4. Un chestionar de calitate a somnului (scor 0-21) a fost aplicat aceluiași grup de 30 de rezidenți după o gardă, după o zi liberă și după o săptămână de concediu.
5. Doi radiologi au clasificat independent 120 de radiografii toracice ca "pneumonie" sau "fără pneumonie". Ce măsură folosiți pentru a evalua acordul lor?
6. Se dorește estimarea riscului de a dezvolta diabet în funcție de scorul FINDRISC, vârstă și sex, într-o cohortă urmărită 10 ani.
7. Într-un studiu, autorii au comparat HbA1c între 3 tipuri de tratament prin trei teste t și au găsit $p = 0,03$, $p = 0,20$ și $p = 0,04$. Calculați riscul global de eroare de tip I pentru trei comparații independente la $\alpha = 0,05$ și pragul corectat Bonferroni. Care comparații rămân semnificative după corecție?
8. Un tabel 2 x 2 are pe rânduri 12 și 48 de pacienți, iar pe coloane 10 și 50. Calculați cea mai mică frecvență așteptată și decideți dacă aplicați chi-pătrat sau Fisher.
9. Pe setul mg2026, rezolvați complet, în formatul a-g, întrebarea: "Diferă colesterolul total între bărbați și femei?" (în JASP, T-Tests > Independent Samples T-Test).
10. Pe setul mg2026, rezolvați în formatul a-g întrebarea: "Este diabetul confirmat prin TTGO asociat cu mediul de rezidență?"
11. Citiți rezumatul unui articol recent din specialitatea care vă interesează (de exemplu prin PubMed) și completați pentru el un rând de tipul celor din tabelul 8.3.

Capcane frecvente

- Alegerea testului după tipul variabilei de grupare, uitând de tipul variabilei rezultat (de exemplu chi-pătrat pentru "glicemia după sex", doar pentru că sexul este nominal).
- Confundarea datelor pereche cu grupurile independente, mai ales când aceiași pacienți apar în două coloane ale foii de calcul.
- Verificarea normalității pe eșantionul total în loc de fiecare grup (sau de diferențe, la datele pereche).
- Trecerea automată la test neparametric când Levene $p < 0,05$: soluția firească este varianta Welch.
- Raportarea testului care "a ieșit" în locul celui stabilit dinainte.

- Testarea diferențelor de la includere între brațele unui studiu randomizat și interpretarea lor ca "eșec al randomizării".
- Interpretarea unei comparații înainte-după din interiorul unui grup ca dovadă a eficacității.
- Folosirea corelației pentru a demonstra acordul dintre două metode.
- Numărarea ochilor, dinților sau vizitelor ca și cum ar fi pacienți independenți.

De reținut

- Ordinea deciziei: scopul (diferență, asociere, acord, predicție), tipul variabilei rezultat, numărul de grupuri sau momente, independent sau pereche, apoi distribuția și mărimea eșantionului.
- Pentru fiecare situație există o variantă parametrică, una neparametrică și una categorială (figura 8.1 și tabelul 8.1).
- Unitatea de analiză este pacientul (sau unitatea randomizată); observațiile multiple de la același pacient nu sunt independente.
- Testul, variabila principală și regula de rezervă se stabilesc în protocol, înainte de a vedea datele.
- Formatul a-g: ipoteza de cercetare, ipotezele statistice, metoda și justificarea, rezolvarea și programul, rezultatul cu grafic, concluzia statistică, concluzia de cercetare.
- Un rezultat complet are statistica testului, gradele de libertate, valoarea p exactă și o mărime a efectului cu interval de încredere; măsura de efect trebuie să se potrivească designului (OR, RR, HR).

Întrebări de autoevaluare

1. (CS) Un studiu măsoară tensiunea arterială la 40 de pacienți înainte și după 8 săptămâni de tratament; diferențele au distribuție normală. Testul potrivit este:
    a) testul t pentru eșantioane independente
    b) testul t pereche
    c) testul Mann-Whitney
    d) testul chi-pătrat
    e) ANOVA unifactorială
2. (CS) Se compară scorul de durere (0-10) între trei scheme analgezice, cu câte 12 pacienți pe schemă. Testul potrivit este:
    a) testul Friedman
    b) testul Wilcoxon pentru date pereche
    c) testul Kruskal-Wallis
    d) trei teste Mann-Whitney fără corecție
    e) testul McNemar
3. (CS) Pentru a evalua acordul dintre glucometrul portabil și analizorul de laborator pe aceleași 80 de probe, metoda cea mai potrivită este:
    a) corelația Pearson
    b) testul t pentru eșantioane independente
    c) analiza Bland-Altman
    d) testul chi-pătrat
    e) regresia logistică
4. (CS) Riscul global de cel puțin o eroare de tip I când se fac 3 comparații independente, fiecare la $\alpha = 0{,}05$, este aproximativ:
    a) 5%
    b) 10%

c) 14%

d) 15%

e) 20%

5. (CS) Într-o cohortă urmărită 5 ani, cu durate de urmărire diferite, efectul unui factor asupra mortalității se exprimă cel mai corect prin:
   a) odds ratio dintr-un tabel 2 x 2
   b) hazard ratio dintr-o regresie Cox
   c) coeficientul de corelație Spearman
   d) V Cramér
   e) statistica H Kruskal-Wallis
6. (CM) Datele sunt pereche în situațiile:
   a) glicemia aceluiași pacient la 0 și la 6 luni
   b) acuitatea vizuală la ochiul drept și la ochiul stâng ale acelorași pacienți
   c) HbA1c la bărbați și la femei
   d) cazuri și martori potriviți individual după vârstă și sex
   e) TAS la fumători și la nefumători
7. (CM) Despre alegerea dintre testul t și varianta neparametrică sunt corecte:
   a) normalitatea se verifică în fiecare grup, nu pe eșantionul total
   b) pentru testul t pereche se verifică normalitatea diferențelor
   c) un Shapiro-Wilk cu $p < 0,05$ la $n = 300$ obligă întotdeauna la test neparametric
   d) variațele inegale (Levene $p < 0,05$) se tratează de regulă cu varianta Welch
   e) testul se alege după ce s-a văzut care variantă dă $p < 0,05$
8. (CM) Un articol descrie un studiu randomizat cu trei brațe și raportează $p < 0,05$ pentru scăderea glicemiei în interiorul fiecărui braț. Sunt corecte:
   a) rezultatul arată că fiecare intervenție produce o schimbare față de valoarea inițială
   b) rezultatul nu arată că o intervenție este superioară celorlalte
   c) comparația potrivită este între brațe (de exemplu ANOVA pe modificări)
   d) valorile p din interiorul brațelor dovedesc eficacitatea tuturor intervențiilor
   e) regresia spre medie poate contribui la scăderile observate
9. (CM) Unitatea de analiză este încălcată (observațiile nu sunt independente) dacă:
   a) se analizează 400 de dinți de la 40 de pacienți ca 400 de observații independente
   b) se analizează 170 de ochi de la 102 pacienți fără a ține cont de pacient
   c) fiecare pacient contribuie cu o singură valoare
   d) se numără internările în loc de pacienți, iar unii pacienți au mai multe internări
   e) pacienții sunt randomizați individual și fiecare apare o singură dată în analiză
10. (CM) În formatul standardizat a-g sunt incluse:

a) ipotezele statistice în cuvinte și matematic

b) justificarea alegerii testului

c) programul și modul de rezolvare

d) doar valoarea p, fără statistici descriptive

e) concluzia de cercetare formulată în cuvinte

Capitolul 9. Analiza relațiilor dintre variabile: corelația și regresia liniară pentru variabile numerice și ordinale

Obiective

La finalul acestui capitol, studentul va putea:

- să recunoască o problemă de analiză bivariată și să numească predictorul (X) și variabila răspuns (Y);
- să construiască și să citească o diagramă de corelație, deosebind variabilele independente, dependente funcțional și corelate;
- să calculeze de mână covarianța, coeficientul de corelație Pearson și coeficientul Spearman pe un set mic de date;
- să interpreteze semnul, intensitatea (după regula lui Colton) și semnificația statistică a unui coeficient de corelație, cu intervalul său de încredere;
- să explice ce înseamnă coeficientul de determinare R^2 și să răspundă la întrebarea "ce proporție din variație se datorează altor factori";
- să estimeze dreapta de regresie prin metoda celor mai mici pătrate, să interpreteze panta și interceptul și să facă o predicție, recunoscând pericolul extrapolării;
- să verifice condițiile regresiei liniare prin analiza reziduurilor;
- să citească un tabel de regresie liniară multiplă sau logistică dintr-un articol medical;
- să argumenteze de ce o corelație, oricât de puternică, nu dovedește singură o relație cauzală și nu măsoară concordanța a două metode.

9.1 Analiza bivariată: două variabile măsurate pe același pacient

Până acum am lucrat aproape mereu cu o variabilă numerică și cu una sau mai multe grupe: tensiunea la femei și la bărbați, glicemia în cele trei brațe ale studiului, glicemia la 0 și la 6 luni. În acest capitol schimbăm întrebarea. Pentru fiecare pacient avem două măsurători numerice, de exemplu indicele de masă corporală (IMC) și tensiunea arterială sistolică (TAS), și vrem să știm dacă ele variază împreună. Pacienții cu IMC mai mare au, în medie, și tensiunea mai mare? Dacă da, cât de strânsă este legătura și cu cât crește tensiunea pentru fiecare unitate de IMC?

Analiza care urmărește simultan comportamentul a două variabile se numește analiză bivariată. Prin convenție, variabila despre care credem că o influențează pe cealaltă sau pe care o folosim ca să o prezicem pe cealaltă se notează cu X și se numește predictor (variabilă explicativă, variabilă independentă). Variabila pe care vrem să o explicăm sau să o prezicem se notează cu Y și se numește variabilă răspuns (variabilă dependentă, variabilă rezultat). În exemplul nostru, IMC este X, iar TAS este Y. Alegerea nu e întotdeauna evidentă: între glicemia a jeun și hemoglobina glicată (HbA1c) nu există un "predictor natural", amândouă reflectă același proces, iar alegerea lui X ține de întrebarea clinică.

Două metode răspund la două întrebări diferite. Corelația măsoară sensul și intensitatea asocierii dintre două variabile numerice, tratate simetric: ambele sunt considerate aleatoare, iar coeficientul de corelație dintre X și Y este același cu cel dintre Y și X. Regresia construiește un model matematic care descrie cum se modifică, în medie, Y atunci când X se modifică, și permite predicția lui Y pentru o valoare dată a lui X. În regresie rolurile nu sunt simetrice: dreapta lui Y în

funcție de X nu este aceeași cu dreapta lui X în funcție de Y. Manualele clasice ale disciplinei prezintă cele două metode împreună, pentru că folosesc aceleași sume de pătrate și se verifică una pe alta [18,42,53].

9.1.1 Diagrama de corelație

Orice analiză bivariată începe cu un grafic, nu cu un număr. Diagrama de corelație (engl. scatter plot) are pe axa orizontală variabila X și pe axa verticală variabila Y; fiecare pacient este un punct. Când se uită la un astfel de grafic, studentul trebuie să-și pună patru întrebări: ce formă are norul de puncte (alungit după o dreaptă, curbat, fără formă), în ce sens merge (crescător sau descrescător), cât de strâns sunt adunate punctele în jurul unei linii imaginare și dacă există puncte izolate, departe de rest.

Figura 9.1 arată cele trei tipare descrise în cursurile disciplinei. În panoul A punctele sunt împrăștiate uniform, fără nicio direcție: cunoașterea lui X nu spune nimic despre Y, variabilele sunt independente. În panoul B toate punctele stau exact pe o curbă: Y este complet determinat de X, avem o dependență funcțională (de tipul relației dintre latura și aria unui pătrat). În biologie și medicină, o astfel de relație perfectă aproape nu apare. Panoul C este situația obișnuită: punctele se adună în jurul unei drepte, dar cu o împrăștiere vizibilă. Y tinde să crească odată cu X, însă pentru aceeași valoare a lui X pacienții au valori diferite ale lui Y. Spunem că variabilele sunt corelate: legătura este statistică (stocastică), nu funcțională.

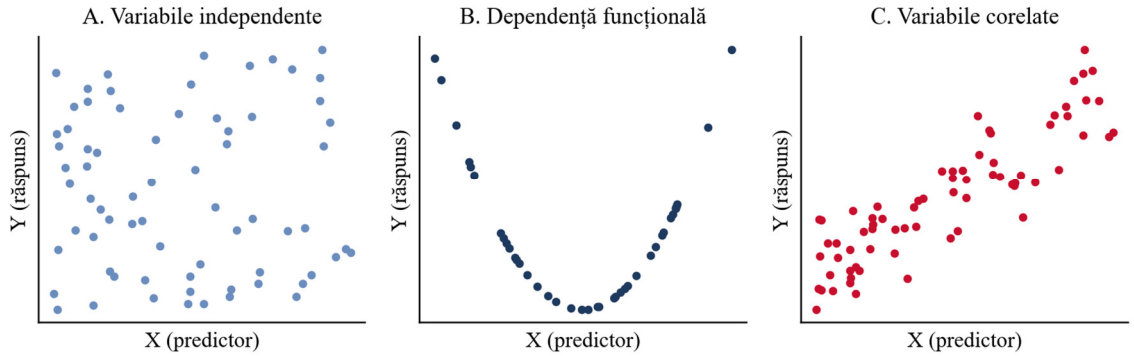


Figura 9.1. Tipare de relație între două variabile numerice. A: variabile independente, norul de puncte nu are nicio direcție. B: dependență funcțională, fiecare valoare a lui X determină exact o valoare a lui Y. C: variabile corelate, Y crește în medie odată cu X, dar cu împrăștiere în jurul tendinței. Date simulate.

9.1.2 Ce poate și ce nu poate spune graficul

Graficul arată lucruri pe care coeficienții le ascund. O relație curbă (de exemplu, mortalitatea în funcție de IMC, crescută atât la subponderali, cât și la obezi, cu forma literei U) poate avea un coeficient Pearson apropiat de zero, deși legătura dintre variabile este puternică. Un singur pacient cu valori extreme poate crea o corelație aparentă sau o poate ascunde. Două subgrupe distincte (bărbați și femei, de exemplu, la relația înălțime-greutate) pot produce împreună un nor alungit, chiar dacă în interiorul fiecărei subgrupe legătura este slabă. De aceea, la laborator, nu acceptăm niciun coeficient de corelație raportat fără diagrama corespunzătoare.

Pe setul didactic mg2026 (160 de adulți cu risc cardiometabolic, cap. 1) vom urmări de-a lungul capitolului mai multe perechi de variabile. Le anunțăm de acum, ca să puteți compara rezultatele: IMC și TAS (corelație slabă, dar clar semnificativă), vârsta și TAS (corelație slabă), glicemia a jeun și HbA1c (corelație moderată, aproape puternică), HDL-colesterolul și trigliceridele (fără corelație), activitatea fizică autoevaluată și IMC (variabilă ordinală, fără corelație).

9.2 Covarianța: primul pas spre o măsură a asocierii

Să presupunem că pentru fiecare pacient calculăm abaterea lui X de la media lui X și abaterea lui Y de la media lui Y. Dacă pacientul are ambele valori peste medie, produsul abaterilor este pozitiv. Dacă are ambele valori sub medie, produsul este tot pozitiv (minus ori minus). Dacă are una peste și una sub medie, produsul este negativ. Însumând produsele pentru toți pacienții, obținem un număr care este mare și pozitiv când variabilele cresc împreună, mare și negativ când una crește în timp ce cealaltă scade și apropiat de zero când nu există o tendință comună.

Împărțind această sumă la n - 1, obținem covarianța de eșantion:

$$\text{cov}(X,\ Y) = \frac{\sum(x_i - m_x)\,(y_i - m_y)}{n-1}$$

unde mx și my sunt mediile celor două variabile. Notăm de acum suma produselor abaterilor cu Sxy, iar sumele pătratelor abaterilor cu Sxx și Syy:

$$S_{\text{xy}} = \sum(x_i - m_x)\,(y_i - m_y);\ S_{\text{xx}} = \sum(x_i - m_x)^2;\ S_{\text{yy}} = \sum(y_i - m_y)^2$$

Observați că Sxx / (n - 1) este chiar varianța lui X din capitolul 2. Covarianța generalizează varianța: varianța este covarianța unei variabile cu ea însăși.

Covarianța are un defect serios pentru interpretare: depinde de unitățile de măsură. Dacă exprimăm greutatea în grame în loc de kilograme, covarianța dintre înălțime și greutate devine de 1000 de ori mai mare, deși relația biologică este aceeași. Nu putem spune, uitându-ne la o covariantă de 43, dacă asocierea este puternică sau slabă. Soluția este standardizarea: împărțim covarianța la produsul abaterilor standard și obținem un număr fără unitate de măsură, coeficientul de corelație.

9.3 Coeficientul de corelație Pearson

9.3.1 Formula și calculul de mână

Coeficientul de corelație liniară Pearson se notează cu r pentru eșantion și cu ρ (rho) pentru populație:

$$\text{r} = \text{cov}\frac{X, Y}{s_x \times s_y} = \frac{S_{\text{xy}}}{\sqrt{S_{\text{xx}} \times S_{\text{yy}}}}$$

Cele două forme sunt echivalente, pentru că factorii n - 1 se simplifică. Pentru calculul de mână, a doua formă este mai comodă.

Să luăm un exemplu mic, cu cifre alese astfel încât calculul să poată fi urmărit fără calculator. Opt pacienți dintr-un cabinet de medicină de familie au IMC și TAS din tabelul 9.1.

Tabelul 9.1. Calculul de mână al coeficientului Pearson pentru 8 pacienți $\left(X = \text{IMC},\ \frac{\text{kg}}{m^2};\ Y = \text{TAS},\ \text{mmHg}\right)$.

Pacient	xi	yi	xi - mx	yi - my	(xi - mx)(yi - my)	$(x_i - m_x)^2$	$(y_i - m_y)^2$
1	23	118	-6	-13	78	36	169
2	25	126	-4	-5	20	16	25
3	26	121	-3	-10	30	9	100
4	28	129	-1	-2	2	1	4
5	29	135	0	4	0	0	16
6	31	131	2	0	0	4	0
7	33	140	4	9	36	16	81
8	37	148	8	17	136	64	289
Total	232	1048	0	0	$S_{\text{xy}} = 302$	$S_{\text{xx}} = 146$	$S_{\text{yy}} = 684$

Pașii sunt următorii. Mediile: $m_x = \frac{232}{8} = \frac{29\ \text{kg}}{m^2}$ și $m_y = \frac{1048}{8} = 131$ mmHg. Abaterile se calculează rând cu rând, iar suma lor trebuie să dea zero pe fiecare coloană; dacă nu dă zero, ați greșit o scădere. Apoi se înmulțesc abaterile și se ridică la pătrat. Covarianța este 302 / 7 = 43,14 kg/m² x mmHg (o unitate de măsură greu de imaginat, ceea ce confirmă discuția de mai sus). Abaterile standard sunt $s_x = \sqrt{\frac{146}{7}} = \frac{4{,}57\ \text{kg}}{m^2}$ și $s_y = \sqrt{\frac{684}{7}} = 9{,}89$ mmHg. În final:

$$r = \frac{302}{\sqrt{146 \times 684}} = \frac{302}{\sqrt{99\,864}} = \frac{302}{316{,}01} = 0{,}956$$

Toate produsele sunt pozitive sau zero: fiecare pacient cu IMC peste medie are și TAS peste medie sau la medie. Coeficientul este aproape de 1, o corelație pozitivă puternică. Pe date reale nu veți obține atât de des asemenea valori; exemplul a fost construit pentru claritate. Vom continua cu aceleași opt perechi la regresie (secțiunea 9.7) și în problema rezolvată 9.1.

9.3.2 Proprietățile coeficientului Pearson

Coeficientul r ia valori între -1 și +1. Semnul indică sensul asocierii: r > 0 înseamnă corelație directă (pozitivă), cele două variabile cresc împreună; r < 0 înseamnă corelație inversă (negativă), când una crește cealaltă tinde să scadă. Valoarea r = +1 corespunde unei corelații directe perfecte (toate punctele pe o dreaptă crescătoare), r = -1 unei corelații inverse perfecte, iar r = 0 lipsei unei relații liniare.

Câteva proprietăți sunt importante la examen. Coeficientul nu are unitate de măsură și nu se schimbă dacă transformăm liniar una dintre variabile (greutatea în grame sau în kilograme, temperatura în grade Celsius sau Fahrenheit dau același r). Este simetric: r(X, Y) = r(Y, X). Măsoară doar relația liniară; o relație curbă puternică poate avea r mic. Este sensibil la valori extreme, pentru că se bazează pe medii și pe pătratele abaterilor. Și, foarte important, depinde de intervalul de valori din eșantion: dacă studiem relația IMC-TAS numai la pacienți cu IMC între 27 și 29, r va fi mult mai mic decât în eșantionul complet, pentru că am tăiat chiar variația care face vizibilă relația (fenomenul se numește restrângerea domeniului).

9.3.3 Cât de puternică este o corelație?

Figura 9.2 arată cum arată norii de puncte pentru câteva valori ale lui r. Merită memorată vizual: un r de 0,5 arată mult mai "împrăștiat" decât se așteaptă de obicei studenții, iar un r de 0,95 apare ca o bandă îngustă.

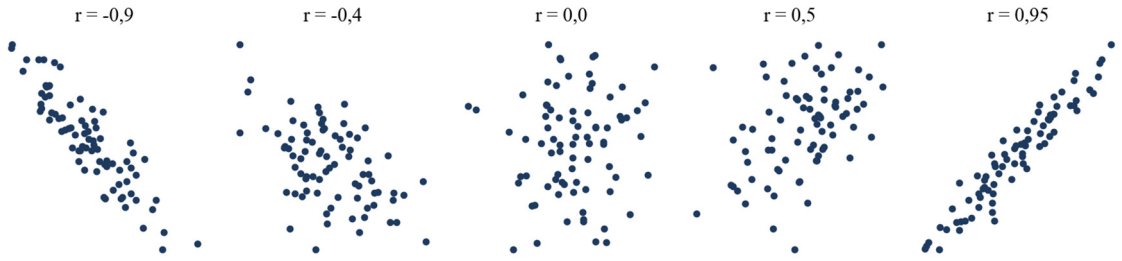


Figura 9.2. Nori de puncte simulați pentru r = -0,9; -0,4; 0; 0,5 și 0,95 (câte 80 de puncte). Intensitatea corelației se citește din cât de strâns sunt adunate punctele în jurul unei drepte, nu din panta acesteia.

Pentru a transforma numărul în cuvinte, cursurile disciplinei folosesc regula empirică propusă de Colton în 1974. Valoarea absolută a lui r se interpretează astfel:

Tabelul 9.2. Regula empirică a lui Colton pentru interpretarea coeficientului de corelație.

Valoarea lui r	Interpretare
-0,25 < r < 0,25	fără corelație (corelație neglijabilă)
0,25 ≤ r < 0,50 sau -0,50 < r ≤ -0,25	corelație slabă
0,50 ≤ r < 0,75 sau -0,75 < r ≤ -0,50	corelație moderată
r ≥ 0,75 sau r ≤ -0,75	corelație puternică
r = 1 sau r = -1	corelație perfectă

Pe mg2026, corelația IMC-TAS (r = 0,405) este slabă, corelația vârstă-TAS (r = 0,256) este slabă, aflată chiar la limita de jos, iar corelația glicemie-HbA1c (r = 0,738) este moderată, foarte aproape de pragul de 0,75. HDL și trigliceridele (r = 0,076) nu sunt corelate.

Regula lui Colton nu este singura. În psihologie și în științele comportamentale se folosește frecvent scara lui Cohen, care consideră r = 0,1 un efect mic, r = 0,3 un efect mediu și r = 0,5 un efect mare [76]. Alte manuale folosesc praguri din 0,2 în 0,2 (foarte slabă, slabă, moderată, puternică, foarte puternică). Cu scara lui Cohen, r = 0,405 dintre IMC și TAS ar fi un efect "mediu spre mare", nu o corelație "slabă". Nu există o scară corectă și una greșită: există convenții. Important este să spuneți ce scară folosiți și să nu schimbați scara în funcție de rezultatul care vă convine. La examenul disciplinei se folosește regula lui Colton.

9.3.4 Condițiile de aplicare ale coeficientului Pearson

Coeficientul Pearson poate fi calculat pentru orice pereche de coloane numerice, dar interpretarea și testul de semnificație cer trei condiții. Prima: relația dintre variabile trebuie să fie liniară, lucru care se verifică pe diagrama de corelație. A doua: pentru test și pentru intervalul de încredere, perechile trebuie să provină dintr-o distribuție normală bivariată; practic, verificăm că fiecare variabilă are o distribuție aproximativ normală (histogramă, grafic Q-Q, test Shapiro-Wilk, cap. 3). A treia: nu trebuie să existe valori extreme care să domine calculul. La acestea se adaugă condiția generală a independenței observațiilor: fiecare pacient apare o singură dată, nu de mai multe ori cu măsurători repetate.

Figura 9.3 arată de ce a treia condiție contează atât de mult. În panoul A, 25 de puncte simulate nu au nicio legătură între ele (r = -0,16). În panoul B s-a adăugat un singur punct, departe de toate celelalte, iar r a sărit la 0,64. Un student care raportează doar numărul ar spune că există o corelație moderată; graficul arată că "corelația" este produsă de un singur pacient.

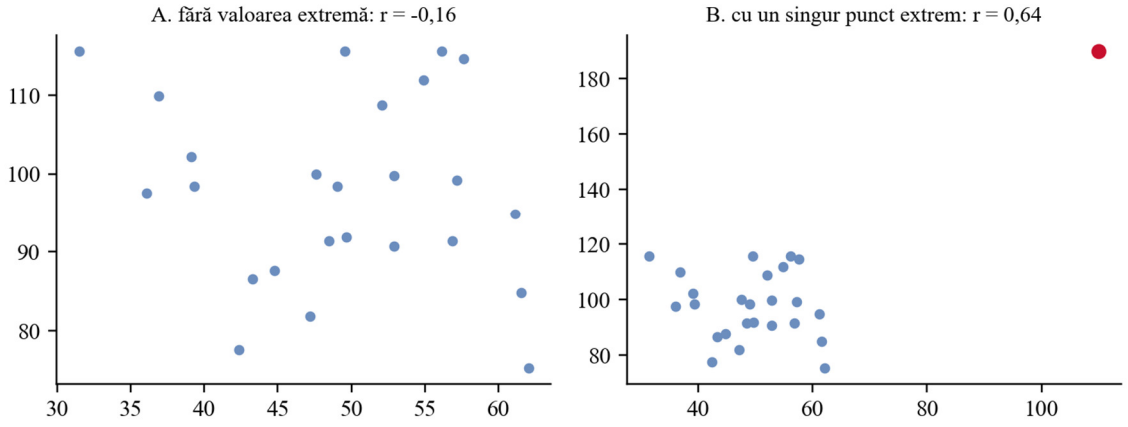


Figura 9.3. Influența unei singure valori extreme asupra coeficientului Pearson. A: 25 de puncte fără relație, r = -0,16. B: aceleași puncte plus un singur punct extrem (roșu), r = 0,64. Date simulate.

Ce facem cu o valoare extremă? Mai întâi verificăm dacă nu este o eroare de introducere (o înălțime de 1,72 introdusă în metri în loc de 172 cm, cum am văzut în fișierul mg2026_brut.csv).

Dacă valoarea este reală, raportăm analiza cu și fără ea și preferăm un coeficient bazat pe ranguri (Spearman), care este mult mai puțin sensibil. Ștergerea pacienților incomozi fără justificare nu este o opțiune.

Pe mg2026, IMC nu respectă normalitatea (asimetrie 1,42; Shapiro-Wilk p < 0,001), din cauza câtorva pacienți cu obezitate severă (maximul este 48,3 kg/m²). TAS este normal distribuită (p = 0,303). Prin urmare, pentru perechea IMC-TAS vom raporta atât Pearson, cât și Spearman, și vom vedea că cele două dau concluzii asemănătoare.

9.3.5 Testarea semnificației coeficientului de corelație

Un r calculat pe un eșantion nu este niciodată exact zero, chiar dacă în populație variabilele nu au nicio legătură. Întrebarea inferențială este: valoarea observată ar putea apărea doar prin variația de eșantionare, dacă în populație ρ = 0? Ipotezele sunt:

H0: ρ = 0 (în populație, cele două variabile nu sunt corelate liniar) H1: ρ ≠ 0 (în populație, variabilele sunt corelate)

Statistica testului urmează, sub H0, o distribuție Student cu n - 2 grade de libertate:

$$t = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}}$$

Pierdem două grade de libertate pentru că am estimat două medii. Pentru exemplul cu opt pacienți: $t = \frac{0{,}956 \times \sqrt{6}}{\sqrt{1-0{,}913}} = \frac{0{,}956 \times 2{,}449}{0{,}294} = 7{,}95$. Valoarea critică pentru 6 grade de libertate și α = 0,05 bilateral este 2,447; cum 7,95 > 2,447, H0 se respinge (p = 0,0002).

Pe mg2026, pentru IMC-TAS: $t = \frac{0{,}405 \times \sqrt{158}}{\sqrt{1-0{,}164}} = \frac{0{,}405 \times 12{,}57}{0{,}914} = 5{,}56$, cu 158 de grade de libertate, p < 0,001. Pentru vârstă-TAS: t = 3,34, p = 0,001. Pentru HDL-trigliceride: t = 0,96, p = 0,34, H0 nu se respinge.

Formula ascunde o lecție pe care o veți întâlni des în articole: semnificația depinde masiv de n. Tabelul 9.3 arată cel mai mic r care devine semnificativ la α = 0,05 pentru diferite dimensiuni ale eșantionului.

Tabelul 9.3. Valoarea minimă a lui |r| semnificativă la α = 0,05 (test bilateral), în funcție de n.

n	10	20	30	50	100	160	867
r minim semnificativ	0,632	0,444	0,361	0,279	0,197	0,155	0,067

Cu 10 pacienți, o corelație moderată de 0,60 nu este semnificativă. Cu 867 de pacienți, o corelație de 0,07, pe care Colton ar numi-o "fără corelație", este semnificativă. Semnificația statistică răspunde doar la întrebarea "este ρ diferit de zero?", nu la întrebarea "este relația importantă?". Vom vedea un exemplu publicat în secțiunea 9.5.

9.3.6 Intervalul de încredere pentru r: transformarea Fisher

Distribuția de eșantionare a lui r nu este simetrică atunci când ρ este departe de zero: r nu poate depăși 1, așa că valorile se "înghesuie" spre capăt. Statisticianul R. A. Fisher a propus o transformare care face distribuția aproximativ normală:

$z = 0{,}5 \times \ln\left[\frac{1+r}{1-r}\right]$, cu eroarea standard $\mathrm{ES}(z) = \frac{1}{\sqrt{n-3}}$

Se calculează intervalul pe scara z, $z \pm 1{,}96 \times \mathrm{ES}(z)$, apoi capetele se transformă înapoi în r cu formula inversă $r = \frac{e^{2z}-1}{e^{2z}+1}$. Pentru IMC-TAS pe mg2026: $z = 0{,}5 \times \ln\left(\frac{1{,}405}{0{,}595}\right) = 0{,}429$; ES =

$\frac{1}{\sqrt{157}} = 0{,}0798$; intervalul pe scara z este 0,429 ± 0,156, adică 0,273 până la 0,586; transformat înapoi, IC 95% pentru ρ este 0,266-0,527.

Intervalul spune mai mult decât valoarea p. Știm acum că, foarte probabil, corelația din populație este cel puțin slabă (peste 0,25) și cel mult moderată (sub 0,53). Pentru glicemie-HbA1c, IC 95% este 0,659-0,802: corelația este cel puțin moderată și poate fi puternică. Pentru HDL-trigliceride, IC 95% este -0,080 până la 0,229 și conține zero, în acord cu p = 0,34. Pentru eșantionul de 8 pacienți, deși r = 0,956, intervalul este larg: 0,769-0,992, pentru că n - 3 = 5 este foarte mic.

9.3.7 Coeficientul de determinare R^2

Ridicând r la pătrat obținem coeficientul de determinare, R² (sau r²). El arată ce proporție din variația lui Y este "explicată", în sens statistic, de relația liniară cu X. Ia valori între 0 și 1 (sau între 0% și 100%) și nu mai păstrează semnul: o corelație de -0,6 și una de +0,6 au același $R^2 = 0{,}36$. De aceea spunem că R² informează despre tăria corelației, nu despre sensul ei.

Pentru IMC și TAS pe mg2026, $R^2 = 0{,}405^2 = 0{,}164$. Aproximativ 16,4% din variabilitatea tensiunii sistolice între pacienți poate fi pusă pe seama diferențelor de IMC; restul de 83,6% ține de alți factori (vârsta, sexul, fumatul, sarea din alimentație, genetica, medicația, variația de la o măsurătoare la alta). Aceasta este exact întrebarea pe care o pune lucrarea practică a disciplinei despre înălțime și greutate: "în ce proporție variația greutății este cauzată de alți factori decât variația înălțimii?". Răspunsul este 1 - R². Pe mg2026, înălțimea și greutatea au r = 0,652, deci $R^2 = 0{,}424$, iar $1 - R^2 = 0{,}576$: aproximativ 57,6% din variația greutății ține de alți factori decât înălțimea.

Pentru glicemie și HbA1c, $R^2 = 0{,}545$: puțin peste jumătate din variația HbA1c se asociază liniar cu variația glicemiei a jeun. Pentru vârstă și TAS, $R^2 = 0{,}066$, adică doar 6,6%. Un r "slab" de 0,26 înseamnă, tradus în R², că vârsta explică foarte puțin din diferențele de tensiune dintre pacienții noștri. Studenții sunt adesea surprinși de cât de mic devine un r moderat după ridicarea la pătrat: r = 0,5 explică doar un sfert din variație.

Ultima precizare ține de limbaj. "Explicat" nu înseamnă "cauzat". Expresia din enunțurile clasice ("variația greutății cauzată de alți factori") trebuie citită statistic: proporția din variație care nu este legată liniar de înălțime. Faptul că înălțimea "explică" 42% din variația greutății nu înseamnă că 42% din kilogramele cuiva provin din înălțime.

Studiu de caz: corelații în elastografia hepatică (colectivul de ultrasonografie de la Timișoara)

Colegii din disciplină care lucrează cu echipa de gastroenterologie au publicat mai multe studii în care o metodă nouă de ultrasonografie este comparată cu o metodă de referință, iar corelația este primul rezultat raportat. Într-un studiu de acuratețe diagnostică pe 64 de persoane (37 cu ciroză, 27 voluntari sănătoși), parametrul obținut prin ultrasonografie cu contrast (CEUS-PAT) a avut o corelație r = 0,81 cu elastografia tranzitorie controlată prin vibrații (VCTE) [2]. Într-un studiu prospectiv pe 165 de pacienți, elastografia armonică în timp a avut, față de aceeași referință, r = 0,82, p < 0,0001 [3]. Pentru steatoză, parametrul UGAP a avut cu CAP r = 0,73, p < 0,0001, pe 179 de subiecți consecutivi [58], iar într-un studiu din 2026 pe 169 de pacienți fracția de grăsime ultrasonografică (USFF) a avut cu CAP r = 0,81, p < 0,001, iar alți doi parametri au avut r = 0,66 și r = 0,83 [65].

Cu regula lui Colton, valorile de 0,81, 0,82 și 0,83 indică o corelație puternică, iar 0,73 și 0,66 o corelație moderată. Rezumatul ultimului studiu numește totuși "moderată" o corelație de 0,83, ceea ce ilustrează exact problema scărilor diferite discutată mai sus: autorii pot folosi alt prag, iar cititorul trebuie să judece numărul, nu adjectivul. A doua lecție este mai importantă. O

> corelație de 0,82 între două metode de măsurare a rigidității hepatice nu înseamnă că ele dau aceleași valori și pot fi folosite una în locul celeilalte. Corelația arată că pacienții sunt ordonați asemănător; concordanța se evaluează altfel (secțiunea 9.6), iar studiul din 2020 a raportat, pentru reproductibilitate, coeficienți de corelație intraclasă de 0,92-0,97 [3]. Ce ar mai fi util într-un astfel de rezumat: intervalul de încredere pentru r și diagrama de corelație, care ar arăta dacă relația rămâne liniară pe tot domeniul de fibroză.

9.4 Corelația pentru date ordinale sau asimetrice: coeficienții Spearman și Kendall

9.4.1 Ideea: lucrăm cu ranguri, nu cu valori

Coeficientul Pearson presupune variabile numerice, relație liniară și distribuție aproximativ normală. Multe variabile medicale nu respectă aceste condiții. Scorurile pe scale Likert, stadiile de boală, clasa NYHA, nivelul de educație sau activitatea fizică autoevaluată (1-5) din mg2026 sunt ordinale: știm ordinea categoriilor, dar distanța dintre "2" și "3" nu este neapărat aceeași cu cea dintre "4" și "5". Alte variabile sunt numerice, dar foarte asimetrice (trigliceridele, IMC în mg2026) sau au valori extreme.

Coeficientul de corelație a rangurilor al lui Spearman (notat rs sau ρ Spearman, engl. Spearman's rank correlation) rezolvă problema printr-un artificiu simplu: înlocuiește fiecare valoare cu rangul ei (locul în ordinea crescătoare) și calculează coeficientul Pearson pe ranguri. Rezultatul măsoară cât de monoton este legătura: dacă Y crește ori de câte ori crește X, chiar și neliniar, coeficientul Spearman este +1. Ca toate metodele bazate pe ranguri din capitolul 7, este robust la valori extreme: cel mai mare IMC primește rangul 160, indiferent dacă este 40 sau 48,3.

9.4.2 Calculul de mână

Când nu există valori egale (ex aequo), Spearman se poate calcula cu o formulă scurtă:

$$r_s = 1 - \frac{6 \times \sum {d_i}^2}{n \times (n^2 - 1)}$$

unde di este diferența dintre rangul lui X și rangul lui Y la pacientul i. Să luăm opt pacienți la care s-au măsurat scorul FINDRISC și glicemia a jeun.

Tabelul 9.4. Calculul coeficientului Spearman pentru scorul FINDRISC și glicemia a jeun la 8 pacienți.

Pacient	FINDRISC	Glicemie (mg/dL)	Rang X	Rang Y	di	d_i^2
1	5	92	1	2	-1	1
2	8	88	2	1	1	1
3	9	101	3	4	-1	1
4	12	99	4	3	1	1
5	14	115	5	6	-1	1
6	15	108	6	5	1	1
7	18	126	7	8	-1	1
8	21	119	8	7	1	1
Total					0	8

$$r_s = 1 - \frac{6 \times 8}{8 \times (64 - 1)} = 1 - \frac{48}{504} = 1 - 0{,}095 = 0{,}905$$

Corelația este pozitivă și puternică: pacienții cu scor FINDRISC mai mare tind să aibă glicemia mai mare. Ordinea nu este perfect respectată (perechile 1-2, 3-4, 5-6, 7-8 își schimbă

locurile), dar abaterile sunt de un singur rang. Pentru n mic, semnificația lui rs se citește din tabele speciale sau din program (aici p = 0,002); pentru n peste aproximativ 10-20 se poate folosi aceeași formulă t ca la Pearson.

9.4.3 Valorile egale (ex aequo)

Variabilele ordinale au aproape întotdeauna valori egale. Pacienții cu aceeași valoare primesc media rangurilor pe care le-ar fi ocupat: dacă doi pacienți au activitatea 2 și ar ocupa locurile 2 și 3, ambii primesc rangul 2,5. Cu multe valori egale, formula scurtă devine inexactă și trebuie calculat Pearson pe ranguri (ceea ce fac automat toate programele). Problema rezolvată 9.2 arată un exemplu în care diferența dintre cele două calcule este mare: -0,85 cu formula scurtă față de -0,91 calculat corect.

9.4.4 Coeficientul tau al lui Kendall

O alternativă, mai puțin folosită în articolele clinice, dar preferată de unii statisticieni pentru eșantioane mici și pentru multe valori egale, este coeficientul tau (τ) al lui Kendall. Ideea lui este numărarea perechilor: pentru oricare doi pacienți ne întrebăm dacă ordinea lor după X este aceeași cu ordinea după Y (pereche concordantă) sau inversă (pereche discordantă). Tau este diferența dintre proporția perechilor concordante și a celor discordante. În exemplul din tabelul 9.4 există 28 de perechi posibile de pacienți; doar cele patru perechi vecine care își schimbă locurile sunt discordante, deci $\tau = (24 - 4) / 28 = 0,71$. Valorile lui τ sunt în general mai mici decât cele ale lui Spearman pentru aceleași date (pe mg2026, pentru IMC-TAS, $\tau = 0,258$ față de rs = 0,372), așa că pragurile lui Colton nu se aplică direct lui τ. Varianta τ-b, raportată de JASP și de alte programe, corectează pentru valorile egale.

9.4.5 Ce arată coeficienții de rang pe mg2026

Activitatea fizică autoevaluată (1-5) este ordinală, așa că relația ei cu IMC se evaluează cu Spearman: rs = -0,012, p = 0,88. Semnul este cel așteptat (mai multă activitate, IMC puțin mai mic), dar valoarea este practic zero: în eșantionul nostru, cât declară pacienții că se mișcă nu spune nimic despre IMC. Explicațiile posibile țin de calitatea măsurătorii (autoevaluarea este imprecisă) și de faptul că eșantionul cuprinde doar persoane supraponderale. Între nivelul de educație și satisfacția la 6 luni, rs = 0,125, p = 0,115: nici corelație după Colton, nici semnificație.

Pentru IMC și TAS, unde IMC este asimetric, Spearman dă rs = 0,372 ($p < 0,001$), foarte aproape de Pearson r = 0,405. Când cele două coeficiente sunt apropiate, valorile extreme nu distorsionează mult rezultatul. Când diferă mult, graficul trebuie privit cu atenție.

> **Studiu de caz: coeficienți de rang în studii de validare și de educație medicală**
>
> Patru articole ale colegilor arată cum se raportează corect corelațiile de rang. Într-un studiu de acuratețe diagnostică pe 157 de pacienți cu hepatită cronică B sau C, cu biopsia hepatică drept standard, rigiditatea măsurată prin două metode a fost corelată cu stadiul de fibroză (o variabilă ordinală, F0-F4) prin coeficientul Spearman: $\rho = 0,826$, $p < 0,001$, cu IC 95% pentru rho 0,769-0,870 [107]. Este unul dintre puținele rezumate care dau intervalul de încredere, iar în problema rezolvată 9.4 vom verifica de mână că intervalul corespunde transformării Fisher. Într-un studiu de validare a unui chestionar de acceptabilitate a tehnologiei mobile în diabet, pe 103 pacienți, autorii au raportat atât coeficienți Spearman ($\rho = -0,429$; $\rho = -0,322$; $\rho = 0,466$; toți cu $P < 0,001$), cât și un coeficient Kendall ($\tau = 0,51$; $P < 0,001$) [54]; observați că tau poate fi mare chiar dacă este, în general, mai mic decât rho.
>
> Un studiu prin chestionar anonim pe 121 de studenți la medicină a folosit mediana și intervalul intercuartilic pentru scoruri pe scala 1-10 și corelația Spearman [16]. Două corelații din rezumat merită puse alături: R = 0,869, $p < 0,001$, și R = 0,107, p = 0,246, aceasta din urmă

> între cunoștințele măsurate și cele autoevaluate. A doua corelație este practic nulă și nesemnificativă: ce cred studenții că știu nu spune mare lucru despre ce știu efectiv. Pentru un curs de biostatistică, rezultatul este o invitație la modestie, iar statistic arată că un rezultat nesemnificativ este și el informativ, dacă eșantionul nu este minuscul. În fine, un studiu pe datele de supraveghere COVID-19 din județul Timiș a găsit corelații Spearman de rho = 0,852 și 0,821 ($p < 0,001$) între incidența la personalul medical și tendința din comunitate [40]. Aici unitatea de analiză nu este persoana, ci perioada de timp: două serii temporale care cresc și scad împreună. Corelația este puternică, dar nu spune de unde s-a infectat un anumit lucrător medical; discutăm această capcană în secțiunea următoare.

9.5 Corelația nu înseamnă cauzalitate

9.5.1 Patru explicații pentru o corelație

Când două variabile sunt corelate, există cel puțin patru explicații posibile, iar cauzalitatea directă este doar una dintre ele. Prima este întâmplarea: cu suficient de multe perechi de variabile testate, unele vor ieși corelate semnificativ doar prin variația de eșantionare (cu $\alpha = 0,05$, una din 20 de perechi fără nicio legătură reală). A doua este cauzalitatea inversă: Y poate influența X, nu invers. Pacienții cu boală coronariană fac mai puțin efort fizic pentru că au angină, nu numai invers. A treia este confuzia: o a treia variabilă (factor de confuzie, engl. confounder) influențează ambele variabile și creează o corelație între ele fără legătură directă. Vânzările de înghețată și numărul de înecuri sunt corelate pentru că amândouă cresc vara. Pe mg2026, vârsta este corelată atât cu TAS, cât și cu riscul de diabet; o corelație brută între TAS și glicemie ar trebui verificată după ajustarea pentru vârstă. A patra explicație este cauzalitatea propriu-zisă, care trebuie argumentată, nu presupusă.

Pentru a cântări argumentele cauzale în studiile observaționale, epidemiologul Austin Bradford Hill a propus în 1965 un set de criterii: forța asocierii, consistența între studii, specificitatea, temporalitatea (cauza precede efectul), gradientul biologic (relația doză-răspuns), plauzibilitatea, coerența cu cunoștințele existente, dovezile experimentale și analogia [108]. Doar temporalitatea este obligatorie. Criteriile sunt discutate pe larg în capitolul 10; aici reținem că un coeficient de corelație, oricât de mare, acoperă cel mult forța asocierii și, uneori, gradientul.

9.5.2 Eroarea ecologică

Corelațiile calculate pe date agregate (țări, județe, luni) pot fi complet diferite de cele de la nivel individual. Dacă țările cu consum mai mare de grăsimi au mortalitate coronariană mai mare, nu rezultă automat că persoanele care mănâncă mai multe grăsimi mor mai des de boală coronariană; poate că în acele țări diferă și alți factori (venitul, fumatul, accesul la servicii medicale). Inferența greșită de la grup la individ se numește eroare ecologică. Corelația dintre incidența COVID-19 la personalul medical și cea din comunitate, discutată mai sus, este o corelație ecologică de acest tip: utilă pentru supravegherea epidemiologică, dar fără să lămurească sursa infecției fiecărui angajat [40].

9.5.3 Regresia spre medie

Un fenomen statistic care a păcălit multe generații de medici este regresia spre medie. Dacă selectăm pacienți pentru că au avut la o măsurătoare o valoare extremă (de exemplu, TAS peste 160 mmHg la o singură vizită), la măsurătoarea următoare valorile lor vor fi, în medie, mai apropiate de media populației, chiar fără niciun tratament. Motivul: orice măsurătoare conține și o componentă întâmplătoare (stres, cafea, manșeta pusă grăbit), iar cei selectați la o valoare extremă au avut, în medie, o componentă întâmplătoare nefavorabilă, care nu se repetă. Cu cât corelația dintre cele două măsurători este mai mică, cu atât regresia spre medie este mai mare. Fenomenul

explică de ce multe tratamente par să funcționeze în studii fără grup de control și de ce randomizarea, cu un grup de control măsurat în același fel (cap. 13), este singura protecție sigură. Termenul "regresie" însuși vine de aici: Francis Galton a observat în secolul al XIX-lea că fiii părinților foarte înalți sunt, în medie, mai puțin înalți decât părinții lor.

9.5.4 Semnificație statistică și importanță clinică

Ultima capcană din această secțiune ține de dimensiunea eșantionului. Am văzut în tabelul 9.3 că, la sute de pacienți, corelații foarte mici devin semnificative. Un p mic arată că ρ este probabil diferit de zero, dar nu că relația contează clinic. Raportați întotdeauna r (sau R^2) împreună cu p și, ideal, cu intervalul de încredere.

> **Studiu de caz: o corelație semnificativă, dar neglijabilă**
>
> Un studiu retrospectiv pe 867 de pacienți cu cancer de colon, operați într-un spital universitar din Timișoara în patru ani (doi înainte și doi în timpul pandemiei COVID-19), a urmărit schimbările din îngrijirea chirurgicală [55]. Rezultatele principale sunt comparații de proporții între perioade: scăderea intervențiilor programate ($p < 0,001$), creșterea intervențiilor de urgență ($p < 0,001$) și creșterea cazurilor în stadiul IV în al doilea an de pandemie ($p = 0,045$). Rezumatul raportează și o corelație inversă între vârsta pacientului și durata intervenției: $r = -0,088$, $p = 0,01$.
>
> Să verificăm. Cu $n = 867$, $t = \frac{0,088 \times \sqrt{865}}{\sqrt{1-0,0077}} = \frac{0,088 \times 29,41}{0,996} = 2,60$, iar pentru 865 de grade de libertate $p = 0,0095$, adică aproximativ 0,01, în acord cu rezumatul. Corelația este deci semnificativă statistic. Însă, după Colton, $|r| = 0,088 < 0,25$ înseamnă "fără corelație", iar $R^2 = 0,0077$: vârsta explică sub 1% din variația duratei operației. Ce spune rezultatul: în această serie, pacienții mai vârstnici nu au avut operații mai lungi, ci, foarte ușor, mai scurte. Ce nu spune: că vârsta ar fi un predictor util al duratei. Rezultatul este un exemplu didactic foarte bun pentru diferența dintre "semnificativ" și "important"; un cititor atent ar vrea să vadă și diagrama de corelație și, eventual, o analiză ajustată pentru tipul intervenției (urgență sau programată), care poate fi un factor de confuzie.

9.6 Corelație și concordanță: metoda Bland-Altman

O greșeală frecventă în literatura medicală, semnalată încă din 1986 de Bland și Altman, este folosirea coeficientului de corelație pentru a arăta că două metode de măsurare "se potrivesc" [102]. Să presupunem că un tensiometru nou dă sistematic cu 10 mmHg mai mult decât cel de referință. Punctele vor sta perfect pe o dreaptă paralelă cu prima bisectoare, r va fi aproape 1, dar metodele nu sunt interschimbabile: un pacient cu 135 mmHg la aparatul de referință ar fi clasificat hipertensiv la cel nou. Corelația măsoară cât de bine sunt ordonați pacienții, nu cât de apropiate sunt valorile.

Mg2026 are un exemplu natural. TAS la includere și TAS la 6 luni au $r = 0,835$ (corelație puternică, IC 95% 0,781-0,877), dar media a scăzut de la 131,1 la 126,4 mmHg. Pacienții își păstrează locul în ierarhie, dar valorile nu sunt aceleași; o corelație mare nu exclude o diferență sistematică.

Metoda Bland-Altman, care evaluează concordanța, are trei pași. Pentru fiecare pacient se calculează diferența dintre cele două metode și media lor. Se calculează media diferențelor, numită deplasare sau eroare sistematică (engl. bias), și abaterea standard a diferențelor. Se calculează limitele de concordanță (engl. limits of agreement): media diferențelor $\pm 1,96 \times$ AS diferențelor. Între aceste limite se află aproximativ 95% dintre diferențele individuale. Graficul Bland-Altman are pe orizontală media celor două metode și pe verticală diferența, cu trei linii orizontale (deplasarea și cele două limite).

Un exemplu numeric scurt: opt pacienți măsurați cu un glucometru portabil și în laborator au diferențele (glucometru minus laborator) 4, -2, 6, 3, 8, 1, 5 și -1 mg/dL. Media diferențelor este 3,0 mg/dL, iar AS a diferențelor este 3,46 mg/dL. Limitele de concordanță sunt $3{,}0 \pm 1{,}96 \times 3{,}46$, adică -3,8 până la 9,8 mg/dL. Dacă o diferență de până la 10 mg/dL este acceptabilă clinic, aparatul poate fi folosit pentru monitorizare; dacă nu, nu. Decizia este clinică, nu statistică, iar coeficientul de corelație nu intră în ea. Pentru variabile categoriale, concordanța se măsoară cu coeficientul kappa [103], discutat împreună cu testele diagnostice.

9.7 Regresia liniară simplă

9.7.1 Modelul

Corelația spune cât de strâns se adună punctele în jurul unei drepte. Regresia găsește dreapta. Modelul de regresie liniară simplă scrie fiecare valoare a lui Y ca o parte explicată de X plus o eroare:

$$y = a + b\,x + \varepsilon$$

În această scriere, a este interceptul (ordonata la origine), adică valoarea prezisă a lui Y atunci când X = 0; b este panta (coeficientul de regresie), adică modificarea medie a lui Y pentru o creștere cu o unitate a lui X; ε este eroarea reziduală, partea din Y pe care modelul nu o explică. Semnul lui b este întotdeauna același cu semnul lui r. Dreapta propriu-zisă, ŷ = a + b x, dă valoarea medie prezisă a lui Y pentru o valoare dată a lui X; un pacient anume se va afla deasupra sau dedesubt, la o distanță egală cu reziduul său, ei = yi - ŷi.

9.7.2 Metoda celor mai mici pătrate

Prin norul de puncte se pot trage nenumărate drepte. Care este "cea mai bună"? Răspunsul clasic, dat de Gauss și Legendre la începutul secolului al XIX-lea, este dreapta pentru care suma pătratelor reziduurilor este minimă:

$\text{SSE} = \sum e_i{}^2 = \sum (y_i - a - b\,x_i)^2 = \text{minim}$

Ridicarea la pătrat face ca reziduurile pozitive și negative să nu se anuleze și penalizează mai puternic abaterile mari. Rezolvând problema de minim, se obțin formulele:

$$b = \frac{S_{\text{xy}}}{S_{\text{xx}}} = r \times \left(\frac{s_y}{s_x}\right) \quad a = m_y - b \times m_x$$

Prima formă a pantei se folosește la calculul de mână, a doua când avem doar statistici rezumative (medii, abateri standard și r), situație frecventă la citirea articolelor. Din a doua formulă rezultă că dreapta trece întotdeauna prin punctul mediilor (mx, my).

Pentru cei opt pacienți din tabelul 9.1: $b = \frac{302}{146} = 2{,}068$ mmHg pentru fiecare kg/m², iar a = 131 - 2,068 x 29 = 71,01 mmHg. Dreapta este:

$$\text{TAS prezisă} = 71{,}01 + 2{,}068 \times \text{IMC}$$

La un IMC de 30 kg/m², TAS prezisă este 71,01 + 2,068 x 30 = 133,1 mmHg.

9.7.3 Interpretarea pantei și a interceptului

Panta este rezultatul clinic al regresiei. "Pentru fiecare kg/m² în plus, TAS este, în medie, mai mare cu 2,07 mmHg" este o propoziție pe care o înțelege orice clinician. Atenție la unități: dacă X ar fi fost exprimat în alte unități (de exemplu, greutatea în grame), panta s-ar fi schimbat corespunzător, deși relația ar fi rămas aceeași. Atenție și la sens: este o diferență între pacienți, observată transversal, nu o promisiune că un pacient care slăbește 1 kg/m² își va scădea tensiunea

cu 2 mmHg. Această a doua afirmație ar fi o afirmație cauzală, pe care doar un studiu de intervenție o poate susține.

Interceptul are adesea o interpretare absurdă. În exemplul nostru, 71 mmHg ar fi tensiunea "unui pacient cu IMC zero", care nu există. Interceptul este un parametru matematic necesar pentru poziționarea dreptei, dar are sens clinic doar dacă X = 0 se află în domeniul datelor (de exemplu, doza zero a unui medicament). Când vreți un intercept interpretabil, puteți centra variabila X (scădeți media din fiecare valoare); interceptul devine atunci valoarea prezisă pentru un pacient cu X egal cu media.

9.7.4 Exemplul mg2026: IMC și TAS

Pe cei 160 de pacienți din mg2026, regresia TAS în funcție de IMC dă rezultatele din tabelul 9.5, iar figura 9.4 arată norul de puncte, dreapta și banda de încredere a dreptei.

Tabelul 9.5. Regresia liniară simplă a TAS (mmHg) în funcție de IMC (kg/m²), mg2026, n = 160.

Coeficient	Estimare	Eroare standard	t	p	IC 95%
Intercept (a)	89,47	7,56	11,84	< 0,001	74,54-104,40
IMC (b)	1,435	0,258	5,56	< 0,001	0,925-1,944

$R^2 = 0{,}164$; eroarea standard a estimării (SEE) = 12,73 mmHg; F(1; 158) = 30,93, p < 0,001.

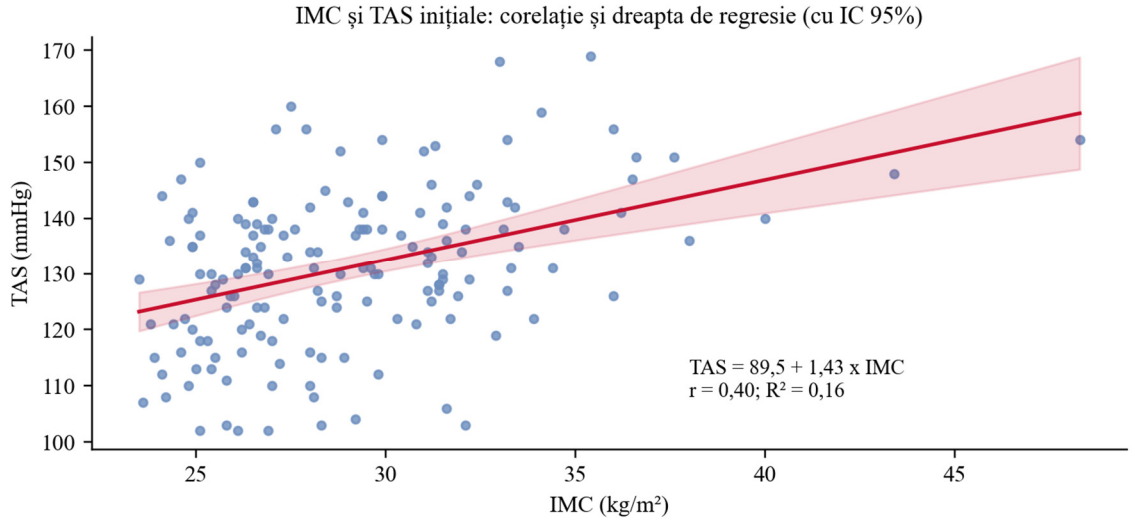


Figura 9.4. Indicele de masă corporală și tensiunea arterială sistolică la includere în setul mg2026 (n = 160): fiecare punct este un pacient, linia roșie este dreapta de regresie TAS = 89,47 + 1,435 x IMC, iar banda colorată este intervalul de încredere 95% pentru valoarea medie prezisă. Date sintetice (setul mg2026).

Ecuația este TAS = 89,47 + 1,435 x IMC. La fiecare kg/m² în plus, TAS este în medie mai mare cu 1,4 mmHg; intervalul de încredere pentru pantă (0,925-1,944) nu conține zero, deci panta este semnificativ diferită de zero. Nu e o coincidență că t pentru pantă (5,56) este exact t de la testul pentru r: în regresia simplă, testul H0: b = 0 și testul H0: ρ = 0 sunt același test, iar $F = t^2 (5{,}56^2 = 30{,}93)$. Predicțiile sunt: la IMC 30, TAS = 89,47 + 1,435 x 30 = 132,5 mmHg; la IMC 35, TAS = 139,7 mmHg. Diferența dintre ele, 7,2 mmHg, este 5 x 1,435.

Ca verificare, panta se poate obține și din statisticile rezumative: $b = \frac{r \times s_y}{s_x} = \frac{0{,}405 \times 13{,}88}{3{,}91} =$ 1,44, iar a = 131,13 - 1,435 x 29,04 = 89,46 (diferențele de la a doua zecimală vin din rotunjiri).

Pentru vârstă, ecuația este TAS = 108,3 + 0,406 x vârsta ($R^2 = 0{,}066$): la fiecare 10 ani în plus, TAS este în medie mai mare cu aproximativ 4 mmHg. Pentru glicemie și HbA1c: HbA1c =

3,38 + 0,0209 x glicemia, adică 0,21 puncte procentuale de HbA1c la fiecare 10 mg/dL de glicemie.

9.7.5 Eroarea standard a estimării și R^2

Două numere descriu cât de bine se potrivește dreapta. Coeficientul de determinare R² este o măsură relativă: proporția din variația lui Y explicată de model. Se poate scrie ca

$$R^2 = 1 - \frac{\text{SSE}}{S_{\text{yy}}}$$

adică 1 minus raportul dintre variația rămasă neexplicată după trasarea dreptei și variația totală a lui Y în jurul mediei sale. Pentru cei opt pacienți, SSE = 59,3 și Syy = 684, deci $R^2 = 1 - \frac{59,3}{684} = 0,913$, exact $r^2 = 0,956^2$.

Eroarea standard a estimării (SEE, engl. standard error of the estimate, numită și abaterea standard reziduală) este o măsură absolută, exprimată în unitățile lui Y:

$$\text{SEE} = \sqrt{\left[\frac{\text{SSE}}{n-2}\right]}$$

Ea arată cât de departe de dreaptă se află, tipic, un pacient. Pentru cei opt pacienți, SEE $= \sqrt{\frac{59,3}{6}} = 3,14$ mmHg. Pentru mg2026, SEE = 12,73 mmHg, foarte aproape de abaterea standard a TAS (13,88 mmHg). Comparația este instructivă: dacă nu am ști IMC-ul, cea mai bună estimare a TAS ar fi media, cu o împrăștiere tipică de 13,9 mmHg; cunoscând IMC, împrăștierea scade doar la 12,7 mmHg. IMC explică o parte reală, dar mică, din variație. Pe măsură ce R² crește, SEE scade, dar cele două spun lucruri diferite: R² răspunde la întrebarea "cât de mult explică modelul?", iar SEE la întrebarea "cât de precisă este o predicție individuală?". Pentru un clinician care vrea să estimeze tensiunea unui pacient, a doua întrebare este adesea mai importantă.

9.7.6 Intervalul de încredere pentru pantă

Panta estimată pe eșantion are propria eroare standard:

$$\text{ES}(b) = \frac{\text{SEE}}{\sqrt{S_{\text{xx}}}}$$

iar intervalul de încredere 95% este $b \pm t(0,975;\ n-2) \times \text{ES}(b)$. Pentru cei opt pacienți, $\text{ES}(b) = \frac{3,14}{\sqrt{146}} = 0,260$, iar t pentru 6 grade de libertate este 2,447; IC 95% $= 2,068 \pm 2,447 \times 0,260 = 2,068 \pm 0,637$, adică 1,43-2,71 mmHg pe kg/m². Pentru mg2026, ES(b) = 0,258, iar cu t ≈ 1,975 pentru 158 de grade de libertate, IC 95% $= 1,435 \pm 0,510 = 0,925 - 1,944$. Intervalul este mult mai util decât p: ne spune că, în populația din care provine eșantionul, diferența de TAS asociată cu un kg/m² se situează plauzibil între aproximativ 0,9 și 1,9 mmHg.

9.7.7 Reziduurile și verificarea condițiilor

Regresia liniară se bazează pe patru condiții, ușor de reținut după inițialele lor în engleză (LINE): relația este liniară (L), observațiile sunt independente (I), reziduurile au distribuție normală (N) și au aceeași varianță pe tot domeniul lui X (E, de la engl. equal variance, homoscedasticitate). Observați că normalitatea se cere reziduurilor, nu variabilei X. X poate fi chiar asimetric, cum este IMC în mg2026.

Condițiile se verifică pe reziduuri. Figura 9.5 arată cele două grafice standard pentru modelul TAS-IMC. În panoul A, reziduurile sunt reprezentate față de valorile prezise: dacă modelul este adecvat, punctele formează o bandă orizontală, fără curbură și fără formă de pâlnie, centrată pe

zero. O curbă ar indica o relație neliniară; o pâlnie (împrăștiere care crește cu valorile prezise) ar indica heteroscedasticitate. În panoul B, histograma reziduurilor trebuie să aibă aproximativ forma unui clopot simetric; pe mg2026, testul Shapiro-Wilk pe reziduuri dă p = 0,878, deci nu există motive să respingem normalitatea.

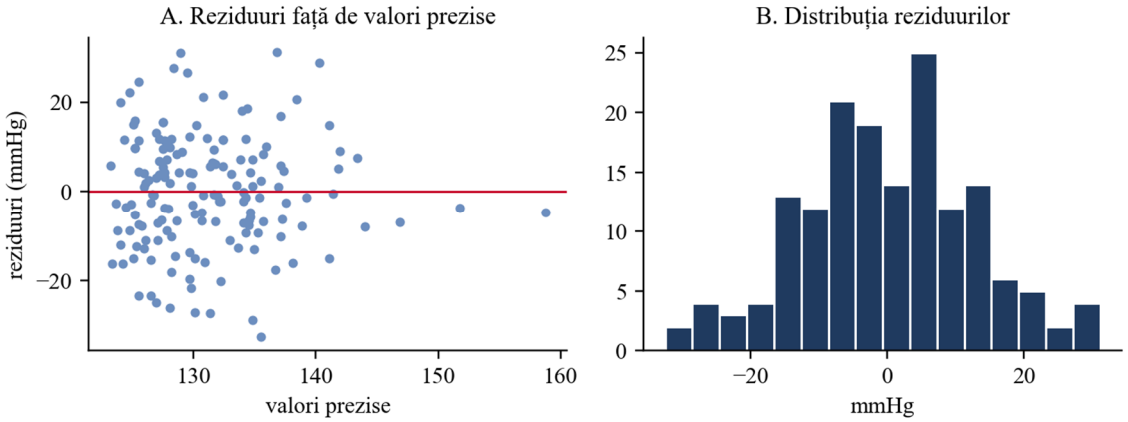


Figura 9.5. Analiza reziduurilor pentru regresia TAS = 89,47 + 1,435 x IMC (mg2026). A: reziduurile față de valorile prezise formează o bandă orizontală, fără curbură sau formă de pâlnie. B: histograma reziduurilor este aproximativ simetrică (Shapiro-Wilk p = 0,878). Date sintetice (setul mg2026).

Independența nu se vede pe aceste grafice, ci din designul studiului: fiecare pacient trebuie să apară o singură dată. Dacă am pune în aceeași regresie TAS la 0 și la 6 luni pentru aceiași pacienți, ca și cum ar fi 320 de observații independente, am încălca această condiție și am obține valori p prea mici. Pentru date ordonate în timp, programele raportează statistica Durbin-Watson (valori apropiate de 2 indică lipsa autocorelației reziduurilor; pe mg2026, în modelul multiplu din secțiunea 9.8, valoarea este 1,97).

9.7.8 Predicția și pericolul extrapolării

Odată estimată dreapta, predicția pare simplă: înlocuim X și obținem Y. Două subtilități separă însă un student bun de unul mediu.

Prima privește tipul de incertitudine. Valoarea 132,5 mmHg prezisă pentru IMC 30 este o estimare a tensiunii medii a tuturor pacienților cu IMC 30. Pentru această medie, intervalul de încredere 95% este îngust: 130,5-134,6 mmHg. Pentru un singur pacient nou cu IMC 30, incertitudinea este mult mai mare, pentru că se adaugă variația individuală (SEE = 12,7 mmHg): intervalul de predicție 95% este 107,3-157,7 mmHg. Cu alte cuvinte, modelul știe destul de bine cât este media, dar nu poate spune dacă un anumit pacient cu IMC 30 va fi normotensiv sau hipertensiv. Pentru IMC 35, intervalul de încredere al mediei este 136,1-143,3 mmHg, iar intervalul de predicție 114,3-165,1 mmHg. Banda de încredere din figura 9.4 se lărgește spre capete, pentru că acolo avem mai puține date.

A doua subtilitate este extrapolarea. Dreapta a fost estimată pe pacienți cu IMC între 23,5 și 48,3 kg/m² și nu avem nicio garanție că relația rămâne liniară în afara acestui interval. Lucrarea practică a disciplinei despre înălțime și greutate întreabă, intenționat, ce greutate ar trebui să aibă o persoană înaltă de 1,78 m și una de 2,35 m. Pe mg2026, dreapta este greutate = -100,40 + 1,087 x înălțimea (cm). La 178 cm, greutatea prezisă este -100,40 + 1,087 x 178 = 93,1 kg, o predicție legitimă, pentru că 178 cm se află în domeniul datelor (145-192 cm). La 235 cm, formula dă 155,0 kg. Numărul se calculează la fel de ușor, dar nu are sprijin în date: nimeni din eșantion nu depășește 192 cm, iar persoanele de peste 2,30 m au adesea afecțiuni (gigantism, acromegalie) care schimbă complet relația. Figura 9.6 face vizibilă problema: linia punctată roșie este extrapolare.

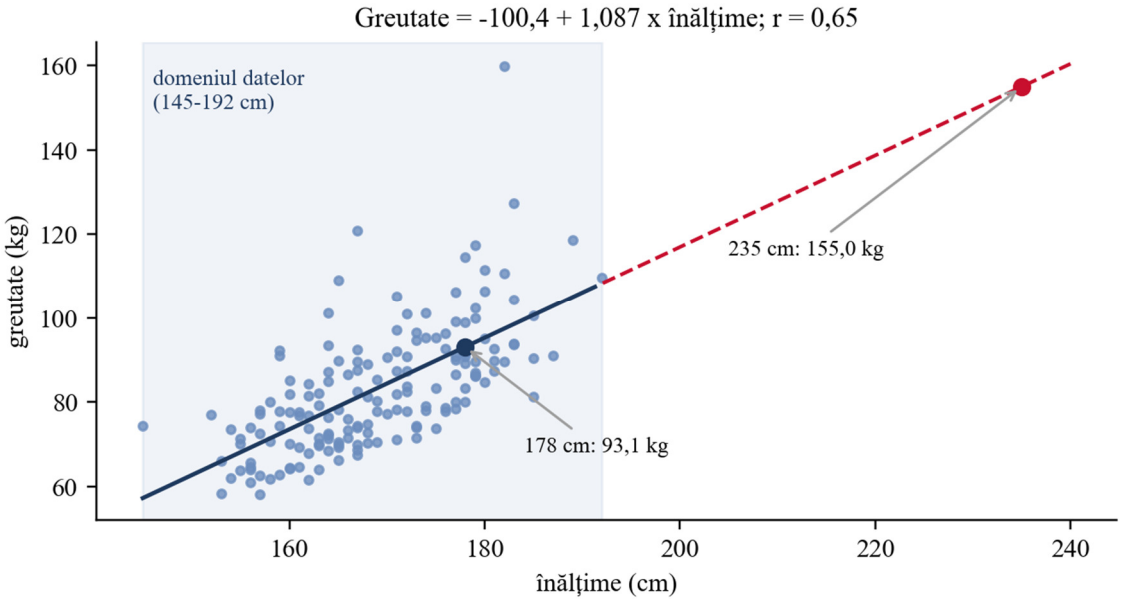


Figura 9.6. Înălțimea și greutatea în setul mg2026 (n = 160), cu dreapta de regresie greutate = -100,4 + 1,087 x înălțime (r = 0,65). Zona colorată este domeniul datelor (145-192 cm). Predicția pentru 178 cm (93,1 kg) este o interpolare; cea pentru 235 cm (155,0 kg) este o extrapolare fără sprijin în date. Date sintetice (setul mg2026).

Un exemplu și mai clar este IMC-TAS: dacă introducem un IMC de 60 kg/m², formula dă 175,6 mmHg. Poate fi adevărat, poate nu; datele noastre nu ne spun nimic despre pacienții cu obezitate extremă. La examen, întrebarea "ce valoare prezice modelul pentru..." trebuie însoțită aproape întotdeauna de un comentariu despre domeniul în care modelul este valabil.

9.8 Regresia liniară multiplă: introducere

9.8.1 De ce avem nevoie de mai mulți predictori

TAS nu depinde doar de IMC. Vârsta, sexul, fumatul, consumul de sare și mulți alți factori contribuie. Mai mult, predictorii sunt legați între ei: pacienții mai vârstnici pot avea și IMC mai mare, iar o parte din asocierea "brută" dintre IMC și TAS ar putea proveni, de fapt, din vârstă. Regresia liniară multiplă estimează efectul fiecărui predictor ținând constanți ceilalți predictori din model:

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k + \varepsilon$$

Fiecare coeficient bj se citește astfel: "modificarea medie a lui Y pentru o creștere cu o unitate a lui xj, la pacienți cu aceleași valori ale celorlalte variabile din model". Expresia "ajustat pentru" din articole înseamnă exact acest lucru. Calculul se face numai cu programul, dar citirea rezultatelor trebuie să o știe orice medic care citește literatură.

9.8.2 Variabilele categoriale în regresie

Sexul și fumatul nu sunt numerice. Ele intră în model prin variabile indicator (engl. dummy variables), codificate 0/1. Pentru sex ajunge o singură variabilă (1 = bărbat, 0 = femeie); coeficientul ei este diferența medie dintre bărbați și femei, la aceleași valori ale celorlalți predictori. Fumatul are trei categorii, deci are nevoie de două variabile indicator; categoria care nu primește indicator devine categoria de referință, cu care se compară celelalte. În modelul de mai jos, referința este "fost fumător". Alegerea referinței nu schimbă modelul, ci doar comparațiile afișate. De obicei se alege categoria cea mai numeroasă sau cea mai naturală clinic (aici ar fi putut fi nefumătorii).

9.8.3 Exemplul mg2026: TAS în funcție de IMC, vârstă, sex și fumat

Tabelul 9.6. Regresia liniară multiplă a TAS la includere (mmHg), mg2026, n = 160.

Predictor	Coeficient (B)	Eroare standard	IC 95%	p
Intercept	64,51	9,39	45,96-83,06	< 0,001
IMC $\left(\frac{\text{kg}}{m^2}\right)$	1,389	0,240	0,915-1,862	< 0,001
Vârsta (ani)	0,449	0,108	0,235-0,662	0,0001
Sex masculin (față de feminin)	-1,481	1,878	-5,19 la 2,23	0,43
Fumător (față de fost fumător)	6,738	2,636	1,53-11,95	0,012
Nefumător (față de fost fumător)	-1,119	2,487	-6,03 la 3,79	0,65

$R^2 = 0{,}308$; R^2 ajustat $= 0{,}286$; F(5; 154) = 13,72, p < 0,001.

Citim tabelul rând cu rând. La aceeași vârstă, același sex și același statut de fumător, fiecare kg/m² în plus se asociază cu o TAS mai mare cu 1,39 mmHg, foarte aproape de panta brută de 1,435. Vârsta nu "înghite" efectul IMC, ceea ce înseamnă că, în mg2026, vârsta nu este un factor de confuzie important pentru relația IMC-TAS. Fiecare an de vârstă adaugă, la același IMC, 0,45 mmHg (4,5 mmHg pe deceniu). Bărbații au TAS cu 1,5 mmHg mai mică decât femeile, dar intervalul (-5,2 la 2,2) conține zero și p = 0,43: nu avem dovezi pentru o diferență între sexe, în acord cu testul t din capitolul 6. Fumătorii actuali au TAS cu 6,7 mmHg mai mare decât foștii fumători (IC 1,5-11,9; p = 0,012), iar nefumătorii nu diferă semnificativ de foștii fumători.

Cu un model estimat putem face și predicții pentru un pacient concret. Un bărbat fumător de 60 de ani, cu IMC 32, are TAS prezisă = 64,51 + 1,389 x 32 + 0,449 x 60 - 1,481 + 6,738 = 141,2 mmHg. O femeie fostă fumătoare cu aceeași vârstă și același IMC are 64,51 + 1,389 x 32 + 0,449 x 60 = 135,9 mmHg.

9.8.4 R^2 ajustat

R² crește ori de câte ori adăugăm un predictor, chiar unul fără nicio legătură cu Y (de exemplu, ultima cifră a codului numeric personal), pentru că orice variabilă "prinde" o mică parte din variația întâmplătoare. R² ajustat penalizează numărul de predictori:

$$R^2 \text{ ajustat } = 1 - \frac{(1 - R^2) \times (n - 1)}{n - k - 1}$$

unde k este numărul de predictori (aici 5, pentru că fumatul contează de două ori). Pentru mg2026: $1 - \frac{(1-0{,}308)\times 159}{154} = 1 - 0{,}692 \times 1{,}0325 = 0{,}286$. Diferența mică dintre R² și R² ajustat arată că modelul nu are predictori în exces față de mărimea eșantionului. O regulă practică frecvent citată cere cel puțin 10-15 observații pentru fiecare predictor; cu 160 de pacienți și 5 predictori suntem departe de limită.

Modelul multiplu explică 30,8% din variația TAS, aproape dublu față de IMC singur (16,4%). Rămâne totuși aproape 70% neexplicat. În cercetarea clinică, valori ale R² de 0,2-0,4 pentru variabile biologice sunt obișnuite; nu înseamnă că modelul este "prost", ci că tensiunea arterială are multe cauze pe care nu le-am măsurat.

9.8.5 Coeficienți nestandardizați și standardizați

Articolele raportează uneori coeficientul nestandardizat B (în unitățile lui Y pe unitatea lui X, ca în tabelul 9.6), alteori coeficientul standardizat β (beta), obținut după transformarea tuturor

variabilelor în scoruri z. Beta se citește "cu câte abateri standard se modifică Y pentru o abatere standard a lui X" și permite compararea predictorilor măsurați în unități diferite. Din păcate, notația nu este uniformă: unii autori notează cu β chiar coeficientul nestandardizat. Când citiți un articol, verificați în secțiunea de metode sau din unități (un β exprimat în cm sau în kg/m² este nestandardizat).

9.8.6 Coliniaritatea

Dacă doi predictori sunt foarte puternic corelați între ei, modelul nu poate separa efectele lor, iar coeficienții devin instabili și au erori standard mari. Pe mg2026, IMC și circumferința taliei au r = 0,885: ambele măsoară adipozitatea. Introducerea lor simultană într-un model pentru TAS ar produce coeficienți greu de interpretat. Programele raportează factorul de inflație a varianței (VIF, engl. variance inflation factor); valori peste 5-10 semnalează coliniaritate.

9.8.7 Cum se citește un tabel de regresie într-un articol

Pentru un tabel de regresie dintr-un articol, o ordine de lectură utilă este următoarea. Identificați variabila dependentă și tipul de regresie (liniară pentru Y numeric, logistică pentru Y binar, Cox pentru timp până la eveniment). Verificați câte observații au intrat efectiv în model (pacienții cu date lipsă sunt de obicei excluși) și câți predictori sunt. Citiți coeficientul fiecărui predictor de interes împreună cu unitatea de măsură și cu intervalul de încredere; un coeficient fără unitate nu poate fi interpretat. Căutați categoriile de referință pentru variabilele categoriale. Uitați-vă la R^2 (sau R^2 ajustat) pentru a judeca cât explică modelul. În fine, întrebați-vă ce variabile lipsesc: un factor de confuzie omis poate face ca un coeficient să pară semnificativ. Ghidurile de raportare pentru studii observaționale cer explicit descrierea variabilelor de ajustare și a modului în care au fost alese [25].

Studiu de caz: regresia multiplă în articolele colegilor

Un studiu transversal pe 220 de adulți cu boli dermatologice cronice a analizat aderența la tratament (scorul MARS-5, 21,50 ± 4,19) în funcție de satisfacția față de serviciile farmaceutice și de sprijinul social [72]. Rezumatul raportează un coeficient nestandardizat B = 0,076 (IC 95% 0,017-0,135; p = 0,011) și o creștere a varianței explicate de la 13,3% la 28,0% după adăugarea altor variabile în model; o asociere a ajuns la limita semnificației (p = 0,053) după ajustare. Lecția: R^2 crește odată cu adăugarea predictorilor relevanți, iar un coeficient se poate atenua când introducem o variabilă care mediază sau confundă relația. Interpretarea lui B cere scara predictorului: un coeficient de 0,076 pare mic, dar dacă predictorul este un scor cu amplitudine mare (scorul de satisfacție PSPSQ-2 are 62,92 ± 9,57), diferența dintre pacienți aflați la capetele scalei poate deveni relevantă. Fără unitate, coeficientul nu spune nimic; cu unitate, devine o mărime clinică.

Un studiu transversal multivariat pe 245 de persoane a comparat stima de sine la purtătorii de proteze auditive inteligente, la nepurtători și la persoane cu auz normal [82]. Pe lângă ANOVA (F = 299,00, p < 0,001, η^2 = 0,712), autorii raportează corelații Pearson: stima de sine se corelează negativ cu handicapul auditiv (r = -0,573, p < 0,001) și cu tinitusul (r = -0,443, p < 0,001), iar cele două scale se corelează între ele pozitiv (r = 0,729, p < 0,001). Modelul de regresie multiplă, cu un coeficient de aproximativ 11,9 pentru purtarea protezei, explică aproximativ 65% din variația stimei de sine. Un R^2 atât de mare este neobișnuit în cercetarea clinică și se explică prin diferența foarte mare dintre grupe. Corelația de 0,729 dintre doi predictori potențiali semnalează și riscul de coliniaritate, dacă ambii intră în același model.

În cardiologie, un studiu caz-control pe 60 de pacienți cu insuficiență cardiacă ischemică și 60 de martori de aceeași vârstă a raportat valori ale osteoprotegerinei mai mari la bolnavi (4,7 ± 0,25 față de 1,3 ± 0,67 ng/mL) și ale vitaminei D mai mici (20,49 ± 7,31 față de

37,09 ± 4,59 ng/mL), apoi corelații Pearson foarte puternice ale markerilor studiați cu volumul telediastolic (r = 0,862), cu fracția de ejecție (r = -0,832) și cu viteza undei pulsului (r = 0,833), toate cu p < 0,001, o corelație moderată cu vitamina D (r = -0,636) și una slabă cu rata de filtrare glomerulară (r = -0,488) [87]. Autorii au continuat cu o regresie multiplă în care vitamina D a fost variabila dependentă. Observați că, pentru că studiul include atât bolnavi, cât și sănătoși, o parte din corelațiile puternice poate proveni din diferența dintre cele două grupe: în interiorul fiecărei grupe, relația ar putea fi mult mai slabă (situația celor două subgrupe din secțiunea 9.1.2).

Un studiu de cohortă retrospectivă pe 186 de pacienți după infarct miocardic acut a folosit regresia liniară multiplă pentru remodelarea ventriculară în funcție de numărul de clase de tratament recomandat de ghiduri (GDMT) [109]. Rezultatele sunt date ca β = 0,120 cm (p = 0,004) și β = 0,204 cm (p < 0,001) pentru fiecare clasă suplimentară. Unitatea (cm) arată că sunt coeficienți nestandardizați, interpretați direct ca diferență de diametru pentru fiecare clasă de tratament în plus. Numai 18,8% dintre pacienți au primit toți cei patru piloni terapeutici, ceea ce limitează precizia la capătul de sus al scalei.

Două studii ale colectivului de nutriție arată regresia liniară în epidemiologia nutrițională. Într-o cohortă prospectivă pe 921 de studenți urmăriți în primul an de facultate, modificarea medie a IMC a fost -0,45 ± 1,07 kg/m² (IC 95% -0,52 la -0,38), iar un comportament de crononutriție s-a asociat cu o diferență ajustată de $\beta = +\frac{0{,}22\ \text{kg}}{m^2} (p = 0{,}004)$ [5]. Într-un studiu transversal pe 693 de studenți, tiparele alimentare s-au asociat cu HbA1c: β = 0,147 (p < 0,001) pentru un tipar și β = -0,109 (p < 0,001) pentru altul, de sens invers [6]. Ambele rezumate raportează coeficienți cu valoarea p, fără interval de încredere; pentru un cititor, intervalul ar fi arătat cât de precis este estimat efectul.

În fine, regresia poate construi un scor clinic. Într-un studiu pe 124 de pacienți cu ciroză compensată, autorii au combinat rigiditatea splinei (SS), rigiditatea ficatului (LS) și dimensiunea splinei (SSZ) într-un scor obținut prin regresie multiplă: 0,053 x SS + 0,054 x LS + 0,059 x SSZ - 1,84, cu un prag > 0,34 și o arie sub curba ROC de 0,93, mai mare decât a fiecărui parametru luat separat (0,84; 0,86; 0,71) [110]. Formula are exact structura b0 + b1 x1 + b2 x2 + b3 x3 din secțiunea 9.8.1. Validitatea ei pe alți pacienți decât cei pe care a fost construită trebuie verificată separat, într-o cohortă de validare.

9.9 Regresia logistică: o punte spre capitolul 10

9.9.1 Când Y are doar două valori

Ce facem când variabila rezultat este binară: diabet da/nu, deces da/nu, scădere în greutate de cel puțin 5% da/nu? Regresia liniară nu se potrivește: ar putea prezice "probabilități" negative sau mai mari de 1, iar reziduurile nu pot fi normale. Regresia logistică modelează nu probabilitatea p direct, ci logaritmul șansei (engl. log odds sau logit):

$$\ln\left[\frac{p}{1-p}\right] = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$$

Șansa (engl. odds) este raportul dintre probabilitatea de a avea evenimentul și probabilitatea de a nu-l avea; o probabilitate de 20% înseamnă o șansă de 0,2 / 0,8 = 0,25, adică "1 la 4". Capitolul 10 explică pe larg această noțiune. Din punct de vedere practic, acum trebuie reținut că exponențiala unui coeficient, e^b, este un raport de șanse (engl. odds ratio, OR): de câte ori se înmulțește șansa evenimentului când predictorul crește cu o unitate, la valori constante ale celorlalți predictori.

9.9.2 Exemplul mg2026: diabetul confirmat prin TTGO

Tabelul 9.7. Regresia logistică a diabetului confirmat prin TTGO (29 de cazuri din 160), mg2026.

Predictor	OR ajustat	IC 95%	p
Obezitate (IMC ≥ 30), față de IMC < 30	6,01	2,35-15,39	0,0002
Vârsta, pentru fiecare an	1,11	1,05-1,18	0,0003
Sex masculin, față de feminin	0,97	0,40-2,38	0,95

Citire: la aceeași vârstă și același sex, șansa de a avea diabet este de aproximativ 6 ori mai mare la pacienții obezi decât la cei neobezi; intervalul larg (2,35-15,39) reflectă numărul mic de cazuri (29). Fiecare an de vârstă înmulțește șansa cu 1,11, adică o creștere de 11% pe an, la același statut ponderal și același sex. Pentru 10 ani, efectele se înmulțesc, nu se adună: 1,11^10 ≈ 2,8, deci un pacient cu 10 ani mai în vârstă are o șansă de aproape trei ori mai mare. Sexul nu contează (OR 0,97, interval care conține 1, p = 0,95). Observați regula care se va repeta în capitolul 10: pentru rapoarte (OR, RR), valoarea "fără efect" este 1, nu 0, iar un interval care conține 1 corespunde unui rezultat nesemnificativ.

În capitolul 10 vom calcula de mână, din tabelul 2 x 2, OR brut dintre obezitate și diabet: 4,32 (IC 95% 1,86-10,03). OR ajustat pentru vârstă și sex (6,01) este mai mare decât cel brut. Diferența nu înseamnă o eroare: ajustarea pentru un factor puternic legat de rezultat (vârsta) poate modifica OR, chiar dacă acel factor este slab legat de expunere. Pentru acum ajunge să rețineți că OR brut și OR ajustat răspund la întrebări diferite și că un articol bun le raportează pe amândouă.

9.9.3 Aceeași întrebare, două modele

Studiul prospectiv pe 921 de studenți citat mai sus arată cum alegerea dintre regresia liniară și cea logistică depinde de forma în care tratăm variabila rezultat [5]. Modificarea IMC, ca variabilă continuă, a fost analizată prin regresie liniară $\left(\beta = +\frac{0{,}22\ \text{kg}}{m^2}\right)$. Aceeași modificare, dihotomizată ca "creștere în greutate semnificativă clinic" (prezentă la 7,2% dintre studenți), a fost analizată prin regresie logistică. Prima abordare folosește toată informația; a doua răspunde la o întrebare clinică mai directă (cine ajunge la o creștere relevantă?), dar pierde informație prin tăierea variabilei continue. Un studiu transversal pe 2976 de studenți din județul Timiș a folosit regresia logistică pentru identificarea factorilor asociați cu fumatul intens [22]; predictorii identificați astfel sunt factori asociați, nu cauze demonstrate, discuție pe care o continuăm în capitolul 10.

9.10 Probleme rezolvate

Problema rezolvată 9.1. Pearson și regresia, de mână, pe opt pacienți

Enunț. Pentru cei opt pacienți din tabelul 9.1 (IMC și TAS): (a) calculați coeficientul Pearson și interpretați-l după Colton; (b) testați semnificația la α = 0,05 și calculați IC 95% pentru ρ; (c) calculați R^2; (d) estimați dreapta de regresie a TAS în funcție de IMC; (e) calculați SEE și IC 95% pentru pantă; (f) preziceți TAS pentru IMC 30 și comentați o predicție pentru IMC 45.

Rezolvare. (a) Din tabelul 9.1: Sxy = 302, Sxx = 146, Syy = 684. $r = \frac{302}{\sqrt{146 \times 684}} = 0{,}956$. După Colton, |r| ≥ 0,75: corelație directă puternică.

(b) Ipoteze: H0: ρ = 0; H1: ρ ≠ 0. $t = \frac{0{,}956 \times \sqrt{6}}{\sqrt{1-0{,}9133}} = \frac{2{,}342}{0{,}2945} = 7{,}95$; gl = 6; t critic = 2,447. Cum 7,95 > 2,447, H0 se respinge (p = 0,0002). IC 95% prin transformarea Fisher: $z = 0{,}5 \times \ln\left(\frac{1{,}956}{0{,}044}\right) = 1{,}893$; ES = $\frac{1}{\sqrt{5}} = 0{,}447$; $z \pm 0{,}877$ dă 1,016 și 2,770; transformat înapoi: 0,769-0,992.

(c) $R^2 = 0{,}956^2 = 0{,}913$. Aproximativ 91% din variația TAS în acest mic grup se asociază liniar cu IMC.

(d) $b = \frac{302}{146} = 2{,}068$; a = 131 - 2,068 x 29 = 71,01. Ecuația: TAS = 71,01 + 2,068 x IMC.

(e) Valorile prezise sunt 118,6; 122,7; 124,8; 128,9; 131,0; 135,1; 139,3 și 147,5, iar reziduurile -0,6; 3,3; -3,8; 0,1; 4,0; -4,1; 0,7 și 0,5. Suma reziduurilor este zero (o verificare utilă). SSE = 59,3; SEE $= \sqrt{\frac{59{,}3}{6}} = 3{,}14$ mmHg. ES$(b) = \frac{3{,}14}{\sqrt{146}} = 0{,}260$; IC $95\% = 2{,}068 \pm 2{,}447 \times 0{,}260 = 1{,}43 - 2{,}71$.

(f) TAS$(30) = 71{,}01 + 2{,}068 \times 30 = 133{,}1$ mmHg. Pentru IMC 45 formula dă 164,1 mmHg, dar valoarea 45 este în afara domeniului observat (23-37 kg/m²): este o extrapolare, iar cu doar opt pacienți nu avem niciun temei să credem că relația rămâne liniară acolo.

Interpretare clinică: în acest grup, pacienții cu IMC mai mare au tensiunea mai mare, cu aproximativ 2 mmHg pe kg/m². Rezultatul este semnificativ, dar eșantionul este foarte mic, iar intervalele pentru ρ (0,77-0,99) și pentru pantă (1,4-2,7) sunt largi.

Problema rezolvată 9.2. Spearman cu valori egale

Enunț. Opt pacienți și-au autoevaluat activitatea fizică pe o scală de la 1 (sedentar) la 5 (foarte activ) și au IMC măsurat: activitate 1, 2, 2, 3, 3, 3, 4, 5; IMC 33,5; 31,0; 29,4; 30,2; 27,8; 28,6; 26,1; 24,9 kg/m². Calculați coeficientul Spearman și comparați rezultatul formulei scurte cu cel exact.

Rezolvare. Activitatea este ordinală, deci alegem Spearman. Rangurile activității: valoarea 1 primește rangul 1; cele două valori 2 ar ocupa locurile 2 și 3, deci primesc 2,5 fiecare; cele trei valori 3 ar ocupa locurile 4, 5 și 6, deci primesc 5 fiecare; 4 primește 7 și 5 primește 8. Rangurile IMC (în ordine crescătoare): 24,9 primește 1; 26,1 primește 2; 27,8 primește 3; 28,6 primește 4; 29,4 primește 5; 30,2 primește 6; 31,0 primește 7; 33,5 primește 8. Pe pacienți, în ordinea din enunț: rangurile activității sunt 1; 2,5; 2,5; 5; 5; 5; 7; 8, iar rangurile IMC sunt 8; 7; 5; 6; 3; 4; 2; 1.

Diferențele di sunt -7; -4,5; -2,5; -1; 2; 1; 5; 7, iar $\sum {d_i}^2 = 49 + 20{,}25 + 6{,}25 + 1 + 4 + 1 + 25 + 49 = 155{,}5$. Cu formula scurtă: $r_s = 1 - \frac{6 \times 155{,}5}{8 \times 63} = 1 - \frac{933}{504} = 1 - 1{,}851 = -0{,}851$.

Calculul exact (Pearson pe ranguri, cum face orice program) dă rs = -0,908, p = 0,002. Diferența (-0,85 față de -0,91) apare din cauza celor cinci valori egale ale activității, care reduc variația rangurilor; formula scurtă presupune că rangurile sunt 1, 2,..., n fără repetiții. Concluzie: corelație inversă puternică în acest mic grup, pacienții mai activi au IMC mai mic. Pe mg2026, aceeași analiză pe 160 de pacienți a dat rs = -0,012, practic zero. Exemplul de opt pacienți a fost construit pentru calcul, iar contrastul ne amintește că un eșantion mic poate produce corelații spectaculoase.

Problema rezolvată 9.3. Înălțimea și greutatea, în formatul temelor disciplinei

Enunț. Pe setul mg2026, investigați asocierea dintre înălțime și greutate. Răspundeți în formatul a-g și precizați: ce proporție din variația greutății este datorată altor factori decât înălțimea? Ce greutate prezice modelul pentru o persoană de 1,78 m și pentru una de 2,35 m?

(a) Ipoteza de cercetare: persoanele mai înalte au, în medie, greutate mai mare.

(b) Ipotezele statistice: H0: ρ = 0, înălțimea și greutatea nu sunt corelate liniar în populație; H1: ρ ≠ 0, cele două variabile sunt corelate.

(c) Metoda: două variabile numerice măsurate pe aceiași pacienți, deci analiză bivariată de corelație și regresie. Înălțimea este aproximativ normală (Shapiro-Wilk p = 0,078);

greutatea este asimetrică (asimetrie 1,29; p < 0,001), de aceea raportăm Pearson și, ca verificare, Spearman. Predictorul X este înălțimea, răspunsul Y este greutatea.

(d) Rezolvare: JASP (Regression > Correlation și Regression > Linear Regression) sau Google Sheets (=CORREL, =SLOPE, =INTERCEPT, =RSQ).

(e) Rezultate: r = 0,652 (IC 95% 0,552-0,733), t = 10,79, gl 158, p < 0,001; Spearman rs = 0,693, p < 0,001. Dreapta: greutate = -100,40 + 1,087 x înălțimea (cm); IC 95% pentru pantă 0,888-1,286 kg/cm; $R^2 = 0{,}424$; SEE = 11,49 kg. Graficul este figura 9.6.

(f) Concluzia statistică: p < 0,001, H0 se respinge; corelația este pozitivă și moderată după Colton ($0{,}50 \leq r < 0{,}75$).

(g) Concluzia de cercetare: persoanele mai înalte cântăresc mai mult, în medie cu 1,09 kg pentru fiecare centimetru. $R^2 = 0{,}424$, deci 42,4% din variația greutății se asociază cu înălțimea, iar 1 - 0,424 = 0,576, adică 57,6%, ține de alți factori (sex, compoziție corporală, alimentație, activitate). Pentru 178 cm, greutatea prezisă este -100,40 + 1,087 x 178 = 93,1 kg (interval de predicție 95% pentru o persoană: 70,2-115,9 kg). Pentru 235 cm, formula dă 155,0 kg, dar este o extrapolare: cea mai mare înălțime din eșantion este 192 cm. Răspunsul corect la a doua întrebare este că modelul nu poate face o predicție credibilă.

Problema rezolvată 9.4. Verificarea unor corelații publicate

Enunț. (a) Un studiu pe 867 de pacienți raportează r = -0,088 și p = 0,01 [55]. Verificați valoarea p. (b) Un studiu pe 121 de studenți raportează R = 0,107, p = 0,246 [16]. Verificați. (c) Un studiu pe 157 de pacienți raportează ρ = 0,826 cu IC 95% 0,769-0,870 [107]. Verificați intervalul prin transformarea Fisher.

Rezolvare. $(a)\ t = \frac{0{,}088 \times \sqrt{865}}{\sqrt{1-0{,}088^2}} = \frac{2{,}588}{0{,}996} = 2{,}60$ (în valoare absolută); cu 865 de grade de libertate, p = 0,0095, adică aproximativ 0,01. Rezultatul este confirmat; corelația este semnificativă, dar neglijabilă ($R^2 = 0{,}8\%$).

(b) $t = \frac{0{,}107 \times \sqrt{119}}{\sqrt{1-0{,}0114}} = \frac{1{,}167}{0{,}994} = 1{,}17$; cu 119 grade de libertate, p ≈ 0,24, foarte aproape de 0,246 raportat (diferența vine din rotunjirea lui R). H0 nu se respinge.

(c) $z = 0{,}5 \times \ln\left(\frac{1{,}826}{0{,}174}\right) = 0{,}5 \times \ln(10{,}494) = 1{,}175$. ES $= \frac{1}{\sqrt{157-3}} = \frac{1}{\sqrt{154}} = 0{,}0806$. $z \pm 1{,}96 \times 0{,}0806 = 1{,}175 \pm 0{,}158$, adică 1,017 și 1,333. Transformăm înapoi cu $r = \frac{e^{2z}-1}{e^{2z}+1}$. Pentru z = 1,017: $e^{2{,}034} = 7{,}64$, deci $r = \frac{6{,}64}{8{,}64} = 0{,}769$. Pentru z = 1,333: $e^{2{,}666} = 14{,}38$, deci $r = \frac{13{,}38}{15{,}38} = 0{,}870$. Intervalul publicat este reprodus exact. Lecția: când un articol dă r și n, puteți reconstrui singuri intervalul de încredere.

Problema rezolvată 9.5. Dreapta de regresie din statistici rezumative

Enunț. Un articol raportează pentru circumferința taliei (X) și TAS (Y): mx = 88,5 cm, sx = 12,1 cm, my = 131,1 mmHg, sy = 13,9 mmHg, r = 0,32, n = 160. (a) Scrieți dreapta de regresie. (b) Preziceți TAS pentru o talie de 100 cm. (c) Ce proporție din variația TAS este asociată cu talia? (d) Este corelația semnificativă?

Rezolvare. $(a)\ b = \frac{r \times s_y}{s_x} = \frac{0{,}32 \times 13{,}9}{12{,}1} = 0{,}368$ mmHg/cm; a = 131,1 - 0,368 x 88,5 = 98,6 mmHg. TAS = 98,6 + 0,368 x talia.

(b) $\text{TAS}(100) = 98{,}6 + 0{,}368 \times 100 = 135{,}3$ mmHg. Valoarea de 100 cm se află în domeniul obișnuit al taliei la adulții supraponderali, deci predicția este o interpolare.

(c) $R^2 = 0{,}32^2 = 0{,}102$; aproximativ 10% din variația TAS se asociază liniar cu talia, iar 90% ține de alți factori.

(d) $t = \frac{0{,}32 \times \sqrt{158}}{\sqrt{1-0{,}1024}} = \frac{4{,}022}{0{,}947} = 4{,}25$; gl 158; p < 0,001. Corelația este semnificativă și slabă după Colton. Pe mg2026, calculul direct dă r = 0,322 și TAS = 98,49 + 0,369 x talia, foarte aproape de rezultatul obținut din statisticile rotunjite.

Problema rezolvată 9.6. Citirea unui tabel de regresie multiplă

Enunț. Folosind tabelul 9.6, răspundeți: (a) Care este diferența medie de TAS între doi pacienți de același sex, cu același statut de fumat și același IMC, dintre care unul are 50 de ani și celălalt 70? (b) Este sexul un predictor semnificativ? (c) De ce coeficientul IMC (1,389) diferă puțin de panta din regresia simplă (1,435)? (d) Interpretați R^2 ajustat.

Rezolvare. (a) 20 de ani $\frac{x\ 0{,}449\ \text{mmHg}}{\text{an}} = 9{,}0$ mmHg; pacientul de 70 de ani are, în medie, TAS cu aproximativ 9 mmHg mai mare. Intervalul pentru 20 de ani se obține înmulțind capetele: 20 x 0,235 = 4,7 și 20 x 0,662 = 13,2 mmHg.

(b) Nu: coeficientul -1,481 are IC 95% de la -5,19 la 2,23, care conține 0, și p = 0,43. Nu avem dovezi că bărbații și femeile diferă ca TAS după ajustarea pentru ceilalți factori.

(c) În regresia simplă, panta IMC preia și partea din asociere care se suprapune cu alți factori corelați cu IMC (vârstă, fumat). După ajustare, coeficientul se referă la pacienți comparabili ca vârstă, sex și fumat. Schimbarea foarte mică (de la 1,435 la 1,389) arată că acești factori nu confundă serios relația IMC-TAS în mg2026.

(d) R^2 ajustat $= 0{,}286$: după penalizarea pentru cei 5 predictori, modelul explică aproximativ 29% din variația TAS. Diferența mică față de R^2 (0,308) arată că nu există predictori inutili în exces.

9.11 Exerciții propuse

1. La șase pacienți s-au măsurat vârsta (ani) și TAS (mmHg): (35; 118), (42; 121), (50; 130), (58; 128), (63; 140), (70; 145). Calculați de mână r, testați-l la α = 0,05, scrieți dreapta de regresie a TAS în funcție de vârstă și preziceți TAS la 60 de ani.
2. Un studiu raportează r = 0,30 între consumul zilnic de sare și TAS. Testați semnificația pentru n = 40 și pentru n = 200. Ce concluzie generală trageți despre relația dintre n și semnificație? Cum clasificați corelația după Colton în ambele cazuri?
3. Pentru r = 0,62 obținut pe 50 de pacienți, calculați IC 95% pentru ρ prin transformarea Fisher. Poate fi corelația din populație "puternică" după Colton?
4. Șapte pacienți au scorul durerii pe scala vizual-analogică (VAS) 2, 5, 3, 8, 6, 9, 4 și scorul de anxietate 10, 14, 9, 20, 18, 17, 12. Calculați de mână coeficientul Spearman și interpretați-l.
5. Pe mg2026, glicemia a jeun și HbA1c au r = 0,738. Calculați R^2 și 1 - R^2 și formulați câte o propoziție pentru fiecare, în stilul temelor disciplinei. Clasificați corelația după Colton.
6. Clasificați după regula lui Colton următorii coeficienți și precizați sensul: 0,81; -0,573; 0,256; -0,088; 0,729; -0,832; 0,107; 0,50.
7. Diferențele dintre un tensiometru automat și cel de referință, la opt pacienți, sunt (mmHg): 4, -2, 6, 3, 8, 1, 5, -1. Calculați deplasarea medie și limitele de concordanță Bland-Altman. De ce coeficientul de corelație dintre cele două aparate nu ar fi răspuns la întrebarea "se pot folosi interschimbabil?"

8. Un model logistic pentru diabet dă pentru vârstă OR = 1,11 pe an. Calculați OR aproximativ pentru o diferență de 5 ani și de 10 ani. Este corect să spunem că diabetul este "de 1,11 ori mai probabil" la fiecare an în plus? Explicați diferența dintre șansă și probabilitate.
9. Folosind tabelul 9.6, calculați TAS prezisă pentru o femeie nefumătoare de 45 de ani cu IMC 27 și pentru un bărbat fumător de 65 de ani cu IMC 35. Ce variabilă contribuie cel mai mult la diferența dintre ei?
10. O regresie a greutății (kg) în funcție de înălțime (cm) pe 50 de studenți dă greutate = -95 + 0,95 x înălțimea, cu $R^2 = 0{,}38$, înălțimile fiind între 155 și 195 cm. (a) Interpretați panta. (b) Ce proporție din variația greutății ține de alți factori? (c) Preziceți greutatea pentru 1,70 m și discutați predicția pentru 1,20 m.

Capcane frecvente

- Raportarea unui coeficient de corelație fără diagrama de corelație; o singură valoare extremă poate crea sau ascunde o corelație (figura 9.3).
- Confuzia dintre semnificație și intensitate: r = -0,088 cu p = 0,01 rămâne o corelație neglijabilă.
- Interpretarea corelației ca dovadă de cauzalitate, ignorând factorii de confuzie, cauzalitatea inversă și eroarea ecologică.
- Folosirea lui r pentru a demonstra concordanța a două metode de măsurare, în loc de metoda Bland-Altman.
- Citirea lui R² ca proporție "cauzată" de X; R² este proporția din variație asociată liniar cu X.
- Interpretarea clinică a interceptului atunci când X = 0 nu are sens (IMC zero, înălțime zero).
- Extrapolarea dreptei de regresie în afara domeniului datelor (greutatea la 2,35 m).
- Verificarea normalității pe variabila X în loc de reziduuri; în regresie, condiția privește reziduurile.
- Folosirea formulei scurte Spearman când există multe valori egale.
- Formularea "se acceptă H0" când p > 0,05; corect este "H0 nu se respinge".

De reținut

- Corelația măsoară sensul și intensitatea asocierii liniare (Pearson) sau monotone (Spearman); regresia descrie cum se modifică Y în medie cu X și permite predicția.
- $r = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}}$, între -1 și 1; testul folosește $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ cu n - 2 grade de libertate; IC se calculează prin transformarea Fisher.
- Regula lui Colton: |r| < 0,25 fără corelație; 0,25-0,50 slabă; 0,50-0,75 moderată; ≥ 0,75 puternică.
- R² = proporția din variația lui Y explicată de model; 1 - R² = proporția datorată altor factori.
- $b = \frac{S_{xy}}{S_{xx}} = \frac{r \times s_y}{s_x}$; a = my - b x mx; dreapta trece prin punctul mediilor.
- SEE (în unitățile lui Y) arată precizia predicției individuale; intervalul de predicție pentru un pacient este mult mai larg decât intervalul de încredere pentru medie.
- Condițiile regresiei (LINE) se verifică pe reziduuri: liniaritate, independență, normalitate, varianță constantă.
- În regresia multiplă, fiecare coeficient este ajustat pentru ceilalți predictori; în regresia logistică, e^b este un OR, iar valoarea fără efect este 1.
- Pe mg2026: IMC-TAS r = 0,405, rs = 0,372, $R^2 = 0{,}164$, TAS = 89,47 + 1,435 x IMC (IC pantă 0,925-1,944), SEE 12,73 mmHg.

Întrebări de autoevaluare

1. (CS) Pentru două variabile numerice, normal distribuite, măsurate pe aceiași 50 de pacienți, cea mai potrivită măsură a intensității asocierii liniare este:
    a) testul t pentru eșantioane independente
    b) coeficientul de corelație Pearson
    c) testul chi-pătrat
    d) coeficientul kappa
    e) raportul de șanse
2. (CS) Un coeficient de corelație r = -0,62 indică, după regula lui Colton:
    a) o corelație inversă puternică
    b) lipsa corelației, deoarece r este negativ
    c) o corelație inversă moderată
    d) o corelație directă moderată
    e) o corelație inversă slabă
3. (CS) Între înălțime și greutate s-a obținut r = 0,60. Proporția din variația greutății datorată altor factori decât înălțimea este:
    a) 40%
    b) 36%
    c) 60%
    d) 64%
    e) 16%
4. (CS) Dreapta de regresie TAS = 89,47 + 1,435 x IMC a fost obținută pe pacienți cu IMC între 23,5 și 48,3. Afirmația corectă este:
    a) un pacient cu IMC 0 are TAS 89,47 mmHg
    b) TAS prezisă pentru IMC 30 este 132,5 mmHg și reprezintă media estimată pentru pacienții cu IMC 30
    c) panta arată că IMC cauzează creșterea TAS cu 1,435 mmHg la fiecare pacient
    d) predicția pentru IMC 70 este la fel de credibilă ca cea pentru IMC 30
    e) r este negativ, deoarece interceptul este pozitiv
5. (CS) Pentru a verifica dacă un glucometru portabil poate înlocui determinarea de laborator, metoda potrivită este:
    a) coeficientul Pearson între cele două determinări
    b) regresia logistică
    c) metoda Bland-Altman, cu deplasarea medie și limitele de concordanță
    d) coeficientul Spearman
    e) testul chi-pătrat de independență
6. (CM) Despre coeficientul de corelație Pearson sunt adevărate:
    a) nu are unitate de măsură
    b) este sensibil la valori extreme
    c) o valoare apropiată de zero exclude orice relație între variabile
    d) se modifică dacă greutatea este exprimată în grame în loc de kilograme
    e) are același semn ca panta dreptei de regresie
7. (CM) Coeficientul de corelație Spearman este preferat celui Pearson atunci când:

a) una dintre variabile este ordinală (de exemplu, activitatea fizică pe scala 1-5)
b) distribuțiile sunt puternic asimetrice sau există valori extreme
c) ambele variabile sunt normale și relația este liniară
d) relația este monotonă, dar neliniară
e) variabilele sunt nominale cu trei categorii

8. (CM) Pe un eșantion de 867 de pacienți s-a obținut, între vârstă și durata intervenției chirurgicale, r = -0,088, p = 0,01. Sunt corecte:
   a) corelația este semnificativă statistic
   b) după Colton, corelația este neglijabilă
   c) vârsta explică aproximativ 8,8% din variația duratei
   d) R^2 este sub 1%
   e) rezultatul demonstrează că vârsta înaintată scurtează intervenția
9. (CM) Condițiile regresiei liniare simple includ:
   a) relația liniară dintre X și Y
   b) independența observațiilor
   c) distribuția normală a reziduurilor
   d) distribuția normală a variabilei X
   e) varianța constantă a reziduurilor pe tot domeniul lui X
10. (CM) Într-o regresie multiplă a TAS, coeficientul pentru "fumător, față de fost fumător" este 6,738 (IC 95% 1,53-11,95; p = 0,012), iar $R^2 = 0{,}308$, R^2 ajustat $= 0{,}286$. Sunt corecte:

a) la aceeași vârstă, același IMC și același sex, fumătorii au TAS în medie cu aproximativ 6,7 mmHg mai mare decât foștii fumători
b) modelul explică aproximativ 31% din variația TAS
c) R^2 ajustat este mai mic decât R^2 pentru că penalizează numărul de predictori
d) fumatul cauzează o creștere a TAS cu exact 6,738 mmHg la fiecare pacient
e) intervalul de încredere care nu conține 0 corespunde unui coeficient semnificativ la $\alpha = 0{,}05$

Capitolul 10. Aplicații statistice în epidemiologie: tipuri de studii observaționale și analiza riscului

Obiective

La finalul acestui capitol, studentul va putea:

- să definească factorul de risc, să-l clasifice și să descrie tipurile de expunere (punctuală, cronică, doză, durată);
- să calculeze și să interpreteze prevalența, incidența cumulativă și rata de incidență cu persoane-timp;
- să recunoască din enunț tipul de studiu (experimental, transversal, de cohortă prospectivă sau retrospectivă, caz-control, ecologic) și să aleagă măsura de asociere potrivită;
- să construiască tabelul de contingență 2 x 2 cu notația a, b, c, d și să calculeze de mână riscul relativ (RR) și raportul de șanse (OR);
- să calculeze intervalele de încredere 95% pentru RR și OR prin metoda logaritmică și să le folosească pentru a decide semnificația;
- să explice de ce OR depășește RR când boala este frecventă și de ce cele două sunt apropiate pentru bolile rare;
- să calculeze riscul atribuibil, fracția atribuibilă și numărul necesar de tratat (NNT) sau pentru a produce un efect advers (NNH);
- să recunoască un factor de confuzie, să explice ideea stratificării (Mantel-Haenszel) și să deosebească confuzia de modificarea efectului;
- să aplice criteriile Bradford Hill și ghidul STROBE la citirea critică a unui studiu observațional.

10.1 Factorul de risc

10.1.1 Ce numim factor de risc

În capitolele anterioare am comparat medii și proporții între grupuri. Epidemiologia pune aceleași instrumente în slujba unei întrebări precise: ce anume crește probabilitatea ca o persoană sănătoasă să facă o anumită boală? Un factor de risc este o caracteristică, un comportament, o expunere sau o condiție asociată cu o probabilitate mai mare de apariție a bolii. Cuvântul cheie este "probabilitate": fumatul nu garantează BPOC, iar mulți pacienți cu BPOC nu au fumat niciodată, dar printre fumători BPOC apare mai des. Factorul de risc este o cauză ipotetică, pe care studiul încearcă să o confirme sau să o infirme.

În medicina de familie lucrăm zilnic cu factori de risc: când calculăm scorul SCORE2 pentru riscul cardiovascular, când aplicăm chestionarul FINDRISC pentru diabet sau când întrebăm un pacient câte țigări fumează pe zi. Fiecare dintre aceste instrumente sintetizează rezultatele a zeci de studii epidemiologice de tipul celor descrise în acest capitol [64,111].

10.1.2 Clasificarea factorilor de risc

Cursurile disciplinei grupează factorii de risc după natura lor. Factorii de mediu cuprind poluanții atmosferici, toxinele, radiațiile, microorganismele infecțioase și expunerile profesionale (pulberi, solvenți, zgomot). Factorii comportamentali țin de obiceiurile persoanei: fumatul, consumul de alcool sau de droguri, alimentația, sedentarismul, nerespectarea măsurilor de protecție a muncii. Factorii sociali includ sărăcia, nivelul de educație, izolarea și evenimentele de viață traumatice

(decesul unui membru al familiei, divorțul). Factorii genetici sunt moșteniți: hipercolesterolemia familială, mutațiile BRCA în cancerul mamar și ovarian, istoricul familial de diabet.

O a doua clasificare, utilă clinic, separă factorii modificabili (fumat, obezitate, hipertensiune, sedentarism) de cei nemodificabili (vârstă, sex, genetică). Primii sunt ținta prevenției; ceilalți ajută la identificarea persoanelor care trebuie urmărite mai atent. În setul mg2026, obezitatea și fumatul sunt modificabile, iar vârsta și sexul nu.

10.1.3 Expunerea la factorul de risc

Expunerea poate fi punctuală (un singur eveniment, limitat în timp: un accident, o înțepătură accidentală cu un ac contaminat, o intoxicație alimentară la o nuntă) sau cronică (prelungită, cel mai frecvent tip în bolile netransmisibile). La expunerea cronică se descriu trei dimensiuni: doza curentă (câte țigări pe zi fumează acum pacientul), doza cumulată (de exemplu, numărul de pachete-an, adică pachete pe zi înmulțite cu anii de fumat) și durata expunerii (de câți ani). Un pacient care fumează un pachet pe zi de 30 de ani are 30 de pachete-an; unul care a fumat două pachete pe zi timp de 15 ani are tot 30 de pachete-an. Măsurarea corectă a expunerii este una dintre cele mai mari dificultăți ale studiilor observaționale, pentru că se bazează adesea pe memoria pacientului.

10.1.4 Factor cauzal și marker de risc

Nu orice factor asociat cu boala o și cauzează. Un factor cauzal participă efectiv la mecanismul bolii: eliminarea lui scade incidența (renunțarea la fumat scade riscul de cancer pulmonar). Un marker de risc (factor favorizant, indicator) este asociat cu boala și ne ajută să identificăm persoanele cu risc mare, dar nu face neapărat parte din lanțul cauzal. Degetele colorate în galben de nicotină sunt asociate cu cancerul pulmonar, dar nimeni nu și-ar propune să prevină cancerul prin spălarea mâinilor. Proteina C reactivă crescută se asociază cu evenimentele coronariene, dar scăderea ei artificială nu garantează scăderea riscului. Diferența are consecințe practice: un marker este bun pentru screening, un factor cauzal este bun și pentru prevenție.

10.2 Măsurarea frecvenței bolii

Înainte de a compara expuși și neexpuși, trebuie să știm cum măsurăm cât de frecventă este o boală. Există două familii de măsuri: prevalența, care descrie câți sunt bolnavi, și incidența, care descrie câți se îmbolnăvesc.

10.2.1 Prevalența

Prevalența este proporția persoanelor dintr-o populație care au boala la un moment dat sau într-o perioadă dată:

Prevalența = număr de persoane cu boala / număr total de persoane din populație

Prevalența punctuală se referă la un moment (de exemplu, ziua examinării), prevalența de perioadă la un interval (de exemplu, proporția persoanelor care au avut cel puțin un episod de lombalgie în ultimul an). Prevalența include atât cazurile noi, cât și pe cele vechi. În setul mg2026, la includere, 29 dintre cei 160 de pacienți aveau diabet confirmat prin TTGO: prevalența este 29 / 160 = 18,1%, cu IC 95% Wilson 12,9-24,8% (cap. 4). Prevalența este măsura de care are nevoie un manager sanitar: câte paturi, câte consultații, câte doze de medicament sunt necesare acum.

10.2.2 Incidența cumulativă (riscul)

Incidența cumulativă este proporția persoanelor inițial fără boală care fac boala într-o perioadă definită:

Incidența cumulativă = număr de cazuri noi în perioadă / număr de persoane fără boală la începutul perioadei

Este echivalentă cu probabilitatea (riscul) de a face boala în acea perioadă. Perioada trebuie specificată întotdeauna: un risc de 8% "în 5 ani" este cu totul altceva decât 8% "pe an". Dacă din 1000 de persoane sănătoase urmărite 5 ani, 80 fac diabet, incidența cumulativă la 5 ani este 80 / 1000 = 8%. Incidența cumulativă cere ca toți participanții să fie urmăriți pe toată perioada; la cei pierduți din urmărire nu știm dacă s-au îmbolnăvit sau nu.

10.2.3 Rata de incidență și persoanele-timp

În studiile reale, participanții intră în studiu la momente diferite, unii sunt pierduți din urmărire, alții decedează din alte cauze. Soluția este numărarea timpului efectiv de observație. O persoană urmărită 3 ani contribuie cu 3 persoane-an; o persoană urmărită 6 luni contribuie cu 0,5 persoane-an. Rata de incidență (densitatea de incidență) este:

Rata de incidență = număr de cazuri noi / suma persoanelor-timp de observație

Exemplu: 200 de pacienți cu diabet tip 2 sunt urmăriți pentru apariția retinopatiei. Din cauza intrărilor eșalonate și a pierderilor, timpul total de observație însumează 850 de persoane-an, iar în acest timp apar 34 de cazuri noi. Rata de incidență este 34 / 850 = 0,040 cazuri pe persoană-an, adică 4,0 cazuri la 100 de persoane-an. Rata nu este o proporție: unitatea ei este "pe timp", iar valoarea poate depăși teoretic 1 (o boală foarte frecventă, cu episoade repetate). Raportul a două rate de incidență se numește raportul ratelor de incidență (IRR, engl. incidence rate ratio) și se interpretează la fel ca riscul relativ.

10.2.4 Legătura dintre prevalență și incidență

Pentru o boală în stare de echilibru (incidență și durată aproximativ constante) și cu prevalență mică, există o relație simplă:

$$P \approx I \times D$$

unde P este prevalența, I rata de incidență și D durata medie a bolii. O boală cronică necurabilă, dar cu supraviețuire lungă (diabetul, hipertensiunea), are prevalență mare chiar dacă incidența este moderată: o incidență de 2 cazuri la 1000 de persoane-an și o durată medie de 10 ani dau o prevalență de aproximativ 0,002 x 10 = 0,02, adică 2%. O boală acută de scurtă durată (gripa, o infecție respiratorie de o săptămână) poate avea incidență foarte mare și prevalență punctuală mică. Relația explică și un paradox aparent: un tratament nou care prelungește supraviețuirea fără să vindece crește prevalența bolii, deși este un progres. De aceea, compararea prevalențelor între regiuni sau perioade cere prudență.

10.3 Tipuri de studii epidemiologice

10.3.1 Studii experimentale și studii observaționale

Diferența fundamentală dintre tipurile de studii este cine decide expunerea. În studiul experimental (de intervenție), cercetătorul alocă expunerea, ideal prin randomizare: el decide cine primește tratamentul nou și cine primește placebo sau tratamentul standard. Randomizarea distribuie în mod egal, în medie, toți factorii cunoscuți și necunoscuți între grupuri, așa că diferența de rezultat poate fi atribuită intervenției. Studiile clinice randomizate oferă cele mai puternice dovezi cauzale, dar nu pot fi folosite pentru factori de risc dăunători: nimeni nu poate randomiza oameni să fumeze sau să lucreze în pulberi. Setul mg2026 este un astfel de studiu experimental simulat (brațele A, B, C), iar capitolul 13 îl discută în detaliu.

În studiul observațional, expunerea s-a produs independent de cercetător (pacientul a ales să fumeze, a lucrat într-o fabrică, s-a născut cu o anumită genă). Cercetătorul doar observă și

măsoară. Dezavantajele principale sunt precizia limitată a măsurării expunerii și faptul că expușii și neexpușii pot diferi și prin alți factori. Studiile observaționale se clasifică după momentul în care se măsoară expunerea și boala. Figura 10.1 prezintă schematic cele trei designuri principale.

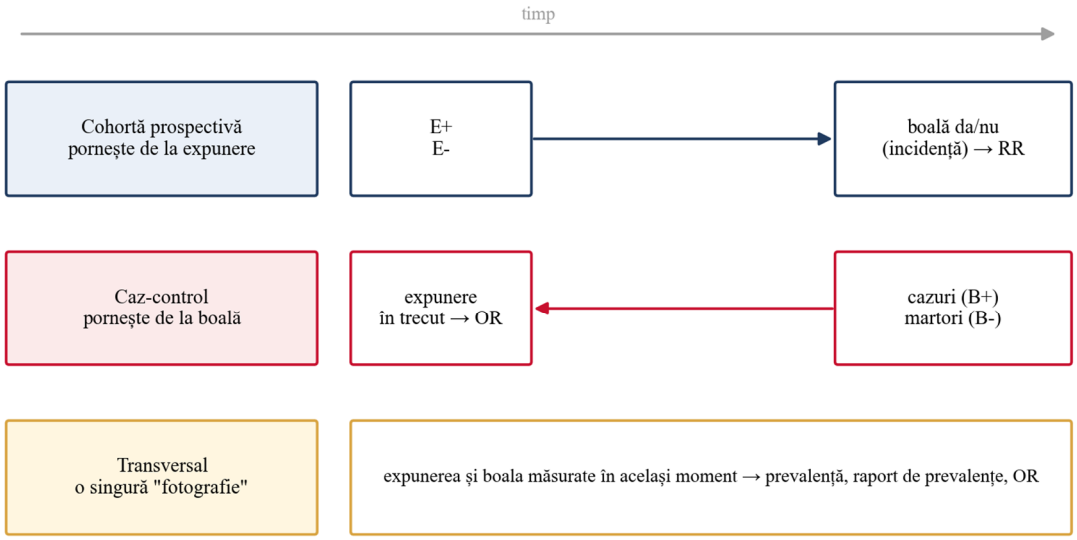


Figura 10.1. Cele trei designuri observaționale principale. Studiul de cohortă pornește de la expunere (E+ și E-) și urmărește în timp apariția bolii, permițând calculul incidenței și al riscului relativ. Studiul caz-control pornește de la boală (cazuri și martori) și caută înapoi expunerea, permițând doar calculul raportului de șanse. Studiul transversal măsoară expunerea și boala în același moment și estimează prevalența, raportul prevalențelor și OR.

10.3.2 Studiul transversal

Studiul transversal (engl. cross-sectional) este o "fotografie" a populației la un moment dat: pentru fiecare participant se înregistrează, în același timp, prezența sau absența expunerii și a bolii. Este un studiu de prevalență și cel mai simplu model de cercetare. Analiza de la includere din mg2026, în care am corelat obezitatea cu diabetul confirmat prin TTGO, este transversală: ambele au fost măsurate la prima vizită.

Măsura de asociere naturală este raportul prevalențelor (RP, engl. prevalence ratio): prevalența bolii la expuși împărțită la prevalența la neexpuși. Formula este identică cu a riscului relativ, dar numele diferă, pentru că nu măsurăm cazuri noi, ci cazuri existente. Se raportează frecvent și OR, mai ales când se folosește regresia logistică.

Avantajele sunt costul redus și rapiditatea, posibilitatea de a studia simultan mai mulți factori și mai multe boli și utilitatea pentru planificarea serviciilor de sănătate. Dezavantajul major este că nu putem stabili succesiunea temporală: dacă obezii au mai des diabet, nu știm, doar din acest studiu, dacă obezitatea a precedat diabetul. Un sondaj transversal pe 2976 de studenți din județul Timiș, dintre care 62,5% fete, a identificat prin regresie logistică factori asociați cu fumatul intens [22]; designul permite descrierea asocierilor, dar nu și stabilirea a ceea ce a apărut întâi, fumatul sau comportamentele asociate. Un al doilea dezavantaj ține de supraviețuirea selectivă: bolile cu evoluție rapidă (deces sau vindecare) sunt subreprezentate, pentru că bolnavii respectivi nu mai sunt prezenți în momentul "fotografiei". Estimarea prevalenței este distorsionată pentru bolile cu evoluție rapidă.

10.3.3 Studiul de cohortă

Într-un studiu de cohortă, participanții sunt selectați inițial fără boală, sunt clasificați după expunere (expuși E+ și neexpuși E-) și sunt urmăriți în timp pentru apariția bolii [112]. Termenul

vine de la cohortele legiunilor romane: un grup care merge împreună înainte. Fiindcă pornim de la persoane sănătoase și numărăm cazurile noi, putem calcula direct incidența la expuși și la neexpuși, deci riscul relativ.

Cohorta prospectivă începe acum și urmărește participanții în viitor (de exemplu, 921 de studenți urmăriți în primul an de facultate [5]). Cohorta retrospectivă (istorică) folosește înregistrări din trecut: cercetătorul stabilește din documente cine a fost expus acum 10 ani și verifică cine s-a îmbolnăvit de atunci. Ambele păstrează logica "de la expunere la boală"; diferă doar momentul în care cercetătorul intră în scenă. Cursurile disciplinei, urmând o tradiție mai veche, reținc că în cohorta retrospectivă se calculează doar OR; în realitate, dacă înregistrările permit identificarea completă a celor expuși și neexpuși și a cazurilor noi, RR se poate calcula și în cohorta retrospectivă. Ce contează este dacă proporțiile de bolnavi din fiecare grup de expunere reflectă riscul real.

Avantajele cohortei sunt succesiunea temporală clară (expunerea precede boala), calculul direct al incidenței și al RR, posibilitatea de a studia mai multe boli produse de aceeași expunere și protecția față de erorile de memorie, pentru că expunerea se înregistrează înainte de apariția bolii. Dezavantajele sunt costul, durata (bolile cu latență lungă, precum cancerul, cer zeci de ani), pierderile din urmărire și ineficiența pentru bolile rare: pentru a observa câteva cazuri de boală rară ar trebui urmărite zeci de mii de persoane.

10.3.4 Studiul caz-control

Studiul caz-control inversează logica: pornește de la boală [113]. Cercetătorul selectează un grup de persoane bolnave (cazurile, B+) și un grup de persoane fără boală (martorii sau controalele, B-), apoi caută în trecutul fiecăruia expunerea la factorul de risc, prin interviu, chestionar sau dosare medicale.

Pentru că cercetătorul decide câte cazuri și câți martori include (de exemplu, un caz la doi martori), proporția de bolnavi din eșantion nu reflectă riscul din populație. De aceea, într-un studiu caz-control nu se poate calcula incidența și nici riscul relativ. Se calculează doar raportul de șanse, care are o proprietate remarcabilă (secțiunea 10.6): OR al expunerii la cazuri față de martori este egal cu OR al bolii la expuși față de neexpuși.

Alegerea martorilor este punctul cel mai delicat. Martorii trebuie să provină din aceeași populație din care provin cazurile: dacă un martor ar fi făcut boala, ar fi fost inclus în studiu drept caz. Pentru cazuri dintr-un spital, martorii se aleg de regulă din același spital, cu alte boli fără legătură cu expunerea, sau din populația deservită de spital. Martorii aleși dintr-o populație diferită (de exemplu, studenți sănătoși pentru cazuri vârstnice internate) produc o eroare de selecție.

Potrivirea (engl. matching) este alegerea martorilor astfel încât să semene cu cazurile la variabile care altfel ar confunda rezultatul (vârstă, sex). Potrivirea poate fi individuală (fiecărui caz i se găsesc unul sau mai mulți martori de aceeași vârstă și același sex) sau de frecvență (distribuția pe vârste și sexe este aceeași în cele două grupuri). Potrivirea elimină confuzia prin variabilele potrivite, dar acele variabile nu mai pot fi studiate ca factori de risc, iar analiza trebuie să țină cont de perechi.

Eroarea de memorie (engl. recall bias) este specifică studiilor caz-control: bolnavii, care și-au căutat explicații pentru boala lor, își amintesc mai bine sau supraestimează expunerile, în timp ce persoanele sănătoase le minimizează. O mamă cu un copil cu malformație congenitală își va aminti în detaliu fiecare medicament luat în sarcină; o mamă cu un copil sănătos poate uita. Rezultatul este un OR artificial mărit.

Avantajele studiului caz-control sunt costul redus, rapiditatea, eficiența pentru bolile rare și pentru cele cu latență lungă și posibilitatea de a studia simultan mai mulți factori de risc pentru aceeași boală. Dezavantajele sunt imposibilitatea calculului incidenței și al RR, erorile de selecție și de memorie și dificultatea de a stabili cu certitudine că expunerea a precedat boala.

O variantă importantă este studiul caz-control "cuibărit" (engl. nested case-control), în care cazurile și martorii sunt extrași dintr-o cohortă deja definită. Combinația păstrează avantajele cohortei (expunerea a fost documentată prospectiv) și eficiența studiului caz-control.

10.3.5 Studiul ecologic

În studiul ecologic, unitatea de analiză nu este persoana, ci grupul: țara, județul, orașul, luna. Se compară frecvența expunerii cu frecvența bolii la nivel de grup, de exemplu consumul mediu de sare pe județ cu mortalitatea prin accident vascular cerebral pe județ. Studiile ecologice sunt ieftine, folosesc date deja existente și sunt utile pentru generarea ipotezelor și pentru evaluarea intervențiilor de sănătate publică. Limita lor majoră este eroarea ecologică, discutată în capitolul 9: o asociere la nivel de grup nu se transferă automat la nivel individual, iar numeroase variabile externe pot acționa ca factori de confuzie. Analiza seriilor de timp întrerupte, în care se compară tendința unei rate înainte și după un eveniment, este o formă modernă de studiu ecologic.

10.3.6 Comparația designurilor

Tabelul 10.1. Comparația principalelor tipuri de studii epidemiologice.

Caracteristica	Transversal	Cohortă	Caz-control	Ecologic	Experimental
Punctul de plecare	populația, la un moment	expunerea	boala	grupuri	alocarea de către cercetător
Direcția	simultan	înainte (spre viitor)	înapoi (spre trecut)	simultan, de obicei	înainte
Măsura frecvenței	prevalența	incidența	nu se poate calcula	rate pe grupuri	incidența
Măsura asocierii	RP, OR	RR, OR, IRR, HR	doar OR	corelație, IRR	RR, diferența de risc, NNT
Succesiunea temporală	nu	da	incertă	nu	da
Boli rare	nepotrivit	nepotrivit	foarte potrivit	posibil	nepotrivit
Cost și durată	mici	mari	mici	foarte mici	foarte mari
Erori tipice	supraviețuire selectivă	pierderi din urmărire	selecție, memorie	eroare ecologică	abateri de la protocol

Tabelul nu trebuie memorat ca o listă; el decurge din întrebarea "de unde pornește studiul?". Dacă pornește de la expunere și numără cazuri noi, putem calcula riscul. Dacă pornește de la boală, nu putem.

Studiu de caz: cinci designuri în articolele colegilor

Un studiu observațional de urmărire pe 110 paciente cu leziuni cervicale de grad înalt tratate prin excizie (LEEP), cu testare HPV la 6, 12 și 18 luni, este o cohortă: pacientele au fost clasificate după vârstă și tipul de HPV și urmărite pentru persistența infecției [114]. Rata de persistență a scăzut de la 40,9% la 6 luni la 20% la 12 luni și 11,8% la 18 luni, iar vârsta, dihotomizată la 36,5 ani, s-a asociat cu persistența: RR = 2,75, IC 95% 1,34-5,64, p = 0,0027. Intervalul nu conține 1, deci asocierea este semnificativă; lățimea lui (de la un risc de 1,3 ori la unul de 5,6 ori mai mare) reflectă cele 110 paciente.

Un studiu transversal pe 101 pacienți HIV-pozitivi sub tratament antiretroviral din vestul României a analizat asocierea dintre severitatea bolii HIV și tulburările psihiatrice prin rapoarte de șanse: OR între 2,81 (p = 0,038) și 8,72 (p < 0,001) pentru diferite tulburări [84]. Designul transversal justifică OR, dar nu permite stabilirea direcției (boala HIV severă duce la

tulburări psihiatrice sau tulburările psihiatrice duc la o aderență mai slabă și deci la boală mai severă?). Faptul că rezumatul dă OR cu p, dar fără intervale de încredere, și că se testează multe tulburări simultan sunt două puncte de lectură critică: cu șapte comparații, probabilitatea ca măcar una să iasă semnificativă întâmplător crește.

Un studiu pe 656 de pacienți cu infarct miocardic cu supradenivelare ST tratați prin angioplastie primară a comparat cei 96 de pacienți cu fenomen de no-reflow cu restul cohortei, privind complicațiile pe termen scurt [115]. Este un design caz-control construit în interiorul unei cohorte. Un studiu retrospectiv pe 435 de pacienți cu colecistită acută litiazică a folosit potrivirea pe vârstă și sex, cu un raport de 1:2 (145 de cazuri cu formă gangrenoasă, 290 de martori cu formă cronică), și a raportat pentru un indice inflamator OR = 1,976, p < 0,001, într-un model de regresie logistică [116]. Potrivirea pe vârstă și sex înseamnă că aceste două variabile nu mai pot fi analizate ca factori de risc, dar nici nu mai pot confunda rezultatul. În fine, un studiu clinic transversal cu analiză caz-control cuibărită a analizat 136 de pacienți (52 de cazuri cu tulburări ale interacțiunii intestin-creier și 84 de martori asimptomatici) și a găsit un OR protector de 0,36 (IC 95% 0,18-0,74; p = 0,009) [71]. Un OR sub 1, cu un interval care nu conține 1, înseamnă că expunerea respectivă era mai rară la cazuri decât la martori.

10.4 Tabelul de contingență 2 x 2

Toate calculele din analiza riscului pornesc de la același tabel. Convenția folosită la disciplină (și în majoritatea manualelor) pune expunerea pe rânduri și boala pe coloane, cu expușii pe primul rând și bolnavii pe prima coloană.

Tabelul 10.2. Notația tabelului de contingență 2 x 2.

	Boală prezentă (B+)	**Boală absentă (B-)**	**Total**
Expuși (E+)	a	b	a + b
Neexpuși (E-)	c	d	c + d
Total	a + c	b + d	$N = a + b + c + d$

Celula a conține expușii bolnavi, b expușii sănătoși, c neexpușii bolnavi, d neexpușii sănătoși. Marginalele pe rânduri (a + b și c + d) sunt numerele de expuși și neexpuși; marginalele pe coloane (a + c și b + d) sunt numerele de bolnavi și sănătoși. La examen, cea mai frecventă greșeală în analiza riscului nu este formula greșită, ci tabelul construit greșit: rânduri și coloane inversate, sau un total pus în locul unei celule. Verificați întotdeauna că a + b + c + d = N și că fiecare total din enunț se regăsește în tabel.

Într-un studiu de cohortă sau experimental, totalurile pe rânduri sunt fixate de design (atâția expuși, atâția neexpuși). Într-un studiu caz-control, fixate sunt totalurile pe coloane (atâtea cazuri, atâția martori). Într-un studiu transversal, fixat este doar N. Această diferență determină ce proporții au sens: într-un studiu caz-control, a / (a + b) nu este riscul bolii la expuși, pentru că numărul de bolnavi a fost ales de cercetător.

10.5 Riscul relativ

10.5.1 Formula și interpretarea

Riscul de boală la expuși este $R_1 = \frac{a}{a+b}$, iar la neexpuși $R_0 = \frac{c}{c+d}$. Riscul relativ (RR, engl. relative risk sau risk ratio) este raportul lor:

$$\mathrm{RR} = \frac{R_1}{R_0} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

RR arată de câte ori este mai mare riscul de boală la expuși decât la neexpuși. Interpretarea are trei situații. RR > 1: expunerea este factor de risc (expușii se îmbolnăvesc mai des). RR = 1: expunerea nu este asociată cu boala (factor indiferent). RR < 1: expunerea este factor de protecție (expușii se îmbolnăvesc mai rar); de exemplu, RR = 0,6 înseamnă un risc cu 40% mai mic la expuși. Ultima interpretare este importantă pentru tratamente și vaccinuri, unde "expunerea" este intervenția.

Desigur, decizia finală nu se ia doar după valoarea punctuală, ci după intervalul de încredere: un RR de 1,8 cu IC 95% 0,9-3,6 nu arată un factor de risc semnificativ, pentru că intervalul conține 1.

10.5.2 Exemplul clasic: fumatul și BPOC

Un studiu urmărește timp de 5 ani 1200 de persoane inițial sănătoase, dintre care 300 sunt fumători. La final, 50 dintre fumători și 70 dintre nefumători au BPOC. Primul pas este tabelul. Fumători: a = 50 bolnavi, b = 300 - 50 = 250 sănătoși. Nefumătorii sunt 1200 - 300 = 900: c = 70 bolnavi, d = 900 - 70 = 830 sănătoși.

Tabelul 10.3. Fumatul și BPOC, studiu de cohortă cu 1200 de participanți.

	BPOC da	**BPOC nu**	**Total**
Fumători	50	250	300
Nefumători	70	830	900
Total	120	1080	1200

R1 = 50 / 300 = 0,1667 (16,7% în 5 ani); $R_0 = \frac{70}{900} = 0{,}0778 (7{,}8\%)$; RR $= \frac{0{,}1667}{0{,}0778} = 2{,}14$. Fumătorii au un risc de BPOC de 2,14 ori mai mare decât nefumătorii în cei 5 ani de urmărire.

O greșeală pe care o vedem an de an: unii studenți împart 70 la 1200 (totalul participanților) în loc de 900 (numărul nefumătorilor) și obțin R0 = 0,0583 și RR = 2,86. Riscul la neexpuși se calculează numai printre neexpuși. Un RR de 2,86 nu are sens în acest tabel, iar o problemă rezolvată greșit la primul pas duce la toate celelalte rezultate greșite.

10.6 Raportul de șanse

10.6.1 Șansa

Șansa (engl. odds) unui eveniment este raportul dintre probabilitatea ca evenimentul să se producă și probabilitatea ca el să nu se producă:

$$\text{șansa } = \frac{p}{1-p}$$

Termenul vine din pariuri: o șansă de "1 la 4" înseamnă că, la fiecare caz care se îmbolnăvește, 4 nu se îmbolnăvesc, adică $p = \frac{1}{5} = 20\%$ și șansa $= \frac{0{,}20}{0{,}80} = 0{,}25$. Pentru evenimente rare, șansa și probabilitatea sunt apropiate (p = 2% dă șansa 0,0204); pentru evenimente frecvente, diferă mult (p = 50% dă șansa 1; p = 80% dă șansa 4).

În tabelul 2 x 2, șansa bolii la expuși este a / b (bolnavi la sănătoși, printre expuși), iar la neexpuși c / d.

10.6.2 Formula și interpretarea

Raportul de șanse (OR, engl. odds ratio) este:

$$\text{OR} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a \times d}{b \times c}$$

Forma a doua, "produsul diagonalei principale împărțit la produsul diagonalei secundare", se reține ușor. Interpretarea urmează aceleași reguli ca la RR: OR > 1 factor de risc, OR = 1 factor indiferent, OR < 1 factor de protecție. Pentru exemplul fumat-BPOC: $\text{OR} = \frac{50 \times 830}{250 \times 70} = \frac{41\,500}{17\,500} = 2{,}37$. Șansa de a face BPOC este de 2,37 ori mai mare la fumători.

OR are o proprietate de simetrie care îl face indispensabil în studiile caz-control. Șansa expunerii la cazuri este a / c, iar la martori b / d. Raportul lor este $\frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a \times d}{b \times c}$, exact același OR. Prin urmare, deși într-un studiu caz-control nu putem calcula riscul bolii, putem calcula OR al expunerii, care este numeric egal cu OR al bolii. Interpretarea corectă într-un studiu caz-control este: "șansa de a fi fost expus este de OR ori mai mare la cazuri decât la martori", iar, prin simetrie, "șansa de boală este de OR ori mai mare la expuși".

10.6.3 De ce OR este mai mare decât RR când boala este frecventă

În exemplul fumat-BPOC, OR = 2,37 și RR = 2,14: apropiate, dar nu egale. Cursurile disciplinei notează "uzual OR > RR". Regula exactă este aceasta: când RR > 1, OR este întotdeauna mai departe de 1 decât RR; când RR < 1, OR este mai mic decât RR (tot mai departe de 1). Diferența depinde de cât de frecventă este boala. Tabelul 10.4 arată, pentru un RR fixat la 2, cum crește OR pe măsură ce riscul la neexpuși crește.

Tabelul 10.4. OR corespunzător unui RR = 2, pentru diferite riscuri la neexpuși.

Risc la neexpuși (R0)	Risc la expuși (R1)	RR	OR
1%	2%	2	2,02
5%	10%	2	2,11
10%	20%	2	2,25
20%	40%	2	2,67
30%	60%	2	3,50
40%	80%	2	6,00

Explicația se vede din formule: $\text{OR} = \frac{\text{RR} \times (1 - R_0)}{1 - R_1}$. Când ambele riscuri sunt mici, raportul (1 - R0) / (1 - R1) este aproape 1, iar OR ≈ RR. Aceasta este "presupunerea bolii rare" (engl. rare disease assumption): pentru boli cu frecvență sub aproximativ 10%, OR obținut într-un studiu caz-control este o bună aproximare a RR. Pentru boli frecvente, OR exagerează efectul dacă este citit ca RR. În setul mg2026, prevalența diabetului la obezi este 33,3%: OR (4,32) este vizibil mai mare decât RP (3,21). Un articol care raportează un OR de 4,32 și îl traduce prin "riscul este de peste 4 ori mai mare" exagerează.

10.7 Intervale de încredere și testarea statistică

10.7.1 Intervalul de încredere pentru RR prin metoda logaritmică

Distribuția de eșantionare a lui RR este asimetrică: RR nu poate fi negativ, dar poate crește oricât. Logaritmul natural al lui RR are însă o distribuție aproximativ normală. Calculul are trei pași: se trece pe scara logaritmică, se construiește intervalul obișnuit (estimare $\pm 1{,}96 \times \text{ES}$), apoi se revine prin exponențiere.

$$\text{ES}[\ln(\text{RR})] = \sqrt{\left[\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}\right]} \quad \text{IC 95\% pentru RR} = \exp[\ln(\text{RR}) \pm 1{,}96 \times \text{ES}]$$

Pentru fumat-BPOC: ln(2,1429) = 0,762. $ES = \sqrt{\frac{1}{50} - \frac{1}{300} + \frac{1}{70} - \frac{1}{900}} = \sqrt{0{,}02 - 0{,}00333 + 0{,}01429 - 0{,}00111} = \sqrt{0{,}02984} = 0{,}1727$. Marja: 1,96 x 0,1727 = 0,339. Pe scara logaritmică: 0,762 - 0,339 = 0,424 și 0,762 + 0,339 = 1,101. Exponențiind: $e^{0{,}424} = 1{,}53$ și $e^{1{,}101} = 3{,}01$. IC 95% pentru RR: 1,53-3,01.

Observați că intervalul nu este simetric în jurul lui 2,14 (2,14 - 1,53 = 0,61, dar 3,01 - 2,14 = 0,87). Asimetria este normală pentru rapoarte și este simetrică pe scara logaritmică; de aceea, diagramele forest folosesc axa logaritmică.

10.7.2 Intervalul de încredere pentru OR (metoda Woolf)

$\mathrm{ES}[\ln(\mathrm{OR})] = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$ IC 95% pentru $\mathrm{OR} = \exp[\ln(\mathrm{OR}) \pm 1{,}96 \times \mathrm{ES}]$

Pentru fumat-BPOC: ln(2,3714) = 0,8635. $ES = \sqrt{\frac{1}{50} + \frac{1}{250} + \frac{1}{70} + \frac{1}{830}} = \sqrt{0{,}02 + 0{,}004 + 0{,}01429 + 0{,}00120} = \sqrt{0{,}03949} = 0{,}1987$. Marja: 1,96 x 0,1987 = 0,3895. Scara logaritmică: 0,474 și 1,253. Exponențiind: 1,61 și 3,50. IC 95% pentru OR: 1,61-3,50.

Formula arată de ce celulele mici produc intervale largi: dacă una dintre celule are valoarea 5, termenul 1/5 = 0,2 domină suma. Dacă o celulă este 0, OR și ES nu pot fi calculate; programele adaugă atunci 0,5 la fiecare celulă (corecția Haldane) sau folosesc metode exacte.

10.7.3 Testul chi-pătrat și testul Fisher

Testul de semnificație al asocierii din tabelul 2 x 2 este testul chi-pătrat de independență din capitolul 7. Ipotezele sunt formulate în termenii măsurii de asociere:

H0: RR = 1 (sau OR = 1), expunerea nu este asociată cu boala H1: RR ≠ 1 (sau OR ≠ 1), expunerea este asociată cu boala

Pentru fumat-BPOC, frecvențele așteptate sub H0 sunt: pentru celula a, $\frac{300 \times 120}{1200} = 30$; b: $\frac{300 \times 1080}{1200} = 270$; c: $\frac{900 \times 120}{1200} = 90$; d: $\frac{900 \times 1080}{1200} = 810$. Toate diferențele observat-așteptat au aceeași valoare absolută, 20 (50 - 30 = 20; 250 - 270 = -20 etc.), ceea ce se întâmplă întotdeauna într-un tabel 2 x 2. $\chi^2 = \frac{20^2}{30} + \frac{20^2}{270} + \frac{20^2}{90} + \frac{20^2}{810} = 13{,}33 + 1{,}48 + 4{,}44 + 0{,}49 = 19{,}75$, cu 1 grad de libertate, p < 0,001. H0 se respinge.

Când o frecvență așteptată este sub 5, se folosește testul exact Fisher. Corecția Yates (de continuitate) reduce valoarea χ^2 și face testul mai conservator; programele o raportează alături de testul necorectat.

10.7.4 Legătura dintre interval și test

Un IC 95% pentru RR sau OR care nu conține 1 corespunde, aproape întotdeauna, unui test semnificativ la α = 0,05; un interval care conține 1 corespunde unui rezultat nesemnificativ. Echivalența nu este perfectă, pentru că intervalul logaritmic și testul chi-pătrat sunt calculate diferit, iar la limită pot da concluzii diferite (problema rezolvată 10.5 arată un astfel de caz). Raportarea intervalului este preferabilă raportării doar a valorii p, pentru că arată și mărimea, și precizia efectului [70].

10.8 Măsuri de impact: riscul atribuibil, fracția atribuibilă, NNT și NNH

10.8.1 Riscul atribuibil (diferența de risc)

RR este o măsură relativă: ne spune de câte ori crește riscul, dar nu și cât de mult în termeni absoluți. Un RR de 2 poate însemna o creștere de la 1 la 2 cazuri la un milion (neglijabilă pentru sănătatea publică) sau de la 20% la 40% (enormă). Riscul atribuibil (RA), numit și diferența de risc (DR, engl. risk difference), este:

$$\text{RA} = R_1 - R_0$$

Pentru fumat-BPOC: RA = 0,1667 - 0,0778 = 0,0889, adică 8,9 puncte procentuale în 5 ani. Din 100 de fumători, aproximativ 9 cazuri de BPOC în 5 ani se datorează fumatului (dacă relația este cauzală), peste cele 7-8 care ar fi apărut oricum. IC 95% pentru diferența de risc se calculează ca la diferența a două proporții (cap. 4): $\text{RA} \pm 1{,}96 \times \sqrt{\left[\frac{R_1(1-R_1)}{a+b} + \frac{R_0(1-R_0)}{c+d}\right]}$. Pentru exemplul nostru: $0{,}0889 \pm 1{,}96 \times \sqrt{\frac{0{,}1667 \times 0{,}8333}{300} + \frac{0{,}0778 \times 0{,}9222}{900}} = 0{,}0889 \pm 1{,}96 \times 0{,}0233 = 0{,}043 - 0{,}135$.

10.8.2 Fracția atribuibilă

Fracția atribuibilă la expuși (FAe, numită și fracția etiologică la expuși) arată ce proporție din cazurile apărute la expuși se datorează expunerii:

$$\text{FAe} = \frac{R_1 - R_0}{R_1} = \frac{\text{RR} - 1}{\text{RR}}$$

Pentru fumat-BPOC: FAe = (2,14 - 1) / 2,14 = 0,533. Aproximativ 53% dintre cazurile de BPOC apărute la fumători ar putea fi atribuite fumatului.

Fracția atribuibilă populațională (FAP) arată ce proporție din toate cazurile din populație s-ar putea evita dacă expunerea ar dispărea. Depinde atât de RR, cât și de cât de răspândită este expunerea (pe = proporția expușilor). Formula lui Levin este:

FAP = pe x (RR - 1) / [1 + pe x (RR - 1)]

Pentru fumat-BPOC, pe = 300 / 1200 = 0,25, deci $\text{FAP} = \frac{0{,}25 \times 1{,}14}{1 + 0{,}25 \times 1{,}14} = \frac{0{,}286}{1{,}286} = 0{,}222$. Aproximativ 22% dintre toate cazurile de BPOC din această populație ar fi evitabile prin eliminarea fumatului. Echivalent, FAP = (Rtotal - R0) / Rtotal = (0,100 - 0,0778) / 0,100 = 0,222, unde Rtotal = 120 / 1200 = 0,100 este riscul în populația totală. Diferența Rtotal - R0 = 2,2 puncte procentuale se numește risc atribuibil populațional. Aceste calcule reprezintă ce fișa disciplinei numește "estimarea riscului la nivelul populației". Ele presupun că asocierea este cauzală și că nu există confuzie; pe date observaționale, sunt estimări optimiste ale beneficiului prevenției.

10.8.3 Numărul necesar de tratat (NNT) și numărul necesar pentru a produce un efect advers (NNH)

Inversul diferenței absolute de risc are o interpretare clinică foarte intuitivă. Într-un studiu de intervenție cu un rezultat favorabil, NNT (engl. number needed to treat) este numărul de pacienți care trebuie tratați pentru a obține un rezultat favorabil în plus față de tratamentul de comparație:

$$\text{NNT} = \frac{1}{|R_1 - R_0|}$$

și se rotunjește întotdeauna în sus, la următorul număr întreg (nu poți trata 5,5 pacienți). Pentru un factor dăunător, același calcul se numește NNH (engl. number needed to harm): câte persoane trebuie expuse pentru a apărea un caz în plus. Pentru fumat-BPOC, $\text{NNH} = \frac{1}{0{,}0889} =$

11,25, adică 12: la fiecare 12 fumători urmăriți 5 ani apare un caz de BPOC în plus față de nefumători.

În studiul mg2026, la 6 luni, 16 din 54 de pacienți din brațul C (consiliere, aplicație și teleconsultație) au slăbit cel puțin 5%, față de 6 din 53 în brațul A (consiliere standard). Proporțiile sunt 29,6% și 11,3%; RR = 2,62 (IC 95% 1,11-6,17); diferența absolută este 18,3 puncte procentuale (IC 95% 3,4-33,2); NNT $= \frac{1}{0{,}183} = 5{,}5$, adică 6 (IC 95% 3,0-29,1). Clinic: pentru fiecare 6 pacienți care primesc pachetul complet în loc de consilierea standard, unul în plus atinge o scădere în greutate relevantă. Intervalul NNT este larg (de la 3 la 29), pentru că diferența de risc este estimată imprecis cu doar 107 pacienți; capătul superior (29) corespunde capătului inferior al diferenței (3,4 puncte procentuale). OR pentru aceeași comparație este 3,30, din nou mai mare decât RR, pentru că rezultatul nu este rar.

Figura 10.2 strânge într-o diagramă forest măsurile de asociere calculate în acest capitol. Fiecare pătrat este estimarea punctuală, iar linia orizontală este intervalul de încredere 95%; linia verticală punctată marchează valoarea 1 (lipsa efectului). Un interval care trece de partea cealaltă a liniei punctate înseamnă un rezultat nesemnificativ.

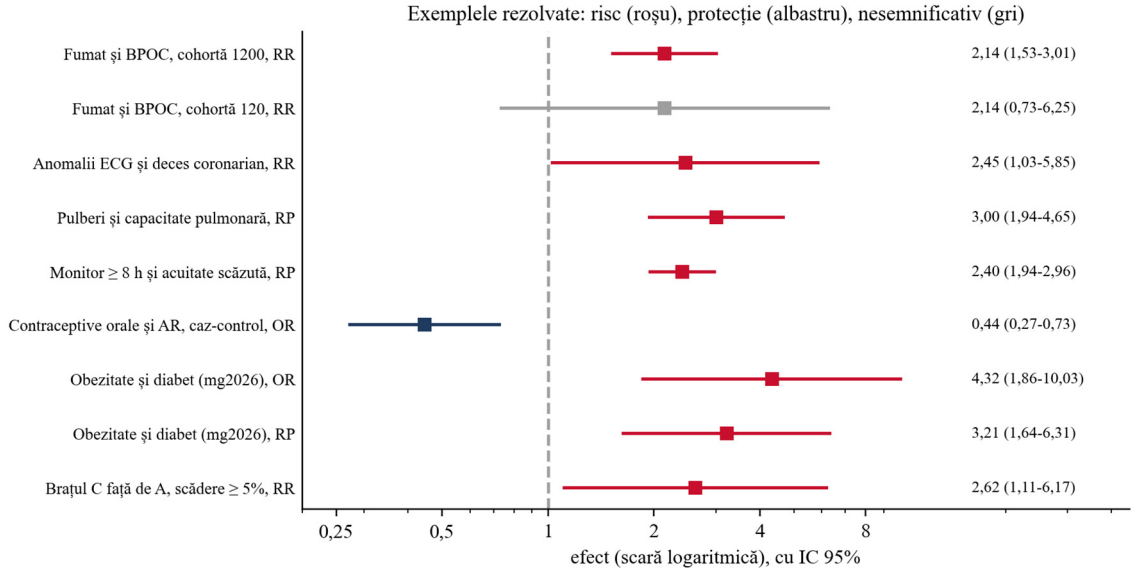


Figura 10.2. Măsurile de asociere din exemplele și problemele rezolvate ale capitolului, cu IC 95% calculat prin metoda logaritmică, pe axă logaritmică. Roșu: factor de risc semnificativ; albastru: factor de protecție semnificativ; gri: interval care conține 1 (nesemnificativ). Același RR = 2,14 pentru fumat și BPOC este semnificativ cu 1200 de participanți și nesemnificativ cu 120. Exercițiile clasice ale disciplinei și setul sintetic mg2026.

10.9 Confuzia, stratificarea și modificarea efectului

10.9.1 Factorul de confuzie

Un factor de confuzie (engl. confounder) este o a treia variabilă care îndeplinește simultan trei condiții: este asociată cu expunerea, este un factor de risc pentru boală (independent de expunere) și nu se află pe calea cauzală dintre expunere și boală. Când există, asocierea brută dintre expunere și boală amestecă efectul expunerii cu efectul factorului de confuzie.

Exemplul didactic clasic este legătura aparentă dintre consumul de cafea și o boală, confundată de fumat. Să presupunem că printre 1000 de persoane, fumătorii beau mult mai des cafea (400 din 500 de fumători, față de 100 din 500 de nefumători) și că fumatul crește riscul bolii (20% la fumători, 5% la nefumători), în timp ce cafeaua nu are niciun efect. Figura 10.3 arată ce se întâmplă.

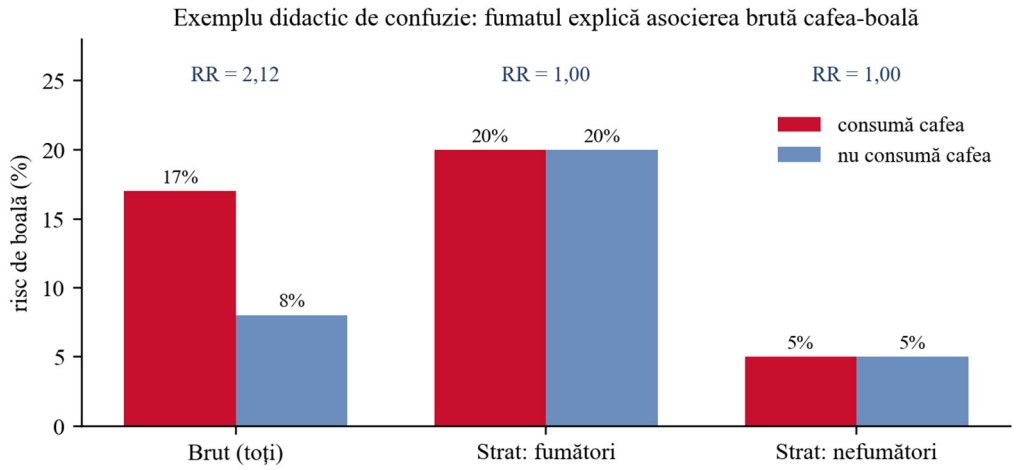


Figura 10.3. Exemplu didactic de confuzie (date construite). În analiza brută, consumatorii de cafea au un risc de 17%, față de 8% la ceilalți (RR = 2,12). În fiecare strat de fumat, riscul este identic la consumatori și neconsumatori (20% la fumători, 5% la nefumători; RR = 1,00). Asocierea brută provine în întregime din faptul că fumătorii beau mai multă cafea.

Analiza brută: printre cei 500 de consumatori de cafea sunt 80 de cazuri la fumători (20% din 400) și 5 la nefumători (5% din 100), deci 85 de cazuri, riscul 17%. Printre cei 500 de neconsumatori sunt 20 de cazuri la fumători (20% din 100) și 20 la nefumători (5% din 400), deci 40 de cazuri, riscul 8%. RR brut = 17% / 8% = 2,125, iar OR brut = $\frac{85\times460}{415\times40} = 2{,}36$. Un cercetător grăbit ar conclude că cafeaua dublează riscul. Analiza stratificată pe fumat arată însă RR = 1 în fiecare strat: printre fumători, 20% față de 20%; printre nefumători, 5% față de 5%. Cafeaua nu are niciun efect; fumatul, asociat cu cafeaua, creează iluzia.

10.9.2 Stratificarea și estimatorul Mantel-Haenszel

Stratificarea înseamnă împărțirea datelor în subgrupe (straturi) după valorile factorului de confuzie și calculul măsurii de asociere în fiecare strat. Dacă măsurile din straturi sunt asemănătoare între ele, dar diferă de măsura brută, avem confuzie. Pentru a obține o singură estimare ajustată, se folosește media ponderată propusă de Mantel și Haenszel:

$$\mathrm{OR(MH)} = \frac{\sum\left(\frac{a_i \times d_i}{n_i}\right)}{\sum\left(\frac{b_i \times c_i}{n_i}\right)}$$

unde indicele i se referă la stratul i, iar ni este totalul stratului. Pentru exemplul cu cafeaua, stratul fumătorilor are a = 80, b = 320, c = 20, d = 80, n = 500, iar stratul nefumătorilor a = 5, b = 95, c = 20, d = 380, n = 500. Numărătorul este $\frac{80\times80}{500} + \frac{5\times380}{500} = 12{,}8 + 3{,}8 = 16{,}6$; numitorul este $\frac{320\times20}{500} + \frac{95\times20}{500} = 12{,}8 + 3{,}8 = 16{,}6$. OR(MH) = 1,00, față de OR brut 2,36.

Pe mg2026, putem verifica dacă sexul confundă asocierea dintre obezitate și diabet. La femei (n = 86): a = 10, b = 19, c = 5, d = 52, OR = 5,47. La bărbați (n = 74): a = 8, b = 17, c = 6, d = 43, OR = 3,37. $\mathrm{OR(MH)} = \frac{\frac{10\times52}{86}+\frac{8\times43}{74}}{\frac{19\times5}{86}+\frac{17\times6}{74}} = \frac{6{,}047+4{,}649}{1{,}105+1{,}378} = \frac{10{,}695}{2{,}483} = 4{,}31$, practic identic cu OR brut de 4,32. Sexul nu este un factor de confuzie pentru această relație, ceea ce se potrivește cu OR al sexului din regresia logistică (0,97, capitolul 9).

Stratificarea funcționează bine pentru unul sau doi factori cu puține categorii. Când vrem să ajustăm simultan pentru mai mulți factori, unii numerici, folosim regresia multiplă (liniară, logistică, Cox), care face, în esență, aceeași operație într-un model. De aceea articolele raportează

frecvent "OR brut" (sau "neajustat") și "OR ajustat" (aOR); comparația lor arată cât de mult a contat ajustarea.

10.9.3 Modificarea efectului

Uneori măsurile de asociere din straturi diferă cu adevărat între ele: expunerea are un efect mai mare într-un subgrup decât în altul. Fenomenul se numește modificarea efectului (sau interacțiune) și nu este o eroare de corectat, ci o informație de raportat. Dacă un medicament reduce riscul de accident vascular cerebral cu 50% la vârstnici și deloc la tineri, o estimare comună ar ascunde un rezultat clinic important. În asemenea situații, se raportează măsura separat pentru fiecare strat.

Diferențele dintre straturi trebuie însă testate, pentru că pe eșantioane mici ele pot apărea doar prin variația de eșantionare. Pe mg2026, dacă stratificăm asocierea obezitate-diabet după vârstă (sub 56 de ani față de 56 de ani și peste), obținem OR = 11,67 la cei mai tineri (doar 6 cazuri de diabet) și OR = 3,99 la cei mai vârstnici. Diferența pare mare, dar testul de omogenitate Breslow-Day dă p = 0,37: nu avem dovezi pentru o modificare reală a efectului, ci doar estimări imprecise într-un strat cu foarte puține cazuri. OR(MH) ajustat pentru vârstă este 5,02 (IC 95% 2,04-12,35), aproape de OR ajustat din regresia logistică (6,01). Un studiu transversal pe 693 de studenți a testat formal un termen de interacțiune în regresia logistică și a raportat o modificare a efectului semnificativă pentru HbA1c (p = 0,023) [6]; acesta este modul corect de a afirma că un efect diferă între subgrupe.

Studiu de caz: ce se întâmplă cu o asociere după ajustare

Patru articole ale colegilor arată, cu cifre, de ce asocierea brută și cea ajustată trebuie citite împreună. Într-o cohortă prospectivă de 278 de pacienți internați cu insuficiență cardiacă decompensată, urmăriți 12 luni, raportul hemoglobină/lățimea distribuției eritrocitare (mediana 0,89) s-a asociat cu supraviețuirea fără evenimente în analiza Kaplan-Meier (log-rank p = 0,027), dar asocierea nu a mai fost semnificativă după ajustarea pentru vârstă, sex, NT-proBNP, fracția de ejecție și rata de filtrare glomerulară într-un model Cox (p = 0,240) [23]. Evenimentul primar a apărut la 167 de pacienți (60,1%). Cea mai probabilă explicație este confuzia: un marker hematologic se asociază cu severitatea bolii, iar după ce severitatea este luată în calcul prin NT-proBNP și fracția de ejecție, informația suplimentară dispare. Un rezultat "pierdut după ajustare" este un rezultat valoros, nu un eșec.

Într-o cohortă prospectivă de 133 de paciente cu carcinom ovarian seros de grad înalt, mutațiile BRCA (prezente la 39,1%) s-au asociat cu un risc de progresie mai mic în analiza univariată: HR = 0,52, IC 95% 0,27-0,99, p = 0,048 [117]. Raportul hazardurilor (HR) se citește ca RR: 0,52 înseamnă un risc instantaneu de progresie cu 48% mai mic. După ajustare, HR = 0,57, p = 0,124. Estimarea punctuală s-a schimbat puțin (de la 0,52 la 0,57), dar intervalul s-a lărgit și a ajuns să conțină 1. Aici problema nu este neapărat confuzia, ci precizia: cu 133 de paciente și mai multe variabile de ajustare, incertitudinea crește. Intervalul brut, care atinge aproape 1 (0,99), anunța deja fragilitatea rezultatului.

Într-o cohortă retrospectivă de 213 pacienți tratați în urgență pentru suspiciune de infecție bacteriană, femeile au fost mai vârstnice decât bărbații (mediana 77,0 față de 67,0 ani, p < 0,001), iar sursa urinară a infecției a fost mai frecventă la femei (51,0% față de 34,5%, p = 0,015) [9]. Rezumatul raportează un OR = 2,14 (IC 95% 1,20-3,81; p = 0,010), care rămâne semnificativ și după corecția Benjamini-Hochberg pentru testări multiple (p ajustat 0,039). Pentru mortalitatea intraspitalicească, 7 femei (7,0%) și 9 bărbați (8,0%), OR = 0,87 (IC 95% 0,31-2,43); rezumatul mai conține OR = 0,80 (IC 95% 0,27-2,35; p = 0,690), tot cu un interval larg care conține 1. Diferența de vârstă dintre sexe este exact situația în care vârsta poate confunda o comparație brută între femei și bărbați. Intervalele foarte largi pentru mortalitate

(de la 0,27 la 2,35) arată că, cu 16 decese, studiul nu poate exclude nici un efect protector important, nici unul dăunător.

Un sondaj online transversal pe 753 de tineri adulți (93,5% femei) a raportat prevalențele comportamentelor alimentare cu intervale Wilson și OR ajustate pentru vârstă și sex: aOR 2,34, 2,49 și 2,73 pentru diferite categorii de IMC, toate cu $p < 0,001$ [20]. Ajustarea pentru sex are însă o putere limitată când 93,5% dintre participanți sunt femei, iar eșantionul de conveniență (cine alege să răspundă online la un chestionar despre alimentație) limitează generalizarea, indiferent de cât de corect este calculat OR.

10.10 De la asociere la cauzalitate

10.10.1 Erorile sistematice ale studiilor observaționale

O asociere observată poate avea patru explicații: întâmplarea, eroarea sistematică (engl. bias), confuzia și cauzalitatea. Întâmplarea este evaluată prin intervalele de încredere și testele statistice. Confuzia a fost discutată mai sus. Erorile sistematice se grupează în două mari categorii. Eroarea de selecție apare când modul de recrutare a participanților face ca grupurile comparate să difere sistematic: martori aleși dintr-o populație diferită de cea a cazurilor, voluntari mai sănătoși decât populația generală, pierderi din urmărire legate atât de expunere, cât și de boală. Eroarea de informare (de măsurare) apare când expunerea sau boala sunt măsurate diferit în grupurile comparate: eroarea de memorie din studiile caz-control, un medic care caută mai atent boala la pacienții despre care știe că sunt expuși, clasificarea greșită a expunerii. Spre deosebire de eroarea întâmplătoare, eroarea sistematică nu scade prin creșterea eșantionului: un studiu mare și deplasat este precis în jurul unei valori greșite.

10.10.2 Criteriile Bradford Hill

În 1965, Austin Bradford Hill, care demonstrase împreună cu Richard Doll legătura dintre fumat și cancerul pulmonar prin studii caz-control și de cohortă, a propus nouă aspecte de luat în considerare înainte de a trage o concluzie cauzală dintr-o asociere observațională [108]. Le aplicăm pe exemplul fumat-BPOC.

Forța asocierii: cu cât RR este mai mare, cu atât este mai puțin probabil ca asocierea să fie produsă în întregime de un factor de confuzie necunoscut (în exemplul nostru, RR = 2,14). Consistența: asocierea a fost observată în mai multe studii, pe populații diferite, cu metode diferite. Specificitatea: expunerea duce la o anumită boală (criteriul cel mai slab, pentru că fumatul produce multe boli). Temporalitatea: expunerea precede boala; este singurul criteriu obligatoriu și este garantat de designul de cohortă. Gradientul biologic: riscul crește cu doza (mai multe pachete-an, risc mai mare). Plauzibilitatea biologică: există un mecanism cunoscut (inflamația cronică a căilor aeriene). Coerența: asocierea nu contrazice ce se știe despre istoria naturală a bolii. Dovezile experimentale: renunțarea la fumat încetinește declinul funcției pulmonare. Analogia: alte expuneri inhalatorii (pulberile profesionale) produc boli asemănătoare.

Bradford Hill nu le-a prezentat ca pe o listă de bifat, ci ca puncte de vedere care, cântărite împreună, fac o explicație cauzală mai mult sau mai puțin credibilă. Un studiu care raportează un OR ajustat cu un gradient doză-răspuns, de exemplu un OR de 1,43 pentru fiecare punct al unui scor de risc alimentar (IC 95% 1,23-1,66), aduce un argument mai solid decât unul care raportează doar o comparație între expuși și neexpuși [4].

10.10.3 Raportarea studiilor observaționale: STROBE

Pentru ca cititorul să poată judeca toate aceste aspecte, studiile observaționale trebuie raportate complet. Ghidul STROBE (Strengthening the Reporting of Observational Studies in

Epidemiology) cuprinde 22 de puncte pentru studiile de cohortă, caz-control și transversale [25]. Printre cerințe: precizarea designului în titlu sau rezumat, descrierea surselor de participanți și a criteriilor de selecție (iar în caz-control, a modului de alegere a martorilor), definirea clară a expunerii, a rezultatului și a factorilor de confuzie, explicarea modului de tratare a datelor lipsă, raportarea numărului de participanți la fiecare etapă, raportarea estimărilor brute și ajustate cu intervalele lor de încredere și discutarea limitelor, inclusiv a direcției probabile a erorilor sistematice. Un rezumat care dă OR fără interval de încredere nu respectă spiritul acestui ghid.

Studiu de caz: rapoarte de șanse în studiile colegilor, citite critic

Un studiu retrospectiv comparativ pe 1418 gravide (279 cu diabet gestațional, 74 cu diabet tip 1, 107 cu diabet tip 2 și 958 de martori) a analizat factorii de risc pentru complicații, cu OR și IC 95% [118]. Vârsta are OR = 1,03 (IC 95% 1,01-1,05; p = 0,002), obezitatea OR = 2,37 (IC 95% 1,42-3,94; p < 0,001), iar hipertensiunea cronică OR = 2,51 (IC 95% 1,26-5,01; p = 0,009). Un OR de 1,03 pare mic lângă 2,37, dar comparația este înșelătoare: OR al vârstei se referă la un an, iar al obezității la prezența față de absența unei condiții. Pentru 10 ani, OR al vârstei devine 1,03^10 ≈ 1,34. Unitatea predictorului trebuie citită întotdeauna înainte de a compara OR-uri.

Un studiu transversal pe 921 de studenți a găsit 24,1% supraponderali sau obezi și, pentru un tipar comportamental, aOR = 7,20 (IC 95% 4,26-12,17), plus aOR = 1,98 (IC 95% 1,26-3,10) pentru alt factor și 1,43 pentru fiecare punct al unui scor de risc alimentar (IC 95% 1,23-1,66; p < 0,001) [4]. Un aOR de 7,20 este o asociere puternică, dar, cu o prevalență a rezultatului de 24%, nu trebuie tradus prin "risc de 7 ori mai mare" (tabelul 10.4). Designul transversal lasă deschisă direcția: comportamentul alimentar a dus la supraponderare sau persoanele supraponderale și-au schimbat comportamentul? Pe aceeași populație, un alt studiu transversal pe 693 de studenți a raportat prevalența prediabetului de 21,1% și un OR = 1,78 pentru fiecare categorie a unui tipar alimentar ordonat (IC 95% 1,38-2,30; p < 0,001), cu prevalențe de 15,4%, 20,0% și 27,3% în cele trei categorii (chi-pătrat p = 0,003) [6]. Creșterea monotonă a prevalenței pe categorii este un gradient, unul dintre argumentele Bradford Hill.

Un studiu retrospectiv de cohortă pe 486 de pacienți cu COVID-19 (200 fără diabet, 62 cu diabet tip 1 și 224 cu diabet tip 2) a raportat pentru internarea în terapie intensivă OR: 2,24, iar pentru ventilația mecanică OR: 2,46 [8]. Rezumatul nu dă intervalele de încredere și nici valorile p pentru aceste OR. Ca exercițiu de lectură critică, întrebarea firească este: cât de precise sunt aceste estimări? Cu 62 de pacienți cu diabet tip 1, un OR de 2,24 ar putea avea un interval larg, eventual incluzând 1; fără interval, cititorul nu poate judeca. Este exact situația pe care ghidul STROBE încearcă să o prevină.

În fine, o analiză retrospectivă a internărilor pentru carcinom scuamos oral și orofaringian în 2018-2024 (248 de internări) a folosit o analiză de serii de timp întrerupte: numărul mediu lunar de internări a scăzut de la 4,9 înainte de pandemie la 2,6 în pandemie și 1,6 după, cu un raport al ratelor de incidență IRR = 0,973 (IC 95% 0,950-0,996; p = 0,024) pentru tendință, iar disecția ganglionară cervicală s-a asociat cu perioada printr-un OR ajustat de 1,87 (IC 95% 1,06-3,29; p = 0,030) [119]. IRR se interpretează ca un RR pentru rate: 0,973 pe lună înseamnă o scădere de aproximativ 2,7% pe lună. Studiul combină o abordare ecologică (rate pe perioade) cu una individuală (regresie logistică pe pacienți).

10.11 Probleme rezolvate

Toate rezultatele numerice de mai jos au fost calculate cu programe și verificate de mână; IC pentru RR și OR sunt calculate prin metoda logaritmică (Katz pentru RR, Woolf pentru OR), aceeași folosită implicit de Epi Info și MedCalc.

Problema rezolvată 10.1. Fumatul și BPOC: studiu de cohortă cu 1200 de participanți

Enunț. Se înrolează 1200 de persoane inițial sănătoase (fumători și nefumători), urmărite 5 ani. Din cei 300 de fumători, 50 fac BPOC; dintre nefumători, 70. (a) Formulați ipotezele de cercetare. (b) Formulați ipotezele statistice. (c) Construiți tabelul de contingență. (d) Este fumatul factor de risc sau de protecție? Calculați RR și OR. (e) Este riscul semnificativ statistic? Ce test folosim? (f) Care este interpretarea clinică?

(a) Ipoteza de cercetare: fumatul crește riscul de apariție a BPOC.

(b) Ipotezele statistice: H0: RR = 1, riscul de BPOC este același la fumători și la nefumători (fumatul nu este asociat cu BPOC); H1: RR ≠ 1, riscul diferă.

(c) Tipul de studiu: cohortă prospectivă (se pornește de la expunere, la persoane sănătoase, și se urmărește apariția bolii). Tabelul este tabelul 10.3: a = 50, b = 250, c = 70, d = 830, N = 1200.

(d) $R_1 = \frac{50}{300} = 16{,}7\%$; $R_0 = \frac{70}{900} = 7{,}8\%$. RR = 2,14 (IC 95% 1,53-3,01, calcul în secțiunea 10.7.1). OR $= \frac{50\times830}{250\times70} = 2{,}37$(; IC 95% 1,61 − 3,50). RR > 1 și intervalul nu conține 1: fumatul este un factor de risc. Fiind un studiu de cohortă, măsura principală este RR; OR se raportează doar ca informație suplimentară.

(e) Testul chi-pătrat (toate frecvențele așteptate sunt mult peste 5): χ^2 = 19,75, gl 1, p < 0,001 (cu corecția Yates 18,78, p < 0,001). H0 se respinge: asocierea este semnificativă statistic. Formularea "acceptăm H1" este tolerată în cursuri, dar corect este "respingem H0".

(f) Interpretarea clinică: în cei 5 ani, fumătorii au avut un risc de BPOC de aproximativ 2,1 ori mai mare decât nefumătorii (plauzibil între 1,5 și 3 ori). Diferența absolută este de 8,9 puncte procentuale (IC 95% 4,3-13,5), NNH = 12: la fiecare 12 fumători urmăriți 5 ani apare un caz de BPOC în plus. Fracția atribuibilă la expuși este 53%, iar fracția atribuibilă populațională 22%. Atenție la calculul greșit cu numitorul 1200 pentru nefumători, care dă un RR fals de 2,86.

Problema rezolvată 10.2. Același risc, eșantion de zece ori mai mic

Enunț. Presupunem că sunt înrolați doar 120 de participanți, dintre care 30 fumători. Se îmbolnăvesc 5 fumători și 7 nefumători. (a) Calculați RR și OR. (b) Este riscul semnificativ? (c) Justificați diferența față de problema 10.1.

Rezolvare. Tabelul: a = 5, b = 25, c = 7, d = 83 (nefumători 120 - 30 = 90; sănătoși 90 - 7 = 83).

(a) $R_1 = \frac{5}{30} = 16{,}7\%$; $R_0 = \frac{7}{90} = 7{,}8\%$; RR = 2,14, exact ca în problema 10.1. OR $= \frac{5\times83}{25\times7} = \frac{415}{175} = 2{,}37$, tot identic.

(b) ES[ln(RR)] $= \sqrt{\frac{1}{5} - \frac{1}{30} + \frac{1}{7} - \frac{1}{90}} = \sqrt{0{,}2 - 0{,}0333 + 0{,}1429 - 0{,}0111} = \sqrt{0{,}2984} = 0{,}546$. IC 95%: $\exp(0{,}762 \pm 1{,}96 \times 0{,}546) = \exp(0{,}762 \pm 1{,}071)$, adică $e^{-0{,}309} = 0{,}73$ și $e^{1{,}833} = 6{,}25$. IC 95% pentru RR: 0,73-6,25. ES[ln(OR)] $= \sqrt{\frac{1}{5} + \frac{1}{25} + \frac{1}{7} + \frac{1}{83}} = \sqrt{0{,}3950} = 0{,}628$; IC 95% pentru OR: 0,69-8,13. Ambele intervale conțin 1. Frecvențele așteptate sunt 3, 27, 9 și 81; una este sub 5, deci testul potrivit este Fisher: p = 0,17 (chi-pătrat necorectat 1,98, p = 0,16). H0 nu se respinge.

(c) Măsurile de asociere sunt identice, dar cu 12 cazuri în total informația este insuficientă pentru a distinge un RR de 2 de un RR de 1. Concluzia corectă nu este "fumatul nu este

factor de risc", ci "studiul nu are suficientă putere statistică pentru a demonstra asocierea". Absența dovezii nu este dovada absenței. Figura 10.2 arată cele două intervale unul sub altul: aceeași estimare, precizii foarte diferite.

Problema rezolvată 10.3. Expunerea la monitor și acuitatea vizuală (studiu transversal)

Enunț. Într-un studiu transversal pe 1000 de persoane, 200 lucrează cel puțin 8 ore pe zi în fața monitorului, dintre care 90 au acuitatea vizuală scăzută; dintre cele 800 care lucrează mai puțin, 150 au acuitatea scăzută. (a) Care este factorul de risc și care ar putea fi influența lui? (b) Calculați măsura de asociere potrivită. (c) Care este șansa (odds) de acuitate scăzută la cei care lucrează sub 8 ore pe zi? (d) Există o asociere semnificativă? (e) Calculați IC 95% pentru OR.

Rezolvare. (a) Factorul de risc investigat este expunerea prelungită la monitor; influența posibilă este creșterea frecvenței acuității vizuale scăzute. Tabelul: a = 90, b = 110, c = 150, d = 650.

(b) Studiul este transversal, deci măsurăm prevalențe: 90 / 200 = 45,0% la expuși și 150 / 800 = 18,75% la neexpuși. Raportul prevalențelor RP $= \frac{45{,}0}{18{,}75} = 2{,}40$(; IC 95% 1,94 − 2,96). OR $= \frac{90\times650}{110\times150} = \frac{58\,500}{16\,500} = 3{,}55$.

(c) Șansa la cei cu expunere sub 8 ore: 150 / 650 = 0,231 (aproximativ 1 persoană cu acuitate scăzută la 4,3 cu acuitate normală). La cei expuși, șansa este 90 / 110 = 0,818. Dacă enunțul cere OR al celor cu expunere scurtă față de cei cu expunere lungă, acesta este inversul: 1 / 3,55 = 0,28.

(d) χ^2 = 60,44, gl 1, p < 0,001; asocierea este semnificativă.

(e) ln(3,545) = 1,266; ES $= \sqrt{\frac{1}{90}+\frac{1}{110}+\frac{1}{150}+\frac{1}{650}} =$ $\sqrt{0{,}01111+0{,}00909+0{,}00667+0{,}00154} = \sqrt{0{,}02841} = 0{,}1685$; 1,266 ± 0,330 dă 0,935 și 1,596; IC 95% pentru OR: 2,55-4,93.

Interpretare: prevalența acuității scăzute este de 2,4 ori mai mare la cei care lucrează cel puțin 8 ore pe zi la monitor. OR (3,55) depășește mult RP (2,40) pentru că afecțiunea este frecventă (24% în total), deci OR nu trebuie citit ca risc. Designul transversal nu permite concluzii despre direcție: e posibil, de exemplu, ca persoanele cu anumite profesii să aibă și alte caracteristici (vârsta) asociate cu acuitatea scăzută.

Problema rezolvată 10.4. Contraceptivele orale și artrita reumatoidă (studiu caz-control)

Enunț. Un cercetător a selectat 100 de femei cu artrită reumatoidă confirmată și 200 de femei tratate în aceeași instituție pentru alte boli musculo-scheletale. Din foile de observație s-a stabilit utilizarea contraceptivelor orale înainte de diagnostic: 40 dintre cazuri și 120 dintre martore le utilizaseră. (a) Ce tip de studiu este? (b) Calculați OR. (c) Calculați IC 95% pentru OR. Se poate calcula RR? (d) Există o asociere semnificativă?

Rezolvare. (a) Studiu caz-control: selecția s-a făcut după boală (cazuri cu artrită, martore fără), iar expunerea a fost căutată în trecut. Martorele provin din aceeași instituție, ceea ce este o alegere rezonabilă. Tabelul: a = 40 (cazuri expuse), b = 120 (martore expuse), c = 60 (cazuri neexpuse), d = 80 (martore neexpuse).

(b) OR $= \frac{40\times80}{120\times60} = \frac{3200}{7200} = 0{,}44$. Șansa de a fi utilizat contraceptive este de 0,44 ori cea a martorelor (40% dintre cazuri față de 60% dintre martore au fost expuse).

(c) $\ln(0{,}444) = -0{,}811$; ES $= \sqrt{\frac{1}{40} + \frac{1}{120} + \frac{1}{60} + \frac{1}{80}} = \sqrt{0{,}025 + 0{,}00833 + 0{,}01667 + 0{,}0125} = \sqrt{0{,}0625} = 0{,}250$; -0,811 ± 0,490 dă -1,301 și -0,321; IC 95% pentru OR: 0,27-0,73. RR nu se poate calcula: proporția de bolnave din eșantion (100 din 300) a fost fixată de cercetător și nu reflectă riscul din populație. Un calcul mecanic al "RR" (40/160 față de 60/140, adică 0,58) nu are interpretare de risc. Dacă artrita reumatoidă este rară în populație, OR = 0,44 aproximează totuși RR.

(d) $\chi^2 = 10{,}71$, gl 1, p = 0,001; intervalul pentru OR nu conține 1. Asocierea este semnificativă. Cum OR < 1, utilizarea contraceptivelor orale apare ca factor de protecție în acest studiu. Interpretarea cauzală cere prudență: martorele cu alte boli musculo-scheletale ar putea avea, din motive legate de bolile lor, un consum diferit de contraceptive față de populația generală (eroare de selecție).

Problema rezolvată 10.5. Anomaliile ECG și decesul de cauză coronariană: un rezultat la limită

Enunț. Au fost urmăriți 20 de ani 47 de bărbați de 40-64 de ani cu anomalii ECG la examinarea inițială și 144 de bărbați de aceeași vârstă fără anomalii. Au decedat prin boală coronariană 8 din primul grup și 10 din al doilea. (a) Tipul de studiu. (b) Testați ipoteza că anomaliile ECG sunt factor de risc. (c) Calculați și interpretați RR. (d) Calculați OR și IC 95%.

Rezolvare. (a) Cohortă prospectivă: grupurile sunt definite după expunere (anomalii ECG) și urmărite pentru deces. Tabelul: a = 8, b = 39, c = 10, d = 134, N = 191.

(b) H0: RR = 1; H1: RR ≠ 1. Frecvențele așteptate: $\frac{a\,47\times18}{191} = 4{,}43$ (sub 5), b 42,57, c 13,57, d 130,43. Chi-pătrat necorectat: 4,22, p = 0,040; cu corecția Yates: 3,12, p = 0,077; testul exact Fisher: p = 0,0485. Pentru că o frecvență așteptată este sub 5, testul recomandat este Fisher, iar p = 0,0485 este chiar sub pragul de 0,05.

(c) $R_1 = \frac{8}{47} = 17{,}0\%$; $R_0 = \frac{10}{144} = 6{,}9\%$; RR = 2,45. ES $= \sqrt{\frac{1}{8} - \frac{1}{47} + \frac{1}{10} - \frac{1}{144}} = \sqrt{0{,}1968} = 0{,}444$; IC $95\% = \exp(0{,}897 \pm 0{,}870) = 1{,}03 - 5{,}85$.

(d) OR $= \frac{8\times134}{39\times10} = \frac{1072}{390} = 2{,}75$; ES $= \sqrt{\frac{1}{8} + \frac{1}{39} + \frac{1}{10} + \frac{1}{134}} = \sqrt{0{,}2581} = 0{,}508$; IC $95\% = \exp(1{,}011 \pm 0{,}996) = 1{,}02 - 7{,}44$.

Interpretare: anomaliile ECG se asociază cu un risc de deces coronarian de aproximativ 2,5 ori mai mare, dar dovada este la limită. Intervalele de încredere încep chiar deasupra lui 1 (1,03 și 1,02), testul necorectat și testul Fisher sunt abia semnificative, iar testul cu corecție Yates nu este. Cu 18 decese în total, o concluzie onestă este: "date compatibile cu un risc crescut, dar imprecise; rezultatul trebuie confirmat pe un eșantion mai mare". Studentul care scrie doar "p < 0,05, deci anomaliile ECG sunt factor de risc" pierde jumătate din informație.

Problema rezolvată 10.6. Expunerea la pulberi și vârsta: doi factori de risc

Enunț. Partea I: sunt înrolați 320 de pensionari de 70-75 de ani, dintre care 80 au lucrat cel puțin 5 ani într-un mediu cu pulberi. La investigații, atât în grupul expus, cât și în cel neexpus, câte 30 de persoane au capacitatea pulmonară scăzută. Partea a II-a: pentru a studia vârsta ca factor de risc independent, se investighează și 200 de persoane de 30-50 de ani, neexpuse la pulberi, dintre care 12 au capacitatea pulmonară scăzută. Analizați ambele părți.

Rezolvare, partea I. Tipul de studiu: transversal, cu expunerea profesională reconstituită din trecut (poate fi privit și ca o cohortă retrospectivă). Tabelul: a = 30, b = 50 (80 - 30), c = 30, d = 210 (240 - 30). Observați capcana enunțului: "30 în ambele grupuri" nu înseamnă risc egal, pentru

că grupurile au mărimi diferite. $R_1 = \frac{30}{80} = 37{,}5\%$; $R_0 = \frac{30}{240} = 12{,}5\%$. RP = 3,00 (IC 95% 1,94-4,65). $\text{OR} = \frac{30\times210}{50\times30} = \frac{6300}{1500} = 4{,}20(;\ \text{IC}\,95\%\ 2{,}32 - 7{,}60)$. χ^2 = 24,62, p < 0,001. Expunerea la pulberi se asociază cu o frecvență de trei ori mai mare a scăderii capacității pulmonare; diferența absolută este de 25 de puncte procentuale, iar fracția atribuibilă la expuși $\frac{\text{RR}-1}{\text{RR}} = \frac{2}{3} = 67\%$.

Partea a II-a. Pentru a izola efectul vârstei, comparăm doar persoane neexpuse, ca expunerea la pulberi să nu confunde comparația: vârstnici neexpuși 30 / 240 = 12,5% față de adulți tineri neexpuși 12 / 200 = 6,0%. Tabelul: a = 30, b = 210, c = 12, d = 188. $\text{RR} = \frac{12{,}5}{6{,}0} = 2{,}08(;\ \text{IC}\,95\%\ 1{,}10 - 3{,}96)$; $\text{OR} = \frac{30\times188}{210\times12} = \frac{5640}{2520} = 2{,}24(;\ \text{IC}\,95\%\ 1{,}11 - 4{,}50)$; χ^2 = 5,34, p = 0,021. Vârsta înaintată este și ea un factor de risc, de aproximativ două ori. Dacă am compara vârstnicii expuși cu tinerii neexpuși, am obține $\text{RR} = \frac{37{,}5}{6{,}0} = 6{,}25$, dar această cifră amestecă efectul vârstei cu cel al pulberilor. Lipsa tinerilor expuși ne împiedică să verificăm dacă efectul pulberilor este același la toate vârstele (modificarea efectului). Lecția de design: alegerea grupului de comparație decide ce întrebare primește răspuns.

Problema rezolvată 10.7. Obezitatea și diabetul în mg2026: RP, OR și fracția atribuibilă

Enunț. La includere în mg2026, dintre cei 54 de pacienți obezi (IMC ≥ 30), 18 au diabet confirmat prin TTGO; dintre cei 106 neobezi, 11. Calculați RP, OR, IC 95%, testul de semnificație și fracția atribuibilă populațională. Comentați diferența dintre OR și RP.

Rezolvare. Tabelul: a = 18, b = 36, c = 11, d = 95, N = 160. Studiul este transversal (analiza de la includere), deci prevalențele sunt 18 / 54 = 33,3% la obezi și 11 / 106 = 10,4% la neobezi.

$\text{RP} = \frac{33{,}3}{10{,}4} = 3{,}21.$ $\text{ES}[\ln(\text{RP})] = \sqrt{\frac{1}{18} - \frac{1}{54} + \frac{1}{11} - \frac{1}{106}} = \sqrt{0{,}0556 - 0{,}0185 + 0{,}0909 - 0{,}0094} = \sqrt{0{,}1185} = 0{,}344;\ \text{IC}\,95\% = \exp(1{,}167 \pm 0{,}675) = 1{,}64 - 6{,}31.$

$\text{OR} = \frac{18\times95}{36\times11} = \frac{1710}{396} = 4{,}32.$ $\text{ES}[\ln(\text{OR})] = \sqrt{\frac{1}{18} + \frac{1}{36} + \frac{1}{11} + \frac{1}{95}} = \sqrt{0{,}1847} = 0{,}430;$ $\text{IC}\,95\% = \exp(1{,}463 \pm 0{,}842) = 1{,}86 - 10{,}03.$

Frecvențele așteptate sunt 9,8; 44,2; 19,2; 86,8 (toate peste 5). χ^2 = 12,70, p < 0,001 (Yates 11,20, p = 0,0008; Fisher p = 0,0008). Asocierea este semnificativă.

Fracția atribuibilă populațională: pe = 54 / 160 = 0,3375; $\text{FAP} = \frac{0{,}3375\times2{,}21}{1+0{,}3375\times2{,}21} = \frac{0{,}746}{1{,}746} = 0{,}43$. Echivalent, (18,1% - 10,4%) / 18,1% = 0,43. Dacă asocierea ar fi cauzală și obezitatea ar putea fi eliminată, aproximativ 43% dintre cazurile de diabet din acest eșantion ar fi evitate. Formularea condițională este obligatorie: datele sunt transversale și nu stabilesc succesiunea temporală.

OR (4,32) este mult mai mare decât RP (3,21) pentru că diabetul este frecvent la obezi (o treime). Raportarea OR ca "risc de peste patru ori mai mare" ar exagera. Ajustarea pentru sex prin Mantel-Haenszel nu schimbă OR (4,31), iar regresia logistică ajustată pentru vârstă și sex dă OR 6,01 (cap. 9). Pe figura 10.2, intervalul OR este mai larg decât al RP, ca de obicei.

10.12 Exerciții propuse

1. În 1985 s-a studiat legătura dintre dispozitivul intrauterin (DIU) și infertilitate. Dintre 283 de femei infertile, 89 utilizaseră un DIU; dintre 3833 de femei martor, 640. (a) Ce tip de

studiu este? (b) Construiți tabelul 2 x 2. (c) Calculați OR cu IC 95% și interpretați. (d) Aplicați testul chi-pătrat la $\alpha = 0{,}05$.

2. Într-un studiu pe 1024 de mame, 23 dintre cele 107 mame fumătoare și 117 dintre cele 917 mame nefumătoare au avut copii cu sindrom Down, cazurile fiind identificate din certificatele de naștere. Calculați OR și IC 95% și discutați dacă OR poate fi aproximat prin RR.
3. În studiul mg2026, la 6 luni, 16 din 53 de pacienți din brațul B și 6 din 53 din brațul A au slăbit cel puțin 5%. Calculați RR, diferența absolută de risc și NNT. Comparați cu rezultatul C față de A (RR 2,62; NNT 6).
4. În mg2026, dintre cei 53 de fumători, 10 au diabet confirmat prin TTGO; dintre cei 107 nefumători și foști fumători, 19. Calculați RP și OR cu IC 95% și formulați concluzia statistică și pe cea clinică.
5. În mg2026, dintre cei 54 de obezi, 21 au TAS $\geq$ 140 mmHg; dintre cei 106 neobezi, 20. Calculați RP, OR, IC 95% și χ^2. De ce diferă OR de RP?
6. O cohortă de muncitori a fost urmărită pentru dermatită de contact. Expușii la un solvent au acumulat 400 de persoane-an și 10 cazuri; neexpușii, 800 de persoane-an și 8 cazuri. Calculați rata de incidență în fiecare grup (la 100 de persoane-an) și raportul ratelor de incidență.
7. O boală cronică are o rată de incidență de 5 cazuri la 1000 de persoane-an și o durată medie de 4 ani. Estimați prevalența. Cum s-ar schimba prevalența dacă un tratament nou ar dubla durata medie a bolii, fără să o vindece?
8. Un medicament produce un efect advers la 6% dintre pacienții tratați, față de 2% la placebo. Calculați diferența de risc, RR și NNH. Ce înseamnă NNH clinic?
9. O expunere are RR = 3 și este prezentă la 40% din populație. Calculați fracția atribuibilă la expuși și fracția atribuibilă populațională.
10. Un studiu este stratificat după un factor de confuzie. Stratul 1: a = 20, b = 80, c = 10, d = 60. Stratul 2: a = 30, b = 20, c = 60, d = 60. Calculați OR în fiecare strat, OR brut (din tabelul combinat) și OR Mantel-Haenszel. Ce concluzie trageți?
11. În mg2026, dintre cei 65 de pacienți din mediul rural, 10 au diabet; dintre cei 95 din mediul urban, 19. Calculați OR cu IC 95%. Este corect să spunem că mediul rural "protejează" împotriva diabetului?

Capcane frecvente

- Construirea greșită a tabelului 2 x 2: rânduri și coloane inversate sau un total pus în locul unei celule; riscul la neexpuși se calculează numai printre neexpuși (fumat-BPOC: RR corect 2,14, nu 2,86).
- Calculul RR într-un studiu caz-control; acolo proporția de bolnavi este fixată de cercetător și se calculează doar OR.
- Citirea unui OR de boală frecventă ca risc relativ: OR exagerează efectul când rezultatul este frecvent (OR 4,32 față de RP 3,21 în mg2026).
- Concluzia "factorul nu are efect" dintr-un rezultat nesemnificativ pe un eșantion mic; corect este "nu am putut demonstra un efect" (problema 10.2).
- Ignorarea frecvențelor așteptate sub 5 și aplicarea mecanică a testului chi-pătrat în loc de Fisher.
- Interpretarea unui OR per unitate (per an, per punct) ca și cum ar fi un OR pentru o variabilă binară.
- Confundarea asocierii cu cauzalitatea în studiile transversale, unde succesiunea temporală nu poate fi stabilită.
- Raportarea OR sau RR fără interval de încredere.

- Confundarea factorului de confuzie cu modificarea efectului: primul se corectează prin ajustare, al doilea se raportează pe straturi.

De reținut

- Notația tabelului: a = expuși bolnavi, b = expuși sănătoși, c = neexpuși bolnavi, d = neexpuși sănătoși.
- $RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$; $OR = \frac{a \times d}{b \times c}$; RR > 1 risc, RR = 1 indiferent, RR < 1 protecție.
- Cohortă și studiu experimental: RR (și OR); transversal: RP (și OR); caz-control: doar OR.
- OR ≈ RR doar pentru boli rare; când boala este frecventă, OR este mai departe de 1 decât RR.
- IC 95%: $ES[\ln(RR)] = \sqrt{\left[\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}\right]}$; $ES[\ln(OR)] = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$; intervalul care conține 1 înseamnă rezultat nesemnificativ.
- Riscul atribuibil (R1 - R0) măsoară impactul absolut; NNT sau $NNH = \frac{1}{|R_1 - R_0|}$, rotunjit în sus.
- Factorul de confuzie se asociază cu expunerea și cu boala; se controlează prin randomizare, potrivire, stratificare (Mantel-Haenszel) sau regresie multiplă.
- Criteriile Bradford Hill cântăresc argumentele cauzale; temporalitatea este singura obligatorie. Studiile observaționale se raportează după ghidul STROBE.
- Pe mg2026: obezitate-diabet RP 3,21 (1,64-6,31), OR 4,32 (1,86-10,03); brațul C față de A: RR 2,62 (1,11-6,17), diferența 18,3 puncte procentuale, NNT 6.

Întrebări de autoevaluare

1. (CS) Un studiu selectează 150 de pacienți cu cancer gastric și 300 de persoane fără cancer și investighează consumul anterior de alimente afumate. Măsura de asociere potrivită este:
   a) riscul relativ
   b) incidența cumulativă
   c) raportul de șanse
   d) raportul prevalențelor
   e) rata de incidență
2. (CS) Într-o cohortă, riscul de boală este 20% la expuși și 10% la neexpuși. Numărul necesar pentru a produce un caz în plus (NNH) este:
   a) 2
   b) 5
   c) 10
   d) 20
   e) 0,1
3. (CS) Un RR = 0,70 cu IC 95% 0,55-0,89 indică:
   a) un factor de risc semnificativ
   b) un factor de protecție semnificativ, cu o reducere a riscului de aproximativ 30%
   c) un rezultat nesemnificativ, deoarece RR este sub 1
   d) o reducere a riscului de 70%
   e) un factor indiferent
4. (CS) Într-un studiu transversal, prevalența bolii este 45% la expuși și 18,75% la neexpuși. Afirmația corectă este:
   a) OR va fi aproximativ egal cu raportul prevalențelor, deoarece boala este rară
   b) raportul prevalențelor este 2,40, iar OR va fi mai mare decât 2,40

c) studiul permite stabilirea succesiunii temporale
d) se poate calcula incidența cumulativă
e) OR va fi mai mic decât raportul prevalențelor

5. (CS) Intervalul de încredere 95% pentru un OR = 2,37 este 0,69-8,13. Concluzia corectă este:
a) expunerea dublează sigur șansa de boală
b) asocierea este semnificativă la $\alpha = 0,05$
c) datele nu permit respingerea H0; studiul nu a demonstrat o asociere
d) expunerea este un factor de protecție
e) intervalul este greșit, deoarece nu este simetric în jurul lui 2,37

6. (CM) Despre studiul de cohortă sunt adevărate:
a) pornește de la persoane fără boală, clasificate după expunere
b) permite calculul direct al incidenței și al riscului relativ
c) este cel mai eficient design pentru bolile foarte rare
d) poate fi prospectiv sau retrospectiv
e) este expus pierderilor din urmărire

7. (CM) Despre raportul de șanse sunt adevărate:
a) se calculează ca (a x d) / (b x c)
b) într-un studiu caz-control, OR al expunerii este egal cu OR al bolii
c) pentru boli rare, OR aproximează riscul relativ
d) OR este întotdeauna egal cu RR
e) OR < 1 indică un factor de protecție

8. (CM) Un factor de confuzie:
a) este asociat cu expunerea
b) este factor de risc pentru boală, independent de expunere
c) se află obligatoriu pe calea cauzală dintre expunere și boală
d) poate fi controlat prin stratificare și estimatorul Mantel-Haenszel
e) poate fi controlat la etapa de design prin randomizare sau potrivire

9. (CM) Despre studiul caz-control sunt adevărate:
a) este potrivit pentru bolile rare și pentru cele cu latență lungă
b) este susceptibil la eroarea de memorie
c) martorii trebuie să provină din aceeași populație din care provin cazurile
d) permite calculul incidenței bolii la expuși
e) potrivirea pe vârstă împiedică studierea vârstei ca factor de risc

10. (CM) Dintre criteriile Bradford Hill pentru evaluarea cauzalității fac parte:
a) temporalitatea
b) gradientul biologic (relația doză-răspuns)
c) valoarea p sub 0,05
d) consistența între studii
e) plauzibilitatea biologică

Capitolul 11. Metode statistice în decizia medicală: analiza bayesiană și medicina bazată pe dovezi

Obiective

La finalul acestui capitol, studentul va putea:

- să explice diferența dintre probabilitatea înțeleasă ca frecvență și probabilitatea înțeleasă ca grad de încredere și să recunoască ambele sensuri în raționamentul clinic;
- să calculeze probabilități necondiționate și condiționate dintr-un tabel boală-simptom și să aplice teorema lui Bayes, inclusiv prin formula probabilității totale;
- să transforme o problemă de diagnostic în frecvențe naturale și să construiască arborele corespunzător;
- să facă trecerea de la probabilitate la șansă și înapoi și să folosească rapoartele de verosimilitate pentru a obține probabilitatea post-test, cu calculul de mână și cu nomograma Fagan;
- să judece când două teste pot fi combinate în serie sau în paralel și ce presupune ipoteza independenței condiționate;
- să rezolve un tabel de diagnostic diferențial cu trei boli și mai multe semne clinice;
- să descrie pragul de testare și pragul de tratament și să le calculeze într-un exemplu simplu;
- să interpreteze un interval credibil și un factor Bayes (BF10) raportat de JASP;
- să aplice cei cinci pași ai medicinei bazate pe dovezi, să citească un grafic forest și să recunoască principalele erori cognitive din raționamentul clinic.

11.1 Incertitudinea în decizia clinică

O doamnă de 58 de ani intră în cabinetul de medicină de familie pentru o rețetă. Are IMC 31, tatăl ei a avut diabet, nu face mișcare. Nu se plânge de nimic. Medicul se întreabă dacă are diabet nediagnosticat. Nu poate ști sigur fără un test, dar nici nu pornește de la zero: știe că, în populația de adulți cu risc cardiometabolic, cam unul din cinci sau șase are diabet nediagnosticat, iar profilul ei îl face să creadă că riscul este chiar mai mare. Din acest moment, fiecare informație nouă (scorul FINDRISC, glicemia a jeun, HbA1c) mută această estimare în sus sau în jos.

Acesta este, pe scurt, subiectul capitolului. Diagnosticul nu este un act de recunoaștere instantanee, ci o succesiune de revizuiri ale unei probabilități. Medicii experimentați fac aceste revizuiri intuitiv, uneori foarte bine, alteori cu erori sistematice pe care le vom descrie în secțiunea 11.11. Statistica oferă instrumentul care face revizuirea explicită și verificabilă: teorema lui Bayes [19,120].

11.1.1 De la simptom la diagnostic: o problemă de probabilități

În capitolele anterioare am folosit probabilitatea mai ales pentru a descrie variabilitatea de eșantionare: cât de probabil este să obținem o diferență atât de mare între două medii dacă H0 ar fi adevărată. În decizia clinică, întrebarea se schimbă. Medicul nu are nevoie de probabilitatea datelor în ipoteza că pacientul este sănătos, ci de probabilitatea ca pacientul să fie bolnav, având în vedere datele. Cele două probabilități sunt diferite și confundarea lor este una dintre cele mai vechi greșeli de raționament din medicină.

Un exemplu clasic: 80% dintre pacienții cu pneumonie au febră. Asta nu înseamnă că 80% dintre pacienții cu febră au pneumonie. Prima afirmație descrie p(febră / pneumonie), a doua ar descrie p(pneumonie / febră). Într-un cabinet de medicină de familie, iarna, majoritatea pacienților

febrili au viroze respiratorii banale, deci a doua probabilitate este mult mai mică decât prima. Teorema lui Bayes este exact formula care leagă cele două probabilități și ne spune ce informație suplimentară este necesară pentru a trece de la una la cealaltă: cât de frecventă este boala în populația din care provine pacientul.

11.1.2 Probabilitatea ca frecvență și probabilitatea ca grad de încredere

Există două feluri de a înțelege cuvântul "probabilitate". În sens frecventist, probabilitatea unui eveniment este limita frecvenței relative a acelui eveniment într-un număr foarte mare de repetări ale aceluiași experiment. Când spunem că prevalența diabetului nediagnosticat în setul mg2026 este 18,1% (29 din 160), descriem o frecvență observată, care estimează o frecvență din populație.

În sens subiectiv sau bayesian, probabilitatea exprimă gradul de încredere al unei persoane într-o afirmație, pe baza informației pe care o are. Când medicul spune "cred că există o probabilitate de 30% ca acest pacient să aibă diabet", el nu se referă la o serie de repetări ale aceluiași pacient (pacientul fie are, fie nu are diabet), ci la cât de convins este. Această probabilitate se poate schimba pe măsură ce apar informații noi, iar teorema lui Bayes spune cum trebuie să se schimbe ca să rămână coerentă.

În practică, cele două sensuri se întâlnesc. Gradul de încredere al medicului înainte de test (probabilitatea pre-test) este bine să fie ancorat într-o frecvență: prevalența bolii la pacienții asemănători. Iar performanța testelor (sensibilitatea, specificitatea) este estimată frecventist, din studii de acuratețe diagnostică, cum vom vedea în capitolul 12. Analiza bayesiană din decizia clinică folosește deci frecvențe pentru a actualiza grade de încredere [63,64].

11.2 Probabilitatea condiționată, revăzută

11.2.1 Tabelul boală-simptom

Cursul disciplinei pornește de la un tabel de contingență în care pe rânduri se află prezența (B+) sau absența (B-) bolii, iar pe coloane prezența (S+) sau absența (S-) simptomului. Același tabel poate fi construit pentru orice pereche boală-simptom, iar într-un sistem expert de diagnostic se construiesc zeci sau sute de astfel de tabele.

Tabelul 11.1. Tabelul boală-simptom, notațiile din curs.

	S+	**S-**	**Total**
B+	a	b	a + b
B-	c	d	c + d
Total	a + c	b + d	$N = a + b + c + d$

Din acest tabel se calculează două tipuri de probabilități. Probabilitățile necondiționate (marginale) se referă la tot eșantionul:

$$p(B^+) = \frac{a+b}{N}$$

$$p(S^+) = \frac{a+c}{N}$$

Probabilitățile condiționate se calculează în interiorul unui rând sau al unei coloane. De exemplu, probabilitatea de a avea simptomul printre cei fără boală este

$$p(S^+/B^-) = \frac{c}{c+d}$$

iar probabilitatea de a avea boala printre cei cu simptom este

$$p(B^+/S^+) = \frac{a}{a+c}$$

Bara oblică se citește "dat fiind" sau "condiționat de". Tot ce se află după bară restrânge populația la care ne referim. Aici apare prima regulă practică: ca să știi ce numitor folosești, uită-te după bară. Dacă după bară scrie B-, numitorul este totalul rândului B-, adică c + d.

Să luăm un exemplu cu numere. Într-un cabinet de medicină de familie au fost evaluați, într-o iarnă, 200 de adulți cu tuse acută. La 30 dintre ei radiografia a confirmat pneumonia. Febra peste 38 °C a fost prezentă la 24 dintre cei cu pneumonie și la 34 dintre cei fără pneumonie (date fictive).

Tabelul 11.2. Pneumonie (B) și febră (S) la 200 de pacienți cu tuse acută (date fictive).

	S+ (febră)	**S-**	**Total**
B+ (pneumonie)	24	6	30
B-	34	136	170
Total	58	142	200

Calculăm:

$$p(B^+) = \frac{30}{200} = 0{,}15$$

$$p(S^+) = \frac{58}{200} = 0{,}29$$

$$p(S^+/B^+) = \frac{24}{30} = 0{,}80$$

$$p(S^+/B^-) = \frac{34}{170} = 0{,}20$$

$$p(B^+/S^+) = \frac{24}{58} = 0{,}414$$

Observați asimetria: febra apare la 80% dintre pacienții cu pneumonie, dar doar 41,4% dintre pacienții febrili au pneumonie. Restul de 58,6% sunt pacienți febrili fără pneumonie, care, fiind mult mai numeroși în populația cu tuse, "diluează" grupul celor cu febră.

11.2.2 Regula înmulțirii și formula probabilității totale

Probabilitatea ca două evenimente să apară simultan se obține cu regula înmulțirii:

$$p(B^+ \text{ și } S^+) = p(S^+/B^+) \times p(B^+) = p(B^+/S^+) \times p(S^+)$$

În exemplul nostru, $p(B^+ \text{ și } S^+) = \frac{24}{200} = 0{,}12$. Verificăm: 0,80 x 0,15 = 0,12 și 0,414 x 0,29 = 0,12. Cele două scrieri sunt egale pentru că descriu aceeași celulă a tabelului, adică celula a.

A doua unealtă este formula probabilității totale. Simptomul poate apărea fie la un bolnav, fie la un nebolnav, iar cele două situații se exclud reciproc. Prin urmare:

$$p(S^+) = p(S^+/B^+) \times p(B^+) + p(S^+/B^-) \times p(B^-)$$

Pentru tabelul 11.2: $p(S^+) = 0{,}80 \times 0{,}15 + 0{,}20 \times 0{,}85 = 0{,}12 + 0{,}17 = 0{,}29$, exact cât am obținut din marginea tabelului. Formula pare banală când avem tabelul complet, dar devine indispensabilă când nu-l avem: de cele mai multe ori, medicul cunoaște sensibilitatea unui test (din literatură) și prevalența bolii (din experiență sau din studii epidemiologice), dar nu are un tabel cu pacienții lui.

Dacă bolile posibile sunt mai multe, de exemplu B1, B2 și B3, mutual exclusive și care acoperă toate posibilitățile, formula se extinde firesc:

$$p(S) = p(S/B_1) \times p(B_1) + p(S/B_2) \times p(B_2) + p(S/B_3) \times p(B_3)$$

Vom folosi această formă în secțiunea 11.6, la diagnosticul diferențial.

11.3 Teorema lui Bayes

11.3.1 Derivarea

Pornim de la cele două scrieri ale regulii înmulțirii pentru aceeași celulă:

$$p(B/S) \times p(S) = p(S/B) \times p(B)$$

Împărțim ambii membri la p(S) și obținem teorema lui Bayes, în forma din cursul disciplinei:

$$p(B/S) = \frac{p(S/B) \times p(B)}{p(S)}$$

Înlocuind numitorul cu formula probabilității totale, obținem forma pe care o vom folosi cel mai des:

$$p(B^+/S^+) = \frac{p(S^+/B^+) \times p(B^+)}{p(S^+/B^+) \times p(B^+) + p(S^+/B^-) \times p(B^-)}$$

Fiecare termen are un nume. p(B+) este probabilitatea a priori sau pre-test: cât de probabilă era boala înainte să știm ceva despre simptom sau test. p(S+ / B+) este verosimilitatea simptomului la bolnavi; când S este rezultatul pozitiv al unui test, aceasta este sensibilitatea. p(S+ / B-) este frecvența simptomului la nebolnavi; pentru un test, este 1 - specificitatea, adică rata falselor pozitive. p(B+ / S+) este probabilitatea a posteriori sau post-test; pentru un test pozitiv, este valoarea predictivă pozitivă.

Aplicat pe tabelul 11.2, fără să ne uităm la celula a:

$$p(B^+/S^+) = \frac{0{,}80 \times 0{,}15}{0{,}80 \times 0{,}15 + 0{,}20 \times 0{,}85} = \frac{0{,}12}{0{,}29} = 0{,}414$$

Am regăsit valoarea calculată direct. Forța formulei apare când schimbăm doar prevalența. Să presupunem că aceeași febră, cu aceeași frecvență la bolnavi și nebolnavi, apare la pacienții cu tuse internați într-o secție de pneumologie, unde pneumonia are o prevalență de 50%. Atunci $p(B^+/S^+) = \frac{0{,}80 \times 0{,}50}{0{,}80 \times 0{,}50 + 0{,}20 \times 0{,}50} = \frac{0{,}40}{0{,}50} = 0{,}80$. Același semn clinic are o valoare predictivă de 41,4% în cabinet și de 80% în spital.

11.3.2 Frecvențele naturale

Formula lui Bayes este corectă, dar puțin intuitivă. Studiile despre raționamentul medicilor au arătat în mod repetat că mulți dintre ei supraestimează mult valoarea predictivă pozitivă când primesc problema sub formă de procente. Aceeași problemă devine mult mai ușoară dacă este formulată în frecvențe naturale: numere întregi de pacienți dintr-un grup imaginar [120].

Figura 11.1 arată metoda pentru un test cu sensibilitatea 80% și specificitatea 90%, aplicat într-o populație cu prevalența bolii de 10%. Luăm 1000 de pacienți. Prevalența de 10% înseamnă 100 de bolnavi și 900 de nebolnavi. Dintre cei 100 de bolnavi, testul este pozitiv la 80 (sensibilitatea 80%) și negativ la 20. Dintre cei 900 de nebolnavi, specificitatea de 90% înseamnă 810 teste negative și 90 de teste pozitive false. În total avem 80 + 90 = 170 de teste pozitive, dintre care doar 80 aparțin bolnavilor.

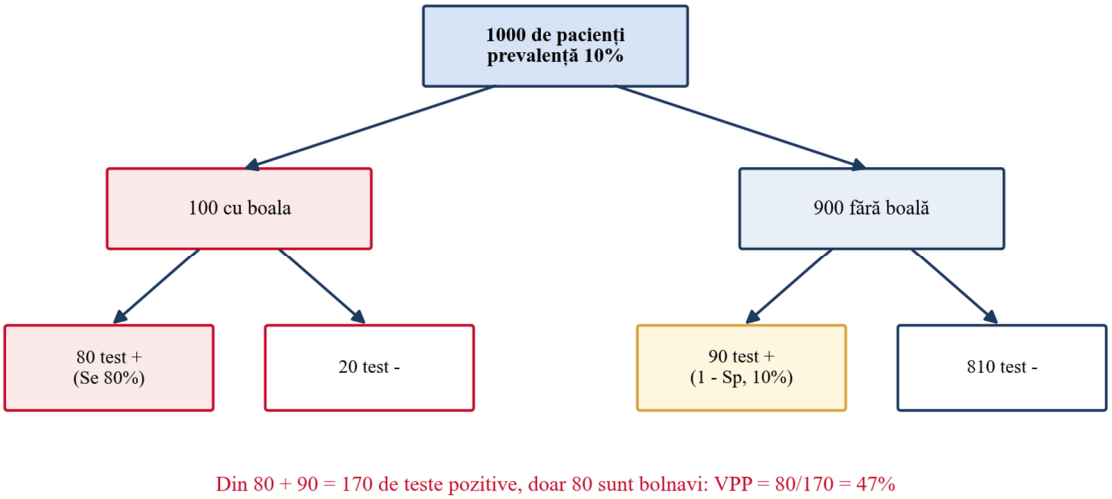


Figura 11.1. Arborele frecvențelor naturale pentru un test cu sensibilitatea 80% și specificitatea 90%, aplicat la 1000 de pacienți cu prevalența bolii de 10%. Din 170 de teste pozitive, doar 80 provin de la bolnavi, deci valoarea predictivă pozitivă este 80/170 = 47%. Exemplu didactic.

Valoarea predictivă pozitivă este 80/170 = 47,1%. Valoarea predictivă negativă este 810/(810 + 20) = 810/830 = 97,6%. Un test care "greșește" la doar 10% dintre sănătoși produce, într-o populație cu boală rară, mai multe rezultate pozitive false decât adevărate. Explicația este pur aritmetică: 10% din 900 (90) este mai mult decât 80% din 100 (80).

La laborator recomand mereu același lucru: când primiți o problemă cu procente, rescrieți-o pentru 1000 sau 10.000 de pacienți și desenați arborele. Durează două minute și elimină aproape toate greșelile de raționament.

11.3.3 Probabilitatea pre-test: prevalența și contextul clinic

Probabilitatea pre-test nu este o proprietate a bolii, ci a pacientului, în contextul în care este văzut. Același diagnostic are probabilități pre-test foarte diferite în medicina de familie, în ambulatoriul de specialitate și în spital. La medicul de familie ajung pacienți nefiltrați, cu simptome vagi și boli în stadii incipiente; prevalența bolilor grave este mică. În spital ajung pacienți deja selectați de alți medici, deci prevalența este mult mai mare. Acesta este motivul pentru care un test cu performanță excelentă într-o clinică universitară poate genera numeroase rezultate fals pozitive când este mutat în asistența primară.

Probabilitatea pre-test se estimează din mai multe surse: prevalența bolii în populația din care provine pacientul (date epidemiologice, registre, studii de prevalență), datele clinice deja cunoscute (vârstă, sex, factori de risc, simptome, examen obiectiv), scoruri clinice validate și, inevitabil, experiența medicului. În setul mg2026, prevalența diabetului confirmat prin TTGO este 18,1% (IC 95% Wilson 12,9-24,8), iar aceasta este probabilitatea pre-test rezonabilă pentru un adult cu risc cardiometabolic dintr-un cabinet asemănător, înainte de orice altă informație. Pentru un adult oarecare, fără factori de risc, probabilitatea ar fi mult mai mică; pentru un pacient obez, de peste 60 de ani, cu antecedente familiale, ar fi mai mare. Analiza transversală din setul mg2026 arată, de exemplu, o prevalență de 33,3% la obezi față de 10,4% la neobezi.

> **Studiu de caz: screeningul cancerului colorectal și raționamentul bayesian [101]**
>
> Iovanescu și colaboratorii (cu M. Frandeș din disciplina noastră) au evaluat fiabilitatea combinării testului imunochimic fecal (FIT) cu sigmoidoscopia flexibilă în screeningul neoplaziei colorectale. Au fost incluși 2201 pacienți simptomatici consecutivi, trimiși pentru FIT; un eșantion de 252 de persoane a efectuat ulterior colonoscopie (standardul de referință),

> iar pe un subeșantion de 57 de pacienți s-a efectuat sigmoidoscopie reală. Colonoscopia a găsit neoplazie avansată la 75 de persoane (29,76%), iar FIT a fost pozitiv în 30,3% dintre cazurile de neoplazie avansată. Rata de complianță la trimiterea pentru colonoscopie a fost de 52% printre cei cu FIT pozitiv, iar comparația pereche a celor două metode s-a făcut cu testul McNemar ($P < 0,001$). Autorii au calculat probabilitățile a priori și a posteriori și au arătat că, la niveluri mici de prevalență, valorile predictive negative se situau între 81,5% și 90,12%. Ce ne învață acest studiu? În primul rând, că prevalența de aproape 30% din eșantionul colonoscopiat nu este prevalența din populația de screening: pacienții erau simptomatici și selectați. În al doilea rând, că valorile predictive trebuie recalculate pentru prevalența populației în care testul va fi folosit, exact cu teorema lui Bayes. În al treilea rând, că testarea în serie (FIT, apoi sigmoidoscopie) și în paralel schimbă compromisul dintre sensibilitate și specificitate, subiect pe care îl reluăm în secțiunea 11.5. În sfârșit, faptul că doar jumătate dintre pacienții FIT pozitivi au acceptat colonoscopia ridică problema biasului de verificare, discutat în capitolul 12.

11.4 Șanse, rapoarte de verosimilitate și forma Bayes cu șanse

11.4.1 Conversia probabilitate-șansă

Șansa (engl. odds) unui eveniment este raportul dintre probabilitatea ca evenimentul să se producă și probabilitatea ca el să nu se producă:

$$\text{șansa} = \frac{p}{1-p}$$

Invers, din șansă se obține probabilitatea:

$$p = \frac{\text{șansa}}{1 + \text{ șansa}}$$

O probabilitate de 0,20 corespunde unei șanse de 0,20/0,80 = 0,25, adică "1 la 4". O probabilitate de 0,50 corespunde șansei 1 ("unu la unu"), iar o probabilitate de 0,90 corespunde șansei 9. Pentru probabilități mici, șansa și probabilitatea sunt aproape egale (p = 0,02 dă șansa 0,0204), dar pe măsură ce probabilitatea crește, șansa crește mult mai repede și nu are limită superioară. Ați întâlnit șansele în capitolul despre analiza riscului, la raportul șanselor (OR); aici le folosim pentru un singur pacient.

Tabelul 11.3. Corespondența dintre probabilitate și șansă.

Probabilitate	Șansă	Citire
0,05	0,053	aproximativ 1 la 19
0,10	0,111	1 la 9
0,181	0,221	aproximativ 1 la 4,5
0,25	0,333	1 la 3
0,50	1	1 la 1
0,75	3	3 la 1
0,90	9	9 la 1

11.4.2 Rapoartele de verosimilitate

Raportul de verosimilitate (RV, engl. likelihood ratio) al unui rezultat de test spune de câte ori este mai probabil acel rezultat la un bolnav decât la un nebolnav [121]. Pentru un test dihotomic avem două rapoarte:

$$\text{RV}^+ = \frac{p(T^+/B^+)}{p(T^+/B^-)} = \frac{\text{Se}}{1-\text{Sp}}$$

$$\text{RV}^- = \frac{p(T^-/B^+)}{p(T^-/B^-)} = \frac{1 - \text{Se}}{\text{Sp}}$$

Un RV egal cu 1 înseamnă că rezultatul este la fel de frecvent la bolnavi și la nebolnavi, deci nu aduce nicio informație. Un RV+ mare (peste 10) face boala mult mai probabilă; un RV- mic (sub 0,1) o face mult mai puțin probabilă. Între 2 și 5 pentru RV+, respectiv între 0,2 și 0,5 pentru RV-, testul mută probabilitatea moderat, uneori suficient, alteori nu. Aceste repere sunt orientative și le folosesc mulți autori [120,121].

Avantajul major al rapoartelor de verosimilitate față de valorile predictive este că, la fel ca sensibilitatea și specificitatea din care provin, nu depind (în principiu) de prevalență. Le putem deci transporta dintr-un studiu publicat în cabinetul nostru și le putem aplica la probabilitatea pre-test a pacientului nostru. Rezerva "în principiu" contează: dacă spectrul de pacienți diferă mult (stadii mai ușoare ale bolii, comorbidități), se schimbă și sensibilitatea, și specificitatea, deci și RV. Vom vedea în studiul de caz din secțiunea 11.8 un exemplu din cercetarea colegilor în care specificitatea unui biomarker se schimbă mult în funcție de momentul recoltării.

Pentru scorul FINDRISC în setul mg2026, la pragul ≥ 12, sensibilitatea este 75,9% (22 din 29), iar specificitatea 74,8% (98 din 131). Atunci:

$$\text{RV}^+ = \frac{0{,}759}{1 - 0{,}748} = \frac{0{,}759}{0{,}252} = 3{,}01$$

$$\text{RV}^- = \frac{1 - 0{,}759}{0{,}748} = \frac{0{,}241}{0{,}748} = 0{,}32$$

Un scor de cel puțin 12 este de 3 ori mai frecvent la persoanele cu diabet decât la cele fără diabet; un scor sub 12 este de aproximativ 3 ori mai rar la persoanele cu diabet.

11.4.3 Teorema lui Bayes în forma șanselor

Dacă scriem teorema lui Bayes o dată pentru B+ și o dată pentru B-, cu același rezultat T al testului, și împărțim cele două relații, numitorul p(T) se simplifică și rămâne o formulă remarcabil de simplă:

$$\text{șansa post-test} = \text{șansa pre-test} \times \text{RV}$$

Algoritmul de calcul are trei pași: (1) transformi probabilitatea pre-test în șansă; (2) înmulțești cu RV-ul rezultatului obținut; (3) transformi șansa post-test înapoi în probabilitate.

Pentru o persoană din setul mg2026 cu FINDRISC ≥ 12, probabilitatea pre-test este 0,181. Șansa pre-test este 0,181/0,819 = 0,221. Șansa post-test este 0,221 x 3,01 = 0,667. Probabilitatea post-test este 0,667/1,667 = 0,40, adică 40,0%. Este exact valoarea predictivă pozitivă calculată direct din tabel, 22/55 = 40,0%, cum era de așteptat. Pentru o persoană cu FINDRISC sub 12: 0,221 x 0,32 = 0,071, deci probabilitatea post-test este 0,071/1,071 = 6,7%.

Figura 11.2 arată relația dintre probabilitatea pre-test și cea post-test pentru câteva valori ale RV. Linia diagonală (RV = 1) este testul inutil. Curbele cu RV mare se ridică repede de la diagonală mai ales în zona probabilităților pre-test mici și medii; curbele cu RV mic coboară mult sub diagonală mai ales la probabilități pre-test mari.

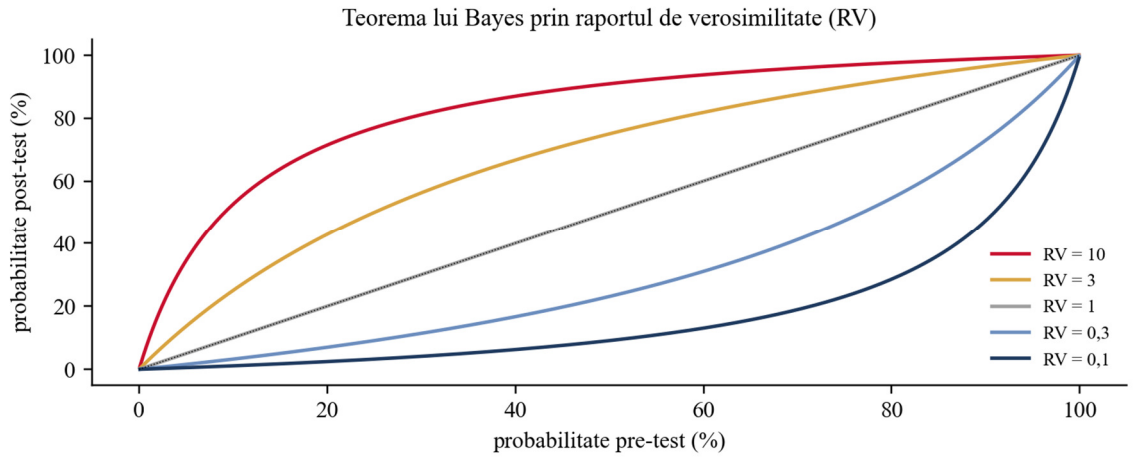


Figura 11.2. Probabilitatea post-test în funcție de probabilitatea pre-test, pentru rapoarte de verosimilitate între 0,1 și 10. Același test mută probabilitatea mult mai mult în zona de mijloc decât la extreme. Exemplu didactic.

Figura arată și ceva important pentru practica de zi cu zi: un test aduce cel mai mult atunci când probabilitatea pre-test este intermediară. Dacă pacientul are o probabilitate pre-test de 2%, nici un RV+ de 10 nu duce probabilitatea mult peste 17%; dacă are 95%, un test negativ cu RV- de 0,1 lasă încă o probabilitate de aproximativ 66%. De aceea, testarea fără o întrebare clinică clară și fără o estimare a probabilității pre-test produce adesea mai multă confuzie decât informație.

11.4.4 Nomograma Fagan

Înainte ca fiecare medic să aibă un calculator în buzunar, Fagan a propus o nomogramă care face calculul grafic [122]. Are trei axe verticale: în stânga probabilitatea pre-test, la mijloc raportul de verosimilitate (pe scară logaritmică), în dreapta probabilitatea post-test. Se trage o linie dreaptă din punctul pre-test, prin valoarea RV, și se citește probabilitatea post-test unde linia întâlnește axa din dreapta. Construcția funcționează pentru că logaritmul șansei post-test este suma dintre logaritmul șansei pre-test și logaritmul RV, iar o sumă de logaritmi se poate reprezenta cu o riglă dreaptă.

Figura 11.3 arată nomograma cu exemplul FINDRISC: din 18,1%, linia care trece prin RV+ = 3,01 ajunge la aproximativ 40%, iar cea care trece prin RV- = 0,32 ajunge la aproximativ 6,7%.

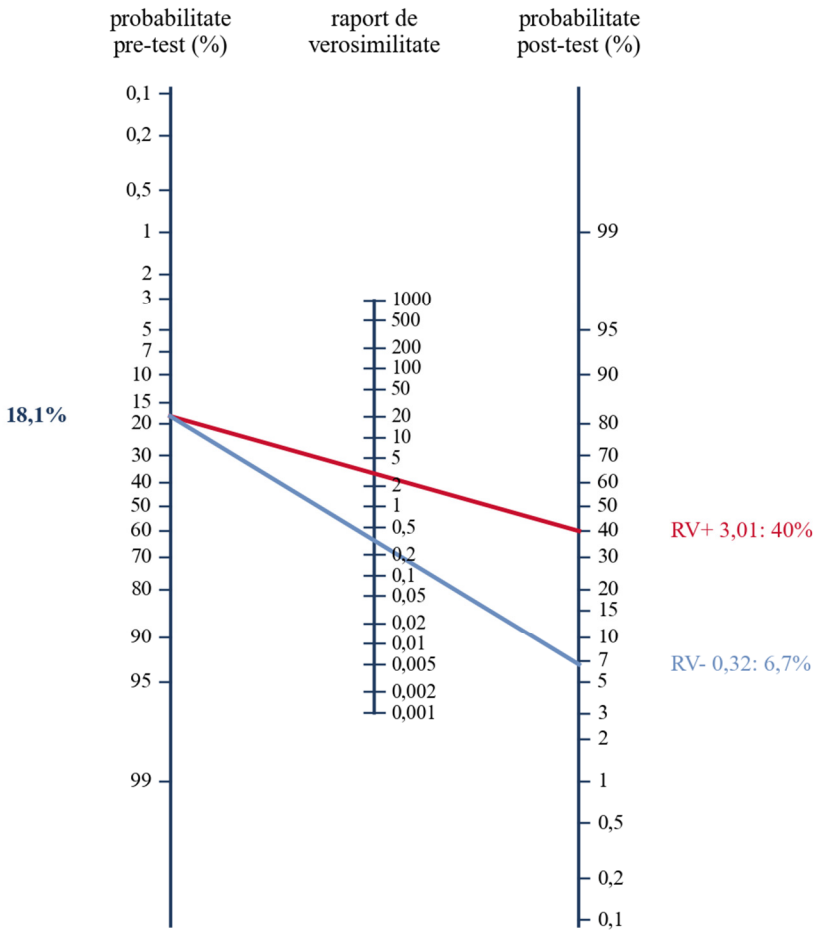


Figura 11.3. Nomograma Fagan aplicată scorului FINDRISC la pragul ≥ 12 în setul mg2026: pornind de la probabilitatea pre-test de 18,1%, un rezultat pozitiv (RV+ 3,01) duce la 40%, iar un rezultat negativ (RV- 0,32) la 6,7%. Date sintetice (setul mg2026).

Nomograma are azi mai ales valoare didactică, dar este utilă și la patul bolnavului pentru o estimare rapidă. Precizia citirii este de câteva procente, suficientă pentru decizie. Pentru calcule exacte vom construi în lucrarea practică 11 un calculator bayesian în Google Sheets.

Studiu de caz: două praguri pentru aceeași elastografie, unul pentru excludere și unul pentru confirmare [3]

Moga și colaboratorii (cu R. Lupușoru din disciplină) au evaluat elastografia time-harmonic pentru fibroza hepatică, cu elastografia tranzitorie controlată prin vibrație (VCTE) ca metodă de referință, la 165 de pacienți. În eșantion, 28,1% aveau ciroză (F4) după VCTE. Autorii nu au raportat un singur prag, ci două. Pragul de excludere a cirozei (engl. rule-out), sub 1,61 m/s, a avut Se 90,7%, Sp 66,6%, VPP 55,7% și VPN 93,9%. Pragul de confirmare (engl. rule-in), peste 1,83 m/s, a avut Se 65,1%, Sp 96,7%, VPP 90,3% și VPN 85,7%. AUROC a fost 0,90 (IC 95% 0,82-0,93). Să traducem în rapoarte de verosimilitate. Pentru pragul de excludere, $RV^- = \frac{1-0,907}{0,666} = 0,14$, adică un rezultat sub prag scade mult șansa de ciroză. Pentru pragul de confirmare, $RV^+ = \frac{0,651}{1-0,967} = 19,7$, adică un rezultat peste prag crește șansa de aproape 20 de ori. Dacă aplicăm aceste RV la prevalența din eșantion (28,1%), obținem prin forma cu șanse VPP de aproximativ 88,5% la pragul de confirmare și VPN de aproximativ 94,8% la pragul de excludere, valori apropiate de cele raportate (diferențele mici provin din rotunjiri și din numărul exact de pacienți analizați la fiecare prag). La o prevalență

> de 10%, cum ar fi într-o populație de screening, aceleași praguri ar da o VPP de aproximativ 68,7% pentru pragul de confirmare. Între cele două praguri rămâne o zonă gri în care testul nu decide, iar pacientul are nevoie de altă investigație. Studiul a raportat și reproductibilitatea (ICC între 0,92 și 0,97), o informație adesea uitată, dar decisivă pentru un test folosit de examinatori diferiți.

11.5 Testarea secvențială și combinarea testelor

11.5.1 Teste în serie și în paralel

Rareori un singur test decide un diagnostic. De obicei combinăm teste, iar modul de combinare schimbă performanța globală. Există două strategii de bază.

În testarea în paralel, se fac ambele teste, iar pacientul este considerat pozitiv dacă cel puțin unul dintre teste este pozitiv (regula "SAU"). Strategia crește sensibilitatea (pierdem mai puțini bolnavi), dar scade specificitatea (crește numărul falselor pozitive). Se folosește când a rata un bolnav este grav, de exemplu la camera de gardă, când pacientul nu va mai reveni.

În testarea în serie, al doilea test se face doar la cei pozitivi la primul, iar pacientul este considerat pozitiv doar dacă ambele teste sunt pozitive (regula "ȘI"). Strategia crește specificitatea și VPP, dar scade sensibilitatea. Este economică, pentru că testul al doilea (de obicei mai scump sau mai invaziv) se aplică doar unei părți dintre pacienți. Screeningul cancerului colorectal (FIT, apoi colonoscopie) și diagnosticul infecției HIV (test de screening, apoi test de confirmare) sunt exemple clasice.

Dacă cele două teste sunt independente condiționat de starea de boală (vom explica imediat ce înseamnă), performanța combinată se calculează simplu:

în serie: Se = Se1 x Se2; Sp = 1 - (1 - Sp1) x (1 - Sp2)

în paralel: Se = 1 - (1 - Se1) x (1 - Se2); Sp = Sp1 x Sp2

Iar în forma cu șanse, testarea secvențială înseamnă înmulțirea succesivă cu rapoartele de verosimilitate:

șansa post-test finală = șansa pre-test x RV1 x RV2

Probabilitatea post-test după primul test devine probabilitatea pre-test pentru al doilea. Este cea mai elegantă proprietate a raționamentului bayesian: informația se acumulează.

11.5.2 Ipoteza independenței condiționate

Două teste sunt independente condiționat dacă, printre bolnavi, rezultatul unui test nu ne spune nimic despre rezultatul celuilalt, și la fel printre nebolnavi. Ipoteza este adesea încălcată, mai ales când testele măsoară aspecte apropiate ale aceleiași fiziopatologii. Scorul FINDRISC și glicemia a jeun, de exemplu, depind amândouă de obezitate, vârstă și sedentarism. O persoană fără diabet, dar cu obezitate abdominală, are șanse mai mari să fie "pozitivă" la ambele.

Setul mg2026 ne permite să verificăm. Considerăm FINDRISC ≥ 12 (Se 75,9%, Sp 74,8%) și glicemia a jeun ≥ 110 mg/dL (Se 82,8%, Sp 70,2%; 24 din 29 de bolnavi pozitivi, 92 din 131 de nebolnavi negativi). Dacă testele ar fi independente condiționat, combinația în serie ar avea:

Se = 0,759 x 0,828 = 0,628; Sp = 1 - 0,252 x 0,298 = 0,925

Datele reale (pacienții pozitivi la ambele teste) arată însă Se 69,0% (20 din 29) și Sp 85,5% (112 din 131). Specificitatea reală este mai mică decât cea prezisă, pentru că printre cei 131 de nebolnavi, 19 sunt pozitivi la ambele teste, mult mai mulți decât ar fi de așteptat dacă testele ar fi independente $\left(\frac{33\times 39}{131} = 9{,}8\right)$. Consecința practică se vede în probabilitatea post-test. Înmulțind succesiv rapoartele de verosimilitate (RV+ 3,01 pentru FINDRISC, RV+ 2,78 pentru glicemie),

am estima pentru un pacient pozitiv la ambele teste o probabilitate post-test de 64,9%. Valoarea observată în date este doar 20/39 = 51,3%. Ipoteza independenței a supraestimat câștigul adus de al doilea test.

Tabelul 11.4. Combinarea FINDRISC ≥ 12 cu glicemia a jeun ≥ 110 mg/dL pentru diabetul confirmat prin TTGO (date sintetice, setul mg2026, n = 160, 29 cu diabet).

Strategie	AP	FP	FN	AN	Se (%)	Sp (%)	VPP (%)	VPN (%)
FINDRISC ≥ 12	22	33	7	98	75,9	74,8	40,0	93,3
Glicemie ≥ 110	24	39	5	92	82,8	70,2	38,1	94,8
În serie (ambele +)	20	19	9	112	69,0	85,5	51,3	92,6
În paralel (cel puțin unul +)	26	53	3	78	89,7	59,5	32,9	96,3

Tabelul confirmă regula generală: seria câștigă specificitate și VPP, paralelul câștigă sensibilitate și VPN. Lecția metodologică este că înmulțirea rapoartelor de verosimilitate este corectă doar pentru teste independente condiționat. Pentru teste corelate, performanța combinației trebuie măsurată direct, pe pacienți la care s-au făcut ambele teste, sau estimată cu un model multivariabil, de exemplu regresia logistică din capitolul 9.

Studiu de caz: reguli clinice care combină mai multe măsurători [110,123]

Două studii ale grupului de hepatologie din Timișoara, cu R. Lupușoru din disciplină, arată cum se combină mai multe măsurători într-o decizie. Barbulescu și colaboratorii au extins și validat criteriile Baveno VI pentru excluderea varicelor esofagiene cu risc înalt, la 839 de pacienți (cohorta de derivare 728, cohorta de validare 111). Criteriile Baveno VI au avut o acuratețe de 86,2% în prima cohortă. Autorii au derivat o regulă de confirmare (trombocite sub 150.000/mm³ și rigiditate hepatică peste 35,3 kPa) și o regulă de excludere (trombocite peste 150.000/mm³ și rigiditate hepatică sub 19,6 kPa), iar pentru zona gri au propus un prag suplimentar al albuminei, sub 3,4 g/dl. În cohorta de validare, pragurile de excludere au avut acuratețe de 100%. Regula "ȘI" din fiecare criteriu este o combinare în serie; lăsarea unei zone gri recunoaște că, între cele două praguri, probabilitatea post-test rămâne prea aproape de mijloc pentru a decide fără endoscopie. Fofiu și colaboratorii au abordat aceeași problemă altfel: la 124 de pacienți analizați cu ciroză compensată, au combinat rigiditatea splinei (prag 13,2 kPa, AUROC 0,84), rigiditatea ficatului (12,1 kPa, AUROC 0,86) și dimensiunea splinei (12,9 cm, AUROC 0,71) într-un scor obținut prin regresie: 0,053 x SS + 0,054 x LS + 0,059 x SSZ - 1,84, cu pragul > 0,34 și AUROC 0,93, superior fiecărui factor luat separat. Modelul de regresie estimează ponderile ținând cont de corelația dintre măsurători, adică exact problema independenței condiționate. Ambele studii trebuie citite cu o rezervă: acuratețea de 100% într-o cohortă de validare de 111 pacienți are un interval de încredere larg, iar un scor derivat pe 124 de pacienți are nevoie de validare externă înainte de folosirea de rutină.

11.6 Diagnosticul diferențial cu mai multe boli

11.6.1 Tabelul Bayes cu trei diagnostice

În practică, medicul nu se întreabă doar "are sau nu are boala X?", ci "care dintre bolile X, Y, Z explică cel mai bine ce văd?". Teorema lui Bayes se extinde direct la mai multe ipoteze, cu condiția ca acestea să fie mutual exclusive (pacientul are una singură) și exhaustive (una dintre ele este adevărată). Pentru fiecare boală Di:

$$p(D_i/S) = \frac{p(S/D_i) \times p(D_i)}{\sum[p(S/D_j) \times p(D_j)]}$$

Numărătorul se calculează pentru fiecare boală, iar numitorul este suma numărătorilor. Practic, probabilitatea posterioară a fiecărei boli este ponderea numărătorului ei în total.

Să luăm un exemplu de medicina de familie, cu cifre fictive, dar plauzibile. Un bărbat de 68 de ani, fost fumător, vine pentru dispnee de efort progresivă în ultimele luni. Pentru simplitate, presupunem că trei diagnostice acoperă aproape toate cazurile de acest tip din cabinetul respectiv: insuficiența cardiacă (IC), bronhopneumopatia obstructivă cronică (BPOC) și anemia. Din experiența cabinetului, la pacienții cu acest profil, probabilitățile a priori sunt 0,35 pentru IC, 0,45 pentru BPOC și 0,20 pentru anemie. Medicul notează trei semne: edeme gambiere bilaterale (E), wheezing la auscultație (W) și paloare tegumentară (P). Frecvența fiecărui semn în fiecare boală este dată în tabelul 11.5.

Tabelul 11.5. Probabilitățile a priori și frecvența semnelor clinice în fiecare diagnostic (date fictive).

Diagnostic	**p(D)**	$p(E/D)$	$p(W/D)$	$p(P/D)$
Insuficiență cardiacă	0,35	0,60	0,20	0,15
BPOC	0,45	0,15	0,60	0,10
Anemie	0,20	0,10	0,05	0,70

Pacientul are edeme (E+), nu are wheezing (W-) și nu este palid (P-). Presupunem că, în interiorul fiecărei boli, semnele apar independent unele de altele (ipoteza independenței condiționate, numită în informatică modelul "Bayes naiv"). Verosimilitatea combinației observate în fiecare boală este produsul probabilităților: pentru un semn prezent folosim p(S / D), pentru un semn absent folosim 1 - p(S / D).

Pentru IC: 0,60 x (1 - 0,20) x (1 - 0,15) = 0,60 x 0,80 x 0,85 = 0,408

Pentru BPOC: 0,15 x (1 - 0,60) x (1 - 0,10) = 0,15 x 0,40 x 0,90 = 0,054

Pentru anemie: 0,10 x (1 - 0,05) x (1 - 0,70) = 0,10 x 0,95 x 0,30 = 0,0285

Înmulțim fiecare verosimilitate cu probabilitatea a priori:

IC: 0,35 x 0,408 = 0,1428

BPOC: 0,45 x 0,054 = 0,0243

Anemie: 0,20 x 0,0285 = 0,0057

Suma este 0,1728. Probabilitățile posterioare sunt:

$$p(\text{IC/date}) = \frac{0{,}1428}{0{,}1728} = 0{,}826$$

$$p(\text{BPOC/date}) = \frac{0{,}0243}{0{,}1728} = 0{,}141$$

$$p(\text{ anemie /date}) = \frac{0{,}0057}{0{,}1728} = 0{,}033$$

Tabelul 11.6. Calculul probabilităților posterioare pentru pacientul cu dispnee (E+, W-, P-).

Diagnostic	**A priori**	**Verosimilitate**	**Produs**	**A posteriori**
Insuficiență cardiacă	0,35	0,408	0,1428	82,6%
BPOC	0,45	0,054	0,0243	14,1%
Anemie	0,20	0,0285	0,0057	3,3%
Total	1,00		0,1728	100%

Ierarhia diagnosticelor s-a inversat. Înainte de examen, BPOC era cel mai probabil diagnostic (45%); după examen, insuficiența cardiacă ajunge la 82,6%. Rezultatul nu înseamnă că diagnosticul este stabilit: o probabilitate de 82,6% justifică un test țintit (de exemplu NT-proBNP și ecocardiografie), nu un tratament definitiv fără confirmare. Calculul se poate face și pas cu pas, semn cu semn, folosind posteriorul de la pasul anterior ca prior pentru pasul următor. După edeme singure, probabilitățile ar fi 70,6% / 22,7% / 6,7%; după adăugarea absenței wheezingului, 78,5% / 12,6% / 8,9%; după absența palorii, 82,6% / 14,1% / 3,3%. Rezultatul final este același, indiferent de ordinea în care adăugăm semnele, cu condiția ca ipoteza de independență să fie valabilă.

11.6.2 Limitele modelului "naiv"

Modelul din exemplul de mai sus are trei slăbiciuni pe care trebuie să le aveți în minte. Prima este exhaustivitatea: dacă pacientul are, de fapt, o a patra boală (de exemplu fibroză pulmonară sau o valvulopatie), nicio cifră din tabel nu o va semnala; probabilitățile vor fi redistribuite între cele trei diagnostice listate. A doua este independența semnelor: edemele și ortopneea apar împreună în insuficiența cardiacă mult mai des decât ar sugera produsul frecvențelor lor, iar înmulțirea lor ca și cum ar fi independente exagerează dovezile. A treia este calitatea numerelor de intrare: probabilitățile a priori și frecvențele semnelor sunt estimări, adesea din studii pe populații diferite.

Cu toate acestea, modelele de tip Bayes naiv au fost printre primele sisteme de sprijin al deciziei clinice și rămân surprinzător de robuste pentru ordonarea diagnosticelor, chiar dacă probabilitățile absolute nu sunt exacte [124]. Cursul disciplinei le prezintă, alături de recunoașterea formelor (engl. pattern recognition), ca metode statistice de clasificare. Ideea de bază, actualizarea probabilităților prin înmulțirea priorului cu verosimilitatea, stă și la baza multor clasificatori moderni, inclusiv a unora folosiți în inteligența artificială medicală.

11.7 Pragurile de decizie: a testa, a trata

Probabilitatea post-test nu este un scop în sine. Ea servește unei decizii: tratăm, investigăm mai departe sau nu facem nimic? Pentru a lega probabilitatea de decizie avem nevoie de pragul de tratament, adică de probabilitatea peste care beneficiul așteptat al tratamentului depășește riscul lui.

Să notăm cu B beneficiul net al tratării unui bolnav (de exemplu complicațiile evitate) și cu R riscul net al tratării unui nebolnav (efecte adverse, costuri, anxietate). Pragul de tratament este

$$p_{\text{tratament}} = \frac{R}{R + B}$$

Dacă tratarea inutilă a unui pacient sănătos "costă" o unitate, iar netratarea unui bolnav "costă" patru unități (B = 4), pragul este 1 / (1 + 4) = 0,20. Peste 20%, merită tratat; sub 20%, nu. Pentru un tratament foarte sigur și foarte eficient (de exemplu un antibiotic oral bine tolerat într-o infecție cu complicații grave) pragul este jos. Pentru un tratament riscant (chimioterapie, anticoagulare la un vârstnic cu risc de sângerare) pragul este sus.

Odată ce avem un test, apar două praguri suplimentare. Pragul de testare (inferior) este probabilitatea pre-test sub care nici un rezultat pozitiv nu ar aduce pacientul peste pragul de tratament; sub el, testul este inutil și pacientul nu se tratează. Pragul de testare-tratament (superior) este probabilitatea pre-test peste care nici un rezultat negativ nu ar coborî pacientul sub pragul de tratament; peste el, se tratează fără test. Între cele două praguri, testul poate schimba decizia și merită făcut [120].

Cu forma în șanse, calculul este direct. Fie pragul de tratament 20% (șansa 0,25) și un test cu RV+ = 6 și RV- = 0,2.

Pragul de testare: șansa pre-test = 0,25 / 6 = 0,0417, deci probabilitatea 0,0417 / 1,0417 = 4,0%.

Pragul superior: șansa pre-test = 0,25 / 0,2 = 1,25, deci probabilitatea 1,25 / 2,25 = 55,6%.

Pacienții cu probabilitate pre-test între 4,0% și 55,6% beneficiază de test. Sub 4,0% nu se testează și nu se tratează; peste 55,6% se tratează direct. Modelul este simplificat (ignoră riscul și costul testului însuși), dar ilustrează ideea principală: decizia de a cere un test depinde de probabilitatea pre-test, de performanța testului și de consecințele erorilor. Un test mai bun (RV+ mai mare, RV- mai mic) lărgește intervalul în care testarea este utilă.

11.8 Screeningul în medicina de familie: exemplul FINDRISC

Scorul FINDRISC (engl. Finnish Diabetes Risk Score) a fost construit pentru a identifica, printr-un chestionar simplu, persoanele cu risc mare de diabet de tip 2 [125]. Ia valori între 0 și 26 și include vârsta, IMC, circumferința taliei, activitatea fizică, consumul zilnic de legume și fructe, tratamentul antihipertensiv, glicemiile crescute în antecedente și diabetul la rudele de gradul I sau II. Este exact tipul de instrument pe care un medic de familie îl poate folosi la un pacient care vine pentru altceva.

În setul mg2026, diabetul a fost confirmat prin TTGO (standardul de referință, conform criteriilor ADA [80]), iar la pragul FINDRISC ≥ 12 avem RV+ 3,01 și RV- 0,32. Pentru a vedea ce înseamnă aceste cifre la pacienți diferiți, aplicăm forma cu șanse la mai multe probabilități pre-test.

Tabelul 11.7. Probabilitatea post-test de diabet după FINDRISC la pragul ≥ 12 (RV+ 3,01; RV- 0,32), pentru diferite probabilități pre-test. Rapoartele de verosimilitate provin din setul mg2026 (date sintetice); probabilitățile pre-test sunt scenarii.

Probabilitate pre-test	Context posibil	Șansa pre-test	Post-test după FINDRISC ≥ 12	Post-test după FINDRISC < 12
2%	adult tânăr, fără factori de risc	0,020	5,8%	0,7%
5%	populația adultă generală	0,053	13,7%	1,7%
10%	adult de vârstă medie, supraponderal	0,111	25,1%	3,5%
18,1%	eșantionul mg2026 (risc cardiometabolic)	0,221	40,0%	6,7%
30%	obez, prediabet cunoscut	0,429	56,3%	12,1%
50%	pacient cu simptome sugestive	1,000	75,1%	24,4%

Tabelul spune mai multe lucruri importante. Un FINDRISC pozitiv nu "pune diagnosticul": chiar și în eșantionul cu risc crescut, doar 4 din 10 pacienți pozitivi au diabet. Un FINDRISC negativ este mai util pentru a liniști decât unul pozitiv pentru a alarma, cel puțin la probabilități pre-test mici și moderate: sub 10% pre-test, un scor sub 12 coboară probabilitatea sub 3,5%. La probabilități pre-test mari (50%), un scor negativ lasă totuși o probabilitate de 24,4%, mult prea mare pentru a renunța la testul de laborator. Rolul FINDRISC în cabinet este, deci, de triaj: decide cine merită glicemie a jeun sau HbA1c, nu cine are diabet. Coloana post-test pentru scorul pozitiv este identică, de altfel, cu tabelul VPP în funcție de prevalență pe care îl veți reîntâlni în capitolul 12: VPP la prevalența p este exact probabilitatea post-test pentru un pacient cu probabilitatea pre-test p.

Când pacientul are deja un rezultat FINDRISC și i se face și glicemia, raționamentul continuă: probabilitatea post-test după FINDRISC devine probabilitate pre-test pentru glicemie. Am văzut în secțiunea 11.5 că, în acest caz particular, ipoteza independenței nu se respectă, deci înmulțirea simplă a rapoartelor de verosimilitate supraestimează probabilitatea finală. În practică, ghidurile cer oricum confirmarea diagnosticului de diabet prin teste de laborator, nu prin calcul probabilistic [80].

Studiu de caz: valorile predictive depind de prevalență și de spectrul pacienților [75,126,127]

Trei studii ale colegilor arată aceeași lecție din unghiuri diferite. Albai și colaboratorii (cu M. Frandeș) au validat scorul ISARIC-4C la 159 de pacienți cu diabet de tip 2 internați pentru infecție SARS-CoV-2, cu o mortalitate de 25,15%. Aria sub curba ROC a fost 0,875 (IC 95% 0,820-0,930), iar la pragul ales scorul a clasificat corect 77,36% din cohortă, cu Se 87,80%, Sp 73,72%, VPP 53,73% și VPN 94,56%. Dacă recalculăm cu teorema lui Bayes VPP la prevalența de 25,15%, obținem aproximativ 52,9%, apropiat de valoarea raportată; aceeași VPP ar scădea mult într-o populație cu mortalitate de 10%. O observație de lectură critică: intervalele de încredere raportate pentru Se și Sp (de exemplu 87,66-87,93) sunt neobișnuit de înguste pentru 159 de pacienți, dintre care doar aproximativ 40 au decedat; un interval Wilson pentru o proporție de 88% calculată pe 40 de cazuri ar avea o lățime de peste 20 de puncte procentuale. Sipos și colaboratorii (cu R. Moleriu) au raportat pentru IL-6 în diagnosticul sepsisului neonatal un prag optim de 181 pg/ml, cu Se 80,1%, Sp 85,7%, VPP 84,6% și VPN 81,8%. Rezumatul nu dă numărul de nou-născuți, dar putem face un exercițiu bayesian invers: din Se și Sp, $\mathrm{RV}^{+} = \frac{0{,}801}{0{,}143} = 5{,}6$; șansa post-test corespunzătoare VPP de 84,6% este 0,846 / 0,154 = 5,49; împărțind, șansa pre-test este aproximativ 0,98, adică o prevalență a sepsisului de aproximativ 49,5% în eșantionul analizat. Printre toți nou-născuții testați într-o maternitate, prevalența sepsisului este mult mai mică, deci în practică VPP ar fi mai mică decât 84,6%. Pavel și colaboratorii (cu A. Tudor) au evaluat H-FABP la 83 de pacienți cu suspiciune de infarct miocardic acut, separat după timpul de la debutul durerii. La sub 4 ore, sensibilitatea și specificitatea au fost 89,67 și 95,65, ceea ce corespunde unui RV+ de aproximativ 20,6; peste 4 ore, sensibilitatea a fost 86,73, iar specificitatea doar 49,84, adică un RV+ de aproximativ 1,73, aproape fără valoare de confirmare. Același biomarker, în aceeași secție, are performanțe complet diferite în funcție de momentul în care sunt testați pacienții. Prin urmare, RV-urile nu sunt constante universale: se aplică pacienților asemănători celor din studiul care le-a estimat.

11.9 Inferența bayesiană față de inferența frecventistă

Până acum am aplicat teorema lui Bayes pacientului individual. Aceeași logică se poate aplica și parametrilor unei populații: o medie, o proporție, o diferență între tratamente. Aceasta este inferența bayesiană, o alternativă la testele de ipoteză și intervalele de încredere frecventiste din capitolele 4-10.

11.9.1 A priori, verosimilitate, a posteriori

În inferența frecventistă, parametrul (de exemplu prevalența diabetului în populație, π) este o constantă necunoscută, iar probabilitatea descrie doar variabilitatea eșantioanelor. În inferența bayesiană, incertitudinea despre parametru se exprimă direct printr-o distribuție de probabilitate. Înainte de date avem distribuția a priori (engl. prior), care rezumă ce știm sau credem despre parametru. Datele contribuie prin verosimilitate (cât de probabile sunt datele observate pentru fiecare valoare posibilă a parametrului). Combinația lor dă distribuția a posteriori (engl. posterior):

a posteriori ∝ verosimilitate x a priori

Semnul ∝ înseamnă "proporțional cu"; o constantă de normalizare face ca aria sub distribuția a posteriori să fie 1. Formula este teorema lui Bayes aplicată unei distribuții întregi, nu unei singure probabilități.

Pentru o proporție, calculul este deosebit de simplu dacă folosim ca a priori o distribuție Beta. O distribuție Beta(α; β) poate fi înțeleasă ca echivalentul a α - 1 "succese" și β - 1 "eșecuri" observate anterior. Dacă observăm k cazuri din n, distribuția a posteriori este Beta(α + k; β + n - k). Pentru diabetul din mg2026, k = 29 și n = 160.

Cu un a priori uniform, Beta(1; 1), care spune că orice prevalență între 0 și 100% este la fel de plauzibilă, a posteriori este Beta(30; 132). Media ei este 30/162 = 18,5%, iar intervalul care conține 95% din probabilitatea a posteriori, numit interval credibil (engl. credible interval), este 12,9-24,8%. Cu un a priori informativ, Beta(10; 40), care exprimă convingerea anterioară că prevalența este în jur de 20% (cu 95% din probabilitate între 10,2% și 32,0%), a posteriori este Beta(39; 171), cu media 18,6% și intervalul credibil 13,6-24,1%. Figura 11.4 arată cele două situații.

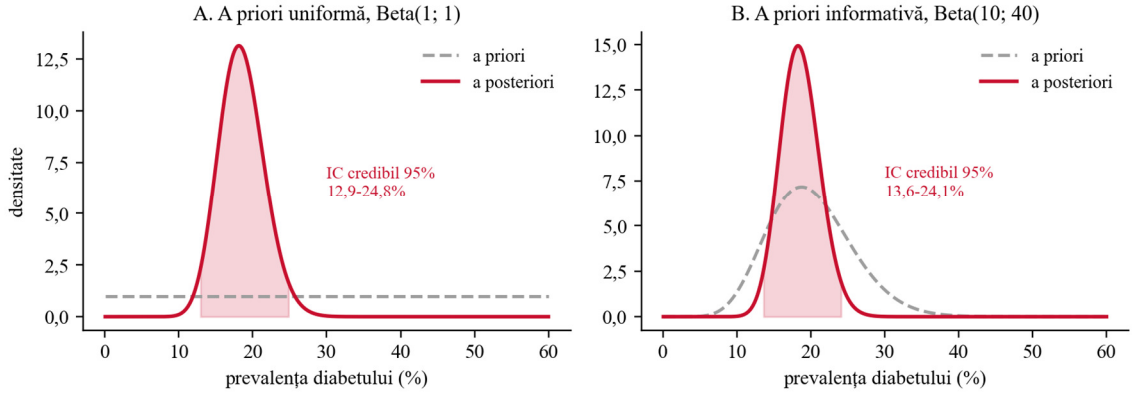


Figura 11.4. Distribuția a priori (linie întreruptă) și a posteriori (linie continuă) pentru prevalența diabetului confirmat prin TTGO, cu 29 de cazuri din 160. A: a priori uniformă, interval credibil 95% 12,9-24,8%. B: a priori informativă centrată pe 20%, interval credibil 13,6-24,1%. Date sintetice (setul mg2026).

Două observații. Cu 160 de pacienți, datele domină: cele două distribuții a posteriori sunt foarte asemănătoare, chiar dacă priorurile diferă mult. Cu 20 de pacienți, priorul ar fi contat mult mai mult. Apoi, intervalul credibil cu a priori uniform (12,9-24,8%) coincide practic cu intervalul de încredere Wilson calculat în capitolul 4 (12,9-24,8%). Pentru eșantioane mari și a priori puțin informative, metodele bayesiene și frecventiste dau rezultate numerice foarte apropiate; diferă însă interpretarea.

11.9.2 Intervalul credibil față de intervalul de încredere

Intervalul de încredere frecventist de 95% are o interpretare incomodă: dacă am repeta studiul de foarte multe ori și am calcula de fiecare dată intervalul, 95% dintre aceste intervale ar conține valoarea adevărată. Nu putem spune că "valoarea adevărată se află cu probabilitatea 95% în intervalul 12,9-24,8%", pentru că, în sens frecventist, parametrul nu este o variabilă aleatoare. Confuzia este una dintre cele mai frecvente interpretări greșite ale statisticii [70].

Intervalul credibil bayesian are exact interpretarea pe care o dorește intuitiv clinicianul: având în vedere priorul și datele, probabilitatea ca prevalența să se afle între 12,9% și 24,8% este 95%. Prețul plătit este alegerea priorului, care introduce un element de judecată ce trebuie justificat și raportat transparent. Un prior ales pentru a obține rezultatul dorit este la fel de neonest ca un test frecventist repetat până iese $p < 0,05$.

11.9.3 Factorul Bayes

Echivalentul bayesian al testului de ipoteză este factorul Bayes. El compară cât de bine prezic datele observate două ipoteze concurente, de obicei H1 (există un efect) și H0 (nu există efect):

$$\mathrm{BF}_{10} = \frac{p(\mathrm{date}/H_1)}{p(\mathrm{date}/H_0)}$$

Un BF10 de 10 înseamnă că datele sunt de 10 ori mai probabile sub H1 decât sub H0. Un BF10 de 0,1 înseamnă că datele sunt de 10 ori mai probabile sub H0; echivalent, $\mathrm{BF}_{01} = \frac{1}{\mathrm{BF}_{10}} = 10$. Spre deosebire de valoarea p, factorul Bayes poate aduce dovezi în favoarea ipotezei nule, nu doar împotriva ei. El este, de fapt, un raport de verosimilitate aplicat ipotezelor:

șansa a posteriori a lui H1 = șansa a priori a lui H1 x BF10

Recunoașteți formula de la diagnostic: ipotezele joacă rolul bolilor, datele joacă rolul rezultatului testului.

Programul JASP, folosit la lucrările practice, oferă pentru multe teste frecventiste o variantă bayesiană (în meniul fiecărui test, la secțiunea Bayesian) și folosește o scală de interpretare larg răspândită [47,49].

Tabelul 11.8. Scala de interpretare a factorului Bayes BF10 folosită frecvent (inclusiv în materialele JASP).

BF10	**Dovezi în favoarea H1**	**BF10**	**Dovezi în favoarea H0**
1-3	anecdotice (slabe)	1/3-1	anecdotice (slabe)
3-10	moderate	1/10-1/3	moderate
10-30	puternice	1/30-1/10	puternice
30-100	foarte puternice	1/100-1/30	foarte puternice
peste 100	extreme	sub 1/100	extreme

Pe setul mg2026, câteva analize din capitolele anterioare, refăcute bayesian (cu a priori implicite din JASP, de exemplu distribuția Cauchy cu lățimea 0,707 pentru testul t), dau următoarele rezultate:

- TAS la includere, femei față de bărbați (testul t pentru eșantioane independente): BF10 = 0,18, adică BF01 = 5,7. Datele sunt de aproximativ 5,7 ori mai probabile dacă nu există diferență între sexe: dovezi moderate pentru H0. Testul frecventist (p = 0,81) spunea doar că nu putem respinge H0; factorul Bayes spune, în plus, că datele susțin activ absența unei diferențe.
- Glicemia la 0 față de 6 luni (testul t pereche): BF10 de ordinul 10^18, dovezi extreme pentru o modificare.
- Corelația IMC-TAS (r = 0,405, n = 160): BF10 de aproximativ 116.000, dovezi extreme.
- Corelația HDL-trigliceride (r = 0,076): BF10 = 0,16, adică BF01 = 6,4, dovezi moderate pentru absența corelației.
- Obezitatea și diabetul (tabel de contingență 2 x 2, eșantionare multinomială comună, a priori implicit): BF10 de aproximativ 85, dovezi foarte puternice pentru asociere.

Analiza bayesiană nu rezolvă problemele de design (un studiu prost proiectat rămâne prost proiectat) și nu înlocuiește judecata clinică despre mărimea efectului. Are însă două avantaje didactice: răspunde direct la întrebarea "cât de mult susțin datele o ipoteză față de alta" și obligă la explicitarea informației anterioare. În literatura medicală, metodele bayesiene apar tot mai des în studiile clinice adaptative și în metaanalize, iar studentul de azi le va citi ca medic [42,63].

11.10 Medicina bazată pe dovezi

Medicina bazată pe dovezi (engl. evidence-based medicine) a fost definită ca utilizarea conștiincioasă, explicită și judicioasă a celor mai bune dovezi disponibile în deciziile despre îngrijirea pacienților individuali, prin integrarea expertizei clinice cu dovezile din cercetarea sistematică [15]. Definiția are trei componente și toate trei contează: dovezile din cercetare, experiența clinicianului și valorile și preferințele pacientului. Medicina bazată pe dovezi nu înseamnă ghiduri aplicate mecanic și nici statistică în locul clinicii.

11.10.1 Cei cinci pași

Practica medicinei bazate pe dovezi se descrie de obicei în cinci pași, ușor de memorat în engleză ca cei cinci "A".

Pasul 1, formularea întrebării (engl. ask). O întrebare clinică bine construită are structura PICO [24]: P (pacientul sau populația), I (intervenția, expunerea sau testul diagnostic), C (comparatorul) și O (rezultatul, engl. outcome). De exemplu: "La adulții cu risc cardiometabolic din medicina de familie (P), folosirea unei aplicații mobile de automonitorizare, adăugată consilierii (I), față de consilierea standard (C), crește proporția celor care pierd cel puțin 5% din greutate la 6 luni (O)?". Pentru o întrebare de diagnostic: "La adulții supraponderali neinvestigați (P), ce sensibilitate și specificitate are scorul FINDRISC $\geq$ 12 (I) pentru diabetul de tip 2 (O), comparat cu TTGO ca standard de referință (C)?".

Pasul 2, căutarea dovezilor (engl. acquire). Se caută în baze de date bibliografice precum PubMed, în colecții de recenzii sistematice și în ghiduri, folosind termenii din PICO. Căutarea eficientă a fost discutată în volumul de informatică medicală din semestrul I [1].

Pasul 3, evaluarea critică (engl. appraise). Studiul găsit este valid (designul, randomizarea, orbirea, pierderile din urmărire, standardul de referință)? Care sunt rezultatele (mărimea efectului, intervalul de încredere)? Se aplică pacientului meu? Pentru studiile diagnostice, evaluarea critică urmează lista STARD, pe care o discutăm în capitolul 12.

Pasul 4, aplicarea (engl. apply). Dovezile se integrează cu situația concretă a pacientului: comorbidități, preferințe, acces la tratament, costuri. Aici intervine raționamentul bayesian din acest capitol, pentru că probabilitatea pre-test a pacientului meu poate fi diferită de prevalența din studiu.

Pasul 5, evaluarea rezultatului (engl. assess). Medicul își evaluează propria practică: a dus decizia la rezultatul dorit? Ce se poate face mai bine data viitoare?

11.10.2 Ierarhia dovezilor și sistemul GRADE

Nu toate studiile oferă dovezi la fel de solide. Pentru întrebările despre eficacitatea tratamentelor, ierarhia clasică plasează la bază opinia experților și seriile de cazuri, urmate de studiile caz-control, de studiile de cohortă, de studiile clinice randomizate și, în vârf, de recenziile sistematice cu metaanaliză ale studiilor randomizate. Pentru întrebările de diagnostic, vârful îl ocupă studiile transversale pe pacienți consecutivi, la care toți primesc atât testul evaluat, cât și standardul de referință; studiile de tip caz-control (bolnavi cunoscuți față de voluntari sănătoși) supraestimează sistematic acuratețea. Pentru întrebările de prognostic, designul preferat este cohorta prospectivă.

Piramida este utilă ca primă orientare, dar simplifică. Un studiu randomizat mic, cu multe pierderi din urmărire, poate oferi dovezi mai slabe decât o cohortă mare și bine condusă. De aceea, abordarea actuală, GRADE (engl. Grading of Recommendations, Assessment, Development and Evaluation), evaluează încrederea în dovezi pentru fiecare rezultat în parte, nu pentru fiecare tip de studiu [128]. Dovezile din studiile randomizate pornesc de la încredere mare, cele din studiile observaționale de la încredere mică. Încrederea poate fi coborâtă pentru risc de

bias, inconsistență între studii, caracter indirect (populația sau rezultatul diferă de cele care ne interesează), imprecizie (intervale de încredere largi) și bias de publicare. Poate fi urcată, mai ales pentru studiile observaționale, când efectul este foarte mare sau există un gradient doză-răspuns. Rezultatul este una dintre patru categorii: încredere mare, moderată, mică sau foarte mică.

11.10.3 Recenzii sistematice și metaanalize

O recenzie sistematică răspunde unei întrebări clare (de obicei de tip PICO), cu un protocol stabilit dinainte, o căutare exhaustivă în mai multe baze de date, criterii explicite de includere și excludere, evaluarea calității fiecărui studiu și o sinteză transparentă. Se deosebește de recenzia narativă, în care autorul alege studiile după criterii nedeclarate. O metaanaliză este componenta statistică a unei recenzii sistematice: combină rezultatele studiilor într-o estimare globală, ponderând fiecare studiu după precizia lui (de regulă, invers proporțional cu varianța).

Rezultatul unei metaanalize se prezintă într-un grafic de tip forest (engl. forest plot), pe care l-ați întâlnit deja în capitolul 10 pentru măsurile de asociere din setul mg2026. Fiecare rând este un studiu: un pătrat pentru estimarea punctuală (cu aria proporțională cu ponderea studiului) și o linie orizontală pentru intervalul de încredere. Linia verticală marchează absența efectului (1 pentru RR și OR, 0 pentru diferențe). Rombul de la bază este estimarea combinată, cu lățimea egală cu intervalul ei de încredere. Dacă rombul nu atinge linia verticală, efectul combinat este semnificativ statistic.

Heterogenitatea descrie cât de mult diferă rezultatele studiilor între ele dincolo de ce ar explica întâmplarea. Se testează cu testul Q al lui Cochran și se cuantifică prin I^2, proporția variabilității totale datorată diferențelor reale dintre studii. Ca reper orientativ, un I^2 în jur de 25% indică heterogenitate mică, 50% moderată și 75% mare. Când heterogenitatea este mare, combinarea studiilor într-o singură cifră trebuie făcută cu prudență, cu un model cu efecte aleatorii (care admite că fiecare studiu are propriul efect adevărat), și trebuie căutate explicațiile: populații diferite, doze diferite, designuri diferite.

Pentru raportarea recenziilor sistematice se folosește ghidul PRISMA 2020 [129], care include o listă de verificare și diagrama de flux a selecției studiilor (câte înregistrări au fost găsite, câte au rămas după eliminarea duplicatelor, câte au fost excluse la citirea titlului și a rezumatului, câte la citirea textului integral și din ce motive). În metaanalizele de acuratețe diagnostică se combină perechi de sensibilitate și specificitate, iar rezultatul se prezintă de obicei printr-o curbă ROC sumară, subiect care depășește nivelul acestui curs.

Spre deosebire de o recenzie sistematică, o recenzie narativă nu declară criteriile de selecție și nu produce o estimare combinată. Un exemplu din cercetarea colegilor este recenzia lui Suciu și colaboratorii despre insuficiența cardiacă cu fracție de ejecție păstrată, care sintetizează narativ studii publicate între 2009 și 2025 [130]. Astfel de lucrări sunt valoroase pentru a înțelege un domeniu, dar se situează mai jos în ierarhia dovezilor decât o recenzie sistematică cu metaanaliză.

11.11 Erori cognitive în raționamentul clinic

Raționamentul clinic folosește două sisteme. Primul este rapid, intuitiv, bazat pe recunoașterea tiparelor ("am mai văzut asta"). Al doilea este lent, analitic și explicit, cum este calculul bayesian din acest capitol. Medicii experimentați folosesc mai ales primul sistem și, de cele mai multe ori, cu succes. Intuiția are însă erori sistematice, numite biasuri cognitive, care pot fi cunoscute și corectate [120].

Neglijarea frecvenței de bază (engl. base-rate neglect) este eroarea centrală a acestui capitol: judecăm probabilitatea unei boli după cât de bine se potrivește tabloul clinic, ignorând cât de rară este boala. Un tânăr cu palpitații, transpirații și cefalee "se potrivește" cu feocromocitomul, dar

boala este atât de rară încât, chiar cu un tablou sugestiv, probabilitatea post-test rămâne mică. Antidotul este să începi întotdeauna cu întrebarea "cât de frecventă este această boală la pacienții ca acesta?".

Ancorarea (engl. anchoring) înseamnă fixarea pe prima impresie și ajustarea insuficientă când apar informații noi. Pacientul trimis de la triaj cu eticheta "atac de panică" va fi privit prin această lentilă, iar o troponină ușor crescută poate fi minimalizată. În termeni bayesieni, ancorarea înseamnă a nu actualiza suficient probabilitatea după un rezultat cu RV mare. Înrudită cu ea este închiderea prematură: oprirea căutării după primul diagnostic plauzibil.

Disponibilitatea (engl. availability) înseamnă supraestimarea probabilității bolilor pe care ni le amintim ușor: cazul dramatic de săptămâna trecută, boala discutată recent la o conferință, diagnosticul ratat care ne-a adus o reclamație. După un caz de embolie pulmonară ratat, mulți medici cer angio-CT la pacienți cu probabilitate pre-test foarte mică. Probabilitatea reală nu s-a schimbat; s-a schimbat doar ușurința cu care ne vine în minte boala.

Alte erori frecvente sunt biasul de confirmare (căutăm doar informațiile care susțin ipoteza noastră), efectul de încadrare (aceeași informație, prezentată ca "90% supraviețuire" sau ca "10% mortalitate", produce decizii diferite) și încrederea excesivă. Nicio listă nu elimină aceste erori, dar două obiceiuri le reduc mult: estimarea explicită a probabilității pre-test înainte de a cere un test și întrebarea "ce altceva ar putea fi?" înainte de a închide diagnosticul.

11.12 Comunicarea riscului către pacient

Pacienții, la fel ca medicii, înțeleg mult mai bine frecvențele naturale decât procentele condiționate. "Din 100 de persoane ca dumneavoastră care au un scor FINDRISC de cel puțin 12, aproximativ 40 au diabet, iar 60 nu au" este o propoziție pe care o înțelege oricine. "Valoarea predictivă pozitivă a testului este 40%" nu este.

A doua regulă este să comunici riscul absolut, nu doar riscul relativ. În studiul simulat mg2026, proporția pacienților care au slăbit cel puțin 5% a fost 29,6% în brațul C și 11,3% în brațul A. Riscul relativ este 2,62 ("de 2,6 ori mai mulți"), dar diferența absolută este 18,3 puncte procentuale, iar numărul necesar de tratat (NNT) este 5,5: aproximativ 6 pacienți trebuie incluși în programul C în loc de A pentru ca unul în plus să atingă obiectivul. Pentru un efect advers rar, diferența dintre relativ și absolut este și mai importantă: un medicament care "dublează riscul" de tromboză poate crește riscul absolut de la 2 la 4 la 10.000 de pacienți pe an. Ambele afirmații sunt adevărate, dar doar a doua permite pacientului să cântărească decizia.

A treia regulă este folosirea aceluiași numitor pentru toate opțiunile ("din 1000 de persoane") și, când se poate, a unui suport vizual: o grilă cu 100 de figurine, dintre care unele colorate. Comunicarea incertitudinii face și ea parte din comunicarea riscului. A spune "testul a ieșit pozitiv, ceea ce înseamnă că probabilitatea să aveți boala a crescut de la aproximativ 1 din 5 la 2 din 5; avem nevoie de un test de confirmare" este mai onest și, pe termen lung, mai liniștitor decât un verdict prematur.

11.13 Probleme rezolvate

Problema rezolvată 11.1. Tabelul boală-simptom și teorema lui Bayes

Într-un serviciu de urgență au fost analizați retrospectiv 400 de pacienți cu durere toracică. La 60 dintre ei s-a confirmat un sindrom coronarian acut (B+). Subdenivelarea segmentului ST pe ECG-ul inițial (S+) a fost prezentă la 36 dintre pacienții cu sindrom coronarian și la 34 dintre ceilalți (date fictive). Calculați p(B+), p(S+ / B+), p(S+ / B-), p(S+), p(B+ / S+) și p(B+ / S-). Verificați p(B+ / S+) cu teorema lui Bayes.

Rezolvare. Construim tabelul cu notațiile din curs. Pe rândul B+: a = 36, b = 60 - 36 = 24. Pe rândul B-: c = 34, d = 340 - 34 = 306.

	S+	S-	Total
B+	36	24	60
B-	34	306	340
Total	70	330	400

Probabilitățile necondiționate: $p(B^+) = \frac{60}{400} = 0{,}15$; $p(S^+) = \frac{70}{400} = 0{,}175$.

Probabilitățile condiționate de boală: $p(S^+/B^+) = \frac{36}{60} = 0{,}60$; $p(S^+/B^-) = \frac{34}{340} = 0{,}10$.

Probabilitățile condiționate de simptom: $p(B^+/S^+) = \frac{36}{70} = 0{,}514$; $p(B^+/S^-) = \frac{24}{330} = 0{,}073$.

Verificarea cu Bayes, fără celula a: $p(B^+/S^+) = \frac{0{,}60\times0{,}15}{0{,}60\times0{,}15+0{,}10\times0{,}85} = \frac{0{,}090}{0{,}090+0{,}085} = \frac{0{,}090}{0{,}175} = 0{,}514$.

Interpretare clinică. Subdenivelarea ST ridică probabilitatea de sindrom coronarian de la 15% la 51,4%, dar nu o face certă. Absența ei o coboară la 7,3%, valoare încă prea mare pentru a externa pacientul fără alte investigații (troponine seriate). Raportul de verosimilitate al semnului este $\mathrm{RV}^+ = \frac{0{,}60}{0{,}10} = 6$, iar $\mathrm{RV}^- = \frac{0{,}40}{0{,}90} = 0{,}44$: semnul este util pentru confirmare, mult mai puțin pentru excludere.

Problema rezolvată 11.2. Frecvențe naturale pentru un test rapid

Un test rapid pentru anticorpii anti-VHC are sensibilitatea 98% și specificitatea 97%. Se propune folosirea lui (a) în populația generală, unde prevalența infecției este 1%, și (b) într-un centru pentru persoane care își injectează droguri, unde prevalența este 20% (cifre ilustrative). Calculați VPP și VPN în cele două situații folosind frecvențe naturale.

Rezolvare (a). Luăm 10.000 de persoane. Prevalența de 1% dă 100 de infectați și 9900 de neinfectați. Dintre cei 100, testul este pozitiv la 98 (Se 98%) și negativ la 2. Dintre cei 9900, specificitatea de 97% dă 9603 negativi și 297 de pozitivi falși. Total pozitivi: 98 + 297 = 395. $\mathrm{VPP} = \frac{98}{395} = 24{,}8\%$. Total negativi: 2 + 9603 = 9605. $\mathrm{VPN} = \frac{9603}{9605} = 99{,}98\%$.

Rezolvare (b). Din 10.000 de persoane, 2000 sunt infectate și 8000 nu. Pozitivi adevărați: 0,98 x 2000 = 1960; negativi falși: 40. Pozitivi falși: 0,03 x 8000 = 240; negativi adevărați: 7760. $\mathrm{VPP} = \frac{1960}{1960+240} = \frac{1960}{2200} = 89{,}1\%$. $\mathrm{VPN} = \frac{7760}{7760+40} = 99{,}49\%$.

Interpretare. Același test, cu o specificitate excelentă de 97%, produce în populația generală trei pozitivi falși pentru fiecare pozitiv adevărat: aproximativ 3 din 4 rezultate pozitive sunt false. În populația cu risc înalt, 9 din 10 rezultate pozitive sunt adevărate. De aceea, un test de screening pozitiv în populația generală trebuie întotdeauna confirmat cu un test diferit (de exemplu determinarea ARN-ului viral), iar comunicarea către pacient trebuie să explice acest lucru înainte de efectuarea testului.

Problema rezolvată 11.3. Același studiu, prevalențe diferite: genotiparea HPV și dubla imunomarcare p16/Ki-67

Moleriu și colaboratorii au urmărit 184 de paciente cu citologie LSIL, dintre care 64 au fost biopsiate [131]. Histologia (standardul de referință) a arătat leziuni de grad înalt la 18 din 64

(28,13%). Genotiparea HPV a avut Se 94,44% (17/18) și Sp 58,7% (27/46); dubla imunomarcare p16/Ki-67 a avut Se 66,67% (12/18) și Sp 84,78% (39/46). Calculați RV+ și RV- pentru fiecare test, apoi VPP și VPN la prevalența din studiu și la o prevalență de 10%.

Rezolvare. Reconstituim tabelele 2 x 2. Pentru HPV: a = 17, b = 1, c = 46 - 27 = 19, d = 27. Pentru p16/Ki-67: a = 12, b = 6, c = 46 - 39 = 7, d = 39. Observăm că 17 + 19 = 36 de paciente au fost HPV pozitive (56,25% din 64), iar 12 + 7 = 19 au fost p16/Ki-67 pozitive (29,69%), ceea ce se potrivește cu proporțiile din rezumat.

Rapoartele de verosimilitate:

HPV: $\mathrm{RV}^{+} = \frac{0{,}9444}{1-0{,}587} = \frac{0{,}9444}{0{,}413} = 2{,}29$; $\mathrm{RV}^{-} = \frac{0{,}0556}{0{,}587} = 0{,}095$.

p16/Ki-67: $\mathrm{RV}^{+} = \frac{0{,}6667}{0{,}1522} = 4{,}38$; $\mathrm{RV}^{-} = \frac{0{,}3333}{0{,}8478} = 0{,}393$.

La prevalența din studiu (28,13%, șansa 0,391): pentru HPV, $\mathrm{VPP} = \frac{17}{36} = 47{,}2\%$ și $\mathrm{VPN} = \frac{27}{28} = 96{,}4\%$; pentru p16/Ki-67, $\mathrm{VPP} = \frac{12}{19} = 63{,}2\%$ și $\mathrm{VPN} = \frac{39}{45} = 86{,}7\%$. Aceleași valori se obțin cu forma cu șanse (de exemplu 0,391 x 2,29 = 0,896, iar 0,896/1,896 = 47,2%).

La prevalența de 10% (șansa 0,111): HPV pozitiv dă 0,111 x 2,29 = 0,254, adică VPP 20,3%; HPV negativ dă 0,111 x 0,095 = 0,0106, adică probabilitate post-test 1,0% (VPN 99,0%). Pentru p16/Ki-67: 0,111 x 4,38 = 0,487, VPP 32,7%; 0,111 x 0,393 = 0,0437, probabilitate post-test 4,2% (VPN 95,8%).

Interpretare. Cele două teste au profiluri opuse. Genotiparea HPV, foarte sensibilă și puțin specifică, este bună pentru excludere: un rezultat negativ coboară mult probabilitatea leziunii de grad înalt. Imunomarcarea p16/Ki-67, mai specifică, este mai bună pentru confirmare. Combinarea lor (HPV ca prim test, p16/Ki-67 la cele pozitive) ar fi o strategie în serie logică, dar performanța combinației nu se poate calcula din rezumat, pentru că nu știm câte paciente au fost pozitive la ambele teste.

Problema rezolvată 11.4. Testare secvențială cu rapoarte de verosimilitate

Golu și colaboratorii au studiat elastografia 2D-SWE în leziunile paratiroidiene, la 22 de pacienți cu hiperparatiroidism și 43 de martori sănătoși [132]. La pragul de 12,5 kPa, sensibilitatea a fost 93% și specificitatea 86% (AUC 0,949). (a) Calculați RV+ și RV-. (b) Care este probabilitatea post-test după un rezultat pozitiv, la prevalența din studiu și la o probabilitate pre-test de 5%? (c) La pacientul cu probabilitatea pre-test de 5% și elastografie pozitivă se efectuează un al doilea test, independent condiționat, cu RV+ = 8,5 (valoare ipotetică). Care este probabilitatea finală?

Rezolvare. (a) $\mathrm{RV}^{+} = \frac{0{,}93}{1-0{,}86} = \frac{0{,}93}{0{,}14} = 6{,}64$. $\mathrm{RV}^{-} = \frac{0{,}07}{0{,}86} = 0{,}081$.

(b) Prevalența din studiu este 22/65 = 33,8%, cu $\frac{\text{șansa } 0{,}338}{0{,}662} = 0{,}512$. Șansa post-test este 0,512 x 6,64 = 3,40, iar probabilitatea 3,40/4,40 = 77,3%. La probabilitatea pre-test de 5%, șansa este 0,0526; șansa post-test 0,0526 x 6,64 = 0,350; probabilitatea 0,350/1,350 = 25,9%.

(c) Probabilitatea de 25,9% devine noua probabilitate pre-test, cu șansa 0,350. După al doilea test pozitiv: 0,350 x 8,5 = 2,97, adică probabilitatea 2,97/3,97 = 74,8%. Echivalent, într-un singur pas: 0,0526 x 6,64 x 8,5 = 2,97.

Interpretare. Prevalența de 33,8% din studiu este artificială, rezultată din designul caz-control (22 de pacienți și 43 de martori aleși). Valoarea predictivă de 77,3% nu se transferă în practică. În plus, designul caz-control, cu martori sănătoși, tinde să supraestimeze sensibilitatea și

specificitatea (biasul de spectru, discutat în capitolul 12), deci RV-urile reale pot fi mai modeste. Calculul secvențial este valabil numai dacă cele două teste sunt independente condiționat.

Problema rezolvată 11.5. Diagnostic diferențial al durerii toracice în cabinetul de medicină de familie

Un pacient de 55 de ani se prezintă la medicul de familie cu durere toracică intermitentă. Medicul consideră trei diagnostice, presupuse mutual exclusive și exhaustive, cu probabilitățile a priori (date fictive, dar plauzibile pentru asistența primară): durere musculo-scheletală 0,50; boală de reflux gastroesofagian (BRGE) 0,30; angină pectorală stabilă 0,20. Trei elemente ale anamnezei și examenului au frecvențele din tabel.

Diagnostic	**p(D)**	$p(\text{ durere la efort } / D)$	$p(\text{ durere reprodusă la palpar } / D)$	$p(\text{ arsură postprandială } / D)$
Musculo-scheletală	0,50	0,20	0,70	0,05
BRGE	0,30	0,10	0,10	0,70
Angină stabilă	0,20	0,80	0,15	0,15

Pacientul descrie durere la efort, durerea nu se reproduce la palpare și nu are arsuri postprandiale. Calculați probabilitățile posterioare, presupunând independența condiționată a celor trei elemente.

Rezolvare. Verosimilitatea combinației (efort +, palpare -, arsură -) în fiecare diagnostic:

Musculo-scheletală: 0,20 x (1 - 0,70) x (1 - 0,05) = 0,20 x 0,30 x 0,95 = 0,0570

BRGE: 0,10 x (1 - 0,10) x (1 - 0,70) = 0,10 x 0,90 x 0,30 = 0,0270

Angină: 0,80 x (1 - 0,15) x (1 - 0,15) = 0,80 x 0,85 x 0,85 = 0,5780

Produsele cu probabilitățile a priori: 0,50 x 0,0570 = 0,0285; 0,30 x 0,0270 = 0,0081; 0,20 x 0,5780 = 0,1156. Suma este 0,1522.

Probabilitățile posterioare: musculo-scheletală 0,0285/0,1522 = 18,7%; $\frac{\text{BRGE } 0{,}0081}{0{,}1522} = 5{,}3\%$; angină 0,1156/0,1522 = 76,0%.

Interpretare. Angina, cel mai puțin probabil diagnostic înainte de anamneză (20%), devine cel mai probabil (76,0%). Rezultatul justifică trimiterea pentru evaluare cardiologică (ECG de efort sau altă testare funcțională) și inițierea măsurilor de prevenție secundară. Rezultatul depinde însă de ipoteza că cele trei elemente sunt independente în fiecare boală și de faptul că lista de diagnostice este completă; un sindrom coronarian acut, de exemplu, nu este inclus și trebuie exclus clinic înainte de orice calcul.

Problema rezolvată 11.6. Pragul de testare și pragul de tratament

Pentru o faringită acută la un adolescent, medicul consideră că merită tratată cu antibiotic dacă probabilitatea etiologiei streptococice depășește 25%. Un test rapid antigenic are, în scenariul nostru, sensibilitatea 86% și specificitatea 95%. Între ce probabilități pre-test este util testul?

Rezolvare. $\text{RV}^+ = \frac{0{,}86}{0{,}05} = 17{,}2$. $\text{RV}^- = \frac{0{,}14}{0{,}95} = 0{,}147$. Pragul de tratament de 25% corespunde șansei 0,25/0,75 = 0,333.

Pragul inferior (de testare): șansa pre-test pentru care un test pozitiv atinge exact pragul de tratament este 0,333 / 17,2 = 0,0194, adică probabilitatea 0,0194/1,0194 = 1,9%.

Pragul superior (de testare-tratament): șansa pre-test pentru care un test negativ coboară exact la pragul de tratament este 0,333 / 0,147 = 2,27, adică probabilitatea 2,27/3,27 = 69,4%.

Interpretare. Pentru probabilități pre-test între 1,9% și 69,4%, rezultatul testului poate schimba decizia, deci testul este util. Sub 1,9%, chiar un test pozitiv nu ar justifica antibioticul (iar probabilitatea ca rezultatul pozitiv să fie fals este mare). Peste 69,4%, chiar un test negativ ar lăsa probabilitatea peste 25%, deci testul nu schimbă decizia. Modelul ignoră costul și disconfortul testului și se bazează pe un prag de tratament ales prin judecată clinică; pragurile reale se stabilesc în ghiduri.

Problema rezolvată 11.7. Inferență bayesiană pentru o proporție pe un eșantion mic

Într-un studiu pilot, 3 din 20 de pacienți tratați cu un medicament nou au prezentat o reacție adversă cutanată. (a) Cu un a priori uniform Beta(1; 1), care este distribuția a posteriori, media ei și intervalul credibil de 95%? (b) Un studiu anterior sugerează o frecvență de aproximativ 10%, rezumată prin a priori Beta(2; 18) (media 10%, interval a priori 95% de la 1,3% la 26,0%). Ce se schimbă?

Rezolvare. (a) A posteriori este Beta(1 + 3; 1 + 17) = Beta(4; 18). Media este 4/22 = 18,2%. Intervalul credibil de 95% (din tabelele distribuției Beta sau din program) este 5,4-36,3%.

(b) A posteriori este Beta(2 + 3; 18 + 17) = Beta(5; 35). Media este 5/40 = 12,5%, iar intervalul credibil de 95% este 4,3-24,2%.

Interpretare. Cu doar 20 de pacienți, priorul contează: estimarea scade de la 18,2% la 12,5%, iar intervalul se îngustează, pentru că priorul informativ adaugă practic informația a încă 18 pacienți observați anterior. Comparați cu exemplul din secțiunea 11.9, unde, cu 160 de pacienți, cele două priori dădeau rezultate aproape identice. Într-un raport științific, ambele analize trebuie prezentate, iar alegerea priorului trebuie justificată, de exemplu prin referința la studiul anterior.

Problema rezolvată 11.8. De la factorul Bayes la probabilitatea ipotezei

În setul mg2026, comparația TAS între sexe dă BF10 = 0,18, iar asocierea obezitate-diabet dă BF10 de aproximativ 85. Calculați probabilitatea a posteriori a ipotezei H1 pentru fiecare analiză, pornind de la (a) o probabilitate a priori de 50% și (b) una de 20%.

Rezolvare. Folosim șansa a posteriori = șansa a priori x BF10.

TAS și sexul: (a) șansa a priori 1; șansa a posteriori 0,18; $p(H_1/\text{date}) = \frac{0{,}18}{1{,}18} = 15{,}3\%$. (b) Șansa a priori 0,25; șansa a posteriori 0,045; $p(H_1/\text{date}) = \frac{0{,}045}{1{,}045} = 4{,}3\%$.

Obezitate și diabet: (a) 1 x 85 = 85; $p(H_1/\text{date}) = \frac{85}{86} = 98{,}8\%$. (b) 0,25 x 85 = 21,25; $p(H_1/\text{date}) = \frac{21{,}25}{22{,}25} = 95{,}5\%$.

Interpretare. Factorul Bayes descrie forța dovezilor din date; probabilitatea finală a ipotezei depinde și de cât de plauzibilă era ea înainte. Asocierea obezitate-diabet este atât de bine susținută încât rămâne foarte probabilă chiar cu un prior sceptic. Pentru TAS și sex, datele împing ipoteza unei diferențe spre improbabil. Observați analogia perfectă cu diagnosticul: BF10 este raportul de verosimilitate, iar probabilitatea a priori este probabilitatea pre-test.

11.14 Exerciții propuse

1. Într-un studiu pe 500 de pacienți, 60 au boala B. Simptomul S este prezent la 45 dintre bolnavi și la 88 dintre nebolnavi. Construiți tabelul boală-simptom și calculați p(B+), p(S+ / B+), p(S+ / B-) și p(B+ / S+). Verificați ultimul rezultat cu teorema lui Bayes.
2. O mamografie de screening are sensibilitatea 90% și specificitatea 93%. Prevalența cancerului mamar nediagnosticat în populația de screening este 0,8%. Folosind 10.000 de

femei și frecvențe naturale, calculați VPP și VPN. Câte femei cu mamografie pozitivă trebuie investigate suplimentar pentru a găsi un cancer?

3. Transformați în șanse probabilitățile 0,40 și 0,05 și transformați în probabilități șansele 3 și 0,25.
4. Danila și colaboratorii (cu F. Cioca) au raportat, pentru dNLR la pacienții cu COVID-19 și diabet, pragul 2,685 cu sensibilitatea 70,312%, specificitatea 65,978% și AUC 0,624, iar pentru ALRI sensibilitatea 76,429% și specificitatea 69,541% (AUC 0,752) [105]. (a) Calculați RV+ și RV- pentru dNLR. (b) Pentru un pacient cu probabilitatea pre-test de 30%, care este probabilitatea post-test după un dNLR peste prag și după unul sub prag? (c) Comentați numărul de zecimale raportat pentru sensibilitate și specificitate.
5. Folosind datele din problema rezolvată 11.4 (2D-SWE, Se 93%, Sp 86%), calculați probabilitatea post-test după un rezultat negativ, pentru o probabilitate pre-test de 33,8% și de 5%.
6. Testul A are Se 0,90 și Sp 0,80; testul B are Se 0,85 și Sp 0,95. Presupunând independența condiționată, calculați sensibilitatea și specificitatea combinației în serie și în paralel. Ce strategie ați alege pentru triajul la camera de gardă și ce strategie pentru confirmarea înaintea unui tratament toxic?
7. La un serviciu de urgență, un pacient cu cefalee acută are trei diagnostice posibile, cu probabilitățile a priori: migrenă 0,70; cefalee de tip tensional 0,25; hemoragie subarahnoidiană (HSA) 0,05. Frecvențele elementelor clinice sunt: debut brusc, "în lovitură de trăsnet" (0,05 / 0,02 / 0,80); redoare de ceafă (0,05 / 0,10 / 0,60); greață (0,70 / 0,10 / 0,70), în ordinea migrenă / tensională / HSA. Pacientul are toate cele trei elemente. Calculați probabilitățile posterioare, presupunând independența condiționată, și comentați decizia.
8. Pragul de tratament pentru o boală este 10%. Testul disponibil are RV+ = 4 și RV- = 0,25. Calculați pragul de testare și pragul de testare-tratament.
9. Ion și colaboratorii (cu F. Cioca) au comparat scoruri clinice pentru predicția COVID-19 sever la pacienții cu diabet [106]. SOFA a avut AUC 0,771 la internare și 0,873 în ziua a cincea; NEWS2 0,892 la internare și 0,729 în ziua a cincea; APACHE II 0,698 la internare și 0,806 în ziua a cincea. (a) Încadrați fiecare valoare pe scala AUC din curs (capitolul 12). (b) Ce scor ați folosi la internare și ce scor după cinci zile? (c) Studiul descrie o regresie logistică, dar raportează rapoarte de risc instantaneu (HR, de exemplu SOFA în ziua a cincea HR 3,07, IC 95% 2,29-4,12). Ce măsură de efect ar fi fost de așteptat dintr-o regresie logistică?
10. O analiză bayesiană dă BF10 = 0,16 pentru o corelație. (a) Exprimați rezultatul ca BF01 și încadrați-l pe scala din tabelul 11.8. (b) Dacă probabilitatea a priori a existenței corelației era 50%, care este probabilitatea a posteriori?
11. Formulați în structura PICO o întrebare clinică despre utilitatea HbA1c față de TTGO pentru diagnosticul diabetului în cabinetul de medicină de familie și identificați tipul de studiu cel mai potrivit pentru a-i răspunde.

Capcane frecvente

- Confundarea p(S / B) cu p(B / S). Sensibilitatea unui test nu este probabilitatea ca un pacient cu test pozitiv să fie bolnav.
- Ignorarea prevalenței (neglijarea frecvenței de bază): un test cu Se și Sp de 95% poate avea VPP sub 50% dacă boala este rară.

- Folosirea VPP sau VPN raportate într-un studiu caz-control ca și cum ar fi valabile în cabinet. Prevalența dintr-un astfel de studiu este fixată de cercetător.
- Înmulțirea rapoartelor de verosimilitate pentru teste corelate (de exemplu FINDRISC și glicemia), ceea ce supraestimează probabilitatea finală.
- Confuzia dintre șansă și probabilitate: o șansă de 0,25 înseamnă o probabilitate de 0,20, nu de 0,25.
- Interpretarea intervalului de încredere frecventist ca interval credibil ("95% probabilitate ca parametrul să fie aici").
- Citirea unui BF10 mic ca "rezultat neconcludent". Un BF10 de 0,1 este dovadă pentru H0, nu absența dovezilor.

De reținut

- Teorema lui Bayes: $p(B/S) = \frac{p(S/B) \times p(B)}{p(S)}$, cu p(S) din formula probabilității totale.
- Forma cu șanse: șansa post-test = șansa pre-test x RV; $\text{RV}^+ = \frac{\text{Se}}{1-\text{Sp}}$; $\text{RV}^- = \frac{1-\text{Se}}{\text{Sp}}$.
- Probabilitatea pre-test depinde de pacient și de context (cabinet, ambulatoriu, spital); valorile predictive depind de ea.
- Frecvențele naturale (1000 de pacienți, arbore) sunt cel mai sigur mod de a rezolva o problemă de diagnostic și de a explica rezultatul unui pacient.
- În serie: crește Sp și VPP; în paralel: crește Se și VPN. Formulele simple presupun independență condiționată.
- Diagnosticul diferențial bayesian: produs prior x verosimilitate pentru fiecare boală, împărțit la suma produselor.
- Inferența bayesiană: a posteriori $\propto$ verosimilitate x a priori; intervalul credibil și factorul Bayes au interpretări directe, dar cer un prior justificat.
- Medicina bazată pe dovezi: cinci pași (PICO, căutare, evaluare critică, aplicare, evaluare), încrederea în dovezi evaluată prin GRADE, recenziile sistematice raportate după PRISMA 2020.

Întrebări de autoevaluare

1. (CS) În tabelul boală-simptom cu notațiile din curs (a, b pe rândul B+; c, d pe rândul B-), probabilitatea condiționată p(S+ / B-) este:
   a) $\frac{c}{N}$
   b) $\frac{c}{a+c}$
   c) $\frac{c}{c+d}$
   d) $\frac{a+c}{N}$
   e) $\frac{d}{c+d}$
2. (CS) Un test cu sensibilitatea 80% și specificitatea 90% este aplicat într-o populație cu prevalența bolii de 10%. Valoarea predictivă pozitivă este aproximativ:
   a) 80%
   b) 47%
   c) 90%
   d) 8%
   e) 72%
3. (CS) O probabilitate pre-test de 0,20 corespunde unei șanse de:
   a) 0,20

b) 0,25
c) 0,80
d) 4
e) 0,16

4. (CS) Un rezultat de test cu raportul de verosimilitate egal cu 1:
   a) confirmă boala
   b) exclude boala
   c) dublează probabilitatea pre-test
   d) nu modifică probabilitatea bolii
   e) are sensibilitatea 100%
5. (CS) O analiză bayesiană raportează BF10 = 0,05. Interpretarea corectă este:
   a) dovezi puternice în favoarea ipotezei H1
   b) p = 0,05, rezultat la limita semnificației
   c) dovezi puternice în favoarea ipotezei H0 (BF01 = 20)
   d) rezultat neconcludent, nu se poate spune nimic
   e) probabilitatea ca H0 să fie adevărată este 5%
6. (CM) Despre testarea în serie (al doilea test se face doar la cei pozitivi la primul, iar pacientul este considerat pozitiv dacă ambele teste sunt pozitive):
   a) crește specificitatea combinației față de fiecare test luat separat
   b) crește sensibilitatea combinației față de fiecare test luat separat
   c) crește valoarea predictivă pozitivă
   d) este mai economică, pentru că al doilea test se aplică doar unei părți dintre pacienți
   e) formula Se = Se1 x Se2 este valabilă doar dacă testele sunt independente condiționat
7. (CM) Despre rapoartele de verosimilitate:
   a) $RV^{+} = \frac{Se}{1-Sp}$
   b) $RV^{-} = \frac{1-Se}{Sp}$
   c) depind puternic de prevalența bolii, la fel ca valorile predictive
   d) șansa post-test = șansa pre-test x RV
   e) un RV mai mic decât 1 crește probabilitatea bolii
8. (CM) Despre medicina bazată pe dovezi:
   a) integrează dovezile din cercetare cu expertiza clinică și cu valorile pacientului
   b) în structura PICO, C desemnează comparatorul
   c) în sistemul GRADE, dovezile din studiile observaționale pornesc de la încredere mare
   d) o recenzie narativă se situează deasupra unei recenzii sistematice în ierarhia dovezilor
   e) heterogenitatea dintr-o metaanaliză se poate cuantifica prin I^2
9. (CM) Despre erorile cognitive din raționamentul clinic:
   a) neglijarea frecvenței de bază înseamnă ignorarea prevalenței bolii
   b) ancorarea înseamnă ajustarea insuficientă a primei impresii când apar date noi
   c) disponibilitatea duce la supraestimarea probabilității bolilor ușor de amintit
   d) biasul de verificare este o eroare cognitivă a medicului la patul bolnavului
   e) închiderea prematură înseamnă continuarea investigațiilor după stabilirea diagnosticului

10. (CM) Despre inferența bayesiană:

a) distribuția a posteriori este proporțională cu verosimilitatea înmulțită cu distribuția a priori

b) pe măsură ce eșantionul crește, influența priorului asupra rezultatului scade

c) un interval credibil de 95% se poate interpreta ca probabilitatea de 95% ca parametrul să se afle în interval, dat fiind priorul și datele

d) factorul Bayes poate aduce dovezi în favoarea ipotezei nule

e) un a priori Beta(1; 1) favorizează valorile mici ale unei proporții

Capitolul 12. Evaluarea performanței testelor diagnostice: sensibilitate, specificitate, valori predictive și curba ROC

Obiective

La finalul acestui capitol, studentul va putea:

- să definească standardul de referință și testul de evaluat și să descrie designul corect al unui studiu de acuratețe diagnostică;
- să construiască tabelul 2 x 2 (T+/T- față de B+/B-) cu notațiile a, b, c, d și să calculeze prevalența, sensibilitatea, specificitatea, valorile predictive, acuratețea, rata erorii de clasificare, indicele Youden, rapoartele de verosimilitate și raportul șanselor de diagnostic;
- să reconstituie un tabel 2 x 2 complet pornind de la câțiva indicatori și câteva celule cunoscute;
- să explice de ce valorile predictive depind de prevalență și să le recalculeze pentru o altă populație;
- să recunoască biasul de spectru, biasul de verificare și biasul de încorporare într-un articol publicat;
- să construiască de mână o curbă ROC dintr-un tabel de praguri, să interpreteze aria de sub curbă și intervalul ei de încredere și să compare două curbe ROC;
- să aleagă un prag de decizie (indicele Youden, prag de excludere, prag de confirmare) în funcție de scopul clinic;
- să calculeze și să interpreteze coeficientul kappa al lui Cohen și să recunoască situațiile în care se folosește coeficientul de corelație intraclasă;
- să aplice aceiași indicatori la evaluarea unui clasificator de inteligență artificială și să folosească listele STARD 2015 și STARD-AI la citirea critică a unui studiu.

12.1 Testul de evaluat și standardul de referință

Orice test diagnostic, de la o întrebare din anamneză la un biomarker sau la un algoritm de inteligență artificială, trebuie să răspundă la o întrebare simplă: cât de bine separă pacienții care au boala de cei care nu o au? Pentru a răspunde, avem nevoie de două lucruri. Primul este testul de evaluat (engl. index test), adică testul nou, mai simplu, mai ieftin sau mai puțin invaziv, a cărui performanță vrem să o cunoaștem. Al doilea este standardul de referință (engl. reference standard, numit tradițional "gold standard"), adică metoda considerată cea mai bună disponibilă pentru a stabili prezența sau absența bolii.

În setul mg2026, testul de evaluat este scorul FINDRISC, iar standardul de referință este testul oral de toleranță la glucoză (TTGO). În cercetarea colegilor din Timișoara, standardele de referință sunt foarte variate: biopsia hepatică cu stadializare Metavir pentru fibroză [107], histologia pentru leziunile cervicale [131], endoscopia digestivă superioară pentru varicele esofagiene [110], diagnosticul final stabilit pe toate datele disponibile pentru obstrucția biliară malignă [57]. Uneori, standardul de referință este el însuși o metodă neinvazivă, aleasă pentru că este mai bine validată: elastografia tranzitorie (VCTE) pentru elastografia time-harmonic [3] sau parametrul de atenuare controlată (CAP) pentru noile metode de cuantificare a steatozei [58,65].

Standardul de referință nu este aproape niciodată perfect. Biopsia hepatică analizează un fragment de câțiva milimetri dintr-un organ întreg; histologia depinde de anatomopatolog; TTGO are o reproductibilitate modestă. Când standardul de referință greșește, testul de evaluat va părea

mai slab decât este, pentru că dezacordurile vor fi atribuite automat testului nou. Acest lucru trebuie avut în minte când un test nou "pierde" în fața unei referințe imperfecte.

12.1.1 Designul unui studiu de acuratețe diagnostică

Designul ideal este transversal: o serie de pacienți consecutivi, reprezentativi pentru cei la care testul va fi folosit în practică (de exemplu adulți cu suspiciune de boală din cabinetul de medicină de familie), la care se efectuează atât testul de evaluat, cât și standardul de referință, independent și fără ca interpretul unuia să cunoască rezultatul celuilalt. Prevalența bolii în eșantion este atunci o estimare a prevalenței reale, iar toți indicatorii, inclusiv valorile predictive, au sens pentru populația respectivă.

Designul caz-control diagnostic (bolnavi cunoscuți comparați cu voluntari sănătoși) este mai ușor de realizat și se folosește frecvent în fazele de început ale evaluării unui test. Are însă două probleme mari. Prima: prevalența este fixată de cercetător (de exemplu 1:2 cazuri și martori), deci valorile predictive calculate nu au semnificație clinică. A doua: comparația dintre bolnavi tipici, adesea în stadii avansate, și persoane perfect sănătoase exagerează acuratețea, pentru că lipsesc tocmai cazurile dificile (boala incipientă, pacienții cu alte boli care imită boala studiată). Vom reveni la acest bias de spectru în secțiunea 12.5.

12.2 Tabelul 2 x 2 și indicatorii de performanță

12.2.1 Notațiile

Cursul disciplinei folosește un tabel cu boala pe rânduri și rezultatul testului pe coloane. Păstrăm aceste notații în tot capitolul.

Tabelul 12.1. Tabelul 2 x 2 pentru evaluarea unui test diagnostic, notațiile din curs.

	T+	**T-**	**Total**
B+	a (adevărat pozitivi, AP)	b (fals negativi, FN)	a + b
B-	c (fals pozitivi, FP)	d (adevărat negativi, AN)	c + d
Total	a + c	b + d	$N = a + b + c + d$

Celula a conține bolnavii cu test pozitiv (adevărat pozitivi), b bolnavii cu test negativ (fals negativi), c persoanele fără boala studiată, dar cu test pozitiv (fals pozitivi), iar d persoanele fără boală și cu test negativ (adevărat negativi). În unele manuale și programe (de exemplu MedCalc) tabelul este transpus, cu boala pe coloane și testul pe rânduri; atunci literele b și c își schimbă locul. De aceea, înainte de orice calcul, verificați întotdeauna ce reprezintă fiecare celulă, nu ce literă are.

12.2.2 Prevalența, sensibilitatea și specificitatea

Prevalența (în eșantion) este proporția bolnavilor:

prevalența = (a + b) / N

Sensibilitatea (Se) este probabilitatea ca un bolnav să aibă testul pozitiv, adică proporția bolnavilor corect identificați:

$$\text{Se} = p(T^+/B^+) = \frac{a}{a+b}$$

Specificitatea (Sp) este probabilitatea ca o persoană fără boala studiată să aibă testul negativ:

$$\text{Sp} = p(T^-/B^-) = \frac{d}{c+d}$$

Sensibilitatea se calculează doar pe rândul bolnavilor, specificitatea doar pe rândul nebolnavilor. Prin urmare, în principiu, nu depind de prevalență: dacă dublăm numărul de sănătoși

din studiu, specificitatea rămâne aceeași, iar sensibilitatea nici nu observă. Spunem "în principiu" pentru că sensibilitatea și specificitatea depind de spectrul pacienților (stadiul bolii, comorbiditățile), iar spectrul se schimbă adesea odată cu prevalența.

Două mnemotehnici englezești sunt utile la patul bolnavului. SnNout (engl. high Sensitivity, Negative result, rules out): un test foarte sensibil, dacă iese negativ, exclude boala, pentru că un test sensibil ratează puțini bolnavi. SpPin (engl. high Specificity, Positive result, rules in): un test foarte specific, dacă iese pozitiv, confirmă boala, pentru că un test specific dă puține pozitive false. Mnemotehnicile sunt aproximări: un test cu Se 95% și Sp 40% "exclude" bine boala doar dacă probabilitatea pre-test nu este foarte mare, cum am văzut în capitolul 11.

12.2.3 Valorile predictive

Valoarea predictivă pozitivă (VPP) este probabilitatea ca o persoană cu test pozitiv să aibă boala, adică proporția bolnavilor printre cei cu test pozitiv:

$$\text{VPP} = p(B^+/T^+) = \frac{a}{a+c}$$

Valoarea predictivă negativă (VPN) este proporția celor fără boală printre cei cu test negativ:

$$\text{VPN} = p(B^-/T^-) = \frac{d}{b+d}$$

Valorile predictive se calculează pe coloane. Ele răspund la întrebarea pe care și-o pune pacientul ("testul meu a ieșit pozitiv; cât de probabil este să am boala?"), dar depind puternic de prevalență, cum vom vedea în secțiunea 12.4. Valorile predictive sunt, de fapt, probabilitățile post-test din capitolul 11, calculate pentru o probabilitate pre-test egală cu prevalența din eșantion [133].

12.2.4 Acuratețea și rata erorii de clasificare

Acuratețea (AC) este proporția tuturor persoanelor clasificate corect de test:

$$\text{AC} = \frac{a+d}{N}$$

Rata erorii de clasificare (RE) este proporția celor clasificați greșit, adică a falselor negative și a falselor pozitive:

$$\text{RE} = \frac{b+c}{N} = 1 - \text{AC}$$

Acuratețea pare indicatorul cel mai intuitiv, dar poate înșela. Ea este o medie ponderată a sensibilității și specificității, cu ponderi date de prevalență: AC = Se x prevalența + Sp x (1 - prevalența). Când boala este rară, acuratețea este dominată de specificitate. Un "test" care declară pe toată lumea sănătoasă are, la o prevalență de 2%, o acuratețe de 98%, deși nu detectează niciun bolnav. Din acest motiv, acuratețea nu se raportează niciodată singură.

> **Regula care previne greșeala**
>
> Rata erorii de clasificare se calculează din celulele care conțin erori: falsele negative (b) și falsele pozitive (c). Formula corectă este $\text{RE} = \frac{b+c}{N} = \frac{\text{FN+FP}}{N}$. Formula (b + d) / N, care apare uneori în notițe, este greșită: d conține adevărații negativi, adică persoane clasificate corect. O verificare rapidă: AC + RE trebuie să fie exact 1. Pentru tabelul 150 / 30 / 4 / 256 din problema rezolvată 12.1, $\frac{a+d}{N} = \frac{406}{440} = 0{,}923$, iar $\frac{b+c}{N} = \frac{34}{440} = 0{,}077$; suma este 1.

12.2.5 Indicele Youden, rapoartele de verosimilitate și raportul șanselor de diagnostic

Indicele Youden combină sensibilitatea și specificitatea într-o singură cifră [134]:

$$J = \text{Se} + \text{Sp} - 1$$

J variază între 0 (testul nu discriminează mai bine decât aruncarea unei monede) și 1 (test perfect). Are avantajul că nu depinde de prevalență și dezavantajul că acordă aceeași importanță falselor negative și falselor pozitive, ceea ce clinic rareori este adevărat. Îl vom folosi la alegerea pragului pe curba ROC.

Rapoartele de verosimilitate, introduse în capitolul 11, se calculează tot din Se și Sp [121]:

$$\text{RV}^+ = \frac{\text{Se}}{1 - \text{Sp}}$$

$$\text{RV}^- = \frac{1 - \text{Se}}{\text{Sp}}$$

Raportul șanselor de diagnostic (engl. diagnostic odds ratio, DOR) este raportul dintre șansa unui test pozitiv la bolnavi și șansa unui test pozitiv la nebolnavi:

$$\text{DOR} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a \times d}{b \times c} = \frac{\text{RV}^+}{\text{RV}^-}$$

Este același calcul ca raportul șanselor din analiza riscului, aplicat tabelului diagnostic. Un DOR de 1 înseamnă un test inutil; valori de ordinul zecilor sau sutelor indică un test puternic. DOR este util pentru compararea testelor în metaanalize, dar este greu de interpretat clinic și nu spune dacă testul este mai bun la confirmare sau la excludere: două teste cu Se 90% / Sp 50% și Se 50% / Sp 90% au exact același DOR (9).

Tabelul 12.2. Rezumatul indicatorilor de performanță diagnostică (notațiile din curs).

Indicator	**Formula**	**Depinde de prevalență?**	**Întrebarea la care răspunde**
Prevalența	$\frac{a+b}{N}$	(este prevalența)	Cât de frecventă e boala în eșantion?
Sensibilitatea	$\frac{a}{a+b}$	nu (în principiu)	Ce proporție din bolnavi detectează testul?
Specificitatea	$\frac{d}{c+d}$	nu (în principiu)	Ce proporție din sănătoși sunt negativi?
VPP	$\frac{a}{a+c}$	da	Un pozitiv, cât de probabil e bolnav?
VPN	$\frac{d}{b+d}$	da	Un negativ, cât de probabil e sănătos?
Acuratețea	$\frac{a+d}{N}$	da	Ce proporție e clasificată corect?
Rata erorii	$\frac{b+c}{N}$	da	Ce proporție e clasificată greșit?
Indicele Youden	Se + Sp - 1	nu	Cât de bine discriminează, global?
RV+	$\frac{\text{Se}}{1-\text{Sp}}$	nu	De câte ori crește șansa la un pozitiv?
RV-	$\frac{1-\text{Se}}{\text{Sp}}$	nu	De câte ori scade șansa la un negativ?
DOR	$\frac{a \times d}{b \times c}$	nu	Raportul șanselor de test pozitiv

12.2.6 Indicatorii pe setul mg2026 și intervalele lor de încredere

La pragul FINDRISC ≥ 12, tabelul pentru diabetul confirmat prin TTGO este: a = 22 (AP), b = 7 (FN), c = 33 (FP), d = 98 (AN), cu N = 160 și 29 de bolnavi (prevalența 18,1%).

Tabelul 12.3. FINDRISC ≥ 12 față de TTGO, setul mg2026 (date sintetice).

	FINDRISC ≥ 12	**FINDRISC < 12**	**Total**
Diabet (TTGO)	22	7	29
Fără diabet	33	98	131
Total	55	105	160

Se $= \frac{22}{29} = 75{,}9\%$; Sp $= \frac{98}{131} = 74{,}8\%$; VPP $= \frac{22}{55} = 40{,}0\%$; VPN $= \frac{98}{105} = 93{,}3\%$; AC $= \frac{120}{160} = 75{,}0\%$; RE $= \frac{40}{160} = 25{,}0\%$; J = 0,759 + 0,748 - 1 = 0,507; RV+ = 3,01; RV- = 0,32; DOR $= \frac{22\times98}{7\times33} = \frac{2156}{231} = 9{,}33$.

Fiecare dintre acești indicatori este o proporție estimată pe un eșantion și trebuie însoțit de un interval de încredere. Pentru proporții se folosește de preferință metoda Wilson, discutată în capitolul 4. Sensibilitatea se bazează doar pe cei 29 de bolnavi, deci intervalul ei este larg: Se 75,9% (IC 95% 57,9-87,8). Specificitatea, calculată pe 131 de nebolnavi, este mai precisă: 74,8% (IC 95% 66,7-81,5). VPP 40,0% (IC 95% 28,1-53,2), VPN 93,3% (IC 95% 86,9-96,7). Pentru rapoartele de verosimilitate se folosește o metodă pe scară logaritmică: RV+ 3,01 (IC 95% 2,10-4,31), RV- 0,32 (IC 95% 0,17-0,62), iar DOR 9,33 (IC 95% 3,65-23,84).

Lecția este una pe care o veți revedea la fiecare articol de acuratețe diagnostică: sensibilitatea este aproape întotdeauna estimată mai imprecis decât specificitatea, pentru că bolnavii sunt mai puțini. Un studiu cu 30 de bolnavi nu poate estima sensibilitatea mai precis de aproximativ ±15 puncte procentuale, oricâte zecimale ar raporta [135].

12.3 Reconstituirea unui tabel 2 x 2

Un tip de problemă specific cursului nostru, frecvent la examen, cere reconstituirea întregului tabel pornind de la câțiva indicatori și câteva celule. Problemele par complicate, dar se rezolvă întotdeauna cu aceeași strategie, în patru pași.

Pasul 1. Scrieți tabelul gol, cu literele a, b, c, d și totalurile marginale, și notați formula fiecărui indicator dat: Se $= \frac{a}{a+b}$, Sp $= \frac{d}{c+d}$, VPP $= \frac{a}{a+c}$, VPN $= \frac{d}{b+d}$, prevalența = (a + b)/N.

Pasul 2. Completați celulele date direct (de exemplu "AP = 90" înseamnă a = 90).

Pasul 3. Alegeți un indicator care conține o singură necunoscută și rezolvați ecuația. De exemplu, dacă VPP $= \frac{9}{12}$ și a = 90, atunci $\frac{90}{90+c} = \frac{9}{12}$, de unde 9(90 + c) = 12 x 90, deci c = 30. Repetați până când toate celulele sunt cunoscute.

Pasul 4. Calculați totalurile și verificați indicatorii dați. Apoi calculați indicatorii ceruți.

O greșeală frecventă este folosirea unui indicator cu două necunoscute înainte de a afla una dintre ele. Un alt tip de greșeală apare când rapoartele sunt date ca fracții (2/3, 6/8): nu le transformați în zecimale rotunjite (0,66) înainte de rezolvare, ci lucrați cu fracția, altfel celulele nu mai ies numere întregi. Problemele rezolvate 12.2-12.5 aplică strategia pas cu pas.

12.4 Dependența valorilor predictive de prevalență

Sensibilitatea și specificitatea sunt proprietăți ale testului (pentru un spectru dat de pacienți). Valorile predictive sunt proprietăți ale testului aplicat unei anumite populații. Teorema lui Bayes, scrisă cu indicatorii din acest capitol, arată exact legătura:

$$\mathrm{VPP} = \frac{\mathrm{Se} \times P}{\mathrm{Se} \times P + (1 - \mathrm{Sp}) \times (1 - P)}$$

$$\mathrm{VPN} = \frac{\mathrm{Sp} \times (1 - P)}{\mathrm{Sp} \times (1 - P) + (1 - \mathrm{Se}) \times P}$$

unde P este prevalența. Pe măsură ce prevalența crește, VPP crește și VPN scade.

Figura 12.1 arată această dependență pentru FINDRISC ≥ 12 (Se 76%, Sp 75%), iar tabelul 12.4 dă valorile exacte calculate pe setul mg2026 pentru câteva prevalențe.

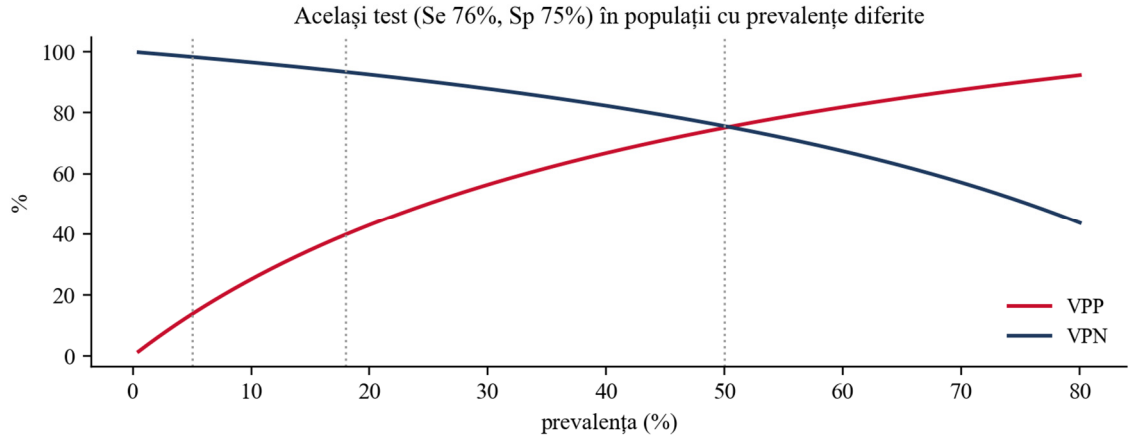


Figura 12.1. Valoarea predictivă pozitivă și negativă a aceluiași test (Se 76%, Sp 75%, FINDRISC ≥ 12) în populații cu prevalențe diferite. Liniile punctate marchează prevalențele de 5%, 18% și 50%. Date sintetice (setul mg2026).

Tabelul 12.4. VPP și VPN ale FINDRISC ≥ 12 (Se 75,9%, Sp 74,8%) pentru diferite prevalențe ale diabetului nediagnosticat. Date sintetice (setul mg2026).

Prevalența	**2%**	**5%**	**10%**	**18%**	**30%**	**50%**
VPP (%)	5,8	13,7	25,1	39,8	56,3	75,1
VPN (%)	99,3	98,3	96,5	93,4	87,9	75,6

La o prevalență de 2% (de exemplu adulți tineri fără factori de risc), din 100 de persoane cu FINDRISC ≥ 12, doar aproximativ 6 au diabet. La o prevalență de 50% (de exemplu pacienți cu poliurie, polidipsie și scădere ponderală), 75 din 100 au diabet. Testul este același; populația este alta. Concluzia practică are două fețe. Valorile predictive raportate într-un articol se aplică doar populațiilor cu prevalență asemănătoare. Pentru a le transfera în altă populație, se recalculează din Se, Sp și noua prevalență (sau, echivalent, cu rapoartele de verosimilitate) [133,136].

Această dependență explică de ce screeningul în populația generală produce atâtea rezultate fals pozitive și de ce testele de confirmare sunt indispensabile. Explică și de ce același biomarker pare excelent într-un studiu de spital și dezamăgitor în cabinet.

12.5 Biasuri în studiile de acuratețe diagnostică

Indicatorii calculați corect dintr-un tabel corect pot fi totuși înșelători dacă tabelul provine dintr-un studiu cu design defectuos. Trei biasuri apar atât de des încât orice medic trebuie să le recunoască.

Biasul de spectru apare când pacienții din studiu nu seamănă cu cei la care testul va fi folosit. Cea mai frecventă formă este designul caz-control cu bolnavi în stadii avansate și martori perfect sănătoși. Bolnavii avansați au valori foarte modificate ale testului (sensibilitate aparent mare), iar sănătoșii nu au nicio boală care să producă rezultate pozitive false (specificitate aparent mare). În practică, testul va întâlni cazuri incipiente și pacienți cu alte boli, iar performanța va scădea. Biasul de spectru nu este doar o problemă de design: el reflectă faptul că sensibilitatea și specificitatea variază real cu tipul de pacient. Studiul lui Pavel și colaboratorii despre H-FABP, discutat în capitolul 11, a arătat sensibilitatea și specificitatea de 89,67 și 95,65 sub 4 ore de la debutul durerii, dar 86,73 și 49,84 peste 4 ore [127]. Sipos și colaboratorii au observat, la IL-6 în sepsisul neonatal, o sensibilitate de 90% într-un subgrup și de 71,4% în altul, după statusul culturilor [126].

Biasul de verificare (engl. verification bias sau work-up bias) apare când standardul de referință se efectuează doar la o parte dintre pacienți, iar selecția depinde de rezultatul testului de evaluat. Tipic, pacienții cu test pozitiv sunt trimiși la biopsie sau colonoscopie, iar cei cu test negativ nu. Atunci falsele negative rămân nedescoperite (celula b este subestimată), iar sensibilitatea pare mai mare decât este. În studiul lui Moleriu și colaboratorii, dintre 184 de paciente cu citologie LSIL, doar 64 au fost biopsiate [131]; în studiul lui Iovanescu și colaboratorii, 252 din 2201 pacienți au efectuat colonoscopia, iar complianța la colonoscopie printre cei FIT pozitivi a fost de 52% [101]. În ambele situații, criteriile după care s-a decis verificarea influențează estimările, iar autorii trebuie să le descrie.

Biasul de încorporare apare când rezultatul testului de evaluat face parte din definiția standardului de referință. Dacă diagnosticul final al pancreatitei se stabilește, printre altele, pe baza lipazei, nu putem evalua apoi corect acuratețea lipazei: testul "se confirmă pe sine". Problema este frecventă când standardul de referință este un diagnostic final clinic, construit din toate datele disponibile, inclusiv testul evaluat. Studiul lui Rațiu și colaboratorii a comparat ERCP, MRCP și histologia cu diagnosticul final al obstrucției biliare maligne [57]; la un astfel de design, cititorul trebuie să verifice dacă diagnosticul final a folosit rezultatele testelor comparate.

Alte surse de bias sunt interpretarea testului cu cunoașterea rezultatului standardului de referință (bias de interpretare, prevenit prin orbire), excluderea rezultatelor neinterpretabile (de exemplu măsurătorile elastografice nevalide) și folosirea aceluiași eșantion atât pentru alegerea pragului, cât și pentru estimarea performanței la acel prag (supraestimare optimistă, corectată prin validare pe un eșantion separat).

Studiu de caz: designul caz-control și performanța aparentă a testelor [2,116,132]

Trei studii ale colegilor folosesc un design de tip caz-control și oferă un bun exercițiu de lectură critică. Lupușoru și colaboratorii au evaluat ecografia cu contrast cu imagistică parametrică a timpului de sosire (CEUS-PAT) pentru ciroza hepatică, la 64 de persoane: 37 cu ciroză și 27 de voluntari sănătoși. Valorile CEUS-PAT au fost 0,83 ± 0,09 la cirotici și 0,49 ± 0,11 la sănătoși ($p < 0{,}0001$), corelația cu VCTE a fost $r = 0{,}81$, iar la pragul > 0,7 aria sub curba ROC a fost 0,98, cu Se 89,19%, Sp 100%, VPP 100% și VPN 87,1%. Cifrele se pot reconstitui exact: 33 din 37 de cirotici au fost pozitivi (89,19%), toți cei 27 de voluntari au fost negativi (100%), deci 31 de negativi, dintre care 27 sănătoși $\left(\frac{\text{VPN } 27}{31} = 87{,}1\%\right)$. Golu și colaboratorii (cu L. Moleriu și A. Tudor) au studiat elastografia 2D-SWE la 22 de pacienți cu hiperparatiroidism și 43 de martori sănătoși; valorile medii ale rigidității raportate în rezumat

au fost 10,2 ± 4,9 kPa și 19,5 ± 7,6 kPa (p = 0,007), iar la pragul de 12,5 kPa, Se 93%, Sp 86% și AUC 0,949. Feier și colaboratorii (cu C. Muntean) au comparat 145 de colecistite gangrenoase cu 290 de colecistite cronice, într-un raport fixat de 1:2, potrivite pe vârstă și sex; cel mai bun indice inflamator a avut AUC 0,889, Se 79,5% și Sp 86,3%. Ce au în comun cele trei studii? Prevalența este fixată de design (57,8%, 33,8% și 33,3%), deci valorile predictive nu se aplică direct în practică. Când martorii sunt voluntari sănătoși (primele două studii), AUC-urile foarte mari (0,98 și 0,949) reprezintă probabil limita superioară a performanței; pacienții cu hepatopatie cronică necirotică sau cu noduli tiroidieni ar fi fost comparatorii relevanți clinic. Studiul lui Feier este mai aproape de situația clinică, pentru că martorii au tot colecistită, dar raportul 1:2 rămâne artificial. Toate trei sunt pași necesari în evaluarea unui test; următorul pas ar fi o cohortă de pacienți consecutivi cu suspiciune clinică.

12.6 Teste cu rezultat continuu și alegerea pragului

12.6.1 Compromisul dintre sensibilitate și specificitate

Multe teste nu au un rezultat "pozitiv" sau "negativ" de la natură, ci o valoare numerică: glicemia, troponina, rigiditatea hepatică, un scor clinic. Pentru a le folosi la decizie, alegem un prag (engl. cut-off): valorile peste prag sunt considerate pozitive. Alegerea pragului schimbă tabelul 2 x 2 și, odată cu el, toți indicatorii.

Mecanismul se vede în panoul B al figurii 12.3, mai jos: distribuția scorului FINDRISC la persoanele cu diabet (media 14,76, AS 3,96) este deplasată spre dreapta față de distribuția la persoanele fără diabet (media 9,40, AS 3,70), dar cele două distribuții se suprapun. Oriunde am pune pragul, unii bolnavi vor rămâne sub el (fals negativi) și unii sănătoși vor fi deasupra (fals pozitivi). Dacă mutăm pragul spre stânga (valori mai mici), prindem mai mulți bolnavi (sensibilitatea crește), dar declarăm pozitivi și mai mulți sănătoși (specificitatea scade). Dacă îl mutăm spre dreapta, se întâmplă invers. Nu există prag care să le crească pe amândouă; există doar compromisuri. Cursul disciplinei descrie această relație ca pe o relație de "invers proporționalitate" între sensibilitate și specificitate; mai exact, este o relație monotonă inversă: când una crește, cealaltă nu poate decât să scadă sau să rămână la fel.

12.6.2 Tabelul pragurilor pentru FINDRISC

Tabelul 12.5 arată, pentru câteva praguri ale scorului FINDRISC, cele patru celule și indicatorii principali, calculați pe setul mg2026.

Tabelul 12.5. Performanța FINDRISC pentru diabetul confirmat prin TTGO, la diferite praguri (n = 160; 29 cu diabet, 131 fără). Date sintetice (setul mg2026).

Prag	AP	FP	FN	AN	Se (%)	Sp (%)	1 - Sp	VPP (%)	VPN (%)	J
≥ 8	28	94	1	37	96,6	28,2	0,718	23,0	97,4	0,248
≥ 10	28	66	1	65	96,6	49,6	0,504	29,8	98,5	0,462
≥ 11	27	44	2	87	93,1	66,4	0,336	38,0	97,8	0,595
≥ 12	22	33	7	98	75,9	74,8	0,252	40,0	93,3	0,507
≥ 14	17	16	12	115	58,6	87,8	0,122	51,5	90,6	0,464
≥ 15	13	11	16	120	44,8	91,6	0,084	54,2	88,2	0,364
≥ 16	12	7	17	124	41,4	94,7	0,053	63,2	87,9	0,360

Citiți tabelul de sus în jos. Pe măsură ce pragul crește, AP scad și FN cresc (sensibilitatea scade de la 96,6% la 41,4%), iar FP scad și AN cresc (specificitatea crește de la 28,2% la 94,7%). VPP crește, VPN scade. Indicele Youden atinge maximul, 0,595, la pragul ≥ 11 (Se 93,1%, Sp 66,4%). Rețineți că la toate pragurile prevalența rămâne 18,1%; se schimbă doar modul în care testul împarte pacienții.

12.7 Curba ROC

12.7.1 Construcția pas cu pas

Curba ROC (engl. receiver operating characteristic, denumire moștenită din teoria semnalelor radar) reprezintă grafic toate compromisurile posibile dintre sensibilitate și specificitate. Pe axa orizontală se pune 1 - specificitatea (proporția falselor pozitive printre sănătoși), pe axa verticală sensibilitatea. Fiecare prag devine un punct. Cursul scrie relația ca Se = f(1 - Sp).

Construcția de mână are cinci pași.

Pasul 1. Ordonați pragurile de la cel mai strict (cel mai mare, care dă cele mai puține rezultate pozitive) la cel mai permisiv.

Pasul 2. Pentru fiecare prag, calculați Se și 1 - Sp din tabelul 2 x 2 corespunzător (tabelul 12.5).

Pasul 3. Adăugați cele două puncte extreme: (0; 0), care corespunde unui prag atât de mare încât nimeni nu este pozitiv, și (1; 1), care corespunde unui prag atât de mic încât toată lumea este pozitivă.

Pasul 4. Puneți punctele pe grafic și uniți-le cu segmente de dreaptă, în ordinea pragurilor.

Pasul 5. Trasați diagonala de la (0; 0) la (1; 1): ea reprezintă un test fără valoare, la care proporția pozitivilor este aceeași la bolnavi și la sănătoși.

Pentru FINDRISC, cu cele șase praguri ≥ 16, ≥ 15, ≥ 14, ≥ 12, ≥ 10 și ≥ 8, punctele (1 - Sp; Se) sunt: (0; 0), (0,053; 0,414), (0,084; 0,448), (0,122; 0,586), (0,252; 0,759), (0,504; 0,966), (0,718; 0,966), (1; 1). Figura 12.2 arată curba obținută, peste curba completă construită cu toate pragurile posibile ale scorului.

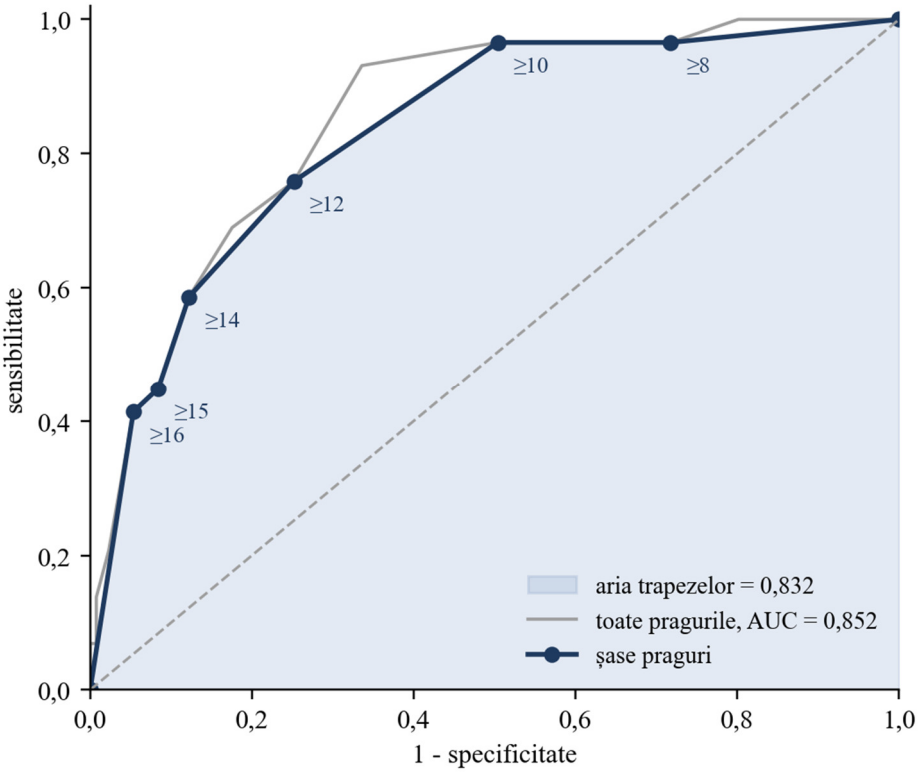


Figura 12.2. Curba ROC a scorului FINDRISC pentru diabetul confirmat prin TTGO, construită de mână din șase praguri (linia cu puncte), peste curba completă cu toate pragurile (linia gri). Aria trapezelor de sub curba cu șase praguri este 0,832; aria sub curba completă este 0,852. Date sintetice (setul mg2026).

Aria de sub curbă se poate calcula de mână prin metoda trapezelor. Între două puncte consecutive (x1; y1) și (x2; y2), aria trapezului este (x2 - x1) x (y1 + y2) / 2. Pentru cele șapte segmente obținem 0,0111 + 0,0132 + 0,0197 + 0,0873 + 0,2172 + 0,2064 + 0,2776 = 0,832. Cu toate pragurile posibile, aria este 0,852, valoarea raportată de programe. Diferența apare pentru că, unind cu linii drepte puncte rare, "tăiem colțurile" curbei; cu cât folosim mai multe praguri, cu atât aria trapezelor se apropie de aria reală.

12.7.2 Ce înseamnă aria de sub curba ROC

Aria de sub curba ROC (engl. area under the curve, AUC), numită în curs și coeficientul c, este un indicator global al capacității testului de a separa bolnavii de sănătoși, independent de prag [137,138]. Are o interpretare probabilistică foarte concretă: AUC este probabilitatea ca un bolnav ales la întâmplare să aibă o valoare a testului mai mare decât un sănătos ales la întâmplare (perechile cu valori egale se numără pe jumătate).

Pentru FINDRISC în setul mg2026 putem verifica direct. Există 29 x 131 = 3799 de perechi bolnav-sănătos. În 3154 de perechi, bolnavul are scorul mai mare; în 169 de perechi, scorurile sunt egale. AUC $= \frac{3154+\frac{169}{2}}{3799} = \frac{3238,5}{3799} = 0,852$. Interpretarea: dacă luăm la întâmplare un pacient cu diabet și unul fără diabet, în 85,2% din cazuri cel cu diabet are scorul FINDRISC mai mare. Calculul este identic cu statistica U a testului Mann-Whitney din capitolul 7, împărțită la produsul volumelor celor două grupuri; AUC este deci o măsură a mărimii efectului pentru diferența dintre distribuțiile unui marker la bolnavi și sănătoși.

Un test perfect are AUC = 1 (curba urcă vertical până la Se = 1 la 1 - Sp = 0, apoi merge orizontal). Un test inutil are AUC = 0,5 (curba se suprapune diagonalei). O AUC sub 0,5 indică de obicei o eroare de codificare: testul funcționează "invers", iar inversarea direcției (valorile mici sunt pozitive) dă AUC = 1 - AUC.

12.7.3 Scale de interpretare a AUC

Cursul disciplinei folosește următoarea scală pentru coeficientul c:

Tabelul 12.6. Scala de interpretare a ariei de sub curba ROC din cursul disciplinei.

AUC	Interpretare
0,91-1	excelent
0,81-0,90	foarte bun
0,71-0,80	bun
0,61-0,70	satisfăcător
sub 0,6	slab

Alte manuale folosesc scale ușor diferite. O variantă frecventă în literatura clinică consideră 0,9-1 excelent, 0,8-0,9 bun, 0,7-0,8 acceptabil (sau moderat), 0,6-0,7 slab și 0,5-0,6 eșec. Diferențele dintre scale arată că pragurile sunt convenții, nu legi. Interpretarea trebuie să țină cont și de contextul clinic: o AUC de 0,75 poate fi foarte utilă pentru un test ieftin de triaj și insuficientă pentru un test care decide o intervenție chirurgicală. Pe scala din curs, FINDRISC (0,852) și glicemia a jeun (0,844) sunt teste "foarte bune" pentru diabetul confirmat prin TTGO în setul mg2026, HbA1c (0,764) este "bun", iar IMC (0,731) și circumferința taliei (0,737) sunt "bune" spre limita inferioară.

Exemple din cercetarea colegilor acoperă toată scala: CEUS-PAT pentru ciroză, AUC 0,98 (excelent) [2]; scorul combinat pentru varicele cu risc înalt, AUROC 0,93 [110]; ISARIC-4C

pentru mortalitate, 0,875 [75]; strain-ul longitudinal global pentru insuficiența cardiacă cu fracție de ejecție păstrată, 0,833 [90]; ALRI la pacienții cu COVID-19 și diabet, 0,752, față de dNLR, doar 0,624 (satisfăcător) [105].

12.7.4 Intervalul de încredere pentru AUC

AUC este o estimare pe eșantion și are o eroare standard. Hanley și McNeil au propus o formulă care depinde doar de AUC și de numărul de bolnavi și sănătoși [137]; metoda DeLong, folosită implicit de multe programe, estimează eroarea standard neparametric, din datele propriu-zise. Pentru FINDRISC în mg2026, IC 95% este 0,762-0,943 cu metoda Hanley-McNeil și aproximativ 0,782-0,923 cu metoda DeLong. Pentru glicemia a jeun, AUC 0,844, cu IC 95% 0,752-0,937 (Hanley-McNeil). Ambele intervale sunt largi, pentru că se bazează pe doar 29 de bolnavi. Limita inferioară a intervalului este cea care contează pentru o afirmație prudentă: putem spune că FINDRISC are cel puțin o performanță "bună" (0,76), nu neapărat "foarte bună".

Un interval de încredere care include 0,5 înseamnă că testul nu discriminează semnificativ mai bine decât întâmplarea. Testul statistic asociat (H0: AUC = 0,5) apare în toate programele; la FINDRISC $p < 0,001$.

12.7.5 Compararea a două curbe ROC

Când două teste sunt efectuate la aceiași pacienți, AUC-urile lor sunt corelate, iar compararea trebuie să țină cont de această corelație, la fel cum testul t pereche ține cont de corelația dintre măsurătorile repetate. Metoda DeLong face exact acest lucru: estimează varianța diferenței dintre două AUC-uri pe aceleași date și calculează o statistică z. Când testele sunt evaluate pe grupuri diferite de pacienți, se folosește o comparație pentru eșantioane independente.

Figura 12.3 compară FINDRISC cu glicemia a jeun în setul mg2026.

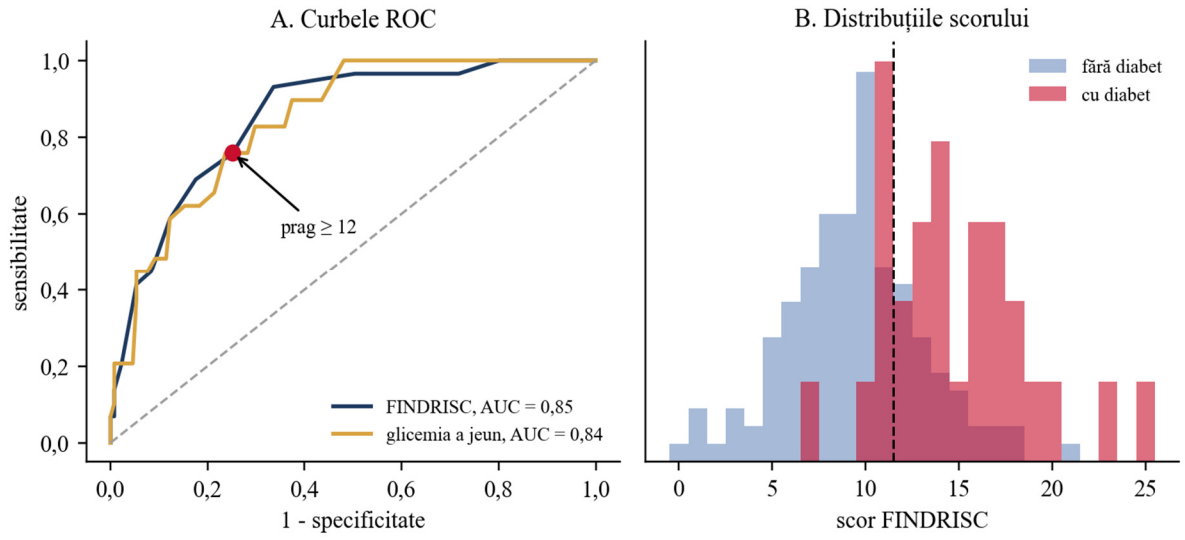


Figura 12.3. A: curbele ROC ale scorului FINDRISC (AUC 0,85) și glicemiei a jeun (AUC 0,84) pentru diabetul confirmat prin TTGO; punctul roșu marchează pragul FINDRISC ≥ 12. B: distribuțiile scorului FINDRISC la persoanele fără diabet și cu diabet; linia punctată marchează pragul dintre 11 și 12. Date sintetice (setul mg2026).

Diferența dintre AUC-uri este 0,852 - 0,844 = 0,008 (IC 95% -0,060 la 0,077), z = 0,24, p = 0,81 (metoda DeLong, corelația dintre cele două AUC-uri fiind aproximativ 0,5). Cele două teste au o capacitate de discriminare practic identică pentru diabetul confirmat prin TTGO. Alegerea dintre ele se face după alte criterii: FINDRISC nu necesită recoltare de sânge și poate fi completat de pacient în sala de așteptare, iar glicemia este o măsurătoare obiectivă, dar cere post și laborator. Observați și forma curbelor: ele se intersectează, deci la unele praguri un test este mai bun, la

altele celălalt. AUC rezumă întreaga curbă, inclusiv zone de prag pe care nu le-am folosi niciodată în practică.

> **Studiu de caz: compararea mai multor teste prin AUROC [57,107]**
>
> Rațiu și colaboratorii (cu R. Lupușoru) au comparat, la 160 de pacienți cu obstrucție biliară (vârsta medie 69,31 ± 10,96 ani; 53,1% bărbați), trei metode față de diagnosticul final de malignitate: AUROC a fost 0,88 pentru colangio-RM (IC 95% 0,75-0,90), 0,94 pentru ERCP (IC 95% 0,85-0,99) și 0,80 pentru histologie (IC 95% 0,70-0,82). Două observații de lectură critică. Prima: intervalele de încredere sunt foarte asimetrice în jurul estimărilor (0,88 cu limita superioară 0,90, 0,80 cu limita superioară 0,82), ceea ce este neobișnuit pentru metodele obișnuite de calcul și merită verificat în textul integral. A doua: intervalele se suprapun, dar suprapunerea intervalelor nu este un test al diferenței; pentru teste efectuate la aceiași pacienți, comparația corectă este metoda DeLong pe date pereche. Bâldea și colaboratorii (cu A. Tudor) au comparat elastografia tranzitorie (TE) cu cuantificarea VTQ prin ARFI la 157 de pacienți cu hepatită cronică B sau C, cu biopsia hepatică drept standard. Distribuția stadiilor de fibroză a fost F0 19,7%, F1 22,3%, F2 27,4%, F3 17,8%, F4 12,7%, iar corelația dintre cele două metode a fost puternică (Spearman ρ = 0,826; IC 95% 0,769-0,870). Pentru pragurile succesive de fibroză, AUROC au fost 0,876 față de 0,832 (p = 0,358), 0,826 față de 0,862 (p = 0,313), 0,907 față de 0,880 (p = 0,434) și 0,981 față de 0,974 (p = 0,423), TE față de VTQ. Niciuna dintre diferențe nu a fost semnificativă. Interpretarea corectă nu este "cele două metode sunt echivalente", ci "studiul nu a găsit diferențe; diferențele observate, de ordinul a 0,03-0,04, sunt compatibile cu întâmplarea". Pentru a demonstra echivalența ar fi fost nevoie de un design de echivalență, cu o marjă stabilită dinainte, sau de intervalele de încredere ale diferențelor.

12.7.6 Alegerea pragului: indicele Youden și pragurile clinice

Pragul cu indicele Youden maxim este punctul de pe curba ROC cel mai îndepărtat vertical de diagonală. Pentru FINDRISC în mg2026, pragul Youden este ≥ 11, cu Se 93,1%, Sp 66,4% și J = 0,595. Pentru glicemia a jeun, pragul Youden este ≥ 110 mg/dL, cu Se 82,8%, Sp 70,2% și J = 0,530. Indicele Youden presupune însă că o falsă negativă și o falsă pozitivă au aceeași importanță și că prevalența nu contează, ceea ce rareori corespunde realității clinice.

În practică, pragul se alege după scopul testului. Pentru un test de triaj, la care un rezultat negativ încheie investigațiile, contează să nu ratăm bolnavi: se alege un prag de excludere (engl. rule-out) cu sensibilitate mare și VPN mare. FINDRISC ≥ 10 are în mg2026 Se 96,6% și VPN 98,5%: un pacient cu scor sub 10 are o probabilitate de diabet de aproximativ 1,5% și poate fi liniștit. Pentru un test care declanșează un tratament invaziv sau costisitor, contează să nu tratăm sănătoși: se alege un prag de confirmare (engl. rule-in) cu specificitate mare și VPP mare. FINDRISC ≥ 16 are Sp 94,7% și VPP 63,2%. Între cele două praguri rămâne o zonă gri, în care decizia cere informații suplimentare.

Exemplul glicemiei arată încă un lucru important. Pragul clinic de diagnostic al diabetului pentru glicemia a jeun, 126 mg/dL [80], are în setul mg2026 Se doar 37,9% și Sp 94,7% (VPP 61,1%, VPN 87,3%) față de TTGO. Pragul este ales pentru confirmarea diagnosticului (specificitate mare), nu pentru screening; folosit singur, ar rata aproape două treimi dintre diabeticii identificați prin TTGO. Pragul "optim statistic" (110 mg/dL) și pragul "clinic" (126 mg/dL) răspund la întrebări diferite.

Același raționament apare în studiile colegilor. Moga și colaboratorii au raportat pentru elastografia time-harmonic un prag de excludere a cirozei (< 1,61 m/s: Se 90,7%, VPN 93,9%) și un prag de confirmare (> 1,83 m/s: Sp 96,7%, VPP 90,3%) [3]. Barbulescu și colaboratorii au construit reguli de excludere și de confirmare a varicelor cu risc înalt, cu o zonă gri rezolvată prin

albumină [123]. Popa și colaboratorii (cu A. Tudor) au raportat pentru GFAP, la 85 de pacienți cu alertă de cod AVC în unitatea de primiri urgențe, o acuratețe de 94,2% în diferențierea AVC hemoragic de cel ischemic, la pragul de 77,15 pg/mL; GFAP a fost $142{,}91 \pm 102{,}19$ pg/mL la pacienții cu AVC acut față de $37{,}76 \pm 19{,}92$ pg/mL la cei fără AVC ($p < 0{,}001$) [61]. Abaterea standard apropiată de medie sugerează o distribuție asimetrică, pentru care mediana și testul Mann-Whitney ar fi fost mai potrivite, iar pragul cu patru cifre semnificative, derivat pe 69 de pacienți cu AVC, are nevoie de validare.

Studiu de caz: praguri ROC pentru boli cu mai multe grade [58,65,90]

Steatoza hepatică are patru grade (S0-S3), iar fibroza cinci stadii (F0-F4). Pentru astfel de boli ordinale, analiza ROC se face separat pentru fiecare prag de severitate: S0 față de S1-S3, S0-S1 față de S2-S3 și așa mai departe. Bende și colaboratorii (cu R. Lupușoru) au evaluat parametrul UGAP față de CAP la 176 de subiecți clasificați: S0 27,2%, S1 31,6%, S2 7,3%, S3 33,9%. Valorile medii ale UGAP au crescut monoton cu gradul ($198{,}3 \pm \frac{25{,}7 d_B}{m}$ la S0 și $270{,}8 \pm \frac{31{,}62 d_B}{m}$ la S3; $p < 0{,}001$), corelația cu CAP a fost $r = 0{,}73$, iar pragurile au fost 192,5 dB/m pentru S1 (AUC 0,83), 231 dB/m pentru S2 (AUC 0,90) și 248 dB/m pentru S3 (AUC 0,91). Nica și colaboratorii (cu R. Lupușoru) au evaluat fracția de grăsime ecografică (USFF) la 169 de pacienți, cu CAP ca referință (praguri CAP 248, 268 și 280 dB/m; 52,6% cu steatoză cel puțin moderată). USFF s-a corelat cu CAP ($r = 0{,}81$), iar pragurile > 5%, > 7,5% și > 10% au avut AUROC 0,90, 0,92 și 0,91 pentru steatoza cel puțin ușoară, moderată și, respectiv, severă. Un detaliu interesant pentru lectura critică: rezumatul calificează drept "moderată" corelația $r = 0{,}83$ dintre parametrul TAI și CAP, deși pe scala Colton din curs ar fi "foarte bună spre puternică". Stoichescu-Hogea și colaboratorii (cu A. Tudor) au comparat strain-ul longitudinal global (GLS) în trei grupuri (62 de pacienți cu insuficiență cardiacă cu fracție de ejecție păstrată, 46 de hipertensivi, 40 de martori): -17,6 ± 2,3% în primul grup, -18,4 ± 3,0% la hipertensivi și -20,2 ± 1,4% la martori, iar pragul optim de -19,35% a avut AUC 0,833. Aici "pozitiv" înseamnă o valoare mai puțin negativă, iar programul trebuie setat corect pentru direcția testului (altfel AUC ar ieși 0,167). În toate trei studiile, standardul de referință (CAP sau diagnosticul clinic) este imperfect, iar pragurile sunt derivate pe același eșantion pe care se estimează performanța, deci au nevoie de validare externă.

12.8 Acordul dintre evaluatori

Un test diagnostic util trebuie să fie nu doar valid (să măsoare ce trebuie), ci și fiabil: să dea același rezultat când este repetat sau interpretat de alt examinator. Pentru rezultate categoriale (de exemplu "hipertrofie ventriculară stângă prezentă sau absentă" pe ECG), acordul se măsoară cu coeficientul kappa al lui Cohen. Pentru rezultate numerice (rigiditatea hepatică în kPa, tensiunea arterială), se folosește coeficientul de corelație intraclasă și metoda Bland-Altman [102].

12.8.1 Coeficientul kappa al lui Cohen

Proporția de acord observat (Po) este insuficientă, pentru că doi evaluatori ar fi de acord de multe ori și din întâmplare, mai ales când una dintre categorii este foarte frecventă. Kappa corectează acordul observat cu acordul așteptat din întâmplare (Pe):

$$\kappa = \frac{\mathrm{Po} - \mathrm{Pe}}{1 - \mathrm{Pe}}$$

Acordul așteptat se calculează din totalurile marginale, la fel ca frecvențele așteptate din testul chi-pătrat: pentru fiecare categorie, produsul proporțiilor cu care cei doi evaluatori au folosit-o, apoi suma. Kappa este 1 la acord perfect, 0 când acordul nu depășește întâmplarea și

negativ când acordul este mai mic decât cel întâmplător. Pentru interpretare se folosește frecvent scala Landis și Koch [103].

Tabelul 12.7. Scala Landis și Koch pentru interpretarea coeficientului kappa.

κ	Acord
sub 0	slab (sub nivelul întâmplării)
0-0,20	ușor
0,21-0,40	acceptabil (modest)
0,41-0,60	moderat
0,61-0,80	substanțial
0,81-1,00	aproape perfect

Exemplul complet se găsește în problema rezolvată 12.8. Kappa are câteva limite: depinde de prevalența categoriilor (la prevalențe extreme, kappa poate fi mic chiar cu acord observat mare) și nu face diferența între dezacordurile mici și cele mari în cazul categoriilor ordonate. Pentru scale ordinale (de exemplu gradele de steatoză) se folosește kappa ponderat.

12.8.2 Coeficientul de corelație intraclasă

Pentru măsurători numerice, coeficientul de corelație Pearson nu este potrivit pentru acord: două aparate care diferă sistematic cu 5 kPa pot avea r = 1, dar nu sunt în acord. Coeficientul de corelație intraclasă (ICC) măsoară proporția din variabilitatea totală datorată diferențelor reale dintre pacienți, față de cea datorată diferențelor dintre măsurători sau evaluatori. Valori peste 0,90 indică fiabilitate excelentă, între 0,75 și 0,90 bună, între 0,50 și 0,75 moderată.

Moga și colaboratorii au inclus în studiul elastografiei time-harmonic un substudiu de reproductibilitate pe 27 de pacienți: acordul global dintre examinatori a fost 0,94 (IC 95% 0,89-0,97), iar ICC-urile pentru diferite comparații au fost 0,92 (IC 95% 0,82-0,96), 0,94 (IC 95% 0,87-0,97) și 0,97 (IC 95% 0,95-0,99) [3]. Un test cu o astfel de reproductibilitate poate fi folosit de examinatori diferiți fără ca diferențele dintre ei să schimbe clasificarea pacienților. ICC se folosește și pentru fiabilitatea test-retest a chestionarelor, de exemplu valori ICC între 0,82 și 0,91 la un interval de 14 zile într-un studiu de validare al colegilor [73].

12.9 Evaluarea clasificatorilor de inteligență artificială

Algoritmii de inteligență artificială care clasifică imagini medicale (radiografii, retinografii, imagini dermatoscopice) sau semnale (ECG) sunt, din punct de vedere statistic, teste diagnostice. Se evaluează cu aceleași instrumente, chiar dacă literatura de informatică folosește alți termeni.

Tabelul 2 x 2 se numește aici matrice de confuzie (engl. confusion matrix). Sensibilitatea se numește recall (sau rata adevăraților pozitivi), VPP se numește precizie (engl. precision), iar specificitatea rămâne specificitate (rareori, rata adevăraților negativi). Un indicator specific domeniului este scorul F1, media armonică a preciziei și a sensibilității:

F1 = 2 x precizia x sensibilitatea / (precizia + sensibilitatea)

F1 este mare doar când ambele componente sunt mari și este folosit mai ales când clasa pozitivă este rară. Spre deosebire de AUC, F1 depinde de prevalență (prin precizie) și de pragul ales.

Problema clasică a evaluării clasificatorilor este paradoxul acurateței la clase dezechilibrate. Dacă doar 8% dintre retinografiile dintr-un program de screening au retinopatie diabetică, un algoritm care declară toate imaginile normale are acuratețea 92%, dar sensibilitatea 0. De aceea,

un articol care raportează doar "acuratețea modelului" nu poate fi evaluat. Alte capcane frecvente sunt evaluarea pe același set de date pe care s-a antrenat modelul (supraajustare, engl. overfitting), absența unei validări externe pe date din alt centru și alegerea pragului după rezultate. Problema rezolvată 12.10 calculează toți acești indicatori pentru un exemplu concret. Pentru modelele de predicție bazate pe regresie sau învățare automată există un ghid de raportare dedicat, TRIPOD+AI [139].

12.10 Raportarea studiilor de acuratețe diagnostică: STARD 2015 și STARD-AI

Ghidul STARD 2015 (engl. Standards for Reporting of Diagnostic Accuracy Studies) cuprinde 30 de puncte pe care un articol de acuratețe diagnostică ar trebui să le raporteze [140]. Câteva contează mai mult pentru evaluarea critică: identificarea studiului ca studiu de acuratețe diagnostică în titlu sau rezumat; descrierea populației, a criteriilor de eligibilitate și a modului de recrutare (consecutiv sau nu); descrierea testului de evaluat și a standardului de referință, cu detalii suficiente pentru a fi reproduse; precizarea dacă interpreții au fost orbi față de rezultatul celuilalt test; precizarea dacă pragurile au fost stabilite dinainte sau derivate din date; diagrama de flux a participanților (câți au fost eligibili, câți au primit fiecare test, câți au fost excluși și de ce); tabelul 2 x 2 complet; estimările de acuratețe cu intervale de încredere; rezultatele neinterpretabile și modul de tratare a lor.

STARD-AI extinde STARD pentru studiile în care testul de evaluat este un algoritm de inteligență artificială [141]. Adaugă cerințe privind sursa și caracteristicile datelor de antrenare, validare și testare, separarea strictă a acestor seturi, versiunea algoritmului, preprocesarea datelor, modul de alegere a pragului, performanța pe subgrupuri (sex, vârstă, centru, tip de aparat) și analiza erorilor.

La lectura unui articol, folosiți lista ca pe un filtru: dacă tabelul 2 x 2 nu poate fi reconstituit, dacă nu știți cum au fost recrutați pacienții sau dacă pragul a fost ales pe aceleași date pe care se raportează performanța, rezultatele trebuie citite cu prudență, oricât de mari ar fi AUC-ul și sensibilitatea.

12.11 Probleme rezolvate

Problema rezolvată 12.1. Toți indicatorii dintr-un tabel complet

Un test a fost evaluat la 440 de pacienți față de standardul de referință. Dintre cei 180 de bolnavi, 150 au avut testul pozitiv; dintre cei 260 de nebolnavi, 4 au avut testul pozitiv. Calculați toți indicatorii de performanță.

Rezolvare. Tabelul, cu notațiile din curs: a = 150, b = 180 - 150 = 30, c = 4, d = 260 - 4 = 256.

	T+	**T-**	**Total**
B+	150	30	180
B-	4	256	260
Total	154	286	440

Prevalența = 180/440 = 0,409 (40,9%).

$\text{Se} = \frac{150}{180} = 0{,}833(83{,}3\%)$. $\text{Sp} = \frac{256}{260} = 0{,}985(98{,}5\%)$.

$\text{VPP} = \frac{150}{154} = 0{,}974(97{,}4\%)$. $\text{VPN} = \frac{256}{286} = 0{,}895(89{,}5\%)$.

$\text{AC} = \frac{150+256}{440} = \frac{406}{440} = 0{,}923(92{,}3\%)$. $\text{RE} = \frac{30+4}{440} = \frac{34}{440} = 0{,}077(7{,}7\%)$; verificare: 0,923 + 0,077 = 1.

$J = 0{,}833 + 0{,}985 - 1 = 0{,}818$.

$\text{RV}^+ = \frac{0{,}833}{1-0{,}985} = \frac{0{,}833}{0{,}0154} = 54{,}2$. $\text{RV}^- = \frac{0{,}167}{0{,}985} = 0{,}17$.

$\text{DOR} = \frac{150\times256}{30\times4} = 38.400/120 = 320$.

Interpretare. Testul are o specificitate foarte mare și un RV+ de peste 50: un rezultat pozitiv practic confirmă boala (SpPin). Sensibilitatea de 83,3% înseamnă însă că 1 din 6 bolnavi are test negativ; un rezultat negativ nu exclude boala, iar RV- de 0,17 o face doar moderat mai puțin probabilă. Valorile predictive foarte bune (97,4% și 89,5%) sunt valabile pentru prevalența de 40,9%; într-o populație de screening, cu prevalență mult mai mică, VPP ar scădea, iar VPN ar crește.

Problema rezolvată 12.2. Reconstituirea tabelului din VPP, VPN și două celule

Pentru un test se cunosc $\text{VPP} = \frac{9}{12}$, $\text{VPN} = \frac{7}{8}$, numărul de adevărat pozitivi AP = 90 și numărul de fals negativi FN = 10. Reconstituiți tabelul și calculați Se, Sp și prevalența.

Rezolvare. Pasul 1: a = AP = 90, b = FN = 10.

Pasul 2: din $\text{VPP} = \frac{a}{a+c} = \frac{9}{12}$ rezultă $\frac{90}{90+c} = \frac{9}{12}$, deci 9 x (90 + c) = 12 x 90, adică 810 + 9c = 1080, de unde c = 30.

Pasul 3: din $\text{VPN} = \frac{d}{b+d} = \frac{7}{8}$ rezultă $\frac{d}{10+d} = \frac{7}{8}$, deci 8d = 70 + 7d, de unde d = 70.

	T+	**T-**	**Total**
B+	90	10	100
B-	30	70	100
Total	120	80	200

Pasul 4 (verificare): $\text{VPP} = \frac{90}{120} = \frac{9}{12}$ și $\text{VPN} = \frac{70}{80} = \frac{7}{8}$.

Indicatori: $\text{Se} = \frac{90}{100} = 0{,}90$; $\text{Sp} = \frac{70}{100} = 0{,}70$; prevalența = 100/200 = 0,50; $\text{AC} = \frac{160}{200} = 0{,}80$; $\text{RE} = \frac{10+30}{200} = 0{,}20$; $\text{RV}^+ = \frac{0{,}90}{0{,}30} = 3{,}0$; $\text{RV}^- = \frac{0{,}10}{0{,}70} = 0{,}14$; J = 0,60.

Interpretare. Testul este mai bun la excludere decât la confirmare. Prevalența de 50% este tipică pentru un design caz-control cu grupuri egale, ceea ce înseamnă că VPP de 75% nu se transferă într-o populație reală cu prevalență mică.

Problema rezolvată 12.3. Reconstituirea tabelului din Se, VPP și două celule

Se cunosc $\text{Se} = \frac{2}{3}$, $\text{VPP} = \frac{6}{8}$, FN = 30 și AN = 80. Reconstituiți tabelul și calculați Sp și VPN.

Rezolvare. b = 30, d = 80.

Din $\text{Se} = \frac{a}{a+b} = \frac{2}{3}$: $\frac{a}{a+30} = \frac{2}{3}$, deci 3a = 2a + 60, de unde a = 60.

Din $\text{VPP} = \frac{a}{a+c} = \frac{6}{8}$: $\frac{60}{60+c} = \frac{6}{8}$, deci 6 x (60 + c) = 480, adică 360 + 6c = 480, de unde c = 20.

	T+	T-	Total
B+	60	30	90
B-	20	80	100
Total	80	110	190

Sp $= \frac{80}{100} = 0{,}80$. VPN $= \frac{80}{110} = 0{,}727$. Prevalența = 90/190 = 0,474; AC $= \frac{140}{190} = 0{,}737$; RE $= \frac{50}{190} = 0{,}263$; $RV^+ = \frac{0{,}667}{0{,}20} = 3{,}33$; $RV^- = \frac{0{,}333}{0{,}80} = 0{,}42$; DOR $= \frac{60\times80}{30\times20} = 8$.

Observație. Sensibilitatea 2/3 este 0,667, nu 0,66. Rotunjirea prematură a fracției (la 0,66) duce la a = 58,2, adică la o celulă cu număr neîntreg, semn sigur de eroare. Lucrați cu fracțiile până la final.

Problema rezolvată 12.4. Construirea tabelului din prevalență, Se și Sp

Un test rapid are Se = 85% și Sp = 90%. Construiți tabelul pentru 500 de pacienți cu prevalența bolii de 20% și calculați VPP și VPN. Refaceți calculul pentru 2000 de persoane cu prevalența 5%.

Rezolvare. Prevalența 20% din 500: a + b = 100 bolnavi, c + d = 400 nebolnavi. a = 0,85 x 100 = 85, b = 15. d = 0,90 x 400 = 360, c = 40.

	T+	T-	Total
B+	85	15	100
B-	40	360	400
Total	125	375	500

VPP $= \frac{85}{125} = 68{,}0\%$; VPN $= \frac{360}{375} = 96{,}0\%$; AC $= \frac{445}{500} = 89{,}0\%$; RE $= \frac{55}{500} = 11{,}0\%$.

La prevalența 5%, din 2000: 100 de bolnavi (a = 85, b = 15) și 1900 de nebolnavi (d = 0,90 x 1900 = 1710, c = 190). VPP $= \frac{85}{85+190} = \frac{85}{275} = 30{,}9\%$; VPN $= \frac{1710}{1725} = 99{,}1\%$.

Interpretare. Sensibilitatea și specificitatea nu s-au schimbat. La scăderea prevalenței de la 20% la 5%, VPP a scăzut de la 68,0% la 30,9%, iar VPN a crescut de la 96,0% la 99,1%. Acuratețea a crescut ușor (de la 89,0% la (85 + 1710)/2000 = 89,8%), pentru că specificitatea, mai mare decât sensibilitatea, cântărește mai mult când predomină sănătoșii; încă un motiv să nu ne bazăm pe acuratețe.

Problema rezolvată 12.5. Reconstituire din Se, VPN, FN și FP

Se cunosc Se = 0,80, VPN = 0,95, FN = 20 și FP = 60. Reconstituiți tabelul.

Rezolvare. b = 20, c = 60. Din Se $= \frac{a}{a+20} = 0{,}80$: a = 0,80a + 16, deci 0,20a = 16, a = 80. Din VPN $= \frac{d}{20+d} = 0{,}95$: d = 19 + 0,95d, deci 0,05d = 19, d = 380.

	T+	T-	Total
B+	80	20	100
B-	60	380	440
Total	140	400	540

Sp $= \frac{380}{440} = 86{,}4\%$; VPP $= \frac{80}{140} = 57{,}1\%$; prevalența = 100/540 = 18,5%; AC $= \frac{460}{540} = 85{,}2\%$; RE $= \frac{80}{540} = 14{,}8\%$; $RV^+ = \frac{0{,}80}{0{,}136} = 5{,}87$; $RV^- = \frac{0{,}20}{0{,}864} = 0{,}23$.

Problema rezolvată 12.6. Evaluarea completă a FINDRISC ≥ 12 pe setul mg2026

Folosind tabelul 12.3, formulați întrebarea, calculați indicatorii cu intervale de încredere și interpretați rezultatul pentru un medic de familie care lucrează cu o populație în care prevalența diabetului nediagnosticat este 5%.

Rezolvare. (a) Întrebarea de cercetare: cât de bine identifică scorul FINDRISC, la pragul ≥ 12, adulții cu diabet nediagnosticat, comparat cu TTGO? (b) Design: studiu transversal de acurateţe diagnostică, la 160 de adulți cu risc cardiometabolic, la care s-au efectuat ambele teste. (c) Indicatori: Se 75,9% (IC 95% 57,9-87,8); Sp 74,8% (IC 95% 66,7-81,5); VPP 40,0% (IC 95% 28,1-53,2); VPN 93,3% (IC 95% 86,9-96,7); acuratețea 75,0%; RV+ 3,01 (IC 95% 2,10-4,31); RV- 0,32 (IC 95% 0,17-0,62); DOR 9,33. (d) La prevalența de 5%: șansa pre-test 0,0526; după un scor pozitiv, 0,0526 x 3,01 = 0,158, deci VPP 13,7%; după un scor negativ, 0,0526 x 0,32 = 0,017, deci probabilitatea post-test 1,7% (VPN 98,3%), exact valorile din tabelul 12.4.

Interpretare clinică. Într-o populație cu prevalența 5%, din 100 de pacienți cu FINDRISC ≥ 12, aproximativ 14 au diabet; restul de 86 vor face un test de laborator și vor primi un rezultat liniștitor. Din 100 de pacienți cu scor sub 12, aproximativ 98 nu au diabet. Scorul este util pentru a selecta pacienții care merită testați, nu pentru a pune diagnosticul. Intervalul de încredere larg al sensibilității (57,9-87,8) arată că, pe un eșantion cu 29 de bolnavi, nu știm precis câți diabetici ratează scorul.

Problema rezolvată 12.7. Curba ROC a glicemiei a jeun, construită de mână

Pe setul mg2026, pentru glicemia a jeun față de diabetul confirmat prin TTGO, la pragurile ≥ 126, ≥ 120, ≥ 115, ≥ 110, ≥ 105 și ≥ 100 mg/dL s-au obținut: AP 11, 14, 19, 24, 26, 29 (din 29 de bolnavi) și FP 7, 15, 28, 39, 54, 68 (din 131 de nebolnavi). Construiți punctele curbei ROC, calculați aria prin metoda trapezelor și comparați-o cu AUC raportată de program (0,844).

Rezolvare. Calculăm $\text{Se} = \frac{\text{AP}}{29}$ și $1 - \text{Sp} = \frac{\text{FP}}{131}$.

Prag (mg/dL)	Se	1 - Sp	Sp
≥ 126	0,379	0,053	0,947
≥ 120	0,483	0,115	0,885
≥ 115	0,655	0,214	0,786
≥ 110	0,828	0,298	0,702
≥ 105	0,897	0,412	0,588
≥ 100	1,000	0,519	0,481

Adăugăm (0; 0) și (1; 1) și calculăm trapezele (lățime x înălțimea medie): 0,053 x 0,190 = 0,0101; 0,061 x 0,431 = 0,0263; 0,099 x 0,569 = 0,0565; 0,084 x 0,741 = 0,0623; 0,115 x 0,862 = 0,0987; 0,107 x 0,948 = 0,1013; 0,481 x 1,000 = 0,4809. Suma: 0,836.

Interpretare. Aria aproximativă, 0,836, este puțin sub valoarea exactă de 0,844, din același motiv ca la FINDRISC: punctele rare "taie colțurile" curbei. Indicele Youden maxim dintre pragurile din tabel este la ≥ 110 mg/dL (0,828 + 0,702 - 1 = 0,530). La pragul de diagnostic ≥ 126 mg/dL, sensibilitatea este doar 37,9%: pragul clinic este un prag de confirmare, nu de screening.

Problema rezolvată 12.8. Coeficientul kappa pentru doi medici de familie

Doi medici de familie au interpretat independent 100 de electrocardiograme pentru criteriile de hipertrofie ventriculară stângă. Ambii au considerat criteriile prezente la 20 de ECG-uri și absente

la 65. Medicul A a spus "prezent", iar medicul B "absent" la 10 ECG-uri; invers, la 5. Calculați kappa.

Rezolvare. Tabelul de acord:

	B: prezent	**B: absent**	**Total**
A: prezent	20	10	30
A: absent	5	65	70
Total	25	75	100

Acordul observat: $\text{Po} = \frac{20+65}{100} = 0{,}85$.

Acordul așteptat din întâmplare: A spune "prezent" în 30% din cazuri, B în 25%; probabilitatea ca ambii să spună "prezent" întâmplător este 0,30 x 0,25 = 0,075. Probabilitatea ca ambii să spună "absent" întâmplător este 0,70 x 0,75 = 0,525. Pe = 0,075 + 0,525 = 0,60.

$\kappa = \frac{0{,}85-0{,}60}{1-0{,}60} = \frac{0{,}25}{0{,}40} = 0{,}625$.

Interpretare. Pe scala Landis și Koch, acordul este substanțial (0,61-0,80). Observați că acordul brut de 85% pare excelent, dar 60 de puncte procentuale din el ar fi apărut și din întâmplare, dată fiind frecvența mare a ECG-urilor normale. Dintre cele 15 dezacorduri, 10 sunt în aceeași direcție (A vede mai des hipertrofie), ceea ce sugerează o diferență sistematică de criteriu între cei doi medici, care merită discutată.

Problema rezolvată 12.9. Reconstituirea tabelului dintr-un articol: CEUS-PAT pentru ciroză

Lupușoru și colaboratorii au raportat, la 37 de pacienți cu ciroză și 27 de voluntari sănătoși, pentru CEUS-PAT la pragul > 0,7: Se 89,19%, Sp 100%, VPP 100%, VPN 87,1% [2]. Reconstituiți tabelul și calculați RV+, RV-, acuratețea și VPP la o prevalență a cirozei de 10% printre pacienții cu hepatopatie cronică.

Rezolvare. a + b = 37, c + d = 27. Din $\text{Se} = \frac{a}{37} = 0{,}8919$, a = 33, deci b = 4. Din $\text{Sp} = \frac{d}{27} = 1$, d = 27, c = 0. Verificare: $\text{VPP} = \frac{33}{33} = 100\%$; $\text{VPN} = \frac{27}{4+27} = \frac{27}{31} = 87{,}1\%$.

	CEUS-PAT > 0,7	**CEUS-PAT ≤ 0,7**	**Total**
Ciroză	33	4	37
Sănătoși	0	27	27
Total	33	31	64

$\text{AC} = \frac{60}{64} = 93{,}8\%$. $\text{RV}^- = \frac{0{,}108}{1} = 0{,}11$. RV+ nu se poate calcula (împărțire la 0, deoarece 1 - Sp = 0); formal este infinit. Tot formal, VPP ar fi 100% la orice prevalență.

Interpretare. O specificitate de 100% observată pe 27 de voluntari sănătoși nu înseamnă că testul nu dă niciodată pozitive false. Intervalul de încredere Wilson pentru 27/27 are limita inferioară în jur de 87,5%, deci specificitatea reală ar putea fi semnificativ mai mică, mai ales la pacienți cu hepatopatii necirotice, absenți din eșantion. Aici, VPP "100% la orice prevalență" este o iluzie produsă de un eșantion mic și de un spectru favorabil. Recomandarea corectă pentru un astfel de rezultat este validarea pe pacienți consecutivi cu boală hepatică cronică.

Problema rezolvată 12.10. Evaluarea unui clasificator de inteligență artificială

Un algoritm pentru detectarea retinopatiei diabetice a fost testat pe 1000 de retinografii, dintre care 80 cu retinopatie (confirmată de doi oftalmologi). Algoritmul a clasificat corect 72 dintre imaginile cu retinopatie și 874 dintre cele fără. Calculați matricea de confuzie, sensibilitatea

(recall), specificitatea, precizia (VPP), scorul F1 și acuratețea. Comparați acuratețea cu cea a unui "algoritm" care declară toate imaginile normale.

Rezolvare. AP = 72, FN = 8, AN = 874, FP = 920 - 874 = 46.

	Algoritm: retinopatie	**Algoritm: normal**	**Total**
Retinopatie	72	8	80
Fără retinopatie	46	874	920
Total	118	882	1000

Sensibilitatea (recall) = 72/80 = 90,0%. Specificitatea = 874/920 = 95,0%. Precizia = 72/118 = 61,0%. $F_1 = \frac{2\times0{,}610\times0{,}900}{0{,}610+0{,}900} = \frac{1{,}098}{1{,}510} = 0{,}727$. Acuratețea = (72 + 874)/1000 = 94,6%.

Un "algoritm" care declară totul normal ar avea AP = 0, AN = 920, acuratețea 920/1000 = 92,0%, sensibilitatea 0 și F1 0.

Interpretare. Acuratețea de 94,6% este doar cu 2,6 puncte procentuale peste cea a unui clasificator inutil; ea nu este un indicator suficient la clase dezechilibrate. Sensibilitatea de 90% și specificitatea de 95% descriu mai bine algoritmul. Precizia de 61,0% spune că aproape 4 din 10 trimiteri la oftalmolog vor fi pozitive false, un cost care trebuie cântărit la planificarea programului de screening.

12.12 Exerciții propuse

1. Un test a fost aplicat la 200 de pacienți: 45 de bolnavi cu test pozitiv, 5 bolnavi cu test negativ, 30 de nebolnavi cu test pozitiv și 120 de nebolnavi cu test negativ. Construiți tabelul și calculați prevalența, Se, Sp, VPP, VPN, AC, RE, J, RV+, RV- și DOR.
2. Pentru un test se cunosc VPP $= \frac{4}{5}$, VPN $= \frac{9}{10}$, AP = 80 și FN = 20. Reconstituiți tabelul și calculați Se, Sp și N.
3. Pentru un test se cunosc Sp $= \frac{3}{4}$, VPN $= \frac{15}{16}$, AN = 150 și AP = 40. Reconstituiți tabelul și calculați Se și VPP.
4. Un test are Se = 75% și Sp = 95%. Construiți tabelul pentru 1000 de pacienți cu prevalența 10% și calculați VPP și VPN.
5. Folosind datele din studiul lui Moleriu și colaboratorii (problema rezolvată 11.3): 18 leziuni de grad înalt și 46 fără, HPV cu 17 pozitive printre bolnave și 27 negative printre celelalte, p16/Ki-67 cu 12 și 39, construiți cele două tabele și calculați pentru fiecare test acuratețea și indicele Youden [131]. Care test este mai potrivit pentru excludere și care pentru confirmare?
6. Un marker are, la cinci praguri, punctele (1 - Sp; Se): (0,05; 0,40), (0,10; 0,60), (0,20; 0,80), (0,40; 0,90), (0,70; 0,98). Calculați aria de sub curba ROC prin metoda trapezelor (adăugați punctele (0; 0) și (1; 1)) și încadrați-o pe scala din curs.
7. Doi radiologi au citit 200 de radiografii pulmonare pentru pneumonie. Ambii au spus "pneumonie" la 30, ambii "fără pneumonie" la 145; radiologul A a spus "pneumonie", iar B "fără" la 15, iar invers la 10. Calculați kappa și interpretați-l pe scala Landis și Koch.
8. Albai și colaboratorii au raportat pentru ISARIC-4C, la 159 de pacienți cu mortalitate 25,15%, Se 87,80% și Sp 73,72% [75]. (a) Reconstituiți aproximativ tabelul 2 x 2 (numărul de decese, AP, FN, FP, AN, rotunjite la întregi). (b) Calculați VPP din tabelul

reconstituit. (c) Calculați intervalul de încredere Wilson aproximativ pentru sensibilitate și comparați-l cu intervalul raportat (87,66-87,93).

9. Un algoritm de clasificare a leziunilor cutanate a fost testat pe 500 de imagini, dintre care 50 de melanoame. Algoritmul a identificat corect 40 de melanoame și 360 de leziuni benigne. Calculați sensibilitatea, specificitatea, precizia, F1 și acuratețea și comentați.
10. Danila și colaboratorii au raportat AUC 0,624 pentru dNLR și 0,752 pentru ALRI la pacienții cu COVID-19 și diabet [105]. Încadrați cele două valori pe scala din curs și explicați de ce o AUC semnificativ diferită de 0,5 ($p < 0,001$) nu garantează utilitatea clinică.
11. Sipos și colaboratorii au raportat pentru IL-6 în sepsisul neonatal Se 80,1% și Sp 85,7% [126]. Calculați VPP și VPN pentru o prevalență a sepsisului de 10% printre nou-născuții testați.
12. În setul mg2026, glicemia a jeun ≥ 126 mg/dL are Se 37,9% și Sp 94,7% față de TTGO. Calculați RV+ și RV- și explicați de ce acest prag este potrivit pentru confirmare, dar nu pentru screening.

Capcane frecvente

- Calcularea ratei erorii cu formula (b + d)/N. Corect: $\text{RE} = \frac{b+c}{N} = \frac{\text{FN}+\text{FP}}{N} = 1 - \text{AC}$.
- Inversarea rândurilor și coloanelor: în unele programe boala este pe coloane, iar b și c își schimbă locul. Verificați ce conține fiecare celulă.
- Raportarea VPP și VPN dintr-un studiu caz-control ca valori clinice; prevalența din eșantion este fixată de design.
- Rotunjirea fracțiilor (2/3 = 0,66) înainte de reconstituirea tabelului.
- Considerarea acurateței ca indicator suficient, mai ales când boala este rară.
- Interpretarea suprapunerii intervalelor de încredere ale AUC drept dovadă de echivalență sau absența suprapunerii drept dovadă de diferență; comparația se face cu metoda DeLong.
- Alegerea pragului pe același eșantion pe care se raportează performanța, fără validare.
- Raportarea unei specificități de 100% fără interval de încredere, la un eșantion mic de martori sănătoși.

De reținut

- Notațiile din curs: a = AP, b = FN, c = FP, d = AN; $\text{Se} = \frac{a}{a+b}$, $\text{Sp} = \frac{d}{c+d}$, $\text{VPP} = \frac{a}{a+c}$, $\text{VPN} = \frac{d}{b+d}$, $\text{AC} = \frac{a+d}{N}$, $\text{RE} = \frac{b+c}{N}$.
- Se și Sp nu depind de prevalență (în principiu), dar depind de spectrul pacienților; VPP și VPN depind de prevalență și se recalculează cu teorema lui Bayes.
- SnNout: un test foarte sensibil, negativ, exclude boala; SpPin: un test foarte specific, pozitiv, o confirmă.
- J = Se + Sp - 1; $\text{RV}^{+} = \frac{\text{Se}}{1-\text{Sp}}$; $\text{RV}^{-} = \frac{1-\text{Se}}{\text{Sp}}$; $\text{DOR} = \frac{a \times d}{b \times c}$.
- Curba ROC: Se în funcție de 1 - Sp pentru toate pragurile; AUC = probabilitatea ca un bolnav ales la întâmplare să aibă valoarea mai mare decât un sănătos ales la întâmplare.
- Scala din curs: 0,91-1 excelent; 0,81-0,90 foarte bun; 0,71-0,80 bun; 0,61-0,70 satisfăcător; sub 0,6 slab. AUC se raportează cu IC 95%.
- Pe mg2026: FINDRISC AUC 0,852 (IC 95% 0,762-0,943), glicemia a jeun 0,844, diferență nesemnificativă ($p = 0,81$); prag Youden ≥ 11 (Se 93,1%, Sp 66,4%).
- Pragul se alege după scop: excludere (Se și VPN mari), confirmare (Sp și VPP mari), cu o zonă gri între ele.

- Kappa măsoară acordul peste întâmplare pentru categorii (scala Landis și Koch); ICC pentru măsurători numerice.
- Clasificatorii AI se evaluează cu aceiași indicatori (recall = Se, precizie = VPP, F1); raportarea urmează STARD 2015 și STARD-AI.

Întrebări de autoevaluare

1. (CS) Rata erorii de clasificare a unui test, cu notațiile din curs (a = AP, b = FN, c = FP, d = AN), este:
   a) $\frac{a+d}{N}$
   b) $\frac{b+d}{N}$
   c) $\frac{b+c}{N}$
   d) $\frac{b}{a+b}$
   e) $\frac{c}{c+d}$
2. (CS) Un test cu sensibilitatea 99% iese negativ la un pacient cu probabilitate pre-test moderată. Concluzia cea mai potrivită este:
   a) boala este confirmată
   b) boala devine foarte puțin probabilă (SnNout)
   c) rezultatul nu aduce nicio informație
   d) trebuie calculată specificitatea pentru a interpreta un rezultat negativ, sensibilitatea fiind irelevantă
   e) valoarea predictivă pozitivă este 99%
3. (CS) Care dintre următorii indicatori depinde cel mai direct de prevalența bolii în populația testată?
   a) sensibilitatea
   b) specificitatea
   c) indicele Youden
   d) valoarea predictivă pozitivă
   e) raportul de verosimilitate pozitiv
4. (CS) Aria de sub curba ROC a scorului FINDRISC pentru diabet este 0,852. Interpretarea corectă este:
   a) 85,2% dintre pacienți sunt clasificați corect
   b) sensibilitatea scorului este 85,2%
   c) un pacient cu diabet ales la întâmplare are un scor mai mare decât unul fără diabet ales la întâmplare în aproximativ 85,2% din perechi
   d) valoarea predictivă pozitivă este 85,2%
   e) 85,2% dintre pacienții cu scor pozitiv au diabet
5. (CS) Doi evaluatori obțin un coeficient kappa de 0,55. Pe scala Landis și Koch, acordul este:
   a) ușor
   b) acceptabil
   c) moderat
   d) substanțial
   e) aproape perfect
6. (CM) Despre biasul de spectru:

a) apare frecvent în studiile caz-control care compară bolnavi tipici cu voluntari sănătoși
b) apare când pacienții din studiu diferă de cei la care testul va fi folosit în practică
c) apare când standardul de referință se aplică doar pacienților cu test pozitiv
d) poate modifica atât sensibilitatea, cât și specificitatea
e) se elimină prin creșterea numărului de pacienți din studiu

7. (CM) Despre curba ROC:
   a) are pe axa verticală sensibilitatea și pe axa orizontală 1 - specificitatea
   b) diagonala corespunde unui test fără valoare discriminativă (AUC = 0,5)
   c) fiecare punct al curbei corespunde unui prag de decizie
   d) aria de sub curbă depinde direct de prevalența bolii în eșantion
   e) punctul cu indicele Youden maxim este cel mai îndepărtat vertical de diagonală
8. (CM) Pentru tabelul AP = 22, FN = 7, FP = 33, AN = 98:
   a) sensibilitatea este 75,9%
   b) valoarea predictivă pozitivă este 40,0%
   c) rata erorii de clasificare este (7 + 98)/160
   d) raportul șanselor de diagnostic este aproximativ 9,33
   e) specificitatea este 98/105
9. (CM) Despre evaluarea clasificatorilor de inteligență artificială:
   a) precizia (engl. precision) este echivalentă cu valoarea predictivă pozitivă
   b) recall este echivalent cu sensibilitatea
   c) acuratețea este un indicator suficient când clasa pozitivă este rară
   d) scorul F1 este media armonică a preciziei și a sensibilității
   e) ghidul STARD-AI cere descrierea separării datelor de antrenare, validare și testare
10. (CM) Despre compararea a două curbe ROC:

a) pentru două teste efectuate la aceiași pacienți, metoda DeLong ține cont de corelația dintre AUC-uri
b) suprapunerea intervalelor de încredere ale AUC dovedește că testele sunt echivalente
c) diferența FINDRISC - glicemie (0,008; p = 0,81) în setul mg2026 arată că nu s-a găsit o diferență semnificativă de discriminare
d) o diferență nesemnificativă între AUC-uri demonstrează echivalența testelor
e) două curbe ROC se pot intersecta, astfel încât fiecare test să fie mai bun la alte praguri

Capitolul 13. Studiile clinice: clasificare, proiectare, protocol, instrumente de culegere a datelor și analiza integrată a datelor biomedicale

Obiective

La finalul acestui capitol, studentul va putea:

- să deosebească un studiu clinic intervențional de unul observațional și să încadreze un studiu într-o fază de dezvoltare (0-IV);
- să recunoască designul unui studiu (grupuri paralele, încrucișat, factorial, pe clustere, pragmatic sau explicativ, pilot cu un singur braț) și consecințele lui asupra analizei statistice;
- să explice randomizarea simplă, pe blocuri și stratificată, ascunderea alocării și orbirea, cu argumentul statistic din spatele fiecăreia;
- să definească criteriul principal și criteriile secundare, să deosebească un criteriu surogat de unul clinic "dur" și să calculeze dimensiunea eșantionului pentru o medie sau o proporție;
- să compare analiza după intenția de tratament cu analiza per protocol și să explice de ce datele lipsă nu se "șterg" pur și simplu;
- să descrie structura unui protocol, obligațiile de înregistrare și cadrul etic și legal (Declarația de la Helsinki, GCP, Regulamentul UE 536/2014, GDPR);
- să proiecteze un formular de culegere a datelor și un chestionar, să aleagă tipul de întrebare potrivit fiecărei variabile și să interpreteze indicatorii de fidelitate și validitate (alfa Cronbach, ICC, KMO, analiza factorială);
- să construiască un formular electronic în Google Forms, cu reguli de validare, secțiuni și ramificare, și să curețe un set de date brut;
- să scrie planul de analiză al unui studiu randomizat și să-l aplice pe setul mg2026, de la "Tabelul 1" până la concluzie și limite, raportând conform CONSORT 2025.

13.1 Ce este un studiu clinic și ce îi cere statisticii

Până acum am lucrat cu date gata culese. Am descris variabile, am estimat intervale de încredere, am testat ipoteze, am măsurat asocieri și am judecat performanța unui test diagnostic. În acest capitol mergem cu un pas înapoi, acolo unde se nasc datele: în proiectul unui studiu. Un rezultat statistic impecabil calculat pe date prost culese rămâne un rezultat prost. Reciproc, un studiu bine gândit face ca analiza să fie aproape mecanică.

Un studiu clinic este o cercetare pe oameni în care investigatorul atribuie, după un plan stabilit dinainte, o intervenție (un medicament, o procedură, un program educațional, o aplicație mobilă) și urmărește efectul ei asupra unor criterii de evaluare a sănătății. Cuvântul care contează este "atribuie". Într-un studiu observațional, cercetătorul doar privește: pacienții fumează sau nu, sunt obezi sau nu, primesc un tratament pentru că așa a decis medicul curant. Într-un studiu intervențional, expunerea este decisă de protocol, ideal prin tragere la sorți. Această diferență schimbă totul în interpretare: din studiile observaționale tragem concluzii de asociere, din studiile randomizate bine făcute ne permitem concluzii cauzale [63,64].

De ce are nevoie un studiu clinic de statistică încă de la început? Pentru că trebuie să știm, înainte de a include primul pacient, câți pacienți ne trebuie (capitolul 4), ce ipoteză testăm și cu ce prag de semnificație (capitolul 5), ce test vom folosi pentru criteriul principal (capitolele 6-8), cum vom exprima efectul (diferență de medii, risc relativ, NNT, capitolul 10) și cum vom trata

pacienții care abandonează. Toate aceste decizii se scriu în protocol și în planul de analiză statistică. Dacă le luăm după ce am văzut datele, riscăm să alegem, conștient sau nu, varianta care "iese" semnificativă. Ghidurile metodologice ale disciplinei insistă exact pe acest punct: planul precede datele [18,111].

În ierarhia dovezilor din medicina bazată pe dovezi, studiul randomizat controlat (engl. randomised controlled trial, RCT) stă deasupra studiilor de cohortă, caz-control și transversale, sub sinteza sistematică a mai multor RCT. Ierarhia nu este însă o scară rigidă. Un RCT mic, cu multe abandonuri și fără orbire, poate oferi dovezi mai slabe decât o cohortă mare și atent ajustată. De aceea medicul care citește un articol trebuie să judece calitatea concretă a studiului, nu doar eticheta de pe copertă. Sistemele moderne de gradare a certitudinii dovezilor pornesc de la design, dar apoi coboară sau urcă nivelul în funcție de riscul de eroare sistematică, imprecizie și inconsecvență [128].

Pe tot parcursul capitolului ne vom sprijini pe setul didactic mg2026. El simulează un studiu randomizat pragmatic din 12 cabinete de medicină de familie: 160 de adulți cu risc cardiometabolic, repartizați în trei brațe (A, consiliere standard; B, consiliere și aplicație mobilă de automonitorizare; C, consiliere, aplicație și teleconsultație lunară), urmăriți la 0, 3 și 6 luni. Îl vom privi, pe rând, cu ochii celui care proiectează studiul, ai celui care culege datele și ai celui care le analizează. La final vom scrie, împreună, planul integrat de analiză.

13.2 Clasificarea studiilor clinice

13.2.1 Intervențional sau observațional

Prima întrebare pe care o pui când deschizi un articol este: cine a decis expunerea? Dacă investigatorul a decis, studiul este intervențional (experimental). Dacă expunerea s-a produs independent de studiu, este observațional: cohortă, caz-control, transversal (capitolul 10, figura 10.1). Studiile observaționale pot fi prospective sau retrospective, analitice sau descriptive, dar niciunul nu are puterea randomizării de a echilibra factorii de confuzie necunoscuți.

Există și forme intermediare. Un studiu nerandomizat cu două grupuri, în care pacienții au ales singuri tipul de monitorizare sau au fost repartizați după disponibilitatea aparaturii, este intervențional în sensul că intervenția e aplicată după un protocol, dar seamănă cu o cohortă în privința comparabilității grupurilor. Un studiu înainte-după fără grup de control este și el intervențional, dar nu poate separa efectul tratamentului de evoluția naturală a bolii, de regresia spre medie sau de efectul placebo. Le vom întâlni pe toate în studiile de caz din acest capitol.

13.2.2 Fazele dezvoltării unui medicament

Pentru medicamente, studiile clinice urmează un traseu standardizat, prezentat în figura 13.1. Înainte de om există etapa preclinică (laborator, modele animale). Uneori se adaugă o fază 0, cu microdoze administrate unui număr foarte mic de voluntari, pentru a verifica dacă substanța se comportă în organismul uman așa cum sugerau studiile preclinice; nu are scop terapeutic și nu răspunde la întrebări de eficacitate.

Faza I include de obicei 20-80 de voluntari (sănătoși sau, în oncologie, pacienți) și urmărește siguranța, tolerabilitatea, farmacocinetica și doza maximă tolerată. Statistica este predominant descriptivă: frecvențe ale efectelor adverse, medii și intervale pentru concentrații. Faza II, cu 100-300 de pacienți, caută semnale de eficacitate și doza optimă; aici apar primele comparații cu un grup de control și primele estimări ale mărimii efectului, folosite apoi la calculul dimensiunii eșantionului pentru faza III. Faza III este marele studiu randomizat, cu sute sau mii de pacienți, care compară noul tratament cu standardul sau cu placebo pe un criteriu clinic relevant; pe rezultatele ei se acordă autorizația de punere pe piață. Faza IV începe după autorizare și urmărește

efecte adverse rare, utilizarea în populații neincluse în studiile anterioare și eficiența în practica reală (farmacovigilență).



Figura 13.1. Etapele dezvoltării unui medicament: preclinic (laborator, animale), faza I (20-80 de voluntari, siguranță și doză), faza II (100-300 de pacienți, eficacitate inițială), faza III (sute-mii de pacienți, RCT față de standard), faza IV (după autorizare, farmacovigilență).

Logica statistică a fazelor este cea a unei pâlnii. Un efect advers cu frecvența de 1 la 1000 are mari șanse să nu apară deloc într-un studiu de faza III cu 600 de pacienți. Regula aproximativă "a lui 3" spune că, dacă nu am observat niciun eveniment la n pacienți, limita superioară a intervalului de încredere de 95% pentru frecvența lui este aproximativ 3/n; pentru n = 600, asta înseamnă 0,5%. Absența unui efect advers în studiile de autorizare nu dovedește că el nu există, ci doar că este, probabil, mai rar de atât. De aici necesitatea fazei IV.

Studiile care nu testează medicamente (programe educaționale, aplicații mobile, intervenții chirurgicale, dispozitive) nu respectă întotdeauna această numerotare, dar urmează aceeași progresie: studiu de fezabilitate sau pilot, apoi studiu comparativ de mari dimensiuni. Scenariul mg2026 ar corespunde unui studiu de tip faza III pentru o intervenție nefarmacologică: compară strategii deja considerate sigure, pe un criteriu de eficacitate, în condiții apropiate de practica reală.

13.2.3 După obiectiv: superioritate, non-inferioritate, echivalență

Majoritatea studiilor pe care le veți citi sunt studii de superioritate: vor să arate că noul tratament este mai bun decât comparatorul. Ipotezele sunt cele din capitolul 5: H0, nu există diferență; H1, există o diferență. Respingerea H0 susține superioritatea (dacă diferența are sensul așteptat).

Uneori nu dorim un tratament mai eficient, ci unul la fel de eficient, dar mai ieftin, mai comod sau cu mai puține efecte adverse. Atunci formulăm un studiu de non-inferioritate. Stabilim dinainte o marjă de non-inferioritate Δ, cea mai mare pierdere de eficacitate pe care o considerăm clinic acceptabilă. Noul tratament este declarat non-inferior dacă limita intervalului de încredere a diferenței (nou minus standard) nu depășește -Δ. Un exemplu: dacă teleconsultația ar înlocui vizitele fizice, am putea accepta o scădere a glicemiei cu cel mult 3 mg/dL mai mică decât prin vizite fizice; dacă IC 95% al diferenței este, de pildă, -2,1 la 4,5 mg/dL, limita inferioară -2,1 este deasupra lui -3 și concluzionăm non-inferioritatea. Într-un studiu de echivalență cerem ca întregul interval să stea între -Δ și +Δ.

Observați că aici intervalul de încredere, nu valoarea p, este cel care decide. Un studiu de superioritate cu rezultat nesemnificativ nu demonstrează echivalența: "absența dovezii nu este dovada absenței" [79]. Dacă studiul a fost mic, intervalul este larg și include atât efecte importante în favoarea tratamentului, cât și efecte importante împotriva lui. Doar un studiu proiectat de la

început pentru non-inferioritate, cu marja justificată clinic și cu un eșantion suficient, poate susține o astfel de concluzie.

13.2.4 După design

Figura 13.2 rezumă patru designuri pe care le întâlniți frecvent. Fiecare are o consecință directă asupra testului statistic ales.

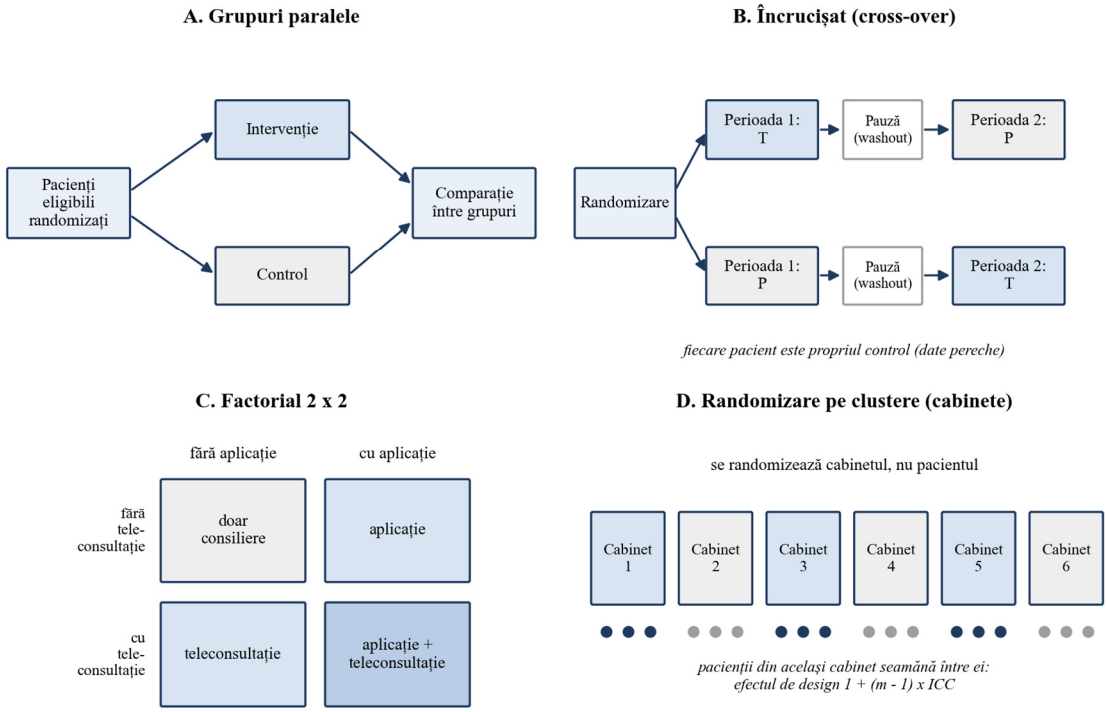


Figura 13.2. Patru designuri de studiu clinic. A: grupuri paralele, fiecare pacient primește o singură intervenție și se compară grupurile între ele. B: design încrucișat (cross-over), fiecare pacient primește ambele tratamente în ordine aleatorie, separate de o perioadă de pauză, deci datele sunt pereche. C: design factorial 2 x 2, două intervenții testate simultan. D: randomizare pe clustere, se randomizează cabinetul, iar pacienții din același cabinet seamănă între ei.

Grupuri paralele. Fiecare pacient este alocat unui singur braț și rămâne acolo până la final. Este designul studiului mg2026 și al majorității RCT. Grupurile sunt independente, deci folosim testul t pentru eșantioane independente sau Mann-Whitney la două brațe, ANOVA sau Kruskal-Wallis la trei sau mai multe brațe, chi-pătrat pentru proporții.

Design încrucișat (cross-over). Fiecare pacient primește, pe rând, ambele tratamente, ordinea fiind tras la sorți; între perioade se lasă o pauză (engl. washout) pentru ca efectul primului tratament să dispară. Fiecare pacient este propriul control, variabilitatea dintre pacienți dispare din comparație și eșantionul necesar scade mult. Fișierul BPOC folosit la lucrările practice ale disciplinei provine dintr-un studiu de acest tip, cu măsurători la începutul studiului și după medicație. Analiza folosește teste pentru date pereche: t pereche sau Wilcoxon. Designul funcționează doar pentru boli cronice stabile și pentru tratamente cu efect reversibil; nu se poate folosi pentru o intervenție care vindecă boala sau pentru un criteriu precum decesul. Riscul specific este efectul rezidual (engl. carry-over): dacă pauza a fost prea scurtă, tratamentul din prima perioadă contaminează a doua.

Design factorial. Testează simultan două (sau mai multe) intervenții. Într-un design 2 x 2 pacienții sunt randomizați în patru grupuri: nimic, doar A, doar B, A și B. Cu același număr de pacienți răspundem la două întrebări. Studiul mg2026 nu este factorial: brațul C adaugă teleconsultația peste aplicație, dar nu există un braț "doar teleconsultație". De aceea nu putem

estima efectul teleconsultației fără aplicație. Un design factorial ar fi permis acest lucru și ar fi permis testarea unei interacțiuni (dacă aplicația și teleconsultația se potențează sau nu).

Randomizare pe clustere. Uneori nu putem randomiza pacienți individuali. Dacă medicul de familie învață să facă teleconsultații, va tinde să le aplice tuturor pacienților lui, inclusiv celor alocați la consiliere standard; apare contaminarea dintre brațe. Soluția este randomizarea cabinetelor (clusterelor): toți pacienții unui cabinet primesc aceeași intervenție. Studiul mg2026 ar fi putut fi proiectat așa, cu cele 12 cabinete randomizate câte 4 în fiecare braț. Prețul este statistic. Pacienții din același cabinet seamănă între ei (același medic, aceeași zonă, același stil de consiliere), deci nu aduc informație independentă. Corelația intraclasă (ICC) măsoară această asemănare, iar efectul de design arată cu cât trebuie mărit eșantionul:

efectul de design = 1 + (m - 1) x ICC

unde m este numărul mediu de pacienți pe cluster. Cu 13 pacienți pe cabinet și ICC = 0,05, efectul de design este 1 + 12 x 0,05 = 1,6: avem nevoie de 60% mai mulți pacienți decât într-un studiu randomizat individual. Analiza trebuie să țină cont de clustere; dacă le ignorăm, erorile standard ies prea mici și valorile p prea optimiste.

Pragmatic sau explicativ. Un studiu explicativ întreabă "poate funcționa intervenția în condiții ideale?": criterii de includere stricte, pacienți motivați, aderență monitorizată atent, adesea un singur centru specializat. Un studiu pragmatic întreabă "funcționează intervenția în practica obișnuită?": criterii largi, cabinete obișnuite, aderență cât se poate de reală, rezultate relevante pentru pacient. Studiul mg2026 este pragmatic, de aceea efectele lui sunt, probabil, mai mici decât ar fi fost într-un studiu explicativ, dar mai aproape de ce va vedea medicul de familie. Majoritatea studiilor reale se plasează undeva între cele două extreme.

Studiul pilot cu un singur braț. Înainte de un RCT costisitor, cercetătorii verifică fezabilitatea: se pot recruta pacienții, acceptă ei intervenția, sunt instrumentele de măsură aplicabile, cât de mare pare efectul? Un studiu pilot cu un singur braț aplică intervenția tuturor și compară valorile din timp. Analiza folosește teste pentru date pereche sau pentru măsurători repetate (t pereche, Wilcoxon, ANOVA cu măsurători repetate, Friedman). Limita este cea a oricărui studiu înainte-după: fără grup de control, nu putem atribui cu certitudine îmbunătățirea intervenției.

Studiu de caz: aplicația mobilă la pacienții cu astm și pilotul cu un singur braț în dermatita atopică

Două studii ale colegilor din disciplină ilustrează extremele de mai sus. Munteanu și colaboratorii au testat o aplicație mobilă pentru aderența la tratament la 93 de pacienți cu astm, într-un studiu pilot controlat cu două grupuri (tratament standard față de tratament plus aplicație) și măsurători repetate pe parcursul unui an [94]. Designul seamănă izbitor cu brațele A și B din mg2026. Analiza respectă structura datelor: grupurile independente au fost comparate cu testul Mann-Whitney ($p < 0,001$), evoluția în timp a fost analizată cu testul Friedman pentru eșantioane dependente, cu corecție Bonferroni pentru comparațiile multiple ($p < 0,001$), iar proporția exacerbărilor a fost comparată cu chi-pătrat: 10,30% față de 46,30%, $\chi^2(1) = 13,707$, $p < 0,001$. Întrebarea pe care ne-o punem la lectură este cum au fost alocați pacienții în cele două grupuri; dacă nu au fost randomizați, diferența poate reflecta și motivația celor care au acceptat aplicația.

La celălalt capăt, Miulescu și colaboratorii au urmărit 21 de copii cu dermatită atopică și disbioză confirmată în laborator, într-un studiu pilot prospectiv cu un singur braț, cu evaluări la includere, la 30 și la 90 de zile [91]. Scorul POEM a scăzut de la 14,67 ± 5,22 la 7,10 ± 3,82, iar SCORAD de la 41,66 ± 15,28 la 17,93 ± 11,68 (ambele $p < 0,001$), analizate cu ANOVA cu măsurători repetate, testul potrivit pentru trei momente de timp la aceiași copii. Aproximativ

> 76% dintre copii au obținut o scădere de cel puțin 50% a POEM. Rezultatul este promițător, dar designul nu permite concluzia cauzală: dermatita atopică are o evoluție fluctuantă, copiii sunt incluși adesea într-un moment de vârf al bolii (regresie spre medie), iar părinții care știu că s-a aplicat o terapie nouă pot raporta simptomele mai favorabil. Concluzia corectă este cea pe care o formulează chiar titlul: un studiu pilot, care justifică un RCT.

13.3 Principiile metodologice ale unui studiu randomizat

13.3.1 Randomizarea

Randomizarea înseamnă alocarea pacienților în brațe printr-un mecanism aleator, pe care nici pacientul, nici medicul nu îl pot influența. Efectul ei este statistic: pe termen lung, toate caracteristicile pacienților, măsurate sau nu, cunoscute sau nu, se distribuie în medie la fel în brațe. Diferențele care rămân între grupuri la includere sunt datorate întâmplării, iar întâmplarea este exact ceea ce testele statistice știu să cuantifice. De aceea un efect semnificativ într-un RCT bine făcut poate fi atribuit intervenției.

Randomizarea simplă alocă fiecare pacient independent, ca la aruncarea unei monede (sau a unui zar, pentru trei brațe). Pe eșantioane mari funcționează bine. Pe eșantioane mici poate produce dezechilibre de mărime: cu 30 de pacienți și două brațe, probabilitatea ca un braț să primească 10 pacienți sau mai puțin nu este neglijabilă. Randomizarea pe blocuri corectează acest lucru. Alegem blocuri de, de exemplu, 6 pacienți pentru trei brațe și, în fiecare bloc, permutăm la întâmplare secvența AABBCC. Un bloc ar putea fi B A C C A B, următorul C B A B A C. La finalul fiecărui bloc, brațele au exact același număr de pacienți. Dezavantajul: dacă medicul știe mărimea blocului, poate ghici ultima alocare din bloc. De aceea se folosesc blocuri de mărime variabilă (3, 6 sau 9, alese și ele la întâmplare).

Randomizarea stratificată se folosește când un factor de prognostic important trebuie să fie echilibrat sigur, nu doar în medie. Se creează câte o listă de randomizare (de obicei pe blocuri) pentru fiecare strat: de exemplu, pentru fiecare cabinet sau pentru diabet confirmat da/nu. În mg2026, o stratificare pe cabinet ar garanta că fiecare medic are pacienți în toate trei brațele, deci "efectul medicului" nu se confundă cu efectul intervenției. Nu stratificăm după prea mulți factori: cu 12 cabinete, două sexe și trei categorii de fumat obținem 72 de straturi, multe cu câte un pacient, iar stratificarea își pierde sensul.

13.3.2 Ascunderea alocării

Secvența de randomizare poate fi perfectă și totuși studiul să fie compromis. Dacă medicul care include pacienții vede lista și știe că următorul pacient va merge în brațul C, poate amâna includerea unui pacient "dificil" până la o alocare care îi convine. Ascunderea alocării (engl. allocation concealment) înseamnă că alocarea unui pacient devine cunoscută doar după ce pacientul a fost inclus definitiv. Metodele sigure sunt randomizarea centralizată, prin telefon sau printr-o platformă web, și plicurile opace, sigilate, numerotate secvențial (mai vulnerabile, dar acceptabile dacă sunt gestionate corect). A nu se confunda cu orbirea: ascunderea alocării protejează momentul includerii și este întotdeauna posibilă; orbirea protejează perioada de după alocare și nu este întotdeauna posibilă.

13.3.3 Orbirea

Orbirea (engl. blinding) înseamnă că unul sau mai mulți participanți la studiu nu știu ce intervenție a primit fiecare pacient: pacientul, medicul care tratează, persoana care evaluează rezultatul, analistul statistic. Termenii "simplu orb" și "dublu orb" sunt folosiți inconsecvent în literatură, de aceea recomandarea actuală este să se precizeze explicit cine a fost orb, nu doar eticheta [142]. Pacientul care știe că primește tratamentul activ poate raporta simptome mai bune;

medicul care știe poate interpreta mai favorabil un rezultat clinic ambiguu; evaluatorul poate măsura cu altă atenție.

În studiul mg2026 pacienții nu pot fi orbi: știu dacă au primit aplicația și teleconsultațiile. Medicii de familie nici ei. Ce se poate face este ca glicemia să fie determinată de un laborator care nu cunoaște brațul, iar analistul să primească datele cu brațele codificate (X, Y, Z). Criteriile obiective (glicemia de laborator, greutatea pe un cântar calibrat) sunt mai puțin vulnerabile la lipsa orbirii decât criteriile subiective (satisfacția, durerea, calitatea vieții). Satisfacția din mg2026 trebuie deci interpretată cu prudență: pacienții din brațul C au primit mai multă atenție și sunt, firesc, mai mulțumiți.

13.3.4 Grupul de control și placeboul

Fără grup de control nu putem separa efectul intervenției de evoluția naturală, de regresia spre medie, de efectul placebo și de efectul simplei participări la un studiu (pacienții observați își schimbă comportamentul, fenomen numit efect Hawthorne). Controlul poate fi placebo (o substanță inactivă identică ca aspect), tratamentul standard, nicio intervenție, sau o listă de așteptare. Etica limitează folosirea placeboului: când există un tratament eficient dovedit, pacienții din grupul de control trebuie să-l primească, iar noul tratament se adaugă peste el sau se compară cu el. În mg2026, brațul A primește consilierea standard; ar fi neetic să lăsăm pacienți cu prediabet fără nicio recomandare de stil de viață.

13.3.5 Criteriile de evaluare

Criteriul principal (engl. primary endpoint) este variabila pe care se bazează concluzia studiului și pentru care s-a calculat dimensiunea eșantionului. Trebuie să fie unul singur, definit precis (ce se măsoară, cum, când) și ales înainte de începerea studiului. În mg2026 el este scăderea glicemiei a jeun între includere și 6 luni (glicemie0 - glicemie6). Criteriile secundare aduc informație suplimentară: scăderea în greutate de cel puțin 5%, scăderea tensiunii, satisfacția. Rezultatele lor sunt considerate de susținere sau generatoare de ipoteze, pentru că, testând multe criterii, crește probabilitatea ca unul să iasă semnificativ doar din întâmplare. Cu 10 criterii independente, testate fiecare la α = 0,05, probabilitatea de a obține cel puțin un rezultat fals pozitiv este 1 - 0,95^10 = 0,40.

Un criteriu surogat este un marker biologic care ține locul unui rezultat clinic important: glicemia sau HbA1c în loc de complicațiile diabetului, colesterolul LDL în loc de infarct, tensiunea arterială în loc de accident vascular cerebral. Surogatele permit studii mai mici și mai scurte, dar istoria medicinei are destule exemple de medicamente care au îmbunătățit surogatul și au înrăutățit rezultatul clinic. Criteriile "dure" (deces, infarct, accident vascular cerebral, spitalizare) sunt cele care contează pentru pacient, dar cer eșantioane mari și urmărire lungă. Un criteriu compus reunește mai multe evenimente (de exemplu, deces cardiovascular, infarct sau AVC) pentru a crește numărul de evenimente; el trebuie citit cu atenție, pentru că efectul poate fi dat aproape în întregime de componenta cea mai puțin gravă.

13.3.6 Dimensiunea eșantionului și puterea

Am învățat în capitolul 4 că dimensiunea eșantionului depinde de patru elemente [42,53]: pragul α (de obicei 0,05), puterea dorită 1 - β (de obicei 80% sau 90%), mărimea efectului pe care vrem să-l detectăm și variabilitatea criteriului. Pentru compararea a două medii, formula aproximativă pe grup este

$$n = \frac{2 \times \left(z\left(1 - \frac{\alpha}{2}\right) + z(1 - \beta) \right)^2 \times \sigma^2}{\Delta^2}$$

unde Δ este diferența minimă relevantă clinic și σ abaterea standard. Exprimat prin mărimea efectului standardizată $d = \frac{\Delta}{\sigma}$, avem $n = \frac{2\times(1{,}96+0{,}84)^2}{d^2}$. Pentru d = 0,5 formula dă aproximativ 63, iar calculul exact cu distribuția t, făcut în programe precum G*Power [77], dă 64 de pacienți pe grup; pentru d = 0,3, 176 pe grup. Pentru două proporții p1 și p2:

$$n = \frac{\left[z\left(1-\frac{\alpha}{2}\right)\times\sqrt{2\times\bar{p}\times(1-\bar{p})} + z(1-\beta)\times\sqrt{p_1\times(1-p_1)+p_2\times(1-p_2)}\right]^2}{(p_1-p_2)^2}$$

cu $\bar{p} = \frac{p_1+p_2}{2}$. La rezultatul obținut adăugăm o rezervă pentru abandonuri: dacă estimăm că 10% dintre pacienți vor fi pierduți, împărțim n la 0,9. Problema rezolvată 13.1 arată calculul complet pentru mg2026.

Un studiu prea mic este inutil și, mai grav, neetic: expune pacienți la riscurile cercetării fără o șansă rezonabilă de a răspunde la întrebare. Un rezultat nesemnificativ într-un studiu cu putere mică nu ne spune aproape nimic despre absența efectului.

13.3.7 Intenţia de tratament şi analiza per protocol

În orice studiu real, unii pacienți nu respectă protocolul: nu instalează aplicația, nu se prezintă la teleconsultații, schimbă brațul, abandonează. Avem două mari opțiuni de analiză. Analiza după intenția de tratament (engl. intention-to-treat, ITT) analizează fiecare pacient în brațul în care a fost randomizat, indiferent ce a primit efectiv. Analiza per protocol (PP) păstrează doar pacienții care au respectat protocolul.

ITT păstrează beneficiul randomizării. Pacienții care nu aderă la tratament nu sunt o submulțime aleatorie: de obicei sunt mai bolnavi, mai puțin motivați sau au efecte adverse. Dacă îi eliminăm din brațul activ, dar nu și pe cei comparabili din brațul de control, grupurile rămase nu mai sunt comparabile. ITT răspunde la întrebarea practică "ce se întâmplă dacă recomand această strategie?", și tinde să subestimeze efectul biologic, pentru că diluează efectul cu pacienții neaderenți. În studiile de superioritate, această subestimare este conservatoare, deci ITT este analiza principală. În studiile de non-inferioritate, diluarea ar favoriza concluzia de non-inferioritate, de aceea se cer ambele analize și se speră ca ele să concorde. În mg2026 nu există pierderi din urmărire (figura 13.3), deci ITT coincide cu analiza tuturor celor 160 de pacienți; pe datele reale lucrurile nu stau niciodată așa.

13.3.8 Datele lipsă şi pierderile din urmărire

Un pacient care nu vine la vizita de 6 luni lasă o celulă goală în tabel. Cea mai comodă soluție, analiza cazurilor complete (ștergerea pacientului), este corectă doar dacă lipsa este complet aleatorie, adică nu are legătură nici cu caracteristicile pacientului, nici cu rezultatul. Rar se întâmplă așa. Pacientul care nu a slăbit și se simte vinovat nu mai vine la control; dacă îl ștergem, brațul lui pare mai eficient decât este.

Se descriu trei mecanisme. Lipsa complet aleatorie: datele lipsesc din motive fără legătură cu studiul (o eprubetă spartă). Lipsa aleatorie condiționată de variabile observate: probabilitatea lipsei depinde de lucruri pe care le-am măsurat (de exemplu, vârstnicii lipsesc mai des), și putem corecta folosindu-le. Lipsa nealeatorie: probabilitatea lipsei depinde de valoarea care lipsește (cei cu glicemia crescută nu mai vin); aici nicio metodă nu corectează complet. Metodele moderne (imputarea multiplă, modelele mixte) depășesc nivelul acestui curs. Ce trebuie să rețineți: raportați câte date lipsesc în fiecare braț și de ce; nu înlocuiți valoarea lipsă cu media grupului (reduce artificial variabilitatea); înlocuirea cu ultima valoare observată este o soluție veche, care presupune că pacientul a rămas "înghețat" și poate distorsiona rezultatul; faceți o analiză de sensibilitate, cu ipoteze diferite despre cei pierduți, și verificați dacă concluzia rămâne aceeași.

13.3.9 Analizele intermediare

În studiile lungi, un comitet independent de monitorizare a datelor și a siguranței analizează datele la momente prestabilite. Poate opri studiul mai devreme pentru beneficiu evident, pentru efecte adverse sau pentru inutilitate (când devine improbabil ca studiul să arate vreodată un efect). Fiecare analiză intermediară este încă un test statistic și, dacă am folosi de fiecare dată pragul 0,05, riscul total de eroare de tip I ar crește. De aceea se folosesc praguri mult mai stricte la analizele intermediare (de ordinul $p < 0,001$), păstrând pentru analiza finală un prag apropiat de 0,05. Studiile oprite timpuriu pentru beneficiu tind să supraestimeze efectul, pentru că s-au oprit tocmai când întâmplarea a produs o diferență mare.

Studiu de caz: un studiu controlat cu placebo raportat după CONSORT și o cohortă nerandomizată cu putere mică

Gherbon și colaboratorii au comparat, timp de 3 luni, 20 de persoane cu obezitate care au primit dietă și Aloe ferox (2 capsule pe zi, 460 mg) cu 20 de persoane potrivite (engl. matched) care au primit dietă și placebo [85]. Studiul este raportat explicit conform CONSORT, ceea ce înseamnă că cititorul găsește în articol fluxul pacienților, modul de alocare și definiția criteriilor. Rezumatul raportează două serii de valori p: pentru IMC ($p = 0,03$), colesterol total ($p = 0,032$), LDL ($p = 0,01$) și glicemia a jeun ($p = 0,018$), respectiv pentru IMC ($p = 0,002$), LDL ($p = 0,039$), glicemia a jeun ($p < 0,001$) și tensiunea diastolică ($p = 0,006$). La lectură, primul lucru de clarificat este care serie corespunde comparației între grupuri la 3 luni și care comparației înainte-după în grupul tratat. Doar comparația între grupuri testează efectul Aloe ferox; comparația înainte-după din grupul activ include și efectul dietei, pe care l-au primit ambele grupuri. Al doilea lucru: cu 20 de pacienți pe grup, studiul are putere pentru efecte mari; multe criterii testate simultan cresc riscul unui rezultat fals pozitiv, deci criteriul principal trebuie identificat în protocol.

Popovici și colaboratorii au comparat, într-o cohortă prospectivă de 100 de pacienți supuși ERCP, sedarea ghidată de monitorizarea entropiei cu monitorizarea standard [78]. Grupurile nu au fost randomizate. Episoadele hipertensive au fost mai rare în grupul cu entropie (6% față de 26%, $p = 0,007$), iar scorurile Aldrete au fost mai bune la 5 minute ($p = 0,003$) și la 15 minute ($p < 0,001$). Autorii raportează onest că puterea statistică a fost limitată (29,3%). Lecția este dublă: o putere de 29,3% înseamnă că, dacă efectul căutat ar exista, studiul ar avea doar aproximativ o șansă din trei să-l detecteze, deci un rezultat nesemnificativ pentru criteriul principal nu exclude efectul; iar absența randomizării lasă loc factorilor de confuzie (cine a decis ce monitorizare primește pacientul?). Rezultatele sunt, corect formulat, generatoare de ipoteze pentru un RCT.

13.4 Protocolul, înregistrarea și cadrul etic

13.4.1 Protocolul

Protocolul este documentul care descrie studiul înainte ca el să înceapă. Funcționează ca un contract între cercetători, pacienți, comisia de etică și cititorii viitori: tot ce se face în studiu trebuie să fi fost anunțat acolo, iar orice abatere trebuie justificată. Pentru protocoalele de studii clinice există un ghid internațional de conținut, SPIRIT, construit în oglindă cu ghidul de raportare CONSORT: ce se promite în protocol se raportează apoi în articol. Colegii din disciplină au publicat și ei protocoale de studiu, de exemplu protocolul de traducere în limba română a trei chestionare de alfabetizare în sănătate din studiul ALSATION [26], ceea ce arată că protocolul nu este rezervat studiilor cu medicamente.

Un protocol complet conține, în esență, următoarele elemente: titlul și identificatorul de înregistrare; justificarea (ce se știe, ce nu se știe, de ce studiul este necesar); obiectivele și

ipotezele, formulate în format PICO (populație, intervenție, comparator, rezultat) [24]; designul (paralel, încrucișat, cluster; superioritate sau non-inferioritate); locul desfășurării; criteriile de includere și de excludere; intervențiile, descrise suficient de precis ca alt centru să le poată reproduce; criteriul principal și criteriile secundare, cu momentul măsurării; calendarul vizitelor; calculul dimensiunii eșantionului, cu toate ipotezele lui; metoda de randomizare, ascunderea alocării și orbirea; metodele de culegere și de management al datelor; planul de analiză statistică (populațiile de analiză ITT și PP, testele, tratarea datelor lipsă, analizele de subgrup prestabilite, analizele intermediare); monitorizarea siguranței; aspectele etice, consimțământul și confidențialitatea; politica de publicare și de partajare a datelor.

Pentru un student de anul I, cea mai utilă parte de exersat este traducerea întrebării clinice în elemente statistice. În mg2026: P, adulți de 35-79 de ani cu risc cardiometabolic din cabinete de medicină de familie; I, consiliere plus aplicație (B) sau consiliere plus aplicație plus teleconsultație (C); C, consiliere standard (A); O principal, scăderea glicemiei a jeun la 6 luni. Ipoteza statistică principală devine H0: $\mu A = \mu B = \mu C$ pentru scăderea glicemiei, iar testul prevăzut este ANOVA unifactorială cu comparații Tukey, respectiv Kruskal-Wallis dacă presupunerile ANOVA nu sunt îndeplinite.

13.4.2 Înregistrarea studiului

Un studiu clinic se înregistrează într-un registru public înainte de includerea primului pacient. Cele mai cunoscute sunt ClinicalTrials.gov (registrul american, folosit și de mulți cercetători europeni) și, pentru studiile cu medicamente din Uniunea Europeană, sistemul CTIS (Clinical Trials Information System), introdus prin Regulamentul UE 536/2014 [143]. Revistele care urmează recomandările ICMJE nu publică rezultatele unui studiu clinic neînregistrat prospectiv [36]. Motivul este statistic, nu birocratic: înregistrarea fixează public criteriul principal. Dacă la final autorii raportează drept "principal" un criteriu care inițial era secundar, pentru că acesta a ieșit semnificativ, comparația cu registrul dezvăluie schimbarea. Studiile cu rezultate negative, care altfel ar rămâne nepublicate, devin și ele vizibile, iar sintezele sistematice nu mai sunt distorsionate de eroarea de publicare.

13.4.3 Etica cercetării pe subiecți umani

Fundamentul etic al cercetării medicale este Declarația de la Helsinki a Asociației Medicale Mondiale, revizuită ultima dată în 2024 [35]. Principiile ei pot fi rezumate astfel: binele și drepturile participantului au prioritate față de interesul științei și al societății; riscurile trebuie să fie proporționale cu beneficiile așteptate; cercetarea trebuie justificată științific și condusă de persoane calificate; protocolul este aprobat de o comisie de etică independentă înainte de începere; participarea este voluntară și se bazează pe consimțământul informat; grupurile vulnerabile beneficiază de protecție suplimentară; folosirea placeboului este permisă doar în condiții stricte; rezultatele, inclusiv cele negative, trebuie făcute publice.

Consimțământul informat este un proces, nu doar o semnătură. Pacientul primește informații despre scopul studiului, procedurile, riscurile și beneficiile posibile, alternativele, dreptul de a se retrage oricând fără consecințe asupra îngrijirii medicale și modul în care vor fi protejate datele lui. Informațiile se dau într-un limbaj pe care pacientul îl înțelege, cu timp de gândire și posibilitatea de a pune întrebări. Pentru minori se cere acordul părinților și, în funcție de vârstă, asentimentul copilului.

Comisia de etică a cercetării (în spitalele și universitățile din România, comisia de etică a instituției) evaluează protocolul, documentele pentru pacienți, competența echipei și justificarea științifică. Din punct de vedere statistic, comisia verifică și dimensiunea eșantionului: un studiu fără putere suficientă expune pacienți la riscuri fără perspectiva unui rezultat util.

Buna practică în studiul clinic (engl. Good Clinical Practice, GCP) este standardul internațional pentru proiectarea, desfășurarea, înregistrarea și raportarea studiilor clinice; versiunea actuală a ghidului ICH E6, revizia 3, pune accent pe calitatea gândită încă din proiectare și pe abordarea bazată pe risc [144]. În Uniunea Europeană, studiile clinice cu medicamente sunt reglementate de Regulamentul 536/2014, care a armonizat procedura de autorizare și a introdus portalul unic CTIS [143]. Prelucrarea datelor personale de sănătate în cercetare intră sub incidența Regulamentului general privind protecția datelor (GDPR), care tratează datele privind sănătatea drept categorie specială, ce poate fi prelucrată în scop de cercetare doar cu garanții adecvate, precum pseudonimizarea și minimizarea datelor [32], completat în România de Legea 190/2018 [33].

Pentru analist, GDPR are consecințe practice. Fișierul de analiză nu conține nume, CNP, adresă sau număr de telefon, ci un cod de studiu (în mg2026, P001-P160). Tabelul de corespondență cod-identitate se păstrează separat, la investigatorul principal, cu acces restricționat. Se culeg doar variabilele necesare întrebării de cercetare: dacă vârsta este suficientă, nu cerem data nașterii. Fișierele nu se trimit pe e-mail necriptat și nu se păstrează pe stick-uri uitate prin buzunare. Un formular Google Forms folosit într-un studiu real cu pacienți trebuie verificat din punctul de vedere al locului unde se stochează datele și al acordului instituției; pentru exercițiile din lucrările practice folosim date fictive.

13.5 Instrumentele de culegere a datelor

13.5.1 Formularul de raportare a cazului

Formularul de raportare a cazului (engl. case report form, CRF) este documentul, pe hârtie sau electronic, în care se notează datele fiecărui pacient la fiecare vizită. Un CRF bun este gândit pornind de la planul de analiză: fiecare câmp corespunde unei variabile din dicționarul de date și fiecare variabilă din planul de analiză are un câmp în CRF. Nimic în plus, nimic lipsă.

Câteva reguli de proiectare previn majoritatea erorilor pe care le vom vedea în secțiunea 13.6. Fiecare câmp numeric are unitatea de măsură tipărită lângă el ("înălțime: ___ cm"), ca să nu primim 1,68 în loc de 168. Câmpurile categoriale au variante predefinite, bifabile, nu text liber; altfel vom avea "f", "F", "Feminin" și "femeie" pentru aceeași valoare. Datele calendaristice au un format unic (zz.ll.aaaa). Pentru fiecare câmp se prevede o modalitate de a nota de ce lipsește valoarea (nu s-a măsurat, pacientul a refuzat, nu se aplică), astfel încât o celulă goală să nu fie ambiguă. Variabilele derivate (IMC, scăderea procentuală a greutății, scorul FINDRISC) nu se calculează de mână în CRF, ci în fișierul de analiză, din variabilele primare. Primul câmp al fiecărei pagini este codul pacientului.

13.5.2 Chestionarul: de la întrebare la variabilă

Multe studii din medicina generală folosesc chestionare: simptome, calitatea vieții, aderența, satisfacția, cunoștințele, comportamentele alimentare. Proiectarea unui chestionar este o meserie, iar prima ei lecție este că tipul întrebării decide tipul variabilei și, prin el, testul statistic (capitolele 2 și 8). Tabelul 13.1 face această legătură.

Tabelul 13.1. Tipuri de întrebări și variabilele pe care le produc.

Tipul întrebării	Exemplu	Variabila rezultată	Descriere și teste uzuale
Răspuns numeric liber	"Ce vârstă aveți? ___ ani"	numerică (continuă sau discretă)	medie și AS sau mediană și IQR; t, ANOVA, corelație
Alegere unică, două variante	"Fumați în prezent? Da / Nu"	nominală dihotomică	frecvențe, %; chi-pătrat, Fisher, RR, OR
Alegere unică, variante neordonate	"Mediul de rezidență: urban / rural"	nominală	frecvențe; chi-pătrat

Tipul întrebării	Exemplu	Variabila rezultată	Descriere și teste uzuale
Alegere unică, variante ordonate	"Educația: gimnaziu / liceu / universitar / postuniversitar"	ordinală	mediană, IQR; Mann-Whitney, Kruskal-Wallis, Spearman
Item Likert	"Sunt mulțumit de program: 1 deloc... 5 foarte"	ordinală	mediană, frecvențe; teste neparametrice
Scor din mai mulți itemi	suma a 10 itemi Likert	aproape numerică	medie sau mediană, după distribuție
Alegere multiplă	"Ce boli cronice aveți? (bifați tot ce se aplică)"	câte o variabilă dihotomică pe variantă	frecvența fiecărei variante
Text liber	"Alte observații"	calitativă	analiză tematică; nu intră în testele statistice

Două greșeli apar an de an la proiectele studenților. Prima: vârsta sau greutatea culese în clase ("18-25 de ani, 26-35 de ani..."). Am transformat o variabilă numerică într-una ordinală și am pierdut informație; nu mai putem calcula media sau corelația. Regula: culegeți valoarea exactă și grupați, dacă e nevoie, la analiză. A doua: întrebări cu două idei ("Faceți sport și mâncați sănătos?"); pacientul care face sport, dar mănâncă prost nu are ce răspunde.

Formularea întrebărilor cere și ea atenție. Întrebările trebuie să fie scurte, neutre (nu "Nu-i așa că aplicația v-a ajutat?"), fără dublă negație, cu o perioadă de referință clară ("în ultimele 7 zile"), cu variante de răspuns exhaustive și care nu se suprapun ("0-1 ore, 1-2 ore" se suprapun la valoarea 1). Întrebările sensibile (venitul, consumul de alcool, viața sexuală) se pun spre final, după ce s-a creat încrederea.

13.5.3 Scalele Likert

O scală Likert propune o afirmație și un număr impar (de obicei 5 sau 7) de variante ordonate de acord sau dezacord, cu un punct neutru la mijloc. Variabila satisfacție din mg2026 este un item Likert cu 5 trepte. Un singur item Likert este o variabilă ordinală: distanța dintre "1, deloc mulțumit" și "2" nu este neapărat egală cu distanța dintre "4" și "5, foarte mulțumit". De aceea îl descriem prin mediană și frecvențe și îl comparăm cu teste neparametrice, cum am făcut în capitolul 7 (Mann-Whitney C față de A: mediane 4 față de 3, $U = 2265{,}5$, $p < 0{,}001$). Un scor obținut prin însumarea mai multor itemi Likert care măsoară același construct are mai multe valori posibile și o distribuție adesea apropiată de cea normală; în practică este tratat ca variabilă numerică, după verificarea distribuției.

Itemii formulați invers ("Mi se pare greu să folosesc aplicația", într-o scală care măsoară ușurința folosirii) se recodifică înainte de însumare: 1 devine 5, 2 devine 4 și așa mai departe. Uitarea recodificării este una dintre cele mai frecvente erori și se recunoaște ușor, pentru că itemul respectiv are corelație negativă cu scorul total.

13.5.4 Fidelitatea unui instrument

Un instrument de măsură este fidel (engl. reliable) dacă dă rezultate consecvente. Pentru chestionare se folosesc două tipuri de fidelitate.

Consistența internă arată dacă itemii unei scale măsoară același lucru. Indicatorul clasic este coeficientul alfa Cronbach:

$$\alpha = \frac{k}{k-1} \times \left(1 - \frac{\sum s^2(\text{item})}{s^2(\text{ total })}\right)$$

unde k este numărul de itemi, s^2(item) varianța fiecărui item, iar s^2(total) varianța scorului total. Dacă itemii sunt puternic corelați între ei, varianța totalului este mult mai mare decât suma varianțelor itemilor, iar α se apropie de 1. Ca reper orientativ, $\alpha \geq 0{,}70$ este considerat acceptabil

pentru cercetare, ≥ 0,80 bun, ≥ 0,90 foarte bun pentru decizii individuale; valori peste 0,95 pot semnala itemi redundanți, aceeași întrebare pusă de mai multe ori cu alte cuvinte. Coeficientul omega McDonald este o alternativă care face presupuneri mai puțin restrictive decât alfa și se raportează tot mai des alături de el. Corelația corectată item-total (corelația fiecărui item cu suma celorlalți) identifică itemii "slabi"; un prag uzual este 0,30.

Fidelitatea test-retest arată dacă același pacient, cu aceeași stare, dă aceleași răspunsuri la două aplicări separate de un interval (de obicei 1-3 săptămâni). Pentru scoruri numerice se folosește coeficientul de corelație intraclasă (ICC), nu coeficientul Pearson. Motivul: r Pearson măsoară doar dacă valorile cresc împreună; dacă la a doua aplicare toți pacienții au cu 5 puncte mai mult, r poate fi 1, deși acordul este slab. ICC penalizează și astfel de diferențe sistematice. Ca reper, ICC sub 0,50 indică fidelitate slabă, 0,50-0,75 moderată, 0,75-0,90 bună și peste 0,90 excelentă. Pentru variabile categoriale, acordul dintre două aplicări sau dintre doi evaluatori se măsoară cu coeficientul kappa [103].

13.5.5 Validitatea unui instrument

Un instrument fidel nu este automat valid. Un cântar care arată constant cu 5 kg în plus este fidel, dar nu valid. Validitatea arată dacă instrumentul măsoară ceea ce pretinde să măsoare.

Validitatea de conținut se judecă de experți și de pacienți: acoperă itemii toate aspectele importante ale construcției, sunt formulați inteligibil? Validitatea de construct verifică dacă scorurile se comportă așa cum prezice teoria, pe mai multe căi. Validitatea convergentă: scorul corelează puternic cu alte măsuri ale aceluiași construct. Validitatea discriminantă: corelează slab cu construcții diferite. Validitatea prin grupuri cunoscute (engl. known-groups): grupuri despre care știm că diferă (de exemplu, dieteticieni față de populația generală, la un chestionar de cunoștințe de nutriție) obțin scoruri diferite. Validitatea de criteriu compară scorul cu un standard de referință, ca în capitolul 12.

Structura internă a unui chestionar se studiază prin analiză factorială. Analiza factorială exploratorie (AFE, engl. EFA) caută grupuri de itemi care corelează între ei și sugerează câți factori (dimensiuni) stau în spatele răspunsurilor. Înainte de a o aplica se verifică dacă datele se pretează: indicele Kaiser-Meyer-Olkin (KMO) trebuie să fie de preferință peste 0,80 (sub 0,50 analiza nu are sens), iar testul de sfericitate Bartlett trebuie să fie semnificativ (respingem ipoteza că itemii sunt necorelați). Numărul de factori se alege cu analiza paralelă, care compară valorile proprii observate cu cele obținute din date aleatorii, sau cu alte criterii. Analiza factorială confirmatorie (AFC, engl. CFA) testează o structură stabilită dinainte și raportează indici de potrivire: CFI și TLI (de dorit peste 0,90, ideal peste 0,95), RMSEA (de dorit sub 0,08, ideal sub 0,06) și SRMR (sub 0,08). Modelarea cu ecuații structurale (SEM) merge mai departe și estimează relațiile dintre factori. Aceste metode depășesc calculul de mână al unui student de anul I, dar trebuie să le recunoașteți și să le puteți citi într-un articol.

Când un chestionar validat în altă limbă este folosit în România, el trebuie tradus și adaptat cultural, nu doar tradus. Procedura standard include două traduceri independente în română, sinteza lor, retraducerea în limba originală de către traducători care nu cunosc originalul, comparația cu originalul, testarea pe un grup mic de pacienți (interviuri cognitive: "ce ați înțeles prin această întrebare?") și apoi validarea psihometrică în populația română. Protocolul ALSATION descrie un astfel de traseu pentru trei chestionare de alfabetizare în sănătate [26].

> **Studiu de caz: validarea unor chestionare în cercetarea colegilor din disciplină**
>
> Marin și colaboratorii au dezvoltat un chestionar bazat pe modelul acceptării tehnologiei (TAM) pentru telemonitorizarea la domiciliu a pacienților cu ciroză hepatică și l-au validat pe 130 de pacienți [10]. Consistența internă a fost foarte bună (α Cronbach = 0,934; ω McDonald = 0,950). AFE a identificat o structură cu doi factori care explică 77,9% din variație, iar AFC a

confirmat-o cu indici acceptabili (CFI = 0,927; TLI = 0,908, ambii peste 0,90, dar sub 0,95). Modelul structural a arătat că utilitatea percepută prezice puternic intenția de utilizare (β = 0,823, $p < 0,001$), iar ușurința percepută influențează utilitatea (β = 0,343, $p < 0,001$); modelul a explicat 69,5% din variația intenției. Întrebarea critică pentru cititor este dacă 130 de pacienți ajung pentru AFE și AFC pe același eșantion; ideal, structura găsită exploratoriu se confirmă pe un eșantion independent.

Dragomir și colaboratorii au construit chestionarul MMBQ-44 pentru povara menopauzei, pe 308 femei de 45-65 de ani [60]. Pașii urmează exact secțiunea 13.5.5: corelațiile corectate item-total au depășit toate 0,30; datele s-au pretat la analiza factorială (KMO = 0,931; Bartlett χ^2 = 9006,2, $p < 0,001$); analiza paralelă a identificat patru factori care explică 48,6% din variația totală; consistența internă a fost α 0,813-0,941 și ω 0,864-0,953. Autorii semnalează și efecte de podea (multe răspunsuri la valoarea minimă) pentru două subscale, 37,0% și 23,4%. Informația contează: o subscală cu efect de podea nu poate detecta ameliorarea la pacientele care pornesc deja de la minim.

Dicu și colaboratorii au validat un instrument multidimensional pentru evaluarea produselor de panificație cu ingrediente alternative pe 504 consumatori, cu o subeșantionare de 30 de persoane reevaluate după 14 zile [73]. Pe lângă AFE (KMO = 0,933) și AFC (CFI = 0,965; TLI = 0,959; RMSEA = 0,061; SRMR = 0,033, toți în zona bună), studiul raportează fidelitatea test-retest prin ICC 0,82-0,91, adică fidelitate bună spre excelentă, și α Cronbach 0,898-0,938. Retestul a fost făcut însă pe doar 30 de persoane, iar intervalele de încredere ale ICC pe un subeșantion atât de mic sunt largi; valorile punctuale se citesc cu prudență.

Alte patru studii completează imaginea. Frandeș și colaboratorii au validat un instrument pentru acceptabilitatea tehnologiei mobile în autogestionarea diabetului la 103 pacienți, cu α Cronbach 0,9 pentru atitudinea față de folosirea tehnologiei mobile și 0,97 pentru o altă dimensiune, și au folosit corelații de rang (Spearman, Kendall), potrivite pentru date ordinale [54]. Putnoky și colaboratorii au adaptat un chestionar de cunoștințe generale de nutriție în patru etape: fidelitatea internă pe 412 persoane (0,878), fidelitatea test-retest pe 46 de persoane (peste 0,880), validitatea prin grupuri cunoscute pe 96 și validitatea convergentă pe 508 [145]. Șerban și colaboratorii au validat chestionarul de alimentație conștientă pe adulți români, cu fidelitate internă 0,72 și externă 0,83 [146]; valoarea 0,72 este la limita acceptabilului și ne amintește că alfa depinde și de numărul de itemi, o subscală scurtă având de regulă un alfa mai mic. În sfârșit, Cîrstea și colaboratorii au tradus și validat versiunea românească a chestionarului PSPSQ 2.0 la 220 de pacienți cu afecțiuni dermatologice cronice [11]: α = 0,978 (subscale 0,961-0,969), o structură care explică 84,0% din variație, CFI = TLI = 0,999, dar RMSEA = 0,109, valoare care indică o oarecare nepotrivire a modelului. Când indicii de potrivire se contrazic, autorii trebuie să discute de ce; un alfa de 0,978 sugerează, în plus, posibila redundanță a unor itemi.

13.5.6 Culegerea electronică a datelor

Introducerea datelor de pe hârtie în calculator este o sursă importantă de erori. Culegerea electronică (engl. electronic data capture, EDC) elimină acest pas și permite verificarea datelor chiar în momentul introducerii. În lucrările practice ale disciplinei folosim Google Forms, legat de o foaie Google Sheets în care se adună răspunsurile [46,147]. Pentru studii clinice reale se folosesc platforme specializate, precum REDCap (răspândită în mediul universitar, cu jurnal de audit și gestionarea drepturilor de acces) sau Epi Info, programul gratuit al CDC care are atât un modul de construcție a formularelor, cât și module de analiză [51].

Un formular Google Forms bine construit folosește patru facilități. Prima este tipul corect de întrebare: "Răspuns scurt" pentru valori numerice, "Variante multiple" pentru alegere unică,

"Casete de selectare" pentru alegere multiplă, "Scară liniară" pentru itemi Likert, "Listă verticală" pentru liste lungi (de exemplu, județul), "Dată" pentru date calendaristice. A doua sunt regulile de validare a răspunsului [45]: pentru vârstă, "Număr", "între 18 și 99"; pentru înălțime, "între 120 și 220" (ar fi oprit valoarea 1,68 din mg2026_brut); pentru glicemie, "Număr", fără text (ar fi oprit "112 mg/dl"); fiecare cu un mesaj de eroare explicit ("introduceți înălțimea în centimetri, de exemplu 168"). A treia sunt întrebările obligatorii pentru variabilele indispensabile (codul pacientului, brațul, criteriul principal), dar nu pentru toate: un formular în care totul este obligatoriu îl obligă pe respondent să inventeze răspunsuri. A patra sunt secțiunile și ramificarea [148]. Prima secțiune este pagina de informare și consimțământ, cu întrebarea "Sunt de acord să particip", iar varianta "Nu sunt de acord" trimite direct la finalul formularului; întrebarea "Fumați?" cu răspunsul "Da" deschide o secțiune despre numărul de țigări pe zi, iar răspunsul "Nu" o sare.

După strângerea răspunsurilor, foaia de calcul se descarcă și se verifică: un rând pe pacient, o coloană pe variabilă, nume de coloane scurte, fără spații și diacritice (glicemie0, nu "Glicemia la includere (mg/dL)"), coduri consecvente pentru categorii. Tabelul pivot din Google Sheets [44] este util pentru verificări rapide de frecvențe.

13.5.7 Dicționarul de date și testarea pilot

Dicționarul de date (engl. data dictionary sau codebook) este tabelul care descrie fiecare variabilă: numele scurt, eticheta completă, tipul, unitatea, valorile permise, codurile pentru categorii, modul de codificare a valorilor lipsă. Tabelul 13.2 reproduce un fragment din dicționarul setului mg2026.

Tabelul 13.2. Fragment din dicționarul de date al setului mg2026.

Variabila	Descriere	Tip	Valori permise / unitate
id	codul pacientului	identificator	P001-P160, unic
brat	brațul de randomizare	nominală	A, B, C
sex	sexul	nominală dihotomică	F, M
varsta	vârsta la includere	numerică	ani, 18-99
educatie	nivelul de educație	ordinală	1 gimnaziu, 2 liceu/profesională, 3 universitar, 4 postuniversitar
inaltime	înălțimea	numerică	cm, 120-220
glicemie0	glicemia a jeun la includere	numerică	mg/dL, 40-400
hba1c0	hemoglobina glicată la includere	numerică	%, 3,5-15, cu o zecimală
diabet_ogtt	diabet confirmat prin TTGO	nominală dihotomică	0 nu, 1 da
satisfactie	satisfacția la 6 luni	ordinală	1-5 (Likert)

Înainte de lansarea studiului, formularul se testează pe câțiva colegi și pe câțiva pacienți (testare pilot). Se cronometrează completarea, se verifică dacă întrebările sunt înțelese la fel de toți, dacă ramificările funcționează, dacă datele ajung corect în foaia de calcul și dacă fișierul descărcat poate fi importat direct în JASP sau MedCalc. Cinci formulare de test completate cu valori extreme (vârsta 18 și 99, câmpuri lăsate goale, răspunsul "Nu sunt de acord") descoperă mai multe probleme decât o oră de recitire.

13.6 Managementul și curățarea datelor

Oricât de bun ar fi formularul, datele brute conțin erori. Fișierul mg2026_brut.csv, pus pe Moodle alături de varianta curată, a fost construit tocmai pentru acest exercițiu: are 161 de rânduri în loc de 160 și o serie de erori tipice. Tabelul 13.3 le prezintă pe toate, împreună cu metoda prin care se detectează.

Tabelul 13.3. Erorile din fișierul mg2026_brut.csv și metoda de detectare.

Pacient	Variabila	Valoarea din fișierul brut	Tipul erorii	Cum se detectează
P005	sex	"f"	codificare inconsecventă	tabel de frecvențe: apar mai mult de 2 categorii
P021	sex	"Feminin"	codificare inconsecventă	idem
P034	tas0	1540	valoare imposibilă (un zero în plus)	minim/maxim, box-plot
P042	varsta	lipsă	valoare lipsă	număr de valori valide sub 160
P057	inaltime	1,72	unitate greșită (m în loc de cm)	minim, histogramă, IMC recalculat aberant
P064	glicemie0	"87 mg/dl"	text într-o coloană numerică	coloana nu este recunoscută ca numerică
P084	fumat	"Fumator "	majusculă și spațiu în plus	tabel de frecvențe: o categorie în plus
P089	hba1c0	"5,2" stocat ca text	separator zecimal diferit	tipul coloanei, valoare nealiniată
P102	greutate0	7,26	virgulă zecimală mutată	minim, IMC recalculat aberant
P121	mediu	"Urban"	majusculă	tabel de frecvențe
P134	glicemie6	lipsă	valoare lipsă la urmărire	număr de valori valide
P091	rând întreg	duplicat (rândurile 91 și 161)	înregistrare dublă	căutarea codurilor duplicate

Procedura de curățare are o ordine logică. Mai întâi se păstrează o copie neatinsă a fișierului brut; nu se lucrează niciodată pe original. Apoi se caută duplicatele după codul pacientului. Se verifică tipul fiecărei coloane (numerică sau text). Se fac tabele de frecvențe pentru toate variabilele categoriale, iar pentru cele numerice minimul, maximul, numărul de valori valide și un box-plot. Se verifică concordanțele logice: IMC recalculat din greutate și înălțime, scăderea în greutate compatibilă cu variabila scadere5, glicemia la 6 luni plauzibilă față de cea inițială.

Corectarea este partea delicată. O eroare de formă ("Urban", "Fumator " cu spațiu, "5,2" stocat ca text) se poate corecta fără dubiu, dacă sensul este evident. O valoare imposibilă nu se "ghicește". Tentația este să credem că 1540 mmHg era 154, că 7,26 kg era 72,6 și că 1,72 m înseamnă 172 cm. În fișierul mg2026 aceste presupuneri se confirmă, dar numai pentru că am verificat fiecare valoare în documentul sursă (fișa pacientului) și, pentru înălțime și greutate, am recalculat IMC-ul păstrat în tabel: $\sqrt{\frac{77,7}{26,3}} = 1,72m$ și 27,0 x 1,64² = 72,6 kg. În datele reale, aceeași cifră deformată poate proveni din mai multe greșeli diferite; 1540 putea fi la fel de bine 150 sau 154, iar o tensiune de 154 mmHg schimbă încadrarea pacientului ca hipertensiv. Corecția se face doar din sursă; dacă sursa nu mai este disponibilă, valoarea se declară lipsă. Valorile lipsă reale (vârsta lui P042, glicemia la 6 luni a lui P134) se recuperează din documentele sursă dacă este posibil; în caz contrar rămân lipsă și se raportează.

Fiecare modificare se notează într-un jurnal de modificări: data, cine a făcut-o, rândul, variabila, valoarea veche, valoarea nouă, motivul și documentul sursă. În platformele profesionale această evidență se ține automat și se numește jurnal de audit (engl. audit trail); ghidul GCP cere ca orice modificare a datelor să fie trasabilă și să nu șteargă valoarea inițială [144]. Pentru un proiect studențesc, o foaie separată în fișierul Google Sheets, numită "modificari", este suficientă.

Analiza statistică începe doar după ce baza de date a fost declarată finală ("înghețată"), iar orice modificare ulterioară se justifică în scris.

13.7 Raportarea: CONSORT 2025

Rezultatele unui RCT se raportează după ghidul CONSORT, actualizat în 2025 [56]. CONSORT este o listă de verificare a elementelor care trebuie să apară în articol (titlul care precizează că studiul este randomizat, înregistrarea, accesul la protocol, designul, participanții, intervențiile, criteriile, dimensiunea eșantionului, randomizarea și ascunderea alocării, orbirea, metodele statistice, fluxul participanților, caracteristicile de bază, rezultatele cu mărimea efectului și precizia ei, efectele adverse, limitele, generalizabilitatea), însoțită de o diagramă de flux. Pentru studiile observaționale ghidul corespunzător este STROBE [25], iar pentru studiile de acuratețe diagnostică STARD [140].

Diagrama de flux CONSORT arată ce s-a întâmplat cu fiecare pacient, de la evaluarea eligibilității până la analiză. Figura 13.3 prezintă diagrama pentru mg2026: 212 persoane evaluate, 52 excluse (31 nu îndeplineau criteriile, 21 au refuzat), 160 randomizate (53 în brațul A, 53 în B, 54 în C), toate urmărite la 3 și 6 luni și analizate după intenția de tratament.

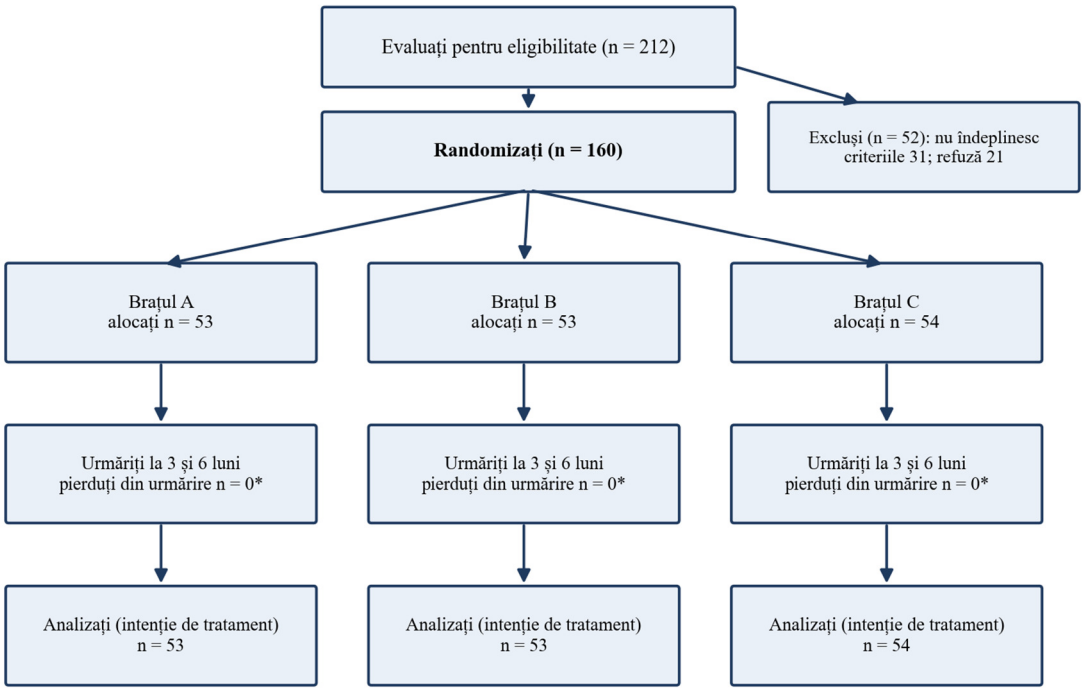


Figura 13.3. Diagrama de flux CONSORT a studiului simulat mg2026: 212 persoane evaluate pentru eligibilitate, 52 excluse (31 nu îndeplinesc criteriile, 21 refuză), 160 randomizate în trei brațe (A 53, B 53, C 54), fără pierderi din urmărire în setul didactic, toate analizate după intenția de tratament.

Diagrama răspunde la întrebări pe care cititorul trebuie să și le pună. Câți pacienți eligibili au refuzat? Dacă refuzul a fost masiv, eșantionul s-ar putea să nu fie reprezentativ: cei care acceptă un studiu cu aplicație mobilă sunt, poate, mai tineri și mai familiarizați cu tehnologia. Câți s-au pierdut în fiecare braț și din ce motive? Pierderi inegale între brațe sunt un semnal de alarmă. Câți au fost analizați, comparativ cu câți au fost randomizați? Dacă numerele diferă, analiza nu mai este ITT. Într-un studiu real, căsuța "pierduți din urmărire" nu conține aproape niciodată zero; în setul didactic am ales această simplificare pentru a nu complica exercițiile, iar nota din figură o spune explicit.

13.8 Analiza integrată a studiului mg2026: un exemplu complet

Am ajuns la partea în care toate capitolele anterioare lucrează împreună. Vom parcurge analiza studiului mg2026 în ordinea în care ar apărea într-un articol: planul, populația, caracteristicile de bază, criteriul principal, criteriile secundare, mărimile efectului, concluzia și limitele, apoi analizele observaționale care se pot face pe aceleași date. Toate cifrele provin din setul mg2026_curat.csv și pot fi reproduse în JASP 0.98.1 [47] sau MedCalc 24.0 [50], cum veți face în lucrarea practică 13, urmând pașii descriși și în culegerile de aplicații practice ale disciplinelor de biostatistică [14,149].

13.8.1 Planul de analiză statistică, scris înainte de date

Planul de analiză al studiului, redactat înainte de deblocarea codurilor brațelor, prevede următoarele. Populația de analiză principală este populația ITT (toți cei 160 de pacienți randomizați, în brațul alocat). Criteriul principal este scăderea glicemiei a jeun la 6 luni (glicemie0 - glicemie6, în mg/dL), comparată între cele trei brațe prin ANOVA unifactorială, după verificarea normalității în fiecare braț (Shapiro-Wilk) și a omogenității variațelor (Levene); comparațiile două câte două se fac cu testul Tukey; dacă presupunerile nu sunt îndeplinite, se folosește Kruskal-Wallis cu comparații Mann-Whitney corectate. Pragul de semnificație este $\alpha = 0,05$, bilateral. Criteriile secundare sunt: proporția pacienților cu scădere în greutate de cel puțin 5% (chi-pătrat, risc relativ, diferență de risc, NNT, cu IC 95%); scăderea în greutate în kg (ANOVA); scăderea TAS la 6 luni (ANOVA); satisfacția la 6 luni (Kruskal-Wallis); evoluția glicemiei la 0, 3 și 6 luni (Friedman și ANOVA cu măsurători repetate). Rezultatele secundare sunt interpretate ca susținătoare, fără corecție formală, dar cu prudență față de comparațiile multiple. Analize observaționale suplimentare, la includere: asocierea obezitate-diabet și performanța scorului FINDRISC față de TTGO.

Un astfel de paragraf nu are nimic spectaculos, dar el este cel care dă credibilitate rezultatelor. Observați că testul alternativ (Kruskal-Wallis) este și el anunțat dinainte, împreună cu regula după care se alege. Alegerea testului după ce am văzut care dă p mai mic este una dintre cele mai frecvente forme de "pescuit" de semnificație.

13.8.2 Caracteristicile de bază: Tabelul 1

Primul tabel al oricărui articol care raportează un RCT descrie pacienții la includere, pe brațe. Variabilele numerice cu distribuție aproximativ normală se descriu prin medie și abatere standard, cele asimetrice prin mediană și interval intercuartil, cele categoriale prin frecvență absolută și procent (capitolul 2). Tabelul 13.4 este "Tabelul 1" al studiului mg2026.

Tabelul 13.4. Caracteristicile pacienților la includere, pe brațe (studiul mg2026, n = 160). Medie (AS), mediană [IQR] sau n (%).

Caracteristica	Brațul A (n = 53)	Brațul B (n = 53)	Brațul C (n = 54)	Total (n = 160)
Vârsta, ani	57,5 (9,2)	56,9 (9,4)	54,5 (7,5)	56,3 (8,8)
Sex feminin	32 (60,4)	26 (49,1)	28 (51,9)	86 (53,8)
Mediu urban	25 (47,2)	36 (67,9)	34 (63,0)	95 (59,4)
Educație universitară sau postuniversitară	26 (49,1)	29 (54,7)	28 (51,9)	83 (51,9)
Fumător actual	18 (34,0)	18 (34,0)	17 (31,5)	53 (33,1)
Fost fumător	5 (9,4)	11 (20,8)	17 (31,5)	33 (20,6)
Nefumător	30 (56,6)	24 (45,3)	20 (37,0)	74 (46,3)
Activitate fizică (1-5)	3 [2-4]	3 [2-3]	3 [2-3]	3 [2-3]
IMC, kg/m²	29,2 [26,6-31,1]	27,9 [26,1-31,4]	27,7 [25,9-31,4]	28,25 [26,2-31,4]
Obezitate (IMC ≥	18 (34,0)	20 (37,7)	16 (29,6)	54 (33,8)

Caracteristica	Brațul A (n = 53)	Brațul B (n = 53)	Brațul C (n = 54)	Total (n = 160)
30)				
Circumferința taliei, cm	89,4 (12,3)	87,8 (12,1)	88,4 (12,1)	88,5 (12,1)
TAS, mmHg	131,5 (13,8)	132,8 (13,8)	129,1 (14,0)	131,1 (13,9)
TAD, mmHg	80,5 (8,7)	80,1 (8,3)	77,1 (8,6)	79,2 (8,6)
Colesterol total, mg/dL	209,3 (39,0)	202,8 (37,5)	215,9 (33,6)	209,4 (36,9)
HDL-colesterol, mg/dL	50,7 (11,5)	51,1 (12,4)	46,9 (11,2)	49,5 (11,8)
Trigliceride, mg/dL	143 [95-204]	144 [112-185]	166,5 [107-213]	151,5 [107-203,25]
Glicemie a jeun, mg/dL	107,8 (15,6)	106,3 (14,7)	102,3 (16,6)	105,4 (15,8)
HbA1c, %	5,61 (0,45)	5,57 (0,43)	5,55 (0,46)	5,58 (0,45)
Scor FINDRISC	10,7 (4,4)	10,1 (4,2)	10,3 (4,2)	10,4 (4,3)
Diabet confirmat prin TTGO	10 (18,9)	9 (17,0)	10 (18,5)	29 (18,1)

Tabelul nu are o coloană cu valori p, și asta este intenționat. CONSORT recomandă să nu se testeze statistic diferențele de la includere într-un RCT [56]. Raționamentul este simplu: dacă randomizarea a fost corectă, orice diferență este, prin definiție, datorată întâmplării, iar un test statistic ar testa o ipoteză despre care știm deja că este adevărată. Cu 20 de rânduri testate la $\alpha = 0,05$, ne așteptăm, în medie, la un rezultat "semnificativ" doar din întâmplare. Ce contează este mărimea diferenței și importanța prognostică a variabilei, nu valoarea p.

Ce citim totuși în tabel? Brațele sunt bine echilibrate pentru vârstă, IMC, tensiune, HbA1c, FINDRISC și prevalența diabetului (18,9%, 17,0% și 18,5%). Câteva diferențe sar în ochi: în brațul C sunt mai mulți foști fumători (31,5% față de 9,4% în A), brațul A are mai mulți pacienți din mediul rural (52,8%), iar glicemia de pornire este cu aproximativ 5 mg/dL mai mică în C decât în A. Ultima diferență contează pentru interpretare: un pacient care pornește de la o glicemie mai mare are, în principiu, mai mult loc de scădere. Faptul că brațul C, care pornește de jos, scade totuși cel mai mult face rezultatul mai convingător, nu mai puțin. Dacă am considera că o diferență de bază poate influența rezultatul, o analiză de sensibilitate cu ajustare (regresie liniară cu glicemia inițială și brațul ca predictori) ar trebui prevăzută în plan, nu decisă după ce am văzut Tabelul 1.

13.8.3 Criteriul principal: scăderea glicemiei la 6 luni

Descrierea pe brațe vine întotdeauna înaintea testului. Scăderea medie a glicemiei a fost 3,0 mg/dL (AS 8,8) în brațul A, 11,1 mg/dL (AS 11,2) în B și 16,3 mg/dL (AS 10,6) în C. Intervalele de încredere de 95% pentru media fiecărui braț, calculate cu distribuția t (capitolul 4), sunt 0,6-5,4 mg/dL pentru A, 8,0-14,2 pentru B și 13,4-19,2 pentru C. Deja aici vedem că intervalul brațului A nu se suprapune cu niciunul dintre celelalte.

Verificăm presupunerile ANOVA. Testul Shapiro-Wilk aplicat scăderii glicemiei în fiecare braț dă $p = 0,575$ (A), $p = 0,052$ (B) și $p = 0,903$ (C); niciunul nu respinge normalitatea, iar cu peste 50 de pacienți pe braț ANOVA este oricum robustă la abateri moderate. Testul Levene dă $p = 0,163$, deci nu respingem omogenitatea varianțelor. ANOVA este, prin urmare, testul potrivit, cum prevedea planul.

Rezultatul: $F(2; 157) = 22,93$, $p < 0,001$. Respingem H0 că scăderea medie a glicemiei este aceeași în cele trei brațe. Mărimea efectului este $\eta^2 = 0,226$: aproximativ 23% din variabilitatea scăderii glicemiei se datorează brațului de tratament. Omega², o variantă mai puțin optimistă, calculată din aceleași sume de pătrate, este 0,215. Comparațiile Tukey arată că toate cele trei perechi diferă semnificativ: B față de A, diferența $8,1$ mg/dL$(\text{IC } 95\% \; 3,4 - 12,8;\ p = 0,0002)$;

C față de A, 13,3 mg/dL (IC 95% 8,6-18,0; p < 0,001); C față de B, 5,2 mg/dL(IC 95% 0,6 − 9,9; $p = 0{,}024$). Testul Kruskal-Wallis, prevăzut ca alternativă, confirmă rezultatul (H = 36,63, p < 0,001). Partea stângă a figurii 13.4 prezintă grafic diferențele Tukey.

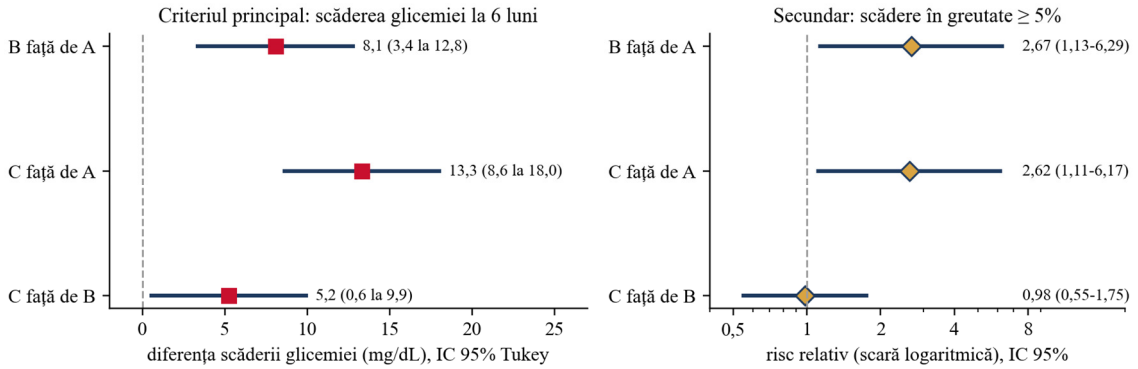


Figura 13.4. Efectele intervențiilor în studiul mg2026. Stânga: diferențele dintre brațe în scăderea glicemiei la 6 luni (mg/dL), cu intervalele de încredere de 95% Tukey; niciun interval nu include 0. Dreapta: riscul relativ de a obține o scădere în greutate de cel puțin 5%, cu IC 95%, pe scară logaritmică; intervalele pentru B și C față de A nu includ 1, cel pentru C față de B include 1.

Cum citim clinic aceste cifre? O scădere suplimentară de 13,3 mg/dL a glicemiei a jeun pentru brațul C față de consilierea standard este considerabilă la o populație cu valori medii de pornire de 105 mg/dL, adică exact în zona prediabetului. Mărimea efectului standardizată (d Cohen, capitolul 6) pentru C față de A este d = 1,37, pentru B față de A d = 0,80, pentru C față de B d = 0,48; după reperele uzuale, efecte mari, respectiv mare și moderat [76]. Diferența C față de B are însă un IC 95% care coboară până la 0,6 mg/dL, adică foarte aproape de zero: datele sunt compatibile atât cu un beneficiu suplimentar substanțial al teleconsultației, cât și cu unul aproape neglijabil. Intervalul, nu valoarea p = 0,024, ne spune cât de precisă este estimarea.

13.8.4 Criteriile secundare

Scăderea în greutate de cel puțin 5%. Au atins acest prag 6 din 53 de pacienți în brațul A (11,3%), 16 din 53 în B (30,2%) și 16 din 54 în C (29,6%). Testul chi-pătrat pentru tabelul 3 x 2 dă χ^2 = 6,77, gl = 2, p = 0,034; cea mai mică frecvență așteptată este 12,6, deci condiția de aplicare (toate frecvențele așteptate cel puțin 5) este îndeplinită. Pentru comparația C față de A, riscul relativ este 2,62 (IC 95% 1,11-6,17): probabilitatea de a slăbi cel puțin 5% a fost de 2,6 ori mai mare în brațul C decât în brațul A. Diferența absolută de risc este 18,3 puncte procentuale (IC 95% 3,4-33,2), iar numărul necesar de tratat este NNT $= \frac{1}{0{,}183} = 5{,}5$, rotunjit în sus la 6 (IC 95% 3,0-29,1): trebuie să includem 6 pacienți în programul C, în loc de consilierea standard, pentru ca încă unul să ajungă la o scădere de 5%. Pentru B față de A, RR = 2,67 (IC 95% 1,13-6,29), cu o diferență de risc de 18,9 puncte procentuale. Pentru C față de B, RR = 0,98 (IC 95% 0,55-1,75): teleconsultația nu a adus niciun câștig suplimentar pe acest criteriu. Partea dreaptă a figurii 13.4 arată aceste trei rapoarte.

Observați lățimea intervalelor. NNT-ul poate fi oriunde între 3 și 29, adică între "foarte eficient" și "modest". Cu doar 22 de evenimente în cele două brațe comparate, precizia este limitată, iar un raport onest spune acest lucru.

Scăderea în greutate în kilograme. A: $0{,}71 \pm 2{,}48$ kg; B: $2{,}84 \pm 2{,}36$ kg; C: $3{,}37 \pm 2{,}39$ kg; ANOVA F = 18,25, p < 0,001. Rezultatul continuu și cel dihotomizat (≥ 5%) spun aceeași poveste, iar varianta continuă păstrează mai multă informație. Dihotomizarea unei variabile numerice pierde putere statistică, motiv pentru care criteriul principal al unui studiu este, de preferință, variabila continuă.

Scăderea tensiunii arteriale sistolice. Aici avem un exemplu instructiv. Scăderea medie a TAS la 6 luni a fost 2,8 mmHg (AS 9,0) în A, 4,5 mmHg (AS 9,6) în B și 7,0 mmHg (AS 8,2) în C. Testul prevăzut în plan, ANOVA, dă F = 2,94, p = 0,056. Kruskal-Wallis, aplicat pe aceleași date, ar da p = 0,046. Ce raportăm? Testul din plan: diferența nu atinge pragul de semnificație statistică (p = 0,056). Să schimbăm testul după ce am văzut rezultatul, ca să "trecem" sub 0,05, ar fi exact practica descrisă mai sus ca pescuit de semnificație. Formularea corectă este că există o tendință de scădere mai mare a TAS în brațele cu aplicație, dar studiul nu a avut putere suficientă pentru acest criteriu secundar și rezultatul trebuie confirmat. Pe ansamblul celor 160 de pacienți, TAS a scăzut în medie cu 4,8 mmHg (AS 9,0; t pereche = 6,68, p < 0,001), dar această scădere globală nu poate fi atribuită intervenției, pentru că nu are comparator: ea include regresia spre medie și efectul participării la studiu.

Satisfacția. Medianele au fost 3 (IQR 2-3) în A, 4 (IQR 3-4) în B și 4 (IQR 3-5) în C; Kruskal-Wallis H = 32,16, p < 0,001; comparația C față de A cu Mann-Whitney: U = 2265,5, p < 0,001, cu o corelație biserială a rangurilor de 0,58 (efect mare). Este un criteriu subiectiv, raportat de pacienți care știu ce intervenție au primit. Diferența poate reflecta atât valoarea reală a programului, cât și simplul fapt că au primit mai multă atenție.

Evoluția glicemiei în timp. Pe întregul eșantion, mediile glicemiei au fost 105,4, 99,4 și 95,2 mg/dL la 0, 3 și 6 luni (mediane 104,5, 99,0 și 92,5); testul Friedman dă $\chi^2 = 74,67$, p < 0,001, iar ANOVA cu măsurători repetate F = 64,97, p < 0,001. Pe brațe, testul Friedman este la limită în brațul A ($\chi^2 = 6,04$, p = 0,049; medii 107,8, 105,7, 104,8 mg/dL) și puternic semnificativ în B ($\chi^2 = 21,07$, p < 0,001; medii 106,3, 101,3, 95,2) și în C ($\chi^2 = 64,63$, p < 0,001; medii 102,3, 91,2, 86,0). Atenție la o capcană: faptul că testul este semnificativ într-un braț și "mai semnificativ" în altul nu dovedește că brațele diferă între ele. Întrebarea "evoluează glicemia diferit în brațe?" este răspunsă direct de criteriul principal (comparația scăderilor între brațe) sau, mai elegant, de termenul de interacțiune timp x braț dintr-o ANOVA mixtă, metodă care depășește nivelul acestui curs.

13.8.5 Comparațiile multiple și ierarhia rezultatelor

Am făcut un singur test principal și cel puțin cinci teste secundare, plus comparațiile post-hoc. Dacă toate H0 ar fi adevărate, probabilitatea ca măcar un test din zece să iasă semnificativ doar din întâmplare ar fi de aproximativ 40%. Studiul mg2026 se apără prin ierarhie: concluzia se sprijină pe criteriul principal, ales înainte și testat o singură dată, cu comparațiile post-hoc controlate prin testul Tukey; criteriile secundare sunt consecvente cu cel principal și sunt prezentate ca susținătoare. Dacă am fi vrut ca un criteriu secundar să fie concluziv, am fi putut folosi corecția Bonferroni (pragul 0,05 împărțit la numărul de teste) sau o testare ierarhică prestabilită, în care al doilea criteriu se testează doar dacă primul a fost semnificativ.

13.8.6 Concluzia și limitele

Un paragraf de concluzie, scris în stilul unui articol, ar suna astfel. "Într-un studiu randomizat pragmatic în 12 cabinete de medicină de familie, adăugarea unei aplicații de automonitorizare la consilierea standard a crescut scăderea glicemiei a jeun la 6 luni cu 8,1 mg/dL (IC 95% 3,4-12,8), iar adăugarea suplimentară a teleconsultației lunare cu 13,3 mg/dL (IC 95% 8,6-18,0) față de consilierea standard. Proporția pacienților care au slăbit cel puțin 5% s-a dublat în brațele cu aplicație (RR 2,62 pentru brațul C față de A, IC 95% 1,11-6,17; NNT 6). Beneficiul suplimentar al teleconsultației față de aplicația singură a fost mai mic și estimat imprecis."

Limitele trebuie spuse la fel de clar. Pacienții și medicii nu au fost orbi, ceea ce afectează mai ales criteriile subiective (satisfacția). Criteriul principal este un surogat (glicemia), nu un rezultat clinic dur (incidența diabetului, evenimentele cardiovasculare). Urmărirea de 6 luni nu spune nimic despre menținerea efectului. Randomizarea a fost individuală în cabinete în care medicul

aplica toate cele trei strategii, deci contaminarea dintre brațe este posibilă și ar fi tins să micșoreze diferențele. Datele sunt complete, situație nerealistă pentru un studiu real. Iar pacienții care au acceptat să participe (160 din 181 de eligibili) pot fi mai motivați decât populația generală de pacienți cu risc cardiometabolic.

13.8.7 Analize observaționale pe aceleași date

Un set de date dintr-un RCT poate fi folosit și pentru întrebări observaționale, cu o condiție pe care mulți o uită: randomizarea protejează doar comparația dintre brațe. Orice altă asociere din același set (obezitate și diabet, vârstă și tensiune) este una observațională, transversală, expusă factorilor de confuzie ca în orice studiu de acest tip (capitolul 10).

La includere, prevalența diabetului confirmat prin TTGO a fost 33,3% la pacienții obezi (18 din 54) și 10,4% la cei neobezi (11 din 106). Raportul prevalențelor este 3,21 (IC 95% 1,64-6,31), iar raportul cotelor OR = 4,32 (IC 95% 1,86-10,03); χ^2 = 12,70, p < 0,001. Fiind un design transversal, măsura naturală este raportul prevalențelor; OR-ul îl supraestimează, cum era de așteptat când boala nu este rară. În regresia logistică ajustată pentru vârstă și sex, OR-ul obezității devine 6,01 (IC 95% 2,35-15,39; p = 0,0002), iar fiecare an de vârstă crește cota de diabet cu 11% (OR 1,11; IC 95% 1,05-1,18). Nu putem spune că obezitatea "cauzează" diabetul pe baza acestor date: nu știm ce a apărut întâi, iar alți factori necontrolați pot explica asocierea.

A doua întrebare observațională: cât de bine identifică scorul FINDRISC pacienții cu diabet confirmat prin TTGO [125]? Aria de sub curba ROC (capitolul 12) este 0,852 (IC 95% 0,762-0,943), apropiată de cea a glicemiei a jeun (0,844). Pragul care maximizează indicele Youden este FINDRISC ≥ 11 (Se 93,1%, Sp 66,4%, J = 0,595). La pragul ≥ 12, tabelul 2 x 2 are 22 adevărat pozitivi, 33 fals pozitivi, 7 fals negativi și 98 adevărat negativi: Se 75,9%, Sp 74,8%, VPP 40,0%, VPN 93,3%. Cu o prevalență de 18,1% în eșantion, un scor pozitiv crește probabilitatea de diabet doar până la 40%, dar un scor negativ o coboară sub 7%. FINDRISC funcționează deci mai bine pentru a exclude decât pentru a confirma, iar TTGO rămâne necesar la cei pozitivi. Figura 13.5 reia curba ROC din capitolul 12.

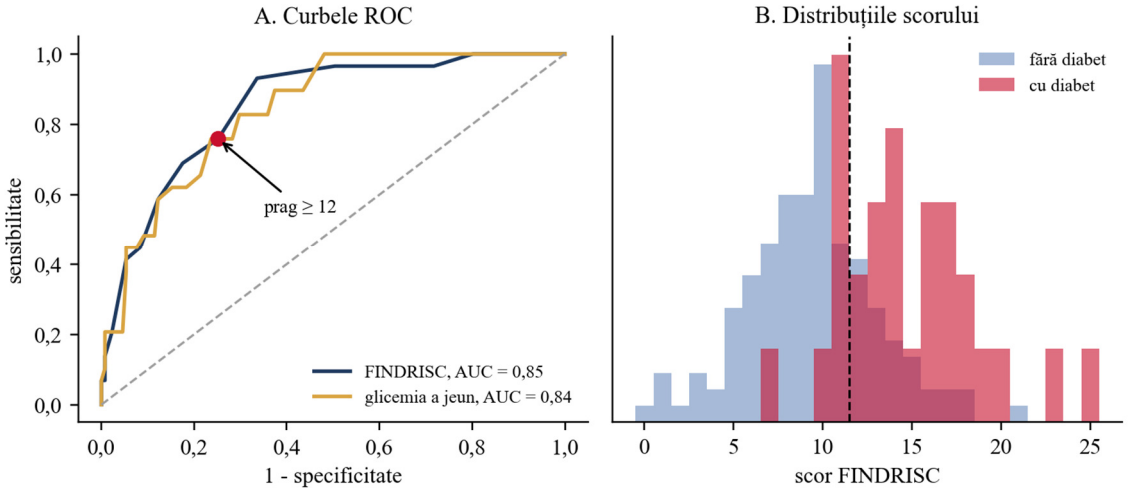


Figura 13.5. Curba ROC a scorului FINDRISC pentru diabetul confirmat prin TTGO la includere în studiul mg2026 (aria de sub curbă 0,852, IC 95% 0,762-0,943), analiză observațională transversală pe datele unui studiu randomizat.

Studiu de caz: confuzia, grupurile paralele nerandomizate și designul înainte-după

Christodorescu și colaboratorii au urmărit timp de 12 luni 278 de pacienți spitalizați cu insuficiență cardiacă decompensată, pentru a vedea dacă raportul hemoglobină/lărgimea distribuției eritrocitare (HRR) prezice evenimentele [23]. Mediana HRR a fost 0,89, iar criteriul principal a apărut la 167 de pacienți (60,1%). În analiza Kaplan-Meier, pacienții cu

HRR scăzut au avut o supraviețuire fără evenimente mai mică (log-rank p = 0,027), dar în modelul Cox ajustat pentru vârstă, sex, NT-proBNP, fracția de ejecție și rata de filtrare glomerulară asocierea nu a mai fost semnificativă (p = 0,240). Este exemplul clasic de confuzie: HRR scăzut însoțește pacienții mai vârstnici, cu funcție renală mai slabă și cu NT-proBNP mai mare, iar aceștia au oricum un prognostic mai rezervat. Într-o cohortă, ajustarea este singura apărare; într-un RCT, randomizarea ar fi echilibrat acești factori. Autorii raportează corect ambele rezultate.

Uivarosan și colaboratorii au comparat, la 70 de pacienți după accident vascular cerebral, două grupuri cu terapie combinată (tehnici moderne de recuperare și medicație neurotrofică, n = 22 și n = 18) cu un grup de control cu tratament standard (n = 30), urmăriți 12 luni [89]. Rezumatul raportează rezultate "semnificativ mai bune" ($p < 0,05$) în grupurile cu terapie combinată și ameliorări la 12 luni față de momentul inițial ($p < 0,05$), fără să precizeze testele. Designul cu grupuri paralele și control este corect ca idee, dar cititorul trebuie să afle din textul complet cum au fost formate grupurile (randomizare sau alocare după alt criteriu?), de ce au mărimi diferite și ce test a fost aplicat. Pentru trei grupuri independente ar fi potrivite ANOVA sau Kruskal-Wallis, iar pentru evoluția în timp un test pentru măsurători repetate.

Buda și colaboratorii au măsurat trombospondina-1 serică la pacienți hipertensivi cu disfuncție endotelială, înainte și după un an de tratament cu perindopril [86]. Valorile au crescut de la 8061,9 ± 3684,80 la 15380 ± 5887 ng/mL($p < 0,001$), comparativ cu $9221,03 \pm 6510,21$ ng/mL la martorii nehipertensivi, iar un alt marker a scăzut de la 0,91 ± 0,51 la $0,50 \pm 0,24$ ng/mL($p < 0,001$). Analiza potrivită este cea pentru date pereche (t pereche sau Wilcoxon, observând că abaterile standard mari față de medii sugerează asimetrie). Grupul martor de persoane sănătoase servește ca reper, nu ca grup de control al tratamentului: el nu a primit nici perindopril, nici placebo. De aceea schimbarea înainte-după nu poate fi atribuită cu certitudine medicamentului, iar numărul de pacienți, nemenționat în rezumat, este informația pe care cititorul trebuie s-o caute prima.

13.9 Probleme rezolvate

Problema rezolvată 13.1. Câți pacienți trebuie incluși în studiul mg2026?

Enunț. Echipa care proiectează studiul mg2026 consideră relevantă clinic o diferență de 6 mg/dL în scăderea glicemiei între un braț de intervenție și brațul de control, iar din studii anterioare estimează o abatere standard a scăderii de 10 mg/dL. Pentru criteriul secundar, se așteaptă ca 10% dintre pacienții din brațul A și 30% dintre cei din brațul C să slăbească cel puțin 5%. Se lucrează cu α = 0,05 (bilateral) și putere 80%. Se estimează 15% pierderi din urmărire. Câți pacienți trebuie incluși pe braț?

Rezolvare. Pentru criteriul principal folosim formula pentru două medii, cu $z\left(1-\frac{\alpha}{2}\right) = 1,96$ și z(1-β) = 0,84 (pentru 80% putere):

$$n = \frac{2 \times (1,96 + 0,84)^2 \times \sigma^2}{\Delta^2} = \frac{2 \times 7,84 \times 100}{36} = 43,6$$

deci 44 de pacienți pe braț. Mărimea efectului standardizată este $d = \frac{6}{10} = 0,6$. Dacă echipa ar fi fost mai prudentă și ar fi presupus d = 0,5, ar fi obținut 64 de pacienți pe braț (calcul exact cu distribuția t).

Pentru criteriul secundar, p1 = 0,10, p2 = 0,30, $\bar{p} = 0,20$:

$\sqrt{2 \times 0,20 \times 0,80} = \sqrt{0,32} = 0,566$; $1,96 \times 0,566 = 1,109\sqrt{0,10 \times 0,90 + 0,30 \times 0,70} = \sqrt{0,30} = 0,548$; $0,84 \times 0,548 = 0,461n = \frac{(1,109+0,461)^2}{(0,30-0,10)^2} = \frac{1,570^2}{0,04} = \frac{2,464}{0,04} = 61,6$

deci 62 de pacienți pe braț. Este mai mult decât pentru criteriul principal: criteriile binare cer, de regulă, eșantioane mai mari.

Corecția pentru pierderi: dacă vrem ca 62 de pacienți să rămână evaluabili în fiecare braț și ne așteptăm la 15% pierderi, includem 62/0,85 = 72,9, adică 73 de pacienți pe braț, aproximativ 219 în total.

Interpretare. Cu 53-54 de pacienți pe braț, studiul mg2026 are putere suficientă pentru criteriul principal (44 necesari), dar este puțin sub necesar pentru criteriul secundar binar. Exact asta se vede în rezultate: criteriul principal este foarte clar, în timp ce intervalele pentru RR sunt largi (1,11-6,17). Formula nu ține cont de faptul că avem trei brațe și trei comparații; o corecție pentru comparațiile multiple ar mări puțin eșantionul.

Problema rezolvată 13.2. Construirea unei liste de randomizare pe blocuri

Enunț. Construiți primele 12 alocări ale unei liste de randomizare pe blocuri de 6 pentru trei brațe (A, B, C), folosind următoarele numere aleatoare (generate, de exemplu, cu funcția =RAND() din Google Sheets): blocul 1: 0,71; 0,15; 0,88; 0,42; 0,07; 0,56; blocul 2: 0,33; 0,91; 0,26; 0,64; 0,48; 0,12. Câte secvențe distincte sunt posibile într-un bloc? De ce este periculoasă o mărime fixă de bloc?

Rezolvare. Fiecare bloc conține literele A, A, B, B, C, C. Asociem fiecărei litere câte un număr aleator, în ordine, și sortăm crescător după număr.

Blocul 1: A 0,71; A 0,15; B 0,88; B 0,42; C 0,07; C 0,56. Ordonat crescător: C (0,07), A (0,15), B (0,42), C (0,56), A (0,71), B (0,88). Secvența: C A B C A B.

Blocul 2: A 0,33; A 0,91; B 0,26; B 0,64; C 0,48; C 0,12. Ordonat: C (0,12), B (0,26), A (0,33), C (0,48), B (0,64), A (0,91). Secvența: C B A C B A.

Primii 12 pacienți: C, A, B, C, A, B, C, B, A, C, B, A. După fiecare 6 pacienți, fiecare braț are exact 2.

Numărul de secvențe distincte într-un bloc este numărul de permutări ale multisetului AABBCC: $\frac{6!}{2!\times 2!\times 2!} = \frac{720}{8} = 90$.

Pericolul mărimii fixe: dacă medicul care include pacienții știe că blocurile au 6 locuri și a văzut deja 5 alocări din bloc, a șasea este complet previzibilă; după 4 alocări, el știe deja în multe cazuri care sunt cele două litere rămase. Poate atunci influența includerea următorului pacient. Soluțiile sunt ascunderea alocării (medicul nu vede lista) și blocurile de mărime variabilă (3, 6 sau 9, în ordine aleatorie). Pentru stratificarea pe cabinet, se generează câte o astfel de listă pentru fiecare dintre cele 12 cabinete.

Problema rezolvată 13.3. De ce nu testăm diferențele din Tabelul 1?

Enunț. Un student propune să adauge în Tabelul 13.4 o coloană cu valori p pentru fiecare dintre cele 20 de rânduri. Observă că pentru fumat chi-pătrat dă p = 0,072, iar pentru mediu p = 0,075, și concluzionează că "randomizarea a eșuat la limită". Ce îi răspundeți?

Rezolvare. Dacă randomizarea a fost corectă, H0 "brațele provin din aceeași populație" este adevărată prin construcție pentru toate caracteristicile de bază. Orice diferență observată este o eroare de eșantionare. Un test statistic în această situație nu ne spune nimic nou: el va da "semnificativ" cu probabilitatea α, adică în aproximativ 5% din cazuri. Cu 15 teste independente la α = 0,05, probabilitatea ca măcar unul să iasă semnificativ doar din întâmplare este

$$1 - 0{,}95^{15} = 1 - 0{,}463 = 0{,}537$$

adică mai mult decât jumătate. Un p = 0,072 nu indică un "eșec la limită" al randomizării; indică o fluctuație obișnuită. Ce trebuie judecat este dacă diferența (de exemplu, 31,5% foști

fumători în C față de 9,4% în A) este suficient de mare și dacă variabila este suficient de puternic legată de rezultat încât să poată influența concluzia. Dacă da, analiza de sensibilitate cu ajustare se prevede în plan. Concluzia corectă: Tabelul 1 se citește clinic, nu prin teste.

Problema rezolvată 13.4. ANOVA calculată de mână din statisticile de grup

Enunț. Pornind doar de la statisticile descriptive ale scăderii glicemiei (A: n = 53, m = 3,02, s = 8,76; B: n = 53, m = 11,09, s = 11,21; C: n = 54, m = 16,33, s = 10,59), calculați F, eta² și d Cohen pentru C față de A.

Rezolvare. Media generală, ponderată cu mărimea grupurilor:

$$m = \frac{53 \times 3{,}02 + 53 \times 11{,}09 + 54 \times 16{,}33}{160} = \frac{160{,}06 + 587{,}77 + 881{,}82}{160} = \frac{1629{,}65}{160} = 10{,}19$$

Suma pătratelor dintre grupuri:

$$\text{SSB} = 53 \times (3{,}02 - 10{,}19)^2 + 53 \times (11{,}09 - 10{,}19)^2 + 54 \times (16{,}33 - 10{,}19)^2$$
$$= 53 \times 51{,}34 + 53 \times 0{,}82 + 54 \times 37{,}76 = 2721{,}1 + 43{,}4 + 2038{,}9 = 4803{,}4$$

Suma pătratelor din interiorul grupurilor, din abaterile standard:

$$\text{SSW} = 52 \times 8{,}76^2 + 52 \times 11{,}21^2 + 53 \times 10{,}59^2 = 3990{,}4 + 6534{,}5 + 5943{,}8 = 16468{,}7$$

Gradele de libertate: între grupuri k - 1 = 2; în interior N - k = 157. Pătratele medii: MSB = $\frac{4803{,}4}{2} = 2401{,}7$; MSW = $\frac{16468{,}7}{157} = 104{,}9$. Raportul $F = \frac{2401{,}7}{104{,}9} = 22{,}90$. Programul, care lucrează cu valorile nerotunjite, dă 22,93; diferența provine doar din rotunjiri.

$\eta^2 = \frac{\text{SSB}}{\text{SSB+SSW}} = \frac{4803{,}4}{21272{,}1} = 0{,}226$.

Pentru d Cohen C față de A, abaterea standard comună este

$$\text{sp} = \sqrt{\left[\frac{53 \times 10{,}59^2 + 52 \times 8{,}76^2}{54 + 53 - 2}\right]} = \sqrt{\frac{9934{,}2}{105}} = \sqrt{94{,}6} = 9{,}73$$

$$d = \frac{16{,}33 - 3{,}02}{9{,}73} = 1{,}37$$

Interpretare. Valoarea critică a lui F pentru 2 și 157 de grade de libertate la α = 0,05 este aproximativ 3,05; F = 22,9 este mult peste ea, deci p < 0,001. Brațul explică 22,6% din variabilitatea scăderii glicemiei, iar diferența C-A este de aproape 1,4 abateri standard, un efect mare.

Problema rezolvată 13.5. Risc relativ, diferență de risc și NNT cu intervale de încredere

Enunț. Calculați RR, diferența de risc și NNT, cu IC 95%, pentru scăderea în greutate de cel puțin 5% în brațul C (16 din 54) față de brațul A (6 din 53).

Rezolvare. Riscurile: $p_C = \frac{16}{54} = 0{,}296$; $p_A = \frac{6}{53} = 0{,}113$. RR = $\frac{0{,}296}{0{,}113} = 2{,}62$.

Intervalul pentru RR se calculează pe scară logaritmică (capitolul 10):

$\text{ES}(\ln \text{RR}) = \sqrt{\frac{1}{16} - \frac{1}{54} + \frac{1}{6} - \frac{1}{53}} = \sqrt{0{,}0625 - 0{,}0185 + 0{,}1667 - 0{,}0189} = \sqrt{0{,}1918} = 0{,}438 \ln 2{,}62 = 0{,}963$; $0{,}963 \pm 1{,}96 \times 0{,}438 = 0{,}963 \pm 0{,}858$, adică 0,105 la 1,821; IC 95% RR = $e^{0{,}105}$ la $e^{1{,}821} = 1{,}11$ la 6,18 (programul, cu valori nerotunjite, dă 1,11-6,17)

Diferența de risc: 0,296 - 0,113 = 0,183 (18,3 puncte procentuale).

$\text{ES(DR)} = \sqrt{\frac{0{,}296\times0{,}704}{54} + \frac{0{,}113\times0{,}887}{53}} = \sqrt{0{,}00386 + 0{,}00189} = 0{,}0759;\ \text{IC}\,95\% = 0{,}183 \pm 1{,}96 \times 0{,}0759 = 0{,}183 \pm 0{,}149$, adică 0,034 la 0,332

$\text{NNT} = \frac{1}{0{,}183} = 5{,}5$, rotunjit în sus la 6. Intervalul pentru NNT se obține inversând limitele intervalului diferenței de risc: 1/0,332 = 3,0 și 1/0,034 = 29 (programul, cu valori nerotunjite, dă 29,1).

Interpretare statistică. IC 95% pentru RR nu include 1, iar cel pentru diferența de risc nu include 0, deci diferența este semnificativă la α = 0,05 (concordant cu testul exact Fisher, p = 0,030). Interpretare clinică. Pentru fiecare 6 pacienți tratați în programul C în loc de consilierea standard, încă unul ajunge la o scădere de 5% a greutății; estimarea este însă imprecisă, compatibilă cu un NNT între 3 și 29.

Problema rezolvată 13.6. Alfa Cronbach calculat de mână

Enunț. O scală de 4 itemi Likert (1-5) care măsoară încrederea în folosirea aplicației mobile a fost aplicată la 40 de pacienți din brațele B și C. Variațele itemilor sunt 1,2; 1,0; 1,4 și 1,1, iar varianța scorului total (suma celor 4 itemi) este 14,2. Calculați α Cronbach și interpretați.

Rezolvare. Suma variațelor itemilor: 1,2 + 1,0 + 1,4 + 1,1 = 4,7.

$$\alpha = \frac{4}{3} \times \left(1 - \frac{4{,}7}{14{,}2}\right) = 1{,}333 \times (1 - 0{,}331) = 1{,}333 \times 0{,}669 = 0{,}892$$

Interpretare. α = 0,89 indică o consistență internă bună: itemii par să măsoare același construct. Gândiți-vă la ce s-ar fi întâmplat dacă itemii ar fi fost necorelați: varianța totalului ar fi fost aproximativ egală cu suma variațelor (4,7), iar α ar fi fost aproape 0. Cu cât itemii covariază mai mult, cu atât varianța totalului crește peste suma variațelor și α se apropie de 1. Un α mare nu demonstrează însă că scala măsoară încrederea și nu altceva (validitatea) și nici că pacienții ar răspunde la fel peste două săptămâni (fidelitatea test-retest, măsurată cu ICC).

Problema rezolvată 13.7. Intenția de tratament față de analiza per protocol

Enunț. Să presupunem, pentru exercițiu, că în brațul C al studiului 10 dintre cei 54 de pacienți nu ar fi folosit niciodată aplicația și nu s-ar fi prezentat la teleconsultații. Dintre cei 44 aderenți, 15 ar fi atins scăderea de 5%, iar dintre cei 10 neaderenți, 1. În brațul A toți pacienții ar fi primit consilierea (6 din 53 cu scădere de 5%). Calculați RR în analiza ITT și în analiza per protocol și comentați.

Rezolvare. ITT: în brațul C rămân toți cei 54 randomizați, cu 15 + 1 = 16 evenimente; $p_C = \frac{16}{54} = 29{,}6\%$; $\text{RR} = \frac{0{,}296}{0{,}113} = 2{,}62$. Per protocol: păstrăm doar cei 44 de aderenți; $p_C = \frac{15}{44} = 34{,}1\%$; $\text{RR} = \frac{0{,}341}{0{,}113} = 3{,}01$.

Comentariu. Analiza per protocol dă un efect mai mare, dar pierde comparabilitatea brațelor. Cei 10 neaderenți nu sunt aleși la întâmplare: sunt probabil mai puțin motivați, mai puțin familiarizați cu telefonul inteligent, poate mai vârstnici. În brațul A există, cu siguranță, pacienți asemănători, dar nu i-am eliminat, pentru că acolo nu exista o intervenție la care să nu adere. Comparăm deci "pacienții motivați din C" cu "toți pacienții din A", iar o parte din diferență ține de motivație, nu de intervenție. ITT răspunde la întrebarea medicului de familie: "dacă recomand acest program pacienților mei, cu aderența lor reală, ce câștig?". Analiza per protocol aproximează efectul la pacientul care aderă, dar cu riscul unei erori sistematice. Când cele două analize concordă ca sens, concluzia este robustă.

Problema rezolvată 13.8. Ce face o singură valoare greșită

Enunț. În fișierul brut, pacientul P034 are TAS0 = 1540 mmHg (valoarea reală, din fișa sursă, este 154). Calculați media, abaterea standard și mediana TAS0 în cele două variante și comentați.

Rezolvare. Pe datele corecte: media 131,13 mmHg, AS 13,88, mediana 131. Înlocuirea valorii 154 cu 1540 adaugă la suma valorilor 1540 - 154 = 1386, deci media crește cu 1386/160 = 8,66 mmHg, la 139,79 mmHg. Abaterea standard explodează la 112,24 mmHg, deoarece abaterea pătratică a acestei singure valori de la medie domină întreaga sumă. Mediana rămâne 131: valoarea extremă a mutat un singur element de sub mediană deasupra ei, ceea ce poate deplasa mediana cel mult cu o poziție, iar în jurul centrului acestui set mai multe valori sunt egale cu 131.

Interpretare. O singură eroare de tastare mută media cu peste 8 mmHg și face ca intervalul de încredere să devină absurd de larg; orice test t aplicat pe aceste date ar fi compromis. Mediana este robustă, exact cum am învățat în capitolul 2. Lecția pentru managementul datelor: minimul, maximul și un box-plot pentru fiecare variabilă numerică, înainte de orice analiză. Și încă o dată: corecția se face din fișa sursă, nu prin ghicire, chiar dacă aici sursa confirmă valoarea de 154 mmHg.

13.10 Exerciții propuse

1. Un studiu compară două programe de exerciții pentru scăderea TAS. Diferența minimă relevantă este 8 mmHg, iar abaterea standard estimată a scăderii este 12 mmHg. Calculați numărul de pacienți pe grup pentru $\alpha = 0,05$ și putere 90% ($z(1-\beta) = 1,28$). Cât devine dacă se estimează 10% pierderi?
2. Într-un studiu de vaccinare antigripală la vârstnici, proporția așteptată de pacienți cu infecție respiratorie severă este 35% fără intervenție și 20% cu un program de vaccinare la domiciliu. Calculați eșantionul pe grup pentru $\alpha = 0,05$ și putere 80%.
3. Un studiu randomizat pe clustere include cabinete cu câte 20 de pacienți. Corelația intraclasă estimată este 0,02. Calculați efectul de design și numărul de pacienți pe braț, dacă un studiu randomizat individual ar fi avut nevoie de 64 pe braț.
4. O scală de 5 itemi are suma varianțelor itemilor 5,5 și varianța scorului total 18,0. Calculați α Cronbach și interpretați.
5. Într-un RCT cu câte 150 de pacienți pe braț, evenimentul nefavorabil (spitalizare în 12 luni) apare la 30 de pacienți tratați cu noul program și la 45 din grupul de control. Calculați riscul relativ, reducerea absolută a riscului și NNT.
6. Un studiu are, pe lângă criteriul principal, 8 criterii secundare, testate fiecare la $\alpha = 0,05$. Care este probabilitatea de a obține cel puțin un rezultat fals pozitiv dacă toate H0 sunt adevărate și testele sunt independente? Ce prag ar impune corecția Bonferroni?
7. Un studiu de non-inferioritate compară teleconsultația cu vizitele fizice, cu marja de non-inferioritate $\Delta = 3$ mg/dL pentru scăderea glicemiei. Ce concluzie trageți dacă IC 95% al diferenței (teleconsultație minus vizite fizice) este: a) -2,4 la 1,8 mg/dL; b) -4,1 la 0,9 mg/dL; c) 0,5 la 4,2 mg/dL?
8. Încadrați designul fiecărui studiu: a) 60 de pacienți cu BPOC primesc, în ordine aleatorie, un bronhodilatator nou și unul standard, câte 4 săptămâni, cu 2 săptămâni de pauză între ele; b) 24 de școli sunt randomizate pentru un program de educație nutrițională; c) 21 de copii primesc o terapie nouă și sunt evaluați la 0, 30 și 90 de zile; d) 400 de pacienți sunt randomizați în patru grupuri: placebo, vitamina D, omega-3, vitamina D și omega-3; e) pacienții cu infarct sunt urmăriți 5 ani după cum au primit sau nu statine la externare, după decizia medicului curant.

9. În coloana "inaltime" a unui fișier brut apar valorile: 168, 172, 1,75, 159, 1680, 165, "1 m 70", 181. Identificați valorile suspecte, tipul erorii și ce faceți cu fiecare.
10. Într-un studiu de faza III cu 300 de pacienți nu a apărut niciun caz de hepatotoxicitate. Folosind regula lui 3, estimați limita superioară a IC 95% pentru frecvența acestui efect advers. Ce înseamnă pentru faza IV?
11. În brațul B al studiului mg2026, 16 din 53 de pacienți au slăbit cel puțin 5%. Calculați proporția și IC 95% Wald. Comparați cu intervalul Wilson.
12. Pentru comparația B față de A în studiul mg2026 (16/53 față de 6/53), calculați RR, diferența de risc și NNT. Comparați cu rezultatele pentru C față de A.

Capcane frecvente

- A confunda ascunderea alocării cu orbirea: prima protejează momentul includerii și este întotdeauna posibilă, a doua protejează evaluarea și nu este întotdeauna posibilă.
- A testa statistic diferențele de bază dintr-un RCT și a declara "eșuată" o randomizare pentru un $p < 0,05$ apărut întâmplător.
- A schimba testul sau criteriul principal după ce s-au văzut datele, pentru a obține $p < 0,05$ (de exemplu, Kruskal-Wallis în loc de ANOVA pentru TAS, $p = 0,046$ în loc de 0,056).
- A interpreta un rezultat nesemnificativ dintr-un studiu de superioritate ca dovadă de echivalență.
- A elimina pacienții neaderenți și a numi analiza rezultată "intenție de tratament".
- A înlocui valorile lipsă cu media grupului sau a șterge pacienții cu date lipsă fără a raporta câți și de ce.
- A culege vârsta sau greutatea în clase, în loc de valoarea exactă.
- A "corecta" o valoare imposibilă prin ghicire (1540 devine 154) fără verificare în documentul sursă.
- A trata un singur item Likert ca variabilă numerică și a-i calcula media fără discuție.
- A considera că un α Cronbach mare dovedește validitatea unui chestionar.
- A analiza un studiu randomizat pe clustere ca și cum pacienții ar fi independenți.
- A citi asocierile observaționale dintr-un RCT (obezitate și diabet) ca și cum ar fi protejate de randomizare.

De reținut

- Studiul intervențional atribuie expunerea; randomizarea este singura metodă care echilibrează și factorii de confuzie necunoscuți.
- Designul decide testul: grupuri paralele, teste pentru eșantioane independente; încrucișat sau înainte-după, teste pentru date pereche; trei momente de timp, Friedman sau ANOVA cu măsurători repetate; clustere, metode care țin cont de corelația intraclasă.
- Criteriul principal, testul, pragul și regula de alegere a testului alternativ se fixează în protocol și în registrul public, înainte de date.
- Dimensiunea eșantionului depinde de α, putere, efectul minim relevant și variabilitate; se adaugă o rezervă pentru pierderi.
- Analiza principală într-un studiu de superioritate este ITT; analiza per protocol este de sensibilitate.
- Etica (Helsinki 2024, consimțământ informat, comisia de etică, GCP, Regulamentul 536/2014, GDPR) nu este o formalitate, ci o condiție a studiului.
- Tipul întrebării din chestionar decide tipul variabilei; fidelitatea (alfa, ICC) și validitatea (conținut, construct, AFE/AFC) se raportează pentru orice instrument nou sau tradus.
- Curățarea datelor se face pe o copie, cu jurnal de modificări, iar corecțiile se fac doar din documentul sursă.

- Un RCT se raportează după CONSORT 2025, cu diagrama de flux, Tabelul 1 fără valori p și efecte exprimate prin mărime și interval de încredere.

Întrebări de autoevaluare

1. (CS) Care dintre următoarele caracteristici definește un studiu intervențional?
    a) urmărirea pacienților în timp
    b) existența unui grup de control
    c) atribuirea expunerii de către investigator, conform protocolului
    d) culegerea prospectivă a datelor
    e) folosirea unui chestionar validat
2. (CS) Într-un studiu încrucișat (cross-over) cu două tratamente, compararea valorilor obținute de același pacient sub cele două tratamente se face, pentru o variabilă numerică cu diferențe normal distribuite, cu:
    a) testul t pentru eșantioane independente
    b) testul t pentru eșantioane pereche
    c) testul chi-pătrat
    d) ANOVA unifactorială
    e) testul Mann-Whitney
3. (CS) Ascunderea alocării (allocation concealment) are rolul de a:
    a) împiedica pacientul să afle ce tratament primește
    b) împiedica evaluatorul să afle brațul pacientului la vizita finală
    c) împiedica anticiparea alocării la momentul includerii pacientului
    d) echilibra numărul de pacienți din brațe
    e) reduce numărul de pacienți pierduți din urmărire
4. (CS) Pentru scăderea în greutate de cel puțin 5% în studiul mg2026, comparația C față de A dă o diferență absolută de risc de 18,3 puncte procentuale. NNT este:
    a) 2
    b) 3
    c) 6
    d) 18
    e) 29
5. (CS) La un chestionar, coeficientul ICC calculat între două aplicări la 14 zile distanță la aceiași pacienți evaluează:
    a) consistența internă
    b) fidelitatea test-retest
    c) validitatea de conținut
    d) validitatea convergentă
    e) structura factorială
6. (CM) Despre analiza după intenția de tratament (ITT) într-un studiu de superioritate, sunt adevărate:
    a) analizează fiecare pacient în brațul în care a fost randomizat
    b) păstrează beneficiul randomizării
    c) tinde să supraestimeze efectul biologic al tratamentului

d) este analiza principală recomandată
e) exclude pacienții care nu au respectat protocolul

7. (CM) Care dintre următoarele sunt criterii surogat?
   a) HbA1c în studiile antidiabetice
   b) decesul de cauză cardiovasculară
   c) colesterolul LDL în studiile cu hipolipemiante
   d) tensiunea arterială în studiile cu antihipertensive
   e) accidentul vascular cerebral
8. (CM) În "Tabelul 1" al unui studiu randomizat:
   a) variabilele numerice cu distribuție normală se descriu prin medie și abatere standard
   b) variabilele asimetrice se descriu prin mediană și interval intercuartil
   c) se recomandă testarea fiecărei diferențe dintre brațe cu un test statistic
   d) variabilele categoriale se descriu prin frecvență absolută și procent
   e) o diferență mare la o variabilă prognostică importantă poate justifica o analiză de sensibilitate prevăzută în plan
9. (CM) În fișierul mg2026_brut.csv, care dintre erori pot fi depistate printr-un tabel de frecvențe al variabilelor categoriale?
   a) sexul codificat "f" și "Feminin"
   b) "Fumator " cu spațiu la final
   c) "Urban" cu majusculă
   d) $\mathrm{TAS}_0 = 1540$ mmHg
   e) înălțimea 1,72 introdusă în metri la un pacient
10. (CM) Despre validarea unui chestionar, sunt adevărate:

a) un indice KMO de 0,93 indică date foarte potrivite pentru analiza factorială
b) testul Bartlett nesemnificativ este o condiție necesară pentru analiza factorială
c) analiza paralelă ajută la alegerea numărului de factori
d) un α Cronbach peste 0,95 poate sugera itemi redundanți
e) un α Cronbach mare demonstrează validitatea de construct

Capitolul 14. Exemple recapitulative: interpretarea integrată a rezultatelor statistice

Obiective

La finalul acestui capitol, studentul va putea:

- să plaseze orice metodă învățată pe harta disciplinei și să explice legătura dintre întrebarea clinică, design, tipul datelor, descriere, estimare, testare, asociere și decizie;
- să citească secțiunile de metode și rezultate ale unui articol medical după o metodă pas cu pas;
- să recunoască testul potrivit pentru o întrebare dată și să depisteze nepotrivirile dintre design, date și analiză;
- să interpreteze împreună valoarea p, mărimea efectului și intervalul de încredere, făcând distincția între semnificația statistică și cea clinică;
- să recalculeze din cifrele publicate indicatori simpli (interval de încredere pentru o proporție, tabel 2 x 2, VPP la altă prevalență, R^2 din r);
- să redacteze corect un paragraf de rezultate și să recunoască formulările greșite;
- să folosească tabelul-sinteză al formulelor din curs și să rezolve grile de tip rezidențiat care combină mai multe capitole.

14.1 Harta disciplinei

Paisprezece capitole par paisprezece subiecte separate. Nu sunt. Toate răspund, din unghiuri diferite, aceleiași succesiuni de întrebări pe care și-o pune orice cercetător și orice medic care citește un articol: ce vreau să aflu, cum am obținut datele, ce fel de date am, cum arată ele, cât de precis estimez ce mă interesează, este diferența sau asocierea reală, cât de mare este, ce decizie clinică iau. Figura 14.1 desenează această succesiune.

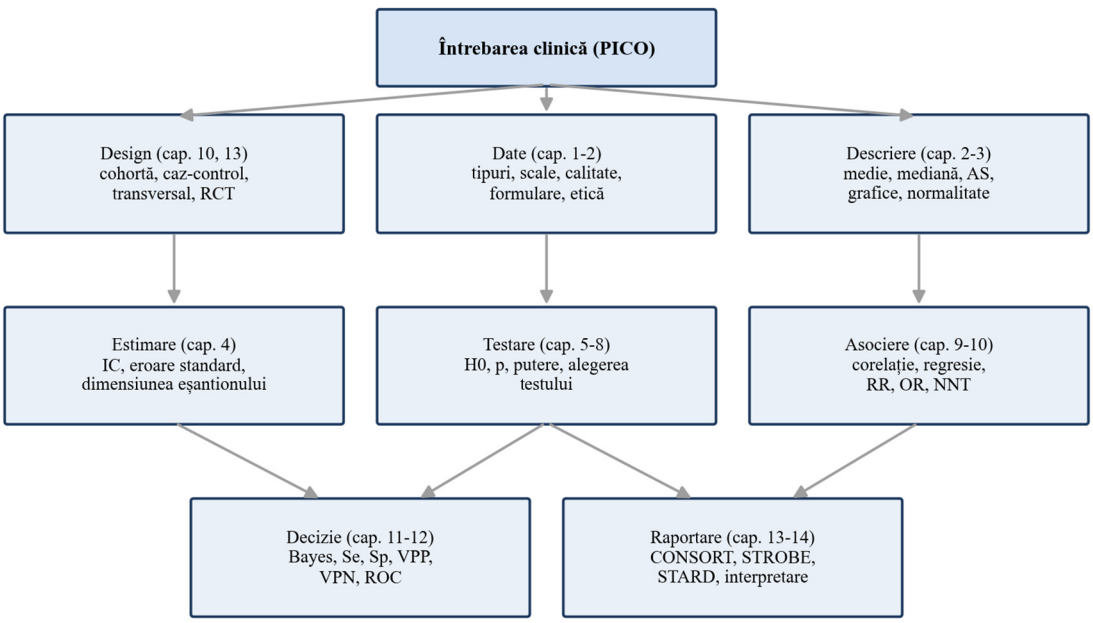


Figura 14.1. Harta disciplinei: de la întrebarea clinică (PICO) la design (capitolele 10 și 13), date (1-2), descriere (2-3), estimare (4), testare (5-8), asociere (9-10), decizie diagnostică (11-12) și raportare (13-14).

Totul pornește de la întrebarea clinică, formulată cât mai precis, de preferință în format PICO [24]. Întrebarea decide designul. Dacă vreau să știu dacă o intervenție funcționează, designul ideal este studiul randomizat (capitolul 13). Dacă vreau să aflu dacă o expunere este factor de risc, folosesc o cohortă sau, pentru boli rare, un studiu caz-control (capitolul 10). Dacă vreau să știu cât de frecventă este o boală, un studiu transversal. Dacă vreau să judec un test diagnostic, un studiu de acuratețe față de un standard de referință (capitolul 12).

Designul și instrumentele de culegere produc date de un anumit tip: numerice, ordinale, nominale (capitolul 2). Tipul datelor decide descrierea: medie și abatere standard pentru variabile numerice simetrice, mediană și interval intercuartil pentru cele asimetrice sau ordinale, frecvențe și procente pentru cele categoriale. Forma distribuției (capitolul 3) decide dacă putem folosi metodele bazate pe distribuția normală sau dacă transformăm datele ori trecem la metode neparametrice.

Din eșantion vrem să spunem ceva despre populație. Estimarea (capitolul 4) dă valoarea cea mai probabilă și intervalul de încredere, adică precizia. Testarea ipotezelor (capitolul 5) ne spune dacă o diferență sau o asociere este compatibilă cu întâmplarea. Alegerea testului (capitolele 6-8) depinde de trei lucruri: tipul variabilei rezultat, numărul de grupuri și dacă grupurile sunt independente sau pereche. Asocierile dintre variabile numerice se măsoară prin corelație și regresie (capitolul 9), iar cele dintre o expunere și o boală prin risc relativ și raport al cotelor (capitolul 10). Pentru decizia la patul bolnavului, teorema lui Bayes (capitolul 11) și indicatorii testelor diagnostice (capitolul 12) traduc rezultatul unui test în probabilitatea ca pacientul să aibă boala. La final, rezultatele se raportează după ghidul potrivit designului: CONSORT pentru studii randomizate [56], STROBE pentru studii observaționale [25], STARD pentru studii de acuratețe diagnostică [140].

Tabelul 14.1 condensează harta într-o formă pe care o puteți folosi la examen.

Tabelul 14.1. De la întrebare la metodă: sinteza disciplinei.

Întrebarea	Designul tipic	Variabila rezultat	Măsura efectului	Testul	Capitolul
Cât de frecventă este boala?	transversal	dihotomică	prevalență cu IC 95%	IC pentru proporție	4, 10
Diferă media între două grupuri?	transversal, cohortă, RCT	numerică	diferența mediilor, d Cohen	t independent / Mann-Whitney	6, 7
S-a schimbat valoarea după tratament?	înainte-după, încrucișat	numerică	diferența medie pereche	t pereche / Wilcoxon	6, 7
Diferă media între 3 sau mai multe grupuri?	RCT cu mai multe brațe, transversal	numerică	eta², diferențe Tukey	ANOVA / Kruskal-Wallis	6, 7
Se schimbă valoarea la 3 momente?	măsurători repetate	numerică sau ordinală	diferențe între momente	ANOVA cu măsurători repetate / Friedman	6, 7
Sunt asociate două variabile categoriale?	orice	nominală	diferența proporțiilor, RR, OR	chi-pătrat / Fisher	7, 10
Variază împreună două variabile numerice?	transversal	numerică	r, rho, R^2, panta	Pearson / Spearman, regresie	9
Este expunerea factor de risc?	cohortă	dihotomică	RR, DR, NNT/NNH	chi-pătrat, IC pentru RR	10
Este expunerea	caz-control	dihotomică	OR	chi-pătrat, IC	10

Întrebarea	Designul tipic	Variabila rezultat	Măsura efectului	Testul	Capitolul
asociată cu boala (boală rară)?				pentru OR	
Cât de bun este testul diagnostic?	acuratețe diagnostică	dihotomică	Se, Sp, VPP, VPN, RV, AUC	ROC, IC pentru proporții	11, 12
Funcționează intervenția?	RCT	oricare	diferență, RR, NNT, cu IC	după tipul variabilei	13

Două observații despre tabel. Prima: nicio coloană nu se poate citi independent; testul depinde și de design (pereche sau independent), nu doar de tipul variabilei. A doua: coloana "măsura efectului" este cel puțin la fel de importantă ca cea a testului. Un articol care raportează doar valori p, fără mărimea efectului și fără interval de încredere, spune doar jumătate din poveste [70,74].

14.2 Cum citim statistica unui articol medical

La seminarii, studenții îmi spun adesea că citesc rezumatul, sar peste metode și se uită direct la concluzie. Este exact ordinea inversă celei corecte. Concluzia este partea cea mai puțin verificabilă a unui articol; metodele și tabelele sunt cele care vă permit să judecați dacă ea este justificată. Manualele de bază ale disciplinei tratează pe larg lectura critică a literaturii medicale [18,19,64,150,151], iar aici o reducem la o metodă scurtă, de folosit la fiecare articol. Propun o metodă în douăsprezece întrebări, pe care o vom aplica apoi în studiile de caz.

1. Titlul și întrebarea. Ce întrebare pune studiul? Este o întrebare de frecvență, de diferență, de asociere, de prognostic, de diagnostic sau de eficacitate? Poate fi scrisă în format PICO?
2. Designul. Cine a decis expunerea? Este un studiu transversal, caz-control, cohortă (prospectivă sau retrospectivă), RCT, înainte-după, de acuratețe diagnostică? Eticheta din titlu corespunde cu ce s-a făcut efectiv?
3. Populația și eșantionul. Cine a fost inclus, de unde, cum a fost selectat, câți au refuzat sau au fost pierduți? Eșantionul este reprezentativ pentru pacienții la care vreau să aplic rezultatul?
4. Variabilele. Care este rezultatul principal și ce tip de variabilă este? Care sunt expunerile sau grupurile? Au fost dihotomizate variabile numerice?
5. Descrierea. Sunt descrise corect variabilele (medie ± AS pentru simetrice, mediană și IQR pentru asimetrice)? O abatere standard aproape egală sau mai mare decât media unei variabile pozitive (durata spitalizării, un marker biologic) semnalează asimetrie.
6. Testele. Se potrivesc testele cu designul și cu tipul datelor (independent sau pereche, două sau mai multe grupuri, parametric sau neparametric)? Sunt numite testele?
7. Mărimea efectului. Cât de mare este diferența sau asocierea: diferența mediilor, RR, OR, HR, r, AUC? Este relevantă clinic?
8. Intervalul de încredere. Cât de precisă este estimarea? Include valoarea de "nul" (0 pentru diferențe, 1 pentru rapoarte)? Este suficient de îngust pentru o decizie?
9. Valoarea p. Ce ipoteză testează? Câte teste s-au făcut? S-a ținut cont de comparațiile multiple?
10. Confuzia și eroarea sistematică. Ar putea un al treilea factor explica asocierea? A fost ajustată analiza? Pentru studii diagnostice, a existat eroare de spectru (bolnavi severi comparați cu voluntari sănătoși)?
11. Limitele. Le recunosc autorii? Lipsesc limite evidente (puterea, lipsa grupului de control, lipsa orbirii, datele lipsă)?
12. Ghidul de raportare. A fost urmat CONSORT, STROBE sau STARD? Elementele obligatorii (diagrama de flux, Tabelul 1, IC pentru efectul principal) sunt prezente?

Metoda nu are nevoie de calculatoare sofisticate. Are nevoie de răbdare și de trei deprinderi pe care le-ați exersat tot semestrul: să recunoașteți tipul de variabilă, să asociați designul cu testul și să citiți un interval de încredere. În multe cazuri, puteți verifica chiar dumneavoastră o cifră din articol: din procent și n calculați intervalul de încredere al unei proporții; din Se, Sp și prevalență reconstituiți tabelul 2 x 2; din r calculați R^2. Studiile de caz de mai jos fac exact acest exercițiu.

Studiile alese provin din cercetarea colegilor din disciplina noastră și din colaborările lor clinice. Folosim doar cifrele raportate în rezumatele publicate. Când semnalăm o problemă metodologică, o facem ca exercițiu de lectură critică: orice articol are limite, iar a le recunoaște este o formă de respect pentru munca autorilor, nu o critică a lor.

Studiu de caz: metoda celor douăsprezece întrebări aplicată unui studiu de hemodializă

Marc și colaboratorii au comparat două cohorte de pacienți hemodializați din 10 centre din România, evaluate în 2015 (n = 1401) și în 2019 (n = 1698), înainte și după tratamentul național cu antivirale cu acțiune directă [39]. Parcurgem rapid metoda. Întrebarea: s-a schimbat prevalența infecțiilor cu virusurile hepatitice B și C? Designul: transversal repetat, două eșantioane independente, nu aceiași pacienți urmăriți în timp. Descrierea: vârsta medie 59,7 ± 12,92 ani în 2015 și 61 ± 12,93 ani în 2019. Testele: compararea mediilor (test t) și a proporțiilor (chi-pătrat) între cohorte. Rezultatele: vârsta diferă semnificativ (p = 0,005), deși diferența este de doar 1,3 ani, irelevantă clinic; cu peste 1400 de pacienți în fiecare cohortă, chiar diferențele mici devin semnificative. Mortalitatea la un an (9,9% față de 10,7%, p = 0,46) și prevalența VHB (4,7% față de 4,8%, p = 0,604) nu diferă, iar prevalența VHC scade de la 16,9% la 10,5% (p < 0,0001), diferență relevantă și clinic. Confuzia: cohortele diferă și prin durata dializei (p < 0,0001) și prin frecvența calcificărilor vasculare (p = 0,015), deci populațiile nu sunt identice; scăderea VHC este compatibilă cu efectul tratamentului antiviral, dar designul nu o poate atribui cu certitudine. Ce ar mai fi util: diferențele de prevalență cu IC 95%. Exercițiul arată că metoda se aplică în câteva minute și că ea separă rezultatul clinic important (VHC) de cel doar statistic (vârsta).

14.3 Studii de caz integrate

14.3.1 Hepatologie: un test diagnostic neinvaziv pentru ciroză

Lupușoru și colaboratorii au evaluat ecografia cu contrast cu imagistică parametrică a timpului de sosire (CEUS-PAT) pentru diagnosticul neinvaziv al cirozei hepatice [2]. Au fost incluși 64 de subiecți: 37 de pacienți cu ciroză și 27 de voluntari sănătoși, cu vârsta medie 58,98 ± 8,90 ani, 56,3% bărbați. Indicele CEUS-PAT a fost 0,83 ± 0,09 la pacienții cu ciroză și 0,49 ± 0,11 la voluntarii sănătoși (p < 0,0001) și a corelat cu elastografia tranzitorie (VCTE) cu r = 0,81. La pragul > 0,7, aria de sub curba ROC a fost 0,98 (p < 0,001), cu Se = 89,19%, Sp = 100%, VPP = 100% și VPN = 87,1%.

Pentru comparație, Moga și colaboratorii au evaluat elastografia time-harmonic pe 165 de pacienți, cu VCTE ca metodă de referință, și au propus două praguri [3]: un prag de excludere a cirozei (< 1,61 m/s), cu Se = 90,7%, Sp = 66,6%, VPP = 55,7%, VPN = 93,9%, și un prag de confirmare (> 1,83 m/s), cu Se = 65,1%, Sp = 96,7%, VPP = 90,3%, VPN = 85,7%, ambele cu AUROC 0,90 (IC 95% 0,82-0,93); în eșantion, 28,1% aveau ciroză (F4). Acordul între examinatori a fost evaluat prin ICC (0,94; IC 95% 0,89-0,97).

Întrebări de interpretare, cu răspunsuri model:

1. Ce tip de design are studiul CEUS-PAT și ce eroare sistematică îl amenință? Răspuns model: este un studiu de acuratețe diagnostică cu design de tip caz-control: s-au comparat pacienți cu boală confirmată cu voluntari sănătoși. Designul expune la eroarea de spectru:

diferența dintre o ciroză constituită și un ficat sănătos este mult mai mare decât diferența dintre o ciroză incipientă și o fibroză avansată, situația reală în care medicul are nevoie de test. Performanța (AUC 0,98, Sp 100%) este, aproape sigur, mai bună decât cea care se va obține la pacienții cu boală hepatică cronică de severitate variată.

2. Reconstituiți tabelul 2 x 2 din cifrele publicate. Răspuns model: 37 de bolnavi, Se 89,19%, deci 33 adevărat pozitivi și 4 fals negativi (33/37 = 0,8919). 27 de sănătoși, Sp 100%, deci 27 adevărat negativi și 0 fals pozitivi. $\text{VPP} = \frac{33}{33} = 100\%$; $\text{VPN} = \frac{27}{27+4} = \frac{27}{31} = 87{,}1\%$, exact valoarea raportată. Reconstituirea confirmă coerența cifrelor.
3. Specificitatea de 100% înseamnă că testul nu dă niciodată rezultate fals pozitive? Răspuns model: nu. Înseamnă că, la cei 27 de voluntari, nu a apărut niciun fals pozitiv. Regula lui 3 dă limita superioară a IC 95% pentru proporția de fals pozitivi aproximativ 3/27 = 11%, deci specificitatea reală ar putea fi și în jur de 89%. Cu 27 de subiecți, "0 din 27" este compatibil cu o specificitate departe de perfecțiune.
4. Prevalența cirozei în eșantion este 37/64 = 57,8%. Cum s-ar schimba valorile predictive la o prevalență de 20%? Răspuns model: presupunând Se 89,2% și o specificitate mai prudentă de 89% (limita de mai sus), la prevalența de 20% VPP ar fi aproximativ 67% și VPN aproximativ 97%; la prevalența din eșantion, aceleași Se și Sp ar da VPP 92% și VPN 86%. Valorile predictive depind de prevalență (capitolele 11-12), deci VPP = 100% din articol nu se transferă automat la populația unui cabinet de gastroenterologie.
5. Ce test s-a folosit, probabil, pentru compararea CEUS-PAT între cele două grupuri și ce informație ar mai fi fost utilă? Răspuns model: două grupuri independente, variabilă numerică cu abateri standard mici față de medii: testul t pentru eșantioane independente (sau Welch, dacă variațele diferă). Utile ar fi fost diferența mediilor cu IC 95% și intervalul de încredere pentru AUC, Se și Sp.
6. Ce înseamnă r = 0,81 cu VCTE? Răspuns model: o corelație pozitivă puternică; $R^2 = 0{,}66$, adică aproximativ două treimi din variabilitatea unei metode este "explicată" liniar de cealaltă. Corelația nu înseamnă acord: două metode pot corela perfect și totuși să dea valori sistematic diferite. Pentru acord se folosesc ICC sau metoda Bland-Altman [102].
7. De ce propune studiul elastografiei time-harmonic două praguri diferite? Răspuns model: un singur prag nu poate avea simultan sensibilitate și specificitate maxime. Pragul de excludere are Se și VPN mari (90,7% și 93,9%): un rezultat sub prag face ciroza puțin probabilă. Pragul de confirmare are Sp și VPP mari (96,7% și 90,3%): un rezultat peste prag face ciroza foarte probabilă. Între praguri rămâne o "zonă gri", în care este nevoie de alte investigații. RV+ la pragul de confirmare este 0,651/(1 - 0,967) = 19,7, iar RV- la pragul de excludere (1 - 0,907)/0,666 = 0,14.

14.3.2 Nutriție: obiceiuri alimentare și greutatea la studenți

Trei studii ale colegilor din disciplină au urmărit studenții din Timișoara la începutul vieții universitare. Moleriu și colaboratorii au analizat transversal 921 de studenți de 18-24 de ani [4]: 24,1% erau supraponderali sau obezi (19,4% supraponderali și 4,7% obezi). Rezumatul raportează, pentru doi dintre factorii comportamentali analizați, rapoarte ale cotelor ajustate de 7,20 (IC 95% 4,26-12,17) și 1,98 (IC 95% 1,26-3,10) pentru supraponderare. Un scor compozit de risc alimentar (DRS, 0-4) a arătat o relație doză-răspuns: proporția supraponderii a crescut de la 19,8% la DRS = 0 la 50,0% la DRS = 4, cu un OR ajustat de 1,43 pentru fiecare punct (IC 95% 1,23-1,66; $p < 0{,}001$).

Piroș și colaboratorii au studiat transversal 693 de studenți, pentru riscul de prediabet [6]. Prevalența prediabetului a fost 21,1% (a diabetului 1,4%). Pe tipare alimentare obținute prin

analiză cluster, prediabetul a fost 15,4% la studenții cu alimentație sănătoasă, 20,0% la cei cu alimentație mixtă și 27,3% la cei orientați spre fast-food (χ^2, p = 0,003), cu un OR de 1,78 pe categorie (IC 95% 1,38-2,30; p < 0,001) și p pentru trend < 0,001.

Muntean și colaboratorii au urmărit prospectiv o cohortă de 921 de studenți în primul an universitar [5]. Diferența medie a IMC a fost -0,45 ± 1,07 kg/m² (IC 95% -0,52 la -0,38; p < 0,001): 48,2% au avut o scădere a IMC, 38,7% au rămas stabili și 13,1% au crescut în greutate; o creștere clinic semnificativă a apărut la 7,2% dintre studenți. Unul dintre comportamentele de crononutriție s-a asociat cu o creștere *a* IMC $\left(\beta = +\frac{0{,}22\ \text{kg}}{m^2};\ p = 0{,}004\right)$.

Întrebări de interpretare, cu răspunsuri model:

1. Calculați IC 95% pentru prevalența supraponderii (24,1% din 921). Răspuns model: ES = $\sqrt{\frac{0{,}241 \times 0{,}759}{921}} = 0{,}0141$; IC 95% = $24{,}1\% \pm 1{,}96 \times 1{,}41\% = 21{,}3\%$ la 26,9%. Cu aproape 1000 de studenți, estimarea este precisă (±2,8 puncte procentuale).
2. Cum interpretați aOR = 7,20 (IC 95% 4,26-12,17)? Răspuns model: după ajustarea pentru ceilalți factori din model, cota de supraponderare este de aproximativ 7 ori mai mare la studenții expuși factorului respectiv. Intervalul este larg (de la 4 la 12), dar este departe de 1, deci asocierea este puternică și semnificativă. Atenție: cu o prevalență a rezultatului de 24%, OR-ul supraestimează raportul prevalențelor; un OR de 7 nu înseamnă că supraponderea este de 7 ori mai frecventă.
3. Ce înseamnă OR = 1,43 pe punct de DRS și cât este OR-ul între DRS = 4 și DRS = 0? Răspuns model: fiecare punct în plus pe scorul de risc crește cota de supraponderare cu 43%. Efectul se înmulțește, nu se adună: între 4 puncte și 0 puncte, OR = $1{,}43^4 = 4{,}18$ (presupunând o relație log-liniară, cum impune modelul logistic).
4. Pot concluziona autorii că alimentația orientată spre comoditate cauzează obezitatea? Răspuns model: nu, pe baza unui design transversal. Expunerea și rezultatul sunt măsurate în același moment, deci nu știm ce a apărut întâi (poate studenții supraponderali își schimbă alimentația); pot exista factori de confuzie nemăsurați (somn, stres, venit). Relația doză-răspuns și forța asocierii susțin, conform criteriilor Bradford Hill [108], o posibilă relație cauzală, dar nu o dovedesc.
5. Verificați IC 95% al diferenței medii a IMC din studiul prospectiv. Răspuns model: ES = $\frac{1{,}07}{\sqrt{921}} = 0{,}0353$; IC 95% = $-0{,}45 \pm 1{,}96 \times 0{,}0353 = -0{,}45 \pm 0{,}069$, adică -0,52 la -0,38, exact ce raportează autorii. Testul potrivit este testul t pentru date pereche (același student, la începutul și la sfârșitul anului).
6. Cum se împacă o scădere medie mică (-0,45 kg/m²) cu faptul că 13,1% dintre studenți au crescut în greutate? Răspuns model: media descrie centrul distribuției schimbărilor, nu fiecare student. Abaterea standard (1,07) este de peste două ori mai mare decât media schimbării, deci schimbările individuale variază mult în ambele sensuri. De aceea autorii au completat analiza liniară a IMC cu o analiză logistică a creșterii clinic semnificative: aceeași variabilă, privită continuu și dihotomizat, răspunde la întrebări diferite.
7. În studiul despre prediabet, de ce se raportează și un "p pentru trend"? Răspuns model: cele trei tipare alimentare au o ordine logică (sănătos, mixt, fast-food), iar proporțiile cresc monoton (15,4%, 20,0%, 27,3%). Chi-pătratul obișnuit (p = 0,003) testează doar dacă proporțiile diferă, fără să țină cont de ordine; testul de trend testează specific o creștere monotonă și are, de regulă, mai multă putere pentru o astfel de ipoteză.

14.3.3 COVID-19 și diabetul: trei studii, trei tipuri de analiză

Cioca și colaboratorii au comparat retrospectiv evoluția COVID-19 la 486 de pacienți: 200 fără diabet, 62 cu diabet de tip 1 și 224 cu diabet de tip 2 [8]. Formele severe au fost 33,5% la diabetul de tip 2 față de 25,8% la tipul 1, mortalitatea 11,6% față de 8,1%, durata spitalizării 11,6 ± 7,0 zile față de 9,1 ± 5,8 zile, glicemia la internare 178,3 ± 34,7 față de 164,8 ± 39,6 mg/dL; rapoartele cotelor au fost 2,24 pentru internarea în terapie intensivă și 2,46 pentru ventilația mecanică.

Albai și colaboratorii au validat scorul ISARIC-4C pentru severitate și mortalitate la 159 de pacienți cu diabet de tip 2 și COVID-19 [75]. Mortalitatea a fost 25,15%. Aria de sub curba ROC a fost 0,875 (IC 95% 0,820-0,930), scorul a clasificat corect 77,36% dintre pacienți, cu Se 87,80% (IC 95% 87,66-87,93), Sp 73,72% (IC 95% 73,48-73,96), VPP 53,73% și VPN 94,56%.

Dănilă și colaboratorii au comparat 168 de pacienți cu COVID-19 și diabet cu 168 de pacienți fără diabet, potriviți pentru sex, IMC și severitate [105]. Pentru indicele dNLR, la pragul 2,685: Se 70,312%, Sp 65,978%, AUC 0,624 ($p < 0,001$); pentru ALRI, la pragul 0,812: Se 76,429%, Sp 69,541%, AUC 0,752. Pentru indicele SII la 7 zile s-a raportat un raport al hazardelor de 2,62 (IC 95% 1,29-5,04).

Întrebări de interpretare, cu răspunsuri model:

1. Ce ne spune durata spitalizării de 11,6 ± 7,0 zile despre distribuția ei și cum ar fi trebuit descrisă? Răspuns model: abaterea standard reprezintă 60% din medie, iar media minus două abateri standard dă o valoare negativă (11,6 - 14,0 = -2,4 zile), imposibilă. Distribuția este asimetrică spre dreapta (câțiva pacienți cu spitalizări foarte lungi). Descrierea potrivită este mediana cu IQR, iar comparația între grupuri, testul Mann-Whitney (sau Kruskal-Wallis pentru trei grupuri).
2. Ce lipsește la raportarea OR = 2,24 pentru internarea în terapie intensivă? Răspuns model: intervalul de încredere (și valoarea p). Fără IC nu știm dacă asocierea este precisă sau compatibilă cu absența efectului. Nu se precizează nici grupul de referință, nici dacă OR-ul este brut sau ajustat.
3. Cu trei grupuri și o variabilă numerică precum glicemia la internare, ce test ați folosi? Răspuns model: ANOVA unifactorială dacă distribuțiile sunt aproximativ normale și variațele omogene (Levene), urmată de comparații post-hoc (Tukey); altfel Kruskal-Wallis cu comparații Mann-Whitney corectate. Compararea doar a tipului 1 cu tipul 2 se face cu testul t independent sau Mann-Whitney.
4. Intervalul de încredere raportat pentru sensibilitatea scorului ISARIC-4C (87,66-87,93) este plauzibil? Răspuns model: nu. Cu o mortalitate de 25,15% la 159 de pacienți, au decedat aproximativ 40 de pacienți. Intervalul Wald pentru o proporție de 87,8% pe 40 de pacienți este $87{,}8\% \pm 1{,}96 \times \sqrt{\frac{0{,}878 \times 0{,}122}{40}} = 87{,}8\% \pm 10{,}1\%$, adică aproximativ 77,7-97,9%. Intervalul publicat, de 0,27 puncte procentuale, ar corespunde unui eșantion de zeci de mii de pacienți. Probabil a fost calculat altfel (de exemplu, prin metode de reeșantionare greșit raportate). Exercițiul arată de ce merită să verificați intervalele.
5. Verificați VPP-ul raportat pentru ISARIC-4C din Se, Sp și prevalență. Răspuns model: $\text{VPP} = \frac{0{,}2515 \times 0{,}878}{0{,}2515 \times 0{,}878 + 0{,}7485 \times 0{,}2628} = \frac{0{,}2208}{0{,}4175} = 0{,}529$, adică aproximativ 53%, apropiat de 53,73% raportat. $\text{VPN} = \frac{0{,}7485 \times 0{,}7372}{0{,}7485 \times 0{,}7372 + 0{,}2515 \times 0{,}122} = 0{,}947$, apropiat de 94,56%. Scorul este util mai ales pentru a identifica pacienții cu risc mic (VPN mare).
6. Comparați AUC 0,624 (dNLR) cu 0,752 (ALRI) și 0,875 (ISARIC-4C). Răspuns model: o AUC de 0,624 indică o discriminare slabă (0,5 ar fi întâmplarea); chiar dacă este semnificativ diferită de 0,5 ($p < 0,001$), dNLR nu este util clinic ca test separat. ALRI are

o discriminare acceptabilă, ISARIC-4C bună. Semnificația statistică a unei AUC nu spune nimic despre utilitatea ei.

7. Ce problemă ridică raportarea unui raport al hazardelor într-un studiu numit transversal și raportarea Se cu trei zecimale? Răspuns model: raportul hazardelor provine din analiza timpului până la eveniment și presupune urmărirea pacienților în timp, deci un design longitudinal; un design pur transversal nu poate produce HR. Cifra 70,312% sugerează o precizie iluzorie: cu 168 de pacienți pe grup, a treia zecimală nu are sens; se raportează 70,3% sau chiar 70%, împreună cu IC 95%.

14.3.4 Medicina de urgență: sexul pacientului, timpii de tromboliză și biomarkerii AVC

Popa și colaboratorii au analizat retrospectiv 213 pacienți (100 de femei, 113 bărbați) tratați în departamentul de urgență pentru suspiciune de infecție bacteriană [9]. Femeile au fost mai vârstnice (mediana 77,0 față de 67,0 ani, $p < 0,001$), iar sursa urinară a infecției a fost mai frecventă la femei (51,0% față de 34,5%, p = 0,015; OR 2,14, IC 95% 1,20-3,81, p = 0,010; p ajustat Benjamini-Hochberg = 0,039). Un rezultat sever a apărut la 14 din 100 de femei (14,0%) și 21 din 113 bărbați (18,6%; OR 0,71, IC 95% 0,34-1,49); mortalitatea intraspitalicească la 7 femei (7,0%) și 9 bărbați (8,0%; OR 0,87, IC 95% 0,31-2,43); după ajustare, OR 0,80 (IC 95% 0,27-2,35; p = 0,690).

În același departament, un studiu cu același prim autor a urmărit pacienții cu AVC ischemic sosiți în primele 4,5 ore [62]: din 355 de pacienți inițiali, 258 au fost eligibili, iar 46 au primit tromboliză intravenoasă (rata 17,83%). Rezumatul raportează, în comparațiile dintre grupuri, o mediană de 100 de minute pentru timpul de la debut la sosirea în urgență ($p < 0,001$) și 15,5 minute pentru timpul de la sosire la CT (p = 0,009), precum și o variabilă clinică cu valori de 9 ± 4,94 față de $13,85 \pm 2,41 (p < 0,001)$. Un al treilea studiu a evaluat biomarkerii GFAP și UCH-L1 la 85 de pacienți cu alertă de AVC (69 cu AVC, 16 fără; 37 ischemice, 32 hemoragice) [61]: GFAP $142,91 \pm 102,19$ pg/mL față de $37,76 \pm 19,92$ pg/mL$(p < 0,001)$, UCH-L1 1307,68 ± 967,54 față de $189,81 \pm 92,69$ pg/mL$(p < 0,001)$; GFAP a diferențiat AVC-ul hemoragic de cel ischemic cu o acuratețe de 94,2% la pragul de 77,15 pg/mL.

Întrebări de interpretare, cu răspunsuri model:

1. Verificați OR-ul pentru rezultatul sever (14/100 femei, 21/113 bărbați). Răspuns model: cotele sunt 14/86 = 0,163 la femei și 21/92 = 0,228 la bărbați; OR $= \frac{0,163}{0,228} = 0,71$, exact valoarea raportată. Analog, pentru mortalitate, (7/93)/(9/104) = 0,075/0,087 = 0,87.
2. Cum interpretați OR 0,71 (IC 95% 0,34-1,49)? Răspuns model: estimarea punctuală sugerează o cotă mai mică a rezultatului sever la femei, dar intervalul include 1 și se întinde de la o reducere de două treimi la o creștere de aproape 50%. Datele sunt compatibile atât cu un efect protector important, cât și cu un risc crescut. Concluzia corectă nu este "sexul nu influențează prognosticul", ci "studiul nu are suficientă precizie pentru a răspunde"; cu doar 35 de evenimente severe, puterea este mică.
3. De ce se raportează un p ajustat Benjamini-Hochberg de 0,039 pe lângă p = 0,015? Răspuns model: autorii au comparat femeile și bărbații pe multe caracteristici. Corecția pentru comparații multiple (aici, controlul proporției de descoperiri false) ridică valoarea p pentru a ține cont de numărul de teste. Faptul că diferența privind sursa urinară rămâne sub 0,05 după corecție o face mai credibilă.
4. De ce s-au raportat timpii prin mediane? Răspuns model: timpii până la un eveniment sunt, aproape întotdeauna, asimetrici spre dreapta (puțini pacienți sosesc foarte târziu).

Mediana nu este influențată de aceste valori extreme. Compararea a două grupuri independente se face atunci cu testul Mann-Whitney.

5. Calculați IC 95% pentru rata de tromboliză. Răspuns model: $p = \frac{46}{258} = 0{,}178$; ES $= \sqrt{\frac{0{,}178 \times 0{,}822}{258}} = 0{,}0238$; IC $95\% = 17{,}8\% \pm 4{,}7\%$, adică 13,2% la 22,5%.
6. Ce sugerează abaterile standard ale biomarkerilor GFAP și UCH-L1 la pacienții cu AVC? Răspuns model: abaterea standard este aproape cât media (102,19 față de 142,91 pentru GFAP; 967,54 față de 1307,68 pentru UCH-L1). Pentru o variabilă care nu poate fi negativă, asta indică o distribuție puternic asimetrică. Testul t poate fi nepotrivit; mai potrivite sunt testul Mann-Whitney sau logaritmarea valorilor (capitolul 3), iar descrierea ar trebui să includă mediana și IQR.
7. Acuratețea de 94,2% pentru diferențierea AVC hemoragic de cel ischemic: ce ar mai trebui raportat? Răspuns model: 94,2% din 69 de pacienți înseamnă 65 clasificați corect. Acuratețea singură nu ajunge: trebuie Se și Sp (cât de bine sunt identificate separat hemoragiile și ischemiile), cu IC 95%, iar pragul de 77,15 pg/mL, derivat pe același eșantion, trebuie validat pe un eșantion independent, pentru că pragurile alese "optim" pe datele proprii supraestimează performanța.

14.3.5 Validarea unui chestionar și analiza unor scoruri

Am discutat în capitolul 13 validarea chestionarului TAM pentru telemonitorizare la pacienții cu ciroză [10] și validarea versiunii românești a PSPSQ 2.0 în dermatologie [11]. Le reluăm aici împreună cu un al treilea studiu, care folosește un chestionar deja validat pentru o întrebare clinică. Frandeș și colaboratorii au studiat implicarea digitală în îngrijirea diabetului la 304 adulți cu diabet [7]. Scala folosită a avut α Cronbach = 0,89; 35,5% dintre participanți au avut o implicare digitală ridicată. Scorurile au fost raportate ca mediane și quartile, de exemplu 1,71 [1,71-2,39] față de 3,86 [3,86-4,43] între grupuri, iar într-o regresie beta multivariabilă, cu regresie liniară ca analiză de sensibilitate, au apărut coeficienți precum β = -0,147 (IC 95% -0,219 la -0,075; p < 0,001), β = 0,996 (IC 95% 0,859-1,134; p < 0,001) și β = 0,069 (IC 95% 0,006-0,132; p = 0,033).

Întrebări de interpretare, cu răspunsuri model:

1. În studiul TAM, α = 0,934 și ω = 0,950. Ce spun aceste cifre și ce nu spun? Răspuns model: spun că itemii scalelor sunt puternic intercorelați (consistență internă foarte bună). Nu spun dacă scala măsoară acceptarea telemonitorizării (validitate) și nici dacă pacienții ar răspunde la fel la o a doua aplicare (fidelitate test-retest).
2. Un model AFC are CFI = 0,999 și RMSEA = 0,109 (PSPSQ 2.0). Cum se poate? Răspuns model: indicii măsoară aspecte diferite ale potrivirii. CFI compară modelul cu un model de independență totală a itemilor; când itemii sunt foarte corelați între ei (α = 0,978), aproape orice model structurat pare mult mai bun decât independența, deci CFI iese foarte mare. RMSEA măsoară eroarea de aproximare pe grad de libertate și penalizează modelele cu puține grade de libertate și eșantioane moderate; 0,109 depășește pragul uzual de 0,08. Raportarea ambilor indici și discutarea contradicției este practica corectă.
3. În modelul TAM, "modelul a explicat 69,5% din variația intenției de utilizare". Cu ce noțiune din capitolul 9 se leagă această cifră? Răspuns model: cu coeficientul de determinare R^2. La o regresie simplă, R^2 este pătratul corelației; aici, într-un model structural cu mai mulți predictori, $R^2 = 0{,}695$ înseamnă că predictorii împreună explică aproximativ 70% din variabilitatea intenției. Restul de 30% ține de factori neincluși și de eroarea de măsurare.

4. Ce înseamnă o mediană egală cu primul quartil, ca în 1,71 [1,71-2,39]? Răspuns model: toate valorile situate între percentila 25 și percentila 50, adică cel puțin un sfert dintre participanții acelui grup, sunt egale cu 1,71, foarte probabil o valoare-limită a scorului. Distribuția are un efect de podea și este puternic asimetrică. Este exact situația în care media și abaterea standard ar descrie prost datele, iar alegerea autorilor de a raporta mediane și de a folosi o regresie potrivită pentru scoruri mărginite (regresia beta) este justificată.
5. Cum interpretați β = 0,069 (IC 95% 0,006-0,132; p = 0,033)? Răspuns model: asocierea este pozitivă și semnificativă la pragul 0,05, iar intervalul nu include 0. Limita inferioară, 0,006, este însă foarte aproape de 0: efectul ar putea fi neglijabil. Comparativ, β = -0,147 (IC -0,219 la -0,075) este estimat mai precis și departe de 0. Mărimea și precizia, nu doar p, diferențiază cele două rezultate.
6. De ce se face o "analiză de sensibilitate" cu regresie liniară? Răspuns model: pentru a verifica dacă concluzia depinde de alegerea modelului. Dacă regresia beta și cea liniară dau coeficienți cu același sens și semnificație, rezultatul este robust la această alegere metodologică.

14.3.6 Insuficiența cardiacă: comparații pe trei grupuri, corelații și confuzie

Stoichescu-Hogea și colaboratorii au comparat deformarea longitudinală globală (GLS) în trei grupuri: 62 de pacienți cu insuficiență cardiacă cu fracție de ejecție păstrată (HFpEF), 46 de pacienți hipertensivi și 40 de martori de aceeași vârstă [90]. GLS a diferit între cele trei grupuri ($p < 0,001$): martorii au avut -20,2 ± 1,4%, hipertensivii -18,4 ± 3,0% ($p = 0,031$ față de martori) și pacienții cu HFpEF -17,6 ± 2,3% ($p < 0,001$ față de martori). GLS a corelat cu o variabilă clinică ordinală cu rho Spearman = 0,423 ($p < 0,001$), iar pragul optim de -19,35% a avut AUC 0,833 ($p < 0,001$).

Buleu și colaboratorii au comparat 60 de pacienți cu insuficiență cardiacă ischemică cu fracție de ejecție redusă (< 40%) cu 60 de martori sănătoși de aceeași vârstă [87]. Osteoprotegerina (OPG) a fost 4,7 ± 0,25 față de 1,3 ± 0,67 ng/mL, iar vitamina 25-OH-D3 20,49 ± 7,31 față de 37,09 ± 4,59 ng/mL (ambele $p < 0,001$). OPG a corelat cu volumul telediastolic al ventriculului stâng ($r = 0,862$), cu fracția de ejecție ($r = -0,832$), cu viteza undei pulsului ($r = 0,833$), cu vitamina D ($r = -0,636$) și cu rata de filtrare glomerulară ($r = -0,488$), toate cu $p < 0,001$.

În sfârșit, studiul HRR discutat în capitolul 13 [23] a arătat o asociere cu prognosticul în analiza Kaplan-Meier (log-rank $p = 0,027$), care a dispărut după ajustarea Cox pentru vârstă, sex, NT-proBNP, fracția de ejecție și rata de filtrare ($p = 0,240$).

Întrebări de interpretare, cu răspunsuri model:

1. Ce test s-a folosit, probabil, pentru compararea GLS în trei grupuri și ce ar trebui verificat? Răspuns model: ANOVA unifactorială, cu comparații post-hoc între perechi. Abaterile standard diferă mult (1,4 la martori, 3,0 la hipertensivi), deci omogenitatea varianțelor (testul Levene) este îndoielnică; o variantă robustă (ANOVA Welch cu post-hoc Games-Howell) sau Kruskal-Wallis ar fi de preferat dacă Levene respinge omogenitatea.
2. De ce se compară grupurile două câte două cu un test post-hoc și nu cu trei teste t? Răspuns model: trei teste t, fiecare la $\alpha = 0,05$, cresc probabilitatea totală a unei erori de tip I la aproximativ $1 - 0,95^3 = 14\%$. Testele post-hoc (Tukey, Bonferroni) controlează această eroare la nivelul întregului set de comparații.

3. Cât din variabilitate explică rho = 0,423 și cum diferă interpretarea de r = 0,862 din studiul OPG? Răspuns model: $\text{rho}^2 \approx 0{,}18$ (aproximativ 18%; pentru Spearman, interpretarea pătratului este doar aproximativă); r = 0,862 dă $R^2 = 0{,}74$. Prima este o corelație moderată, a doua una puternică. Ambele sunt semnificative, dar forța asocierii diferă mult.
4. OPG corelează puternic cu fracția de ejecție (r = -0,832). Putem spune că OPG scade fracția de ejecție? Răspuns model: nu. Corelația măsoară asocierea liniară, nu direcția cauzală. Este posibil ca disfuncția ventriculară să crească OPG, ca ambele să fie determinate de un al treilea factor (ateroscleroza, inflamația, funcția renală, ea însăși corelată cu OPG) sau ca relația să fie bidirecțională.
5. Studiul OPG are un design caz-control cu martori sănătoși. Ce implicații are pentru corelațiile calculate pe întregul eșantion? Răspuns model: dacă corelațiile sunt calculate pe toți cei 120 de subiecți, ele combină două grupuri foarte diferite (OPG 4,7 față de 1,3). Diferența dintre grupuri poate crea singură o corelație puternică pe eșantionul combinat, chiar dacă în interiorul fiecărui grup corelația este slabă (capitolul 9, efectul grupurilor distincte asupra lui r). Merită verificat dacă corelațiile au fost calculate doar la pacienți.
6. Ce lecție comună dau studiul HRR și cel al OPG? Răspuns model: o asociere brută (log-rank p = 0,027; corelații puternice) poate fi produsă sau amplificată de factori de confuzie. Ajustarea multivariabilă (Cox, regresie multiplă) este pasul care separă asocierea proprie a markerului de cea moștenită de la alți factori. În studiul HRR, după ajustare, asocierea nu a mai fost semnificativă.

14.3.7 Farmacologie la vârstnici: prescripții potențial inadecvate

Baneu și colaboratorii au analizat transversal 1498 de prescripții electronice emise pentru vârstnici din mediul urban, în județele Timiș și Arad, după criteriile STOPP/START versiunea 2 [38]. Dintre prescripții, 57% au fost pentru femei; vârsta medie a pacienților a fost 74,1 ± 6,95 ani, iar numărul mediu de medicamente pe prescripție 4,7 ± 1,51. Dintre clasele de medicamente, rezumatul menționează statinele, cu 47,4%. Prevalența medicamentelor potențial inadecvate a fost 18,58%, iar a omisiunilor potențiale de prescriere 49,2%.

Ca termen de comparație pentru date pereche din farmacologie, Gherbon și colaboratorii au urmărit retrospectiv 200 de pacienți cu diabet de tip 2 un an după inițierea unui inhibitor SGLT-2 [59]: glicemia a scăzut de la 180,00 mg% (IQR 154,50-207,00) la 130,00 mg% (IQR 117,50-150,00), HbA1c de la 8,40% (IQR 7,98-9,15) la 7,30% (IQR 6,90-7,95), ambele p < 0,001, iar circumferința abdominală de la 107,05 (± 16,39) la $102{,}50(\pm 15{,}11)$ cm$(p = 0{,}042)$.

Întrebări de interpretare, cu răspunsuri model:

1. Care este unitatea de analiză în studiul STOPP/START și de ce contează? Răspuns model: prescripția, nu pacientul. Dacă un pacient are mai multe prescripții, observațiile nu sunt independente, iar intervalele de încredere calculate ca pentru 1498 de pacienți independenți vor fi prea înguste. Interpretarea corectă este "18,58% dintre prescripții", nu neapărat "18,58% dintre vârstnici".
2. Calculați IC 95% pentru prevalența medicamentelor potențial inadecvate. Răspuns model: $\text{ES} = \sqrt{\frac{0{,}1858 \times 0{,}8142}{1498}} = 0{,}0100$; $\text{IC}\,95\% = 18{,}6\% \pm 2{,}0\%$, adică 16,6% la 20,5% (cu rezerva independenței de la întrebarea 1).
3. Este potrivită descrierea "4,7 ± 1,51 medicamente pe prescripție"? Răspuns model: numărul de medicamente este o variabilă numerică discretă (0, 1, 2...). Cu o medie de 4,7 și o AS de 1,5, distribuția nu este, probabil, foarte asimetrică, deci media și AS sunt

acceptabile; mediana și IQR ar completa util descrierea. Media minus două AS (1,7) rămâne plauzibilă.

4. Ce tip de studiu este și ce concluzii permite? Răspuns model: un studiu descriptiv transversal. Permite estimarea prevalențelor și descrierea practicii de prescriere. Nu permite concluzii despre consecințele clinice ale prescripțiilor inadecvate (spitalizări, căderi), pentru că nu urmărește pacienții în timp.
5. În studiul SGLT-2, de ce se raportează mediane cu IQR și ce test este potrivit? Răspuns model: datele pereche (același pacient înainte și după un an) cu distribuție asimetrică se descriu prin mediane și se compară cu testul Wilcoxon pentru eșantioane pereche. Rezumatul folosește cuvântul "mean" pentru valori care sunt de fapt mediane; la lectură, cifrele cu IQR trebuie citite ca mediane.
6. Lipsa unui grup de control în studiul SGLT-2: ce explicații alternative are scăderea glicemiei? Răspuns model: regresia spre medie (pacienții încep un nou tratament tocmai când glicemia lor este cea mai ridicată), alte modificări de tratament sau de stil de viață în cursul anului, aderența mai bună a celor rămași în evidență. Designul înainte-după arată o schimbare, dar nu o poate atribui cu certitudine medicamentului.

14.3.8 Oncologie: pandemia și cancerul colorectal, mutațiile BRCA și supraviețuirea

Feier și colaboratorii au comparat îngrijirea cancerului de colon înainte și în timpul pandemiei COVID-19, pe 867 de pacienți operați într-un spital universitar, în patru ani [55]. În pandemie au scăzut semnificativ intervențiile elective ($p < 0,001$), au crescut intervențiile de urgență ($p < 0,001$), iar în al doilea an pandemic a crescut proporția cazurilor în stadiul IV ($p = 0,045$). Între vârsta pacientului și durata intervenției s-a raportat o corelație inversă ($r = -0,088$, $p = 0,01$).

Petrușan și colaboratorii au urmărit prospectiv, timp de 24 de luni, 133 de paciente cu carcinom ovarian seros de grad înalt [117]. Mutații BRCA patogene au fost identificate la 39,1% dintre paciente. Rezumatul raportează diferențe de proporții între grupuri (50% față de 77,77%, $p = 0,001$; 32,69% față de 51,85%, $p = 0,017$). În analiza Cox univariată, statusul BRCA s-a asociat cu supraviețuirea fără progresie (HR 0,52, IC 95% 0,27-0,99, $p = 0,048$), dar în modelul multivariabil asocierea nu a mai fost semnificativă (HR 0,57, $p = 0,124$).

Întrebări de interpretare, cu răspunsuri model:

1. Corelația $r = -0,088$ este semnificativă ($p = 0,01$). Este și importantă? Răspuns model: $r^2 = 0,0077$, adică vârsta explică mai puțin de 1% din variabilitatea duratei intervenției. Semnificația vine din eșantionul mare (867 de pacienți): cu n mare, chiar și corelații neglijabile devin semnificative. Este exemplul clasic de diferență între semnificația statistică și cea clinică.
2. Ce test s-a folosit, probabil, pentru compararea proporției intervențiilor de urgență între perioade? Răspuns model: testul chi-pătrat pentru tabele de contingență (perioadă x tip de intervenție), sau Fisher dacă unele frecvențe așteptate sunt mici. Pentru o comparație pe mai mult de două perioade, chi-pătrat pe un tabel k x 2.
3. Rezultatul pentru stadiul IV ($p = 0,045$) trebuie citit cu prudență. De ce? Răspuns model: este unul dintre multe teste (mai multe perioade, mai multe variabile), iar valoarea este aproape de prag. Fără corecție pentru comparațiile multiple, un astfel de rezultat poate fi fals pozitiv. Mărimea diferenței și IC ar ajuta interpretarea.
4. Cum se citește HR = 0,52 (IC 95% 0,27-0,99)? Răspuns model: în orice moment al urmăririi, rata instantanee de progresie la pacientele cu mutație BRCA este aproximativ jumătate din cea a pacientelor fără mutație, adică o reducere relativă de 48%. Limita

superioară a intervalului, 0,99, este însă foarte aproape de 1: datele sunt compatibile și cu o reducere de doar 1%. Iar p = 0,048 este la limită.

5. De ce dispare semnificația în modelul multivariabil (HR 0,57, p = 0,124)? Răspuns model: două explicații, care nu se exclud. Prima, confuzia: o parte din asocierea brută se datorează altor factori (stadiul, tipul de chirurgie, chimioterapia) legați atât de statusul BRCA, cât și de prognostic. A doua, precizia: HR-ul ajustat (0,57) este foarte apropiat de cel brut (0,52), deci efectul nu a dispărut, dar adăugarea covariabilelor a lărgit intervalul de încredere. Cu 133 de paciente, modelul multivariabil are putere limitată. Formularea corectă: asocierea nu își păstrează semnificația după ajustare, iar mărimea ei rămâne incertă.
6. Ce metodă descriptivă și ce test corespund analizei supraviețuirii fără progresie? Răspuns model: curbele Kaplan-Meier pentru descriere, testul log-rank pentru compararea a două curbe și modelul Cox pentru estimarea HR cu ajustare. Toate țin cont de pacientele cenzurate (care nu au progresat până la finalul urmăririi sau au ieșit din studiu), lucru pe care o simplă comparație de proporții nu îl face.

14.3.9 Un studiu de tip RCT: Aloe ferox și o aplicație pentru astm

Am prezentat în capitolul 13 studiul controlat cu placebo privind Aloe ferox la persoanele obeze, raportat după CONSORT [85], și studiul pilot controlat al aplicației mobile pentru astm [94]. Le folosim acum pentru o recapitulare a principiilor studiilor clinice. Reamintim datele principale: în primul studiu, 20 de persoane cu obezitate au primit dietă și Aloe ferox (2 capsule pe zi, 460 mg), iar 20 de martori potriviți au primit dietă și placebo, timp de 3 luni; rezumatul raportează valori p pentru IMC (0,03 și 0,002), colesterol total (0,032), LDL (0,01 și 0,039), glicemia a jeun (0,018 și < 0,001) și tensiunea diastolică (0,006). În al doilea, 93 de pacienți cu astm au fost comparați cu Mann-Whitney ($p < 0,001$), Friedman cu corecție Bonferroni ($p < 0,001$) și chi-pătrat pentru proporția exacerbărilor (10,30% față de 46,30%, $\chi^2(1) = 13,707$, $p < 0,001$).

Întrebări de interpretare, cu răspunsuri model:

1. Ce elemente ale CONSORT ar trebui să găsiți în articolul despre Aloe ferox? Răspuns model: modul de generare a secvenței de randomizare, ascunderea alocării, cine a fost orb (pacienții, evaluatorii), definiția criteriului principal, calculul dimensiunii eșantionului, diagrama de flux cu pierderile, Tabelul 1 pe grupuri, efectul principal cu mărime și IC 95%, efectele adverse și înregistrarea studiului [56].
2. Ce înseamnă "martori potriviți" într-un studiu cu placebo și cum influențează analiza? Răspuns model: martorii au fost aleși să semene cu pacienții tratați pe anumite caracteristici (vârstă, sex, IMC). Dacă alocarea s-a făcut prin potrivire, nu prin randomizare, comparabilitatea este garantată doar pentru variabilele potrivite. Dacă potrivirea s-a făcut pe perechi, analiza ar trebui să țină cont de structura pereche.
3. De ce două serii de valori p pentru aceleași variabile (IMC 0,03 și 0,002)? Răspuns model: o serie corespunde, probabil, comparației între grupuri, iar cealaltă comparației înainte-după într-un grup. Doar comparația între grupuri la 3 luni (sau a schimbărilor între grupuri) testează efectul Aloe ferox, pentru că dieta a fost comună. O scădere semnificativă în grupul tratat, fără comparație cu placebo, poate fi efectul dietei.
4. Cu 20 de pacienți pe grup, ce mărime de efect putea detecta studiul cu putere 80%? Răspuns model: din formula pentru două medii, $d = (1,96 + 0,84) \times \sqrt{\frac{2}{20}} = 2,8 \times 0,316 = 0,89$. Studiul avea putere suficientă doar pentru efecte mari (aproape o abatere standard). Efecte moderate, clinic relevante, puteau trece nedetectate.

5. Calculați riscul relativ pentru exacerbări din studiul aplicației, dacă 10,30% este proporția din grupul cu aplicație. Răspuns model: RR $= \frac{10,30}{46,30} = 0,22$, adică o reducere relativă a riscului de 78% și o reducere absolută de 36 de puncte procentuale (NNT ≈ 3). Interpretarea trebuie însă însoțită de IC 95% și de verificarea modului de alocare: într-un pilot nerandomizat, pacienții care acceptă aplicația pot fi mai aderenți de la început.
6. De ce a fost necesară corecția Bonferroni la testul Friedman? Răspuns model: testul Friedman global spune doar că măcar un moment diferă de celelalte. Pentru a afla care, se fac comparații între perechi de momente (de exemplu, Wilcoxon); cu mai multe comparații, pragul fiecăreia se împarte la numărul lor, pentru a menține eroarea de tip I totală la 0,05.

14.3.10 Reanaliza completă a setului mg2026

Încheiem studiile de caz cu setul pe care l-ați folosit tot semestrul. Întrebările de mai jos parcurg întreaga disciplină pe aceleași date; răspunsurile reiau cifrele calculate în capitolele anterioare, astfel încât să vedeți cum se leagă.

1. Descrieți vârsta, IMC-ul și trigliceridele, justificând alegerea indicatorilor. Răspuns model: vârsta are o distribuție simetrică (asimetrie 0,01; Shapiro-Wilk p = 0,687): media 56,29 ani, AS 8,77. IMC-ul este asimetric spre dreapta (asimetrie 1,42; exces 3,68; Shapiro-Wilk W = 0,904, p < 0,001): mediana 28,25 kg/m², IQR 26,2-31,4. Trigliceridele sunt asimetrice (asimetrie 0,95; p < 0,001): mediana 151,5 mg/dL, IQR 107-203,25; după logaritmare distribuția devine aproape normală (asimetrie -0,28; p = 0,200), iar media geometrică este 144,0 mg/dL (IC 95% 132,95-156,02).
2. Estimați media TAS în populație și interpretați. Răspuns model: media 131,13 mmHg, AS 13,88, ES 1,097; IC 95% 128,96-133,30 mmHg. Cu 95% încredere, media TAS a populației din care provine eșantionul se află între aceste limite. Intervalul pentru valorile individuale este altceva: aproximativ 95% dintre pacienți au TAS între 103,9 și 158,3 mmHg.
3. Diferă TAS între femei și bărbați? Răspuns model: femei 131,37 (AS 14,47; n 86), bărbați 130,85 (AS 13,25; n 74); Levene p = 0,529; testul t Welch t = -0,24, gl 157,4, p = 0,81; diferența M-F -0,5 mmHg (IC 95% -4,9 la 3,8); d = -0,04. Nu respingem H0. Intervalul arată că o diferență mai mare de aproximativ 5 mmHg în oricare sens este puțin probabilă, informație mai utilă decât simplul "nesemnificativ".
4. A scăzut glicemia în cele 6 luni? Răspuns model: de la 105,4 la 95,2 mg/dL; diferența medie 10,19 mg/dL (AS 11,57; IC 95% 8,38-11,99); diferențele sunt normal distribuite (Shapiro-Wilk p = 0,725), deci t pereche: t = 11,14, gl 159, p < 0,001; dz = 0,88. Atenție: scăderea globală nu se poate atribui intervenției; pentru efectul intervenției comparăm brațele.
5. Diferă scăderea glicemiei între brațe? Răspuns model: ANOVA F(2; 157) = 22,93, p < 0,001, $\eta^2 = 0,226$; Tukey: B-A 8,1 mg/dL (IC 3,4-12,8), C-A 13,3 (8,6-18,0), C-B 5,2 (0,6-9,9). Toate brațele diferă între ele; efectul cel mai mare îl are combinația aplicație-teleconsultație.
6. Este IMC-ul asociat cu TAS? Răspuns model: r = 0,405, p < 0,001 (Spearman rho 0,372, preferabil dat fiind asimetria IMC); $R^2 = 0,164$. Dreapta de regresie: TAS = 89,47 + 1,435 x IMC (IC 95% pentru pantă 0,925-1,944): fiecare kg/m² în plus se asociază cu o TAS mai mare cu 1,4 mmHg; la IMC 30, TAS prezisă 132,5 mmHg. În regresia multiplă cu vârstă, sex și fumat, coeficientul IMC rămâne aproape neschimbat (1,389, p < 0,001), iar modelul explică 30,8% din variabilitatea TAS (R² ajustat 0,286); fumătorii au, la

același IMC, vârstă și sex, o TAS mai mare cu 6,7 mmHg decât foștii fumători (p = 0,012).

7. Este obezitatea asociată cu diabetul? Răspuns model: prevalența diabetului 33,3% la obezi față de 10,4% la neobezi; raportul prevalențelor 3,21 (IC 95% 1,64-6,31); OR 4,32 (IC 95% 1,86-10,03); χ^2 = 12,70, p < 0,001. Ajustat pentru vârstă și sex (regresie logistică), OR 6,01 (IC 95% 2,35-15,39). Design transversal: asociere, nu cauzalitate.
8. Cât de util este FINDRISC și cum se schimbă probabilitatea de diabet după rezultatul lui? Răspuns model: AUC 0,852 (IC 95% 0,762-0,943). La pragul ≥ 12: Se 75,9%, Sp 74,8%, RV+ 3,01, RV- 0,32. Pornind de la prevalența de 18,1% (cota pre-test 0,181/0,819 = 0,221), un scor pozitiv dă o cotă post-test 0,221 x 3,01 = 0,665, adică o probabilitate de 0,665/1,665 = 40%, identică cu VPP; un scor negativ dă cota 0,221 x 0,32 = 0,071, probabilitate 6,6%, adică 1 - VPN. Teorema lui Bayes (capitolul 11) și tabelul 2 x 2 (capitolul 12) spun același lucru. Într-o populație cu prevalență de 5%, același test ar avea VPP de doar 13,7%.

14.4 Cum se scrie paragraful de rezultate

Un rezultat statistic bine raportat răspunde la patru întrebări: ce s-a comparat (grupuri, variabila, unitatea), cât de mare este diferența sau asocierea, cât de precisă este estimarea și cât de compatibilă este cu întâmplarea. Ordinea firească este: descrierea în fiecare grup, mărimea efectului cu IC 95%, testul cu statistica lui, gradele de libertate și valoarea p. Concluzia clinică vine apoi, în discuție, nu amestecată cu cifrele.

Iată câteva perechi de exemple, toate pe datele mg2026.

Exemplul 1, slab: "Glicemia a scăzut semnificativ în brațul C (p < 0,05)." Problemele: nu spune față de ce (față de momentul inițial sau față de alt braț?), nu dă nicio cifră pentru mărimea scăderii, "p < 0,05" ascunde o valoare mult mai mică, iar testul nu este numit.

Exemplul 1, corect: "La 6 luni, scăderea medie a glicemiei a jeun a fost 3,0 mg/dL (AS 8,8) în brațul A, 11,1 mg/dL (AS 11,2) în brațul B și 16,3 mg/dL (AS 10,6) în brațul C (ANOVA, $F(2;\ 157) = 22{,}93$, $p < 0{,}001$; $\eta^2 = 0{,}23$). Față de brațul A, scăderea a fost mai mare cu 8,1 mg/dL în brațul B (IC 95% 3,4-12,8) și cu 13,3 mg/dL în brațul C (IC 95% 8,6-18,0) (testul Tukey)."

Exemplul 2, slab: "Nu există diferență de TAS între sexe (p = 0,81), deci sexul nu influențează tensiunea." Problemele: un rezultat nesemnificativ nu dovedește absența diferenței, iar a doua parte a frazei face o afirmație cauzală pe care un studiu transversal nu o poate susține.

Exemplul 2, corect: "TAS medie a fost 131,4 mmHg (AS 14,5) la femei și 130,9 mmHg (AS 13,3) la bărbați; diferența de -0,5 mmHg (IC 95% -4,9 la 3,8; testul t Welch, p = 0,81) nu a fost semnificativă statistic. Datele sunt incompatibile cu diferențe între sexe mai mari de aproximativ 5 mmHg."

Exemplul 3, slab: "Obezitatea crește riscul de diabet de 4,32 ori (p < 0,001)." Problemele: OR nu este risc relativ, "crește" sugerează cauzalitate, lipsește IC.

Exemplul 3, corect: "Prevalența diabetului confirmat prin TTGO a fost 33,3% (18/54) la pacienții obezi și 10,4% (11/106) la cei neobezi (raportul prevalențelor 3,21; IC 95% 1,64-6,31; χ^2 = 12,70, p < 0,001). Asocierea s-a menținut după ajustarea pentru vârstă și sex (OR 6,01; IC 95% 2,35-15,39)."

Exemplul 4, slab: "Corelația dintre IMC și TAS a fost foarte semnificativă (p < 0,001), deci IMC-ul determină TAS." Problemele: "foarte semnificativă" vorbește despre p, nu despre forța corelației; corelația nu arată cauzalitate.

Exemplul 4, corect: "IMC-ul a corelat pozitiv, moderat, cu TAS($r = 0{,}41$, $p < 0{,}001$; $R^2 = 0{,}16$). Fiecare creștere a IMC cu 1 kg/m² s-a asociat cu o TAS mai mare cu 1,4 mmHg (IC 95% 0,9-1,9)."

Exemplul 5, slab: "FINDRISC are o acuratețe de 75%, deci este un test bun." Problemele: acuratețea globală depinde de prevalență și ascunde compromisul dintre Se și Sp; "bun" nu are sens fără scopul testului.

Exemplul 5, corect: "Aria de sub curba ROC a scorului FINDRISC pentru diabetul confirmat prin TTGO a fost 0,852 (IC 95% 0,762-0,943). La pragul ≥ 12, sensibilitatea a fost 75,9% și specificitatea 74,8%; la prevalența din eșantion (18,1%), VPP a fost 40,0% și VPN 93,3%."

Câteva reguli de formă completează exemplele. Raportați valorile p exacte (p = 0,034), cu excepția celor foarte mici (p < 0,001); nu scrieți "p = 0,000", pe care unele programe îl afișează. Folosiți un număr rezonabil de zecimale: un procent dintr-un eșantion de 50 nu se scrie cu trei zecimale. Numele testului apare o dată, la prima folosire. Unitatea de măsură apare de fiecare dată când apare o valoare. Media este însoțită de AS când descrieți datele și de IC când estimați un parametru.

> **Studiu de caz: un rezumat cu multe rapoarte ale cotelor și fără intervale de încredere**
>
> Ursoiu și colaboratorii au analizat asocierea dintre severitatea clinică a infecției HIV și tulburările psihiatrice la 101 pacienți cu tratament antiretroviral din vestul României [84]. Rezumatul raportează șapte rapoarte ale cotelor, toate semnificative: OR = 5,81 (p < 0,001), 3,11 (p = 0,016), 4,31 (p = 0,024), 2,81 (p = 0,038), 4,79 (p = 0,006), 8,72 (p < 0,001) și 4,66 (p = 0,0013). Ca exercițiu de redactare, observați ce lipsește și cum s-ar putea completa. Lipsesc intervalele de încredere: cu 101 pacienți, un OR de 8,72 are, aproape sigur, un interval foarte larg, iar cititorul nu știe dacă limita inferioară este 1,2 sau 5. Lipsește precizarea dacă OR-urile sunt brute sau ajustate. Se testează multe asocieri pe același eșantion, deci riscul ca una dintre ele să fie fals pozitivă nu este neglijabil; câteva valori p (0,024; 0,038) nu ar rezista unei corecții Bonferroni pentru șapte teste (prag 0,05/7 = 0,007). Iar designul transversal nu permite să spunem dacă boala avansată duce la tulburări psihiatrice sau invers. Un paragraf de rezultate rescris ar da, pentru fiecare asociere, frecvențele în cele două grupuri, OR cu IC 95%, valoarea p și, la final, o mențiune despre comparațiile multiple. Nu este o critică a autorilor: limitele de spațiu ale unui rezumat obligă la compromisuri, iar textul complet poate conține informațiile lipsă. Pentru cititor, însă, regula rămâne: un OR fără interval de încredere este o informație incompletă.

14.5 Patruzeci de afirmații: adevărat sau fals

Afirmațiile de mai jos acoperă tot cursul. Încercați să decideți singuri înainte de a citi justificarea; la examenul grilă, multe variante greșite sunt construite exact pe aceste confuzii.

1. Valoarea p este probabilitatea ca H0 să fie adevărată. Fals. Valoarea p este probabilitatea de a obține un rezultat cel puțin la fel de extrem ca cel observat, presupunând că H0 este adevărată; nu este o probabilitate a ipotezei [70].
2. Un IC 95% pentru o diferență de medii care nu include 0 corespunde unui test bilateral semnificativ la α = 0,05. Adevărat. Intervalul și testul folosesc aceeași eroare standard și aceeași valoare critică.
3. Mediana este mai puțin influențată de valorile extreme decât media. Adevărat. În mg2026, o singură valoare TAS de 1540 mmHg (în loc de 154) mută media cu 8,66 mmHg, iar mediana rămâne 131.
4. Abaterea standard scade când crește dimensiunea eșantionului. Fals. Abaterea standard estimează variabilitatea din populație și rămâne, în medie, aceeași; eroarea standard este cea care scade.

5. Eroarea standard a mediei scade proporțional cu $\sqrt{n}$. Adevărat. ES = $\frac{s}{\sqrt{n}}$; pentru a o înjumătăți, n trebuie să crească de 4 ori.
6. Dacă dublăm eșantionul, lățimea intervalului de încredere se înjumătățește. Fals. Lățimea scade cu factorul √2, adică cu aproximativ 29%; înjumătățirea cere un eșantion de 4 ori mai mare.
7. Testul t pereche este potrivit pentru compararea TAS între femei și bărbați. Fals. Femeile și bărbații sunt grupuri independente; se folosește testul t pentru eșantioane independente (sau Welch, Mann-Whitney).
8. O ANOVA semnificativă arată care dintre grupuri diferă între ele. Fals. ANOVA spune doar că cel puțin o medie diferă; perechile se identifică prin teste post-hoc (Tukey, Bonferroni).
9. Testul chi-pătrat cere frecvențe așteptate suficient de mari, de regulă cel puțin 5 în fiecare celulă. Adevărat. Altfel se folosește testul exact Fisher (la tabele 2 x 2).
10. Un coeficient de corelație r = 0,9 arată că X cauzează Y. Fals. Corelația măsoară asocierea liniară; direcția cauzală sau un al treilea factor nu pot fi excluse.
11. $R^2 = 0{,}164$ pentru regresia TAS pe IMC înseamnă că 16,4% din variabilitatea TAS este explicată liniar de IMC. Adevărat. Restul de 83,6% ține de alți factori și de variabilitatea individuală.
12. Coeficientul Spearman este potrivit pentru variabile ordinale. Adevărat. Lucrează pe ranguri și nu cere normalitate.
13. Într-un studiu caz-control putem calcula direct riscul relativ. Fals. Proporția de bolnavi este fixată de cercetător (numărul de cazuri și de martori), deci riscurile nu se pot estima; se calculează OR.
14. Când boala este rară, OR aproximează bine RR. Adevărat. Când a și c sunt mici față de b și d, cotele sunt apropiate de riscuri.
15. NNT este inversul reducerii relative a riscului. Fals. NNT = 1 / reducerea absolută a riscului.
16. Un RR = 1,02 cu IC 95% 0,80-1,30 dovedește că expunerea nu are niciun efect. Fals. Datele sunt compatibile cu o reducere de 20% sau cu o creștere de 30%; nu avem dovezi de efect, dar nici dovada absenței lui [79].
17. La sensibilitate și specificitate constante, VPP crește când crește prevalența. Adevărat. Pentru FINDRISC ≥ 12, VPP crește de la 5,8% la prevalența de 2% la 75,1% la prevalența de 50%.
18. Sensibilitatea unui test se modifică direct cu prevalența bolii. Fals. Se și Sp sunt proprietăți ale testului în raport cu bolnavii, respectiv sănătoșii; se pot schimba cu spectrul bolii, nu cu prevalența ca atare.
19. Un rezultat negativ la un test cu sensibilitate foarte mare ajută la excluderea bolii. Adevărat. Testul ratează puțini bolnavi, deci un negativ are VPN mare.
20. $RV^+ = \frac{Se}{1-Sp}$. Adevărat. Pentru FINDRISC ≥ 12: 0,759 / 0,252 = 3,01.
21. Rata de eroare a clasificării este (FP + FN) / N. Adevărat. Pentru FINDRISC ≥ 12: (33 + 7)/160 = 25%, complementul acurateței de 75%.
22. O arie sub curba ROC de 0,5 corespunde unui test fără valoare discriminativă. Adevărat. Testul nu face mai mult decât aruncarea unei monede.
23. Randomizarea garantează că brațele sunt identice la includere. Fals. Le echilibrează în medie; diferențe întâmplătoare rămân (în mg2026, 9,4% foști fumători în A față de 31,5% în C).
24. Orbirea și ascunderea alocării sunt același lucru. Fals. Ascunderea alocării protejează includerea, orbirea protejează tratamentul și evaluarea.
25. Analiza după intenția de tratament include toți pacienții randomizați, în brațul alocat. Adevărat. Așa se păstrează comparabilitatea dată de randomizare.
26. Un studiu transversal poate stabili ce a apărut întâi, expunerea sau boala. Fals. Expunerea și boala sunt măsurate în același moment.

27. Puterea unui test crește cu dimensiunea eșantionului. Adevărat. Eroarea standard scade, deci efectele reale sunt detectate mai ușor.
28. Eroarea de tip I înseamnă respingerea unei H0 adevărate. Adevărat. Probabilitatea ei este α.
29. Dacă p = 0,06, H0 este adevărată. Fals. Spunem doar că nu respingem H0 la pragul de 0,05; H0 poate fi falsă, iar studiul fără putere suficientă.
30. Un test Shapiro-Wilk nesemnificativ dovedește că datele sunt normale. Fals. Nu respingem normalitatea; la eșantioane mici testul are putere mică și poate rata abateri importante.
31. Logaritmarea poate face simetrică o distribuție cu asimetrie pozitivă. Adevărat. Trigliceridele din mg2026 trec de la asimetrie 0,95 la -0,28.
32. Media ± 1,96 AS este intervalul de încredere pentru medie. Fals. Este intervalul în care se găsesc aproximativ 95% dintre valorile individuale; IC pentru medie folosește eroarea standard.
33. Testul Kruskal-Wallis este alternativa neparametrică a ANOVA unifactoriale. Adevărat. Compară trei sau mai multe grupuri independente pe baza rangurilor.
34. Testul Friedman se folosește pentru grupuri independente. Fals. Este pentru trei sau mai multe măsurători pe aceiași subiecți (date dependente).
35. Într-un eșantion foarte mare, o diferență neglijabilă clinic poate ieși semnificativă statistic. Adevărat. Corelația r = -0,088 dintre vârstă și durata intervenției, la 867 de pacienți, are p = 0,01 [55].
36. Un α Cronbach de 0,93 dovedește validitatea unui chestionar. Fals. Alfa măsoară consistența internă (fidelitatea), nu validitatea.
37. Panta unei drepte de regresie are aceeași unitate de măsură ca variabila dependentă. Fals. Are unitatea lui Y pe unitatea lui X (de exemplu, mmHg pe kg/m²).
38. Coeficientul r nu se schimbă dacă exprimăm înălțimea în metri în loc de centimetri. Adevărat. r nu are unitate de măsură și este invariant la transformări liniare ale variabilelor.
39. În Tabelul 1 al unui RCT, diferențele dintre brațe trebuie testate obligatoriu cu valori p. Fals. CONSORT recomandă să nu se testeze; diferențele sunt, prin construcție, întâmplătoare [56].
40. Testul chi-pătrat semnificativ pentru scădere ≥ 5% pe cele trei brațe ale mg2026 (p = 0,034) arată că toate brațele diferă între ele. Fals. Arată că măcar unul diferă; B și C au proporții aproape identice (RR C față de B 0,98, IC 95% 0,55-1,75).

14.6 Tabelul-sinteză al formulelor

Tabelul 14.2 reunește formulele de lucru din curs, cu trimitere la capitolul în care au fost explicate. Notațiile: n mărimea eșantionului, m media, s abaterea standard, p proporția, a, b, c, d celulele tabelului 2 x 2 (a expuși bolnavi sau test pozitiv bolnavi, b expuși sănătoși sau test pozitiv sănătoși, c neexpuși bolnavi sau test negativ bolnavi, d neexpuși sănătoși sau test negativ sănătoși).

Tabelul 14.2. Formulele principale ale cursului.

Indicator	Formula	Capitolul
Media	$m = \frac{\sum x}{n}$	2
Abaterea standard	$s = \sqrt{\left[\frac{\sum (x-m)^2}{n-1}\right]}$	2
Coeficientul de variație	CV = 100 x s / m (%)	2
Scorul z	$z = \frac{x-m}{s}$	3
Intervalul pentru valori individuale (95%)	$m \pm 1{,}96 \times s$	3

Indicator	Formula	Capitolul
Eroarea standard a mediei	$\text{ES} = \frac{s}{\sqrt{n}}$	4
IC 95% pentru medie	$m \pm \frac{t(n-1;\ 0{,}975) \times s}{\sqrt{n}}$	4
IC 95% pentru o proporție (Wald)	$p_i \pm 1{,}96 \times \sqrt{\left[\frac{p_i(1-p_i)}{n}\right]}$	4
Eșantion pentru o medie cu marja E	$n = \left(\frac{1{,}96 \times \sigma}{E}\right)^2$	4
Eșantion pentru o proporție cu marja E	$n = \frac{1{,}96^2 \times p_i(1-p_i)}{E^2}$	4
Eșantion pe grup, două medii	$n = \frac{2 \times \left(z\left(1-\frac{\alpha}{2}\right) + z(1-\beta)\right)^2 \times \sigma^2}{\Delta^2}$	4, 13
Testul t pentru eșantioane independente	$t = \frac{m_1 - m_2}{\sqrt{s^2 p \times \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$	6
Testul t pereche	$t = \frac{\text{md}}{\frac{\text{sd}}{\sqrt{n}}}$, gl = n - 1	6
d Cohen	$d = \frac{m_1 - m_2}{\text{sp}}$	6
ANOVA	$F = \frac{\frac{\text{SSB}}{k-1}}{\frac{\text{SSW}}{N-k}}$	6
Eta pătrat	$\eta^2 = \frac{\text{SSB}}{\text{SSB} + \text{SSW}}$	6
Frecvența așteptată (chi-pătrat)	E = total rând x total coloană / N	7
Chi-pătrat	$\chi^2 = \frac{\sum(O-E)^2}{E}$, gl = (r - 1)(c - 1)	7
Testul pentru r	$t = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}}$	9
Coeficientul de determinare	$R^2 = r^2$ (regresie simplă)	9
Dreapta de regresie	Y = a + b x X; $b = \frac{\sum(x-m_x)(y-m_y)}{\sum(x-m_x)^2}$; a = my - b x mx	9
Riscul relativ (cohortă, RCT)	$\text{RR} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$	10
Raportul cotelor	$\text{OR} = \frac{a \times d}{b \times c}$	10
IC 95% pentru RR	$\exp\left[\ln \text{RR} \pm 1{,}96 \times \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}\right]$	10
IC 95% pentru OR	$\exp\left[\ln \text{OR} \pm 1{,}96 \times \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\right]$	10
Diferența de risc	$\text{DR} = \frac{a}{a+b} - \frac{c}{c+d}$	10
NNT (NNH)	$\text{NNT} = \frac{1}{\text{DR}}$, rotunjit în sus	10, 13
Fracția atribuibilă la expuși	$\text{FA} = \frac{\text{RR} - 1}{\text{RR}}$	10
Cota și probabilitatea	cota = P / (1 - P); P = cota / (1 + cota)	11

Indicator	Formula	Capitolul
Teorema lui Bayes (VPP)	$\text{VPP} = \frac{\text{Se} \times P}{\text{Se} \times P + (1 - \text{Sp}) \times (1 - P)}$	11
Cota post-test	cota post-test = cota pre-test x RV	11
Sensibilitatea	$\text{Se} = \frac{a}{a + c}$	12
Specificitatea	$\text{Sp} = \frac{d}{b + d}$	12
Valoarea predictivă pozitivă	$\text{VPP} = \frac{a}{a + b}$	12
Valoarea predictivă negativă	$\text{VPN} = \frac{d}{c + d}$	12
Acuratețea	$\frac{a + d}{N}$	12
Rata de eroare	$\frac{b + c}{N}$	12
Rapoartele de verosimilitate	$\text{RV}^+ = \frac{\text{Se}}{1-\text{Sp}}$; $\text{RV}^- = \frac{1-\text{Se}}{\text{Sp}}$	12
Indicele Youden	$J = \text{Se} + \text{Sp} - 1$	12
Alfa Cronbach	$\alpha = \frac{k}{k-1} \times \left[1 - \frac{\sum s^2(\text{item})}{s^2(\text{total})}\right]$	13
Efectul de design (clustere)	1 + (m - 1) x ICC	13

Tabelul nu se învață pe de rost ca o listă. Fiecare formulă are o logică: eroarea standard este variabilitatea împărțită la rădăcina din n; un test statistic este un efect împărțit la eroarea lui standard; un interval de încredere este estimarea plus-minus o valoare critică înmulțită cu eroarea standard; indicatorii testelor diagnostice sunt proporții calculate pe rândurile sau pe coloanele aceluiași tabel 2 x 2. Cine înțelege aceste patru idei poate reconstitui aproape tot tabelul.

14.7 Probleme rezolvate

Problema rezolvată 14.1. Tabelul de frecvențe cu valori lipsă

Enunț. Într-un chestionar aplicat la 200 de pacienți dintr-un cabinet de medicină de familie, întrebarea despre nivelul de educație a primit următoarele răspunsuri: gimnaziu 30, liceu sau școală profesională 74, universitar 62, postuniversitar 22, fără răspuns 12. Construiți tabelul de frecvențe, cu procente din total și procente valide, și precizați categoria mediană.

Rezolvare. Numărul de răspunsuri valide este 200 - 12 = 188. Procentul din total se calculează la 200, procentul valid la 188, iar procentul cumulat pe procentele valide.

Tabelul 14.3. Distribuția nivelului de educație (n = 200, 12 valori lipsă).

Educația	Frecvența	% din total	% valid	% valid cumulat
Gimnaziu	30	15,0	16,0	16,0
Liceu sau școală profesională	74	37,0	39,4	55,3
Universitar	62	31,0	33,0	88,3
Postuniversitar	22	11,0	11,7	100,0
Lipsă	12	6,0	-	-
Total	200	100,0	100,0	

Categoria mediană este cea în care procentul valid cumulat depășește pentru prima dată 50%: liceu sau școală profesională (55,3%). Variabila este ordinală, deci are sens o mediană, nu o

medie. Procentele valide descriu respondenții; dacă lipsa răspunsului nu este întâmplătoare (de exemplu, persoanele cu educație mai redusă evită întrebarea), procentele valide pot fi distorsionate, iar numărul valorilor lipsă trebuie raportat întotdeauna.

Problema rezolvată 14.2. Un studiu caz-control: contraceptivele orale și artrita reumatoidă

Enunț. Un studiu caz-control include 120 de femei cu artrită reumatoidă și 240 de femei fără boală, de vârstă comparabilă. Au folosit contraceptive orale în trecut 42 dintre cazuri și 54 dintre martori. Calculați măsura potrivită a asocierii, cu IC 95%, și testați asocierea.

Rezolvare. (a) Ipoteza de cercetare: folosirea contraceptivelor orale este asociată cu artrita reumatoidă. (b) H0: OR = 1 (nu există asociere); H1: OR ≠ 1. (c) Design caz-control, deci măsura este OR; testul este chi-pătrat pentru tabel 2 x 2. Tabelul: a = 42 (cazuri expuse), b = 54 (martori expuși), c = 78 (cazuri neexpuse), d = 186 (martori neexpuși).

$$\text{OR} = \frac{42\times186}{54\times78} = \frac{7812}{4212} = 1{,}85;\ \text{ES}(\ln \text{OR}) = \sqrt{\frac{1}{42}+\frac{1}{54}+\frac{1}{78}+\frac{1}{186}} =$$

$$\sqrt{0{,}0238 + 0{,}0185 + 0{,}0128 + 0{,}0054} = \sqrt{0{,}0605} = 0{,}246 \ln 1{,}85 = 0{,}618;\quad 0{,}618 \pm 1{,}96 \times 0{,}246 = 0{,}618 \pm 0{,}482,$$ adică 0,136 la 1,100; $\text{IC}\,95\% = e^{0{,}136}$ la $e^{1{,}100} = 1{,}15$ la 3,00

Chi-pătrat (fără corecție) = 6,39, gl = 1, p = 0,011. (e) Rezultat: OR 1,85 (IC 95% 1,15-3,00). (f) Concluzia statistică: respingem H0; intervalul nu include 1. (g) Concluzia de cercetare: cota de folosire a contraceptivelor orale este de aproximativ 1,9 ori mai mare la femeile cu artrită reumatoidă decât la martori. Un studiu caz-control nu permite calculul riscului relativ, iar asocierea poate fi influențată de eroarea de memorie (cazurile își amintesc altfel expunerile trecute) și de factori de confuzie.

Problema rezolvată 14.3. O cohortă: timpul petrecut la monitor și acuitatea vizuală

Enunț. O cohortă de 400 de angajați este urmărită 3 ani. Dintre cei 150 care lucrează cel puțin 6 ore pe zi la monitor, 45 prezintă o scădere a acuității vizuale; dintre cei 250 care lucrează mai puțin, 30. Calculați RR, DR, NNH și fracția atribuibilă la expuși.

Rezolvare. Riscul la expuși: 45/150 = 0,30; la neexpuși: 30/250 = 0,12.

$$\text{RR} = \frac{0{,}30}{0{,}12} = 2{,}5;\ \text{ES}(\ln \text{RR}) = \sqrt{\frac{1}{45}-\frac{1}{150}+\frac{1}{30}-\frac{1}{250}} =$$

$$\sqrt{0{,}0222 - 0{,}0067 + 0{,}0333 - 0{,}0040} = \sqrt{0{,}0449} = 0{,}212;\ \text{IC}\,95\%\ \text{RR} = \exp(\ln 2{,}5 \pm 1{,}96 \times 0{,}212) = \exp(0{,}916 \pm 0{,}415) = 1{,}65 \text{ la } 3{,}79$$ DR = 0,30 - 0,12 = 0,18; $\text{NNH} = \frac{1}{0{,}18} = 5{,}6$, adică 6; $\text{FA} = \frac{2{,}5-1}{2{,}5} = 0{,}60$

Chi-pătrat = 19,94, p < 0,001. OR-ul ar fi $\frac{(45\times220)}{105\times30} = 3{,}14$, mai mare decât RR, cum se întâmplă când evenimentul este frecvent. Interpretare: riscul de scădere a acuității este de 2,5 ori mai mare la cei expuși; la fiecare 6 angajați expuși apare un caz în plus față de neexpuși; 60% din cazurile apărute la expuși ar putea fi atribuite expunerii, dacă relația este cauzală. Fiind o cohortă observațională, angajații expuși pot diferi și prin vârstă sau prin alte caracteristici, deci o analiză ajustată este necesară înainte de o concluzie cauzală.

Problema rezolvată 14.4. Reconstituirea tabelului unui test diagnostic

Enunț. Un test de screening este aplicat la 500 de persoane, dintre care 20% au boala. Sensibilitatea testului este 85%, iar VPP 68%. Reconstituiți tabelul 2 x 2 și calculați Sp, VPN, acuratețea, rata de eroare și rapoartele de verosimilitate.

Rezolvare. Bolnavi: 0,20 x 500 = 100; sănătoși: 400. Adevărat pozitivi: 0,85 x 100 = 85; fals negativi: 15. Din $\text{VPP} = \frac{\text{AP}}{\text{AP+FP}} = 0{,}68$ rezultă $\text{AP} + \text{FP} = \frac{85}{0{,}68} = 125$, deci FP = 40. Adevărat negativi: 400 - 40 = 360.

Tabelul 14.4. Tabelul reconstituit al testului de screening.

	Bolnavi	**Sănătoși**	**Total**
Test pozitiv	85	40	125
Test negativ	15	360	375
Total	100	400	500

$\text{Sp} = \frac{360}{400} = 90\%$; $\text{VPN} = \frac{360}{375} = 96\%$; acuratețea = (85 + 360)/500 = 89%; rata de eroare = (40 + 15)/500 = 11%; $\text{RV}^+ = \frac{0{,}85}{0{,}10} = 8{,}5$; $\text{RV}^- = \frac{0{,}15}{0{,}90} = 0{,}17$. Verificare prin Bayes: cota pre-test 0,20/0,80 = 0,25; cota post-test pentru un rezultat pozitiv 0,25 x 8,5 = 2,125; probabilitatea 2,125/3,125 = 68%, adică exact VPP-ul din enunț.

Problema rezolvată 14.5. Alegerea testului pe setul mg2026

Enunț. Pentru fiecare întrebare, alegeți testul, justificați și dați rezultatul din setul mg2026: a) diferă glicemia inițială între bărbați și femei? b) este nivelul de educație asociat cu satisfacția? c) diferă TAS inițială între fumători, foști fumători și nefumători? d) s-a schimbat greutatea în brațul A? e) este fumatul asociat cu HTA (TAS ≥ 140) la includere?

Rezolvare. a) Variabilă numerică, două grupuri independente; glicemia are o ușoară abatere de la normalitate (SW p = 0,037), dar cu 74 și 86 de pacienți testul t Welch este robust: 106,2 față de 104,8 mg/dL, t = 0,54, p = 0,589; nu respingem H0. b) Două variabile ordinale: corelația Spearman, rho = 0,125, p = 0,115; nu există o asociere semnificativă. c) Variabilă numerică, trei grupuri independente; normalitatea este acceptabilă în fiecare grup (SW p 0,37; 0,86; 0,23), Levene p = 0,114: ANOVA, F = 7,85, p < 0,001, $\eta^2 = 0{,}09$; Tukey: fumătorii au TAS mai mare cu 9,5 mmHg decât foștii fumători (IC 95% 2,5-16,5) și cu 8,5 mmHg decât nefumătorii (IC 95% 2,9-14,2). d) Date pereche (aceiași pacienți la 0 și 6 luni), un singur grup; testul Wilcoxon: mediana diferențelor 1,1 kg, p = 0,052, rezultat la limită, care nu permite respingerea H0 la α = 0,05. e) Două variabile dihotomice, design transversal: chi-pătrat, 21/53 (39,6%) dintre fumători față de 20/107 (18,7%) dintre nefumători și foști fumători au HTA; $\chi^2 = 8{,}15$, p = 0,004; raportul prevalențelor 2,12 (IC 95% 1,26-3,55), OR 2,85 (IC 95% 1,37-5,95).

14.8 Exerciții propuse

1. Într-un studiu pe 250 de adulți, statutul de fumător a fost: nefumători 120, fumători 70, foști fumători 40, fără răspuns 20. Construiți tabelul de frecvențe cu procente din total și procente valide.
2. Un studiu caz-control include 80 de pacienți cu BPOC și 160 de martori. Sunt fumători 56 dintre cazuri și 64 dintre martori. Calculați OR cu IC 95% și chi-pătrat.
3. Într-o cohortă, din 200 de expuși la un factor profesional se îmbolnăvesc 30, iar din 300 de neexpuși 15. Calculați RR cu IC 95%, DR și NNH.
4. Un test cu Se 90% și Sp 95% este aplicat în screening la 10.000 de persoane, cu o prevalență a bolii de 2%. Reconstituiți tabelul 2 x 2 și calculați VPP și VPN. Comentați.

5. La 36 de pacienți, media colesterolului este 220 mg/dL, iar AS 18 mg/dL. Calculați eroarea standard și IC 95% pentru medie (t(35; 0,975) = 2,03).
6. La 27 de pacienți, corelația dintre două variabile numerice este r = 0,6. Calculați R^2, statistica t și precizați dacă r este semnificativ la α = 0,05 (valoarea critică t(25) ≈ 2,06).
7. Folosind dreapta de regresie din mg2026 (TAS = 89,47 + 1,435 x IMC), calculați TAS prezisă la un pacient cu IMC 27 kg/m^2. De ce nu ar fi corectă o predicție la IMC 15?
8. Probabilitatea pre-test a unei boli este 30%, iar RV+ al unui test este 6. Calculați probabilitatea post-test după un rezultat pozitiv.
9. Alegeți testul pentru: a) compararea HbA1c între trei brațe ale unui RCT, cu distribuții normale; b) compararea scorului de durere (1-10) înainte și după o intervenție, la aceiași pacienți; c) asocierea dintre sex și prezența diabetului; d) relația dintre vârstă și TAS; e) compararea unei variabile asimetrice între două grupuri independente; f) evoluția unui scor la 0, 3 și 6 luni la aceiași pacienți, distribuție asimetrică.
10. Scrieți paragraful de rezultate pentru următoarea analiză din mg2026: HDL la femei $53{,}92$ mg/dL$(\text{AS } 11{,}70;\ n = 86)$, la bărbați $44{,}41$ mg/dL$(\text{AS } 9{,}66;\ n = 74)$; Levene p = 0,034; testul t Welch t = 5,63, gl 157,7, p < 0,001; diferența 9,5 mg/dL (IC 95% 6,2-12,9).
11. Rescrieți corect afirmația: "Testul chi-pătrat a arătat că obezitatea cauzează diabet (p < 0,001), riscul fiind de 4,32 ori mai mare."
12. Într-un articol se raportează: "durata spitalizării 6,2 ± 5,9 zile în grupul A și 4,1 ± 4,4 zile în grupul B (testul t, p = 0,04)". Ce observați și ce ați fi recomandat?

Capcane frecvente

- A sări peste metode și a citi doar concluzia articolului.
- A interpreta "nesemnificativ" ca "nu există efect", fără a privi intervalul de încredere.
- A confunda semnificația statistică cu importanța clinică, mai ales în eșantioanele mari.
- A citi un OR ca pe un risc relativ când evenimentul este frecvent.
- A transfera VPP și VPN dintr-un studiu cu prevalență mare la o populație de screening cu prevalență mică.
- A crede o specificitate de 100% obținută pe câteva zeci de voluntari sănătoși.
- A nu verifica dacă intervalele de încredere publicate sunt plauzibile pentru n raportat.
- A trage concluzii cauzale din studii transversale sau din corelații.
- A descrie prin medie ± AS variabile vizibil asimetrice (AS apropiată de medie pentru o variabilă pozitivă).
- A schimba testul după rezultat sau a raporta doar comparațiile semnificative.
- A uita că un HR presupune urmărire în timp și nu poate proveni dintr-un design pur transversal.

De reținut

- Harta disciplinei: întrebare, design, date, descriere, estimare, testare, asociere, decizie, raportare.
- Alegerea testului depinde de tipul variabilei rezultat, de numărul de grupuri și de independența lor.
- Un rezultat complet are trei părți: mărimea efectului, intervalul de încredere și valoarea p.
- Intervalul de încredere arată precizia; un interval larg înseamnă un studiu care nu a decis mare lucru.
- Valorile predictive depind de prevalență; Se, Sp și RV sunt mai stabile între populații.
- Confuzia se combate prin randomizare (în proiectare) sau prin ajustare (în analiză).
- Multe cifre publicate pot fi verificate cu formulele simple ale cursului; faceți acest exercițiu de fiecare dată când ceva pare prea frumos.

Întrebări de autoevaluare

Grila recapitulativă de mai jos are 30 de întrebări în stilul examenului de rezidențiat: întrebările 1-15 au un singur răspuns corect (CS), întrebările 16-30 au între două și patru răspunsuri corecte (CM). Răspunsurile se află în fișierul de răspunsuri al capitolului.

1. (CS) Pentru descrierea IMC în setul mg2026 (asimetrie 1,42, Shapiro-Wilk $p < 0,001$), cei mai potriviți indicatori sunt:
    a) media și abaterea standard
    b) mediana și intervalul intercuartil
    c) modul
    d) amplitudinea
    e) coeficientul de variație
2. (CS) Intervalul de încredere de 95% pentru media TAS în mg2026 este 128,96-133,30 mmHg. Afirmația corectă este:
    a) 95% dintre pacienți au TAS între 128,96 și 133,30 mmHg
    b) media eșantionului se află în interval cu probabilitatea de 95%
    c) procedura folosită produce intervale care conțin media populației în 95% dintre eșantioanele posibile
    d) TAS este semnificativ diferită de 130 mmHg
    e) distribuția TAS este normală
3. (CS) Pentru compararea glicemiei la includere și la 6 luni la aceiași 160 de pacienți, cu diferențe normal distribuite, se folosește:
    a) testul t pentru eșantioane independente
    b) testul t pentru eșantioane pereche
    c) ANOVA unifactorială
    d) testul chi-pătrat
    e) testul Mann-Whitney
4. (CS) Satisfacția (item Likert 1-5) se compară între cele trei brațe ale studiului mg2026 cu:
    a) ANOVA unifactorială
    b) testul Kruskal-Wallis
    c) testul t pereche
    d) testul Friedman
    e) corelația Pearson
5. (CS) Un studiu urmărește timp de 10 ani 500 de fumători și 500 de nefumători, înregistrând apariția BPOC. Designul este:
    a) caz-control
    b) transversal
    c) cohortă prospectivă
    d) studiu randomizat controlat
    e) studiu ecologic
6. (CS) Într-un studiu caz-control, măsura potrivită a asocierii dintre expunere și boală este:
    a) riscul relativ
    b) raportul cotelor
    c) numărul necesar de tratat

d) prevalența
e) incidența cumulată

7. (CS) Corelația r = -0,832 dintre osteoprotegerină și fracția de ejecție indică:
   a) o corelație slabă
   b) o corelație inversă puternică
   c) faptul că osteoprotegerina scade fracția de ejecție
   d) un coeficient de determinare de 0,832
   e) o asociere nesemnificativă
8. (CS) Reducerea absolută a riscului unui eveniment este 10 puncte procentuale. NNT este:
   a) 1
   b) 5
   c) 10
   d) 18
   e) 100
9. (CS) Un test cu Se 90% și Sp 80% este aplicat la 1000 de persoane, cu o prevalență a bolii de 15%. VPP este aproximativ:
   a) 15%
   b) 44%
   c) 80%
   d) 90%
   e) 98%
10. (CS) Rata de eroare a clasificării unui test diagnostic, cu notațiile tabelului 2 x 2, este:

a) $\frac{a+d}{N}$
b) $\frac{b+c}{N}$
c) $\frac{a}{a+c}$
d) $\frac{d}{b+d}$
e) $\frac{a}{a+b}$

11. (CS) O ANOVA cu trei grupuri dă p < 0,001. Pasul următor corect este:

a) teste t între toate perechile, fără corecție
b) teste post-hoc, de exemplu Tukey
c) testul chi-pătrat
d) concluzia că toate grupurile diferă între ele
e) repetarea ANOVA pe alt eșantion înainte de orice interpretare

12. (CS) Într-un studiu randomizat de superioritate, analiza principală recomandată este:

a) analiza per protocol
b) analiza după intenția de tratament
c) analiza doar a pacienților care au terminat studiul
d) analiza doar a pacienților aderenți
e) analiza pe subgrupuri

13. (CS) Pentru criteriul secundar TAS din mg2026, ANOVA prevăzută în plan dă p = 0,056. Formularea corectă este:

a) se acceptă H0, intervenția nu influențează TAS

b) nu se respinge H0; datele nu oferă dovezi suficiente pentru o diferență între brațe

c) există un efect semnificativ al intervenției

d) se folosește Kruskal-Wallis, care dă p < 0,05, și se raportează acesta

e) rezultatul dovedește echivalența brațelor

14. (CS) Într-o cohortă, 18 din 150 de expuși și 10 din 250 de neexpuși dezvoltă boala. Riscul relativ este:

a) 1,8

b) 3,0

c) 3,27

d) 0,33

e) 8,0

15. (CS) La pragul FINDRISC ≥ 12, VPP este 40,0% la o prevalență de 18,1%. Într-o populație cu prevalența 5%, la aceleași Se și Sp, VPP va fi:

a) mai mare de 40%

b) mai mic de 40%

c) egal cu 40%

d) 95%

e) imposibil de estimat

16. (CM) Despre medie și mediană, sunt adevărate:

a) mediana este robustă la valori extreme

b) într-o distribuție cu asimetrie pozitivă, media este de obicei mai mare decât mediana

c) într-o distribuție normală, media și mediana coincid

d) mediana se calculează doar pentru variabile nominale

e) media este indicatorul recomandat pentru un singur item Likert

17. (CM) Despre valoarea p, sunt adevărate:

a) este probabilitatea de a obține un rezultat cel puțin la fel de extrem dacă H0 este adevărată

b) este probabilitatea ca H0 să fie adevărată

c) depinde de mărimea eșantionului

d) o valoare foarte mică garantează un efect important clinic

e) se compară cu pragul de semnificație α

18. (CM) Despre erorile statistice, sunt adevărate:

a) eroarea de tip I înseamnă respingerea unei H0 adevărate

b) α este probabilitatea erorii de tip I

c) puterea testului este 1 - β

d) creșterea eșantionului scade puterea

e) eroarea de tip II înseamnă nerespingerea unei H0 false

19. (CM) Condițiile de aplicare ale testului t pentru eșantioane independente sunt:

a) normalitatea în fiecare grup sau eșantioane suficient de mari

b) independența observațiilor

c) varianțe egale, altfel se folosește varianta Welch
d) variabile nominale
e) aceiași pacienți măsurați de două ori

20. (CM) Despre testul chi-pătrat, sunt adevărate:
a) se aplică variabilelor categoriale
b) cere frecvențe așteptate suficient de mari
c) la frecvențe așteptate mici în tabele 2 x 2 se folosește testul Fisher
d) compară medii
e) frecvența așteptată = total rând x total coloană / N

21. (CM) Pentru dreapta de regresie TAS = 89,47 + 1,435 x IMC din mg2026, sunt adevărate:
a) TAS crește în medie cu 1,435 mmHg pentru fiecare kg/m² de IMC
b) la IMC 30, TAS prezisă este aproximativ 132,5 mmHg
c) interceptul 89,47 mmHg are o interpretare clinică directă
d) cu $R^2 = 0{,}164$, IMC explică liniar 16,4% din variabilitatea TAS
e) ecuația dovedește că IMC-ul crescut cauzează hipertensiunea

22. (CM) Despre RR și OR, sunt adevărate:
a) RR se calculează în studiile de cohortă și în RCT
b) OR este măsura potrivită în studiile caz-control
c) când RR > 1 și evenimentul este frecvent, OR este mai mare decât RR
d) RR = 1 indică absența asocierii
e) un IC 95% care include 1 indică o asociere semnificativă

23. (CM) Despre indicatorii unui test diagnostic, sunt adevărate:
a) $\text{Se} = \frac{a}{a+c}$
b) $\text{Sp} = \frac{d}{b+d}$
c) VPP depinde de prevalență
d) sensibilitatea crește direct cu prevalența
e) $\text{RV}^- = \frac{1-\text{Se}}{\text{Sp}}$

24. (CM) Despre curba ROC, sunt adevărate:
a) reprezintă sensibilitatea în funcție de 1 - specificitate
b) o arie de 0,5 indică lipsa discriminării
c) indicele Youden este Se + Sp - 1
d) pragul care maximizează indicele Youden maximizează VPP
e) aria 0,852 a scorului FINDRISC indică o discriminare bună

25. (CM) Despre raționamentul bayesian, sunt adevărate:
a) cota post-test = cota pre-test x RV
b) probabilitatea = cota / (1 + cota)
c) un RV+ mare crește probabilitatea post-test după un rezultat pozitiv
d) probabilitatea pre-test nu influențează probabilitatea post-test
e) un RV- mai mic decât 1 scade probabilitatea bolii după un rezultat negativ

26. (CM) Pentru scăderea în greutate ≥ 5% în mg2026 ($\chi^2 = 6{,}77$, gl 2, p = 0,034), sunt adevărate:

a) H0 se respinge la α = 0,05
b) cel puțin un braț diferă de celelalte
c) toate cele trei brațe diferă între ele
d) RR C față de B de 0,98 (IC 95% 0,55-1,75) nu arată o diferență între C și B
e) frecvențele așteptate (minimum 12,6) permit aplicarea testului

27. (CM) Despre designurile de studiu, sunt adevărate:
a) randomizarea echilibrează în medie și factorii de confuzie necunoscuți
b) studiul de cohortă permite estimarea incidenței
c) studiul transversal estimează prevalența
d) studiul caz-control estimează direct incidența bolii
e) ajustarea multivariabilă poate corecta parțial confuzia în studiile observaționale

28. (CM) Despre distribuția normală, sunt adevărate:
a) aproximativ 68% dintre valori se află în intervalul $m \pm s$
b) aproximativ 95% dintre valori se află în intervalul $m \pm 1{,}96s$
c) este simetrică față de medie
d) un test Shapiro-Wilk cu p < 0,05 duce la respingerea ipotezei de normalitate
e) toate variabilele biologice au distribuție normală

29. (CM) Despre corelație, sunt adevărate:
a) r ia valori între -1 și 1
b) coeficientul Spearman este potrivit pentru variabile ordinale sau asimetrice
c) r = 0 arată că între variabile nu există nicio relație de niciun fel
d) o singură valoare extremă poate modifica mult valoarea lui r
e) într-un eșantion foarte mare, o corelație slabă poate fi semnificativă statistic

30. (CM) Despre raportarea studiilor, sunt adevărate:
a) CONSORT se folosește pentru studiile randomizate
b) STROBE se folosește pentru studiile observaționale
c) STARD se folosește pentru studiile de acuratețe diagnostică
d) Tabelul 1 al unui RCT trebuie să conțină valori p pentru fiecare caracteristică
e) efectul principal se raportează cu mărimea lui și intervalul de încredere de 95%

Partea a II-a. Lucrări practice

Lucrările practice urmează, săptămână de săptămână, temele de curs și folosesc instrumentele din fișa disciplinei: Google Forms și Google Sheets (cont Google al universității), JASP versiunea 0.98.1 sau mai nouă [47], MedCalc 24.0 sau calculatoarele lui online [50] și Epi Info 7.2.7 [51], care rulează și fără rețea. CDC a anunțat oprirea dezvoltării și a asistenței tehnice pentru Epi Info în septembrie 2025; programul rămâne utilizabil, dar nu mai primește versiuni noi, de aceea lucrările indică de fiecare dată și varianta în JASP sau MedCalc. Fișierele de lucru se descarcă de pe platforma Moodle a disciplinei (Emediqual, moodle.umft.ro): setul didactic "mg2026" în două variante, mg2026_brut.csv și mg2026_curat.csv, plus fișierele clasice ale disciplinei folosite la exercițiile suplimentare. Dicționarul setului mg2026 se află în anexa C. Lucrările predate se încarcă pe Moodle, cu numele LPx_Nume_Prenume.

În configurarea regională România, Google Sheets folosește virgula ca separator zecimal și punctul și virgula ca separator al argumentelor funcțiilor. Formulele din lucrări sunt scrise în această formă, de exemplu =AVERAGE(M2:M161). JASP și MedCalc afișează rezultatele cu punct zecimal; în rapoarte se scrie cu virgulă.

Fiecare lucrare indică rezultatele așteptate pe setul mg2026. Ele nu sunt date pentru a fi copiate, ci pentru ca studentul să-și poată verifica singur pașii: dacă media tensiunii sistolice nu iese 131,13 mmHg, ceva s-a întâmplat la import sau la curățare, și merită găsit înainte de a trece mai departe. Pentru teme și pentru examenul practic se folosește formatul de raportare a-g: ipoteza de cercetare, ipotezele statistice, metoda aleasă cu justificarea ei, modul de rezolvare, rezultatul cu graficul potrivit, concluzia statistică și concluzia de cercetare în cuvinte.

Lucrarea practică 1. Introducere în biostatistică și instrumente digitale: Google Sheets, JASP și MedCalc

Obiectiv

Prima lucrare practică are un scop foarte concret: la sfârșitul ei, fiecare student trebuie să aibă un fișier de date curat, organizat corect, însoțit de un dicționar de date, și să știe să deschidă acest fișier în programele pe care le va folosi tot semestrul. Calculele statistice propriu-zise sunt deocamdată simple. Ce contează acum este disciplina de lucru: un rând pentru fiecare pacient, o coloană pentru fiecare variabilă, valori coerente, nicio eroare ascunsă. Experiența ultimilor ani arată că majoritatea rezultatelor greșite de la examenul practic nu vin dintr-un test ales greșit, ci dintr-un fișier importat prost sau necurățat.

La final, studentul trebuie să poată:

- să construiască în Google Sheets un set de date medicale demonstrativ, cu reguli de validare și cu un dicționar de date;
- să importe corect un fișier CSV în Google Sheets, ținând cont de configurarea regională (virgula zecimală și separatorul ";");
- să detecteze și să corecteze erorile de introducere dintr-un fișier brut, documentând fiecare modificare;
- să deschidă setul de date în JASP și în MedCalc, să stabilească tipul variabilelor și să obțină statisticile descriptive de bază;
- să recunoască rolul programului Epi Info și să încarce rezultatele pe platforma Moodle.

Instrumente

Programe: Google Sheets (cont Google al universității), JASP 0.98.1 [47], MedCalc 24.0 sau calculatoarele lui online [50], Epi Info 7.2.7 [51]. Fișiere de pe Moodle: mg2026_curat.csv (160 de pacienți) și mg2026_brut.csv (161 de rânduri, cu erori lăsate intenționat). Dicționarul setului mg2026 se găsește în anexa C și, pe scurt, în tabelul 1.1 din capitolul 1. Timp orientativ: 100 de minute.

În fișierele mg2026, coloanele sunt în această ordine: A id, B brat, C sex, D varsta, E mediu, F educatie, G fumat, H activitate, I inaltime, J greutate0, K imc0, L talie, M tas0, N tad0, O fc0, P colesterol, Q hdl, R trigliceride, S glicemie0, T glicemie3, U glicemie6, V hba1c0, W findrisc, X diabet_ogtt, Y greutate6, Z tas6, AA scadere5, AB satisfactie. Datele ocupă rândurile 2-161 în fișierul curat și 2-162 în cel brut; rândul 1 conține numele variabilelor.

Desfășurare

L1.1 Crearea unui set de date medicale demonstrativ în Google Sheets

Primul exercițiu nu folosește setul mg2026, ci un set mic, construit de voi, ca să înțelegeți de la început cum se proiectează un tabel de date. Imaginați-vă că sunteți medic rezident și că vreți să notați, pentru 12 pacienți fictivi consultați într-o zi de gardă în cabinetul de medicină de familie, câteva date de bază.

1. Deschideți Google Sheets din contul universitar și creați o foaie nouă, numită LP1_Nume_Prenume. Verificați configurarea regională: Fișier > Setări > Setări regionale (File > Settings > Locale) trebuie să fie "România". Denumirile meniurilor pot varia ușor în funcție de limba interfeței.

2. Redenumiți prima filă "date" și scrieți în rândul 1 numele variabilelor, fără spații și fără diacritice: id, sex, varsta, inaltime, greutate, imc, tas, glicemie, fumat, educatie.
3. Creați a doua filă, numită "dictionar", cu coloanele: nume, descriere, tip, unitate sau coduri, interval plauzibil. Completați câte un rând pentru fiecare variabilă, după modelul din tabelul 2.2 din capitolul 2. De exemplu: tas; tensiunea arterială sistolică, măsurată în poziție șezândă după 5 minute de repaus; numeric întreg; mmHg; 70-260. Pentru educatie: ordinal; 1 gimnaziu, 2 liceu sau școală profesională, 3 universitar, 4 postuniversitar; 1-4.
4. Adăugați reguli de validare, care împiedică introducerea valorilor absurde. Selectați coloana B (sex), de la B2 în jos, apoi Date > Validarea datelor > Adăugați o regulă (Data > Data validation > Add rule). Alegeți criteriul "Drop-down" (listă derulantă) cu elementele F și M și bifați opțiunea de respingere a valorilor nevalide ("Reject the input"). Pentru coloana G (tas) alegeți criteriul "Is between" (este între) 70 și 260; pentru H (glicemie), între 40 și 400; pentru D (inaltime), între 130 și 210; pentru E (greutate), între 30 și 250; pentru I (fumat), o listă derulantă cu nefumator, fost fumator, fumator; pentru J (educatie), o listă cu 1, 2, 3, 4. Regulile de validare transpun în foaia de calcul intervalele plauzibile din dicționar.
5. Calculați IMC-ul cu o formulă, nu de mână. În celula F2 scrieți =ROUND(E2/(D2/100)^2;1) și copiați formula în jos, trăgând de colțul celulei. Observați separatorul ";" dintre argumente, specific configurării regionale România.
6. Introduceți datele pentru 12 pacienți fictivi, cu coduri P01-P12. Încercați intenționat să introduceți o tensiune de 1540 sau sexul "f": regula de validare trebuie să respingă valoarea. Dacă nu o respinge, verificați setarea regulii.
7. Calculați, sub date, câteva statistici simple: numărul de pacienți =COUNT(G2:G13), media TAS =AVERAGE(G2:G13), mediana =MEDIAN(G2:G13), abaterea standard =STDEV(G2:G13), minimul și maximul =MIN(G2:G13) și =MAX(G2:G13), numărul de femei =COUNTIF(B2:B13;"F").

Ce trebuie să obțineți: o foaie cu două file, "date" și "dictionar", în care nicio valoare din afara intervalelor plauzibile nu poate fi introdusă și în care IMC-ul se recalculează automat când modificați greutatea sau înălțimea.

L1.2 Importul fișierului mg2026_curat.csv

Un fișier CSV (engl. comma-separated values) este un fișier text în care valorile sunt separate prin virgulă, iar zecimalele sunt scrise cu punct (73.4). În configurarea regională România, Google Sheets așteaptă virgula zecimală, iar o valoare ca 5.1 poate fi interpretată greșit, de exemplu ca dată calendaristică (5 ianuarie). De aceea importul se face într-o ordine anume.

1. Creați o foaie nouă, mg2026_curat. Înainte de import, setați temporar Fișier > Setări > Setări regionale la "Statele Unite" (United States).
2. Fișier > Importați > Încărcare (File > Import > Upload) și alegeți mg2026_curat.csv descărcat de pe Moodle. La "Locație de import" alegeți "Înlocuiți foaia de calcul" sau "Inserați foi noi", la tipul separatorului "Detectare automată" sau "Virgulă" și lăsați bifată conversia textului în numere.
3. Verificați importul: foaia are 161 de rânduri (1 rând de antet și 160 de pacienți) și 28 de coloane (A-AB). Numerele trebuie să fie aliniate la dreapta; o coloană aliniată la stânga conține text.
4. Reveniți la Setări regionale "România". Valorile numerice rămân numere, dar sunt afișate acum cu virgulă (73,4), iar formulele noi folosesc separatorul ";". Fixați primul rând

(Vizualizare > Înghețare > 1 rând; View > Freeze > 1 row), ca numele variabilelor să rămână vizibile când derulați.

Ce trebuie să obțineți (autoverificare): =COUNTA(A2:A161) dă 160; =AVERAGE(M2:M161) dă 131,13 (media TAS); =STDEV(M2:M161) dă 13,88; =MEDIAN(M2:M161) dă 131; =AVERAGE(D2:D161) dă 56,29 (vârsta medie); =AVERAGE(K2:K161) dă 29,04 (IMC); =AVERAGE(S2:S161) dă 105,44 (glicemia); =COUNTIF(C2:C161;"F") dă 86; =COUNTIF(G2:G161;"fumator") dă 53; =COUNTIF(E2:E161;"urban") dă 95. Dacă media TAS nu este 131,13, nu treceți mai departe până nu găsiți cauza: cel mai probabil, o parte din valori au fost importate ca text sau ca date calendaristice.

L1.3 Curățarea fișierului mg2026_brut.csv

Fișierul brut simulează ceea ce primiți în realitate de la un coleg care a introdus datele de pe fișele de observație: aceleași variabile, dar cu greșeli. El conține 11 erori lăsate intenționat și un rând duplicat. Sarcina voastră este să le găsiți, să le corectați și să documentați fiecare corecție. Lucrați întotdeauna pe o copie: păstrați o filă "brut_original" neatinsă și faceți corecțiile într-o filă "brut_lucru".

1. Importați mg2026_brut.csv într-o foaie nouă, cu aceeași procedură ca la L1.2 (setări regionale Statele Unite la import, apoi România). Verificați: datele ocupă rândurile 2-162, adică 161 de rânduri de date, cu unul mai mult decât ar trebui.
2. Adăugați o filă "jurnal", cu coloanele: nr., id pacient, variabila, valoarea găsită, cum a fost detectată, valoarea corectată, sursa corecției. Fiecare modificare făcută în date trebuie să apară aici. Acesta este principiul trasabilității: oricine trebuie să poată reface drumul de la fișierul brut la cel curat.
3. Căutați erorile în variabilele calitative. Într-o celulă liberă scrieți =SORT(UNIQUE(C2:C162)) pentru sex, apoi la fel pentru E (mediu) și G (fumat). Funcția UNIQUE afișează fiecare valoare distinctă o singură dată și face diferența între litere mari și mici. Ar trebui să vedeți, în loc de două categorii pentru sex, patru: F, f, Feminin, M; pentru mediu, urban, Urban, rural; pentru fumat, pe lângă cele trei categorii corecte, și "Fumator " (cu majusculă și un spațiu la final). Atenție: funcția COUNTIF nu face diferența între litere mari și mici, deci =COUNTIF(C2:C162;"F") numără și valoarea "f" fără să vă avertizeze; JASP, în schimb, le-ar trata ca pe două categorii diferite. Pentru spațiile invizibile, funcția =LEN(G85) arată un caracter în plus față de lungimea așteptată.
4. Căutați erorile în variabilele numerice. Pentru fiecare coloană numerică, calculați =MIN(...), =MAX(...), =COUNT(...) și =COUNTBLANK(...). Comparați minimul și maximul cu intervalul plauzibil din dicționar. Veți găsi o tensiune sistolică maximă de 1540 mmHg (coloana M), o înălțime minimă de 1,72 (coloana I), o greutate minimă de 7,26 kg (coloana J). Funcția COUNT numără doar valorile numerice: pentru glicemia inițială (coloana S) dă 160 în loc de 161, iar pentru HbA1c (coloana V) la fel, semn că există câte o valoare stocată ca text; =ISNUMBER(S65) dă FALSE pentru celula cu "87 mg/dl", iar =ISNUMBER(V90) dă FALSE pentru HbA1c "5,2", scrisă cu virgulă într-un fișier în care toate celelalte zecimale au punct. Funcția COUNTBLANK dă 1 pentru vârstă (coloana D) și 1 pentru glicemia la 6 luni (coloana U).
5. Căutați duplicatele. =COUNTUNIQUE(A2:A162) dă 160, deși există 161 de rânduri: un cod de pacient apare de două ori. Într-o coloană ajutătoare scrieți în AC2 formula =COUNTIF(A$2:A$162;A2) și copiați-o în jos; rândurile cu valoarea 2 sunt duplicatele. Veți găsi pacientul P091 de două ori, cu valori identice în toate coloanele. Ștergeți al doilea rând (ultimul rând al fișierului) sau folosiți Date > Curățarea datelor > Eliminați dublurile (Data > Data cleanup > Remove duplicates).

6. Corectați erorile urmând regula din capitolul 1: o valoare suspectă nu se corectează după intuiție, ci se verifică în documentul sursă. În această lucrare, rolul documentului sursă îl joacă fișierul mg2026_curat.csv, în care valorile au fost preluate din fișele de observație. Tabelul L1.1 rezumă cele 12 probleme și corecțiile lor.

Tabelul L1.1. Erorile din mg2026_brut.csv, modul de detectare și corecția.

Nr.	Pacient	Variabila	Valoarea găsită	Detectare	Corecția și sursa
1	P005	sex	f	UNIQUE	F (uniformizare; confirmat în fișa sursă)
2	P021	sex	Feminin	UNIQUE	F (uniformizare; confirmat în fișa sursă)
3	P034	tas0	1540	MAX	154 (cifră în plus; confirmat în fișa sursă)
4	P042	varsta	lipsă	COUNTBLANK	70 (fișa sursă)
5	P057	inaltime	1,72	MIN	172 (metri transformați în centimetri; verificat din IMC și greutate)
6	P064	glicemie0	"87 mg/dl"	COUNT, ISNUMBER	87 (se șterge unitatea; confirmat în fișa sursă)
7	P084	fumat	"Fumator "	UNIQUE, LEN	fumator (TRIM și litere mici; confirmat în fișa sursă)
8	P089	hba1c0	"5,2" (text, cu virgulă)	COUNT, ISNUMBER	5,2 ca număr (confirmat în fișa sursă)
9	P102	greutate0	7,26	MIN	72,6 (virgulă mutată; verificat din IMC și înălțime)
10	P121	mediu	Urban	UNIQUE	urban (uniformizare)
11	P134	glicemie6	lipsă	COUNTBLANK	135 (fișa sursă)
12	P091	toate	rând duplicat	COUNTUNIQUE, COUNTIF	se șterge al doilea rând

7. Discutați cu colegul de lângă voi ce ați fi corectat dacă nu ați fi avut documentul sursă. Fiecare eroare din fișier este o deformare a valorii adevărate, iar pentru majoritatea dintre ele există o ipoteză foarte probabilă: "f" și "Feminin" înseamnă F, "Urban" înseamnă urban, "Fumator " devine fumator după eliminarea spațiului (funcția TRIM) și trecerea la litere mici (funcția LOWER), "87 mg/dl" devine 87 după ștergerea unității, "5,2" devine numărul 5,2. Pentru două valori numerice, ipoteza se poate verifica independent, pentru că fișierul conține o variabilă derivată redundantă, IMC-ul: pentru P057, înălțimea este $\sqrt{\frac{77,7}{26,3}} = \sqrt{2,954} = 1,719m$, adică 172 cm, exact valoarea 1,72 trecută din metri în centimetri; pentru P102, greutatea este 27,0 x 1,64² = 27,0 x 2,6896 = 72,6 kg, deci 7,26 este valoarea corectă cu virgula mutată. Valoarea 1540 este cel mai probabil 154 cu o cifră dublată, dar la fel de bine ar fi putut fi 150 sau 140 tastate greșit; nimic din fișier nu

permite verificarea, deci aici confirmarea în fișa sursă este obligatorie. Cele două valori lipsă (vârsta pacientului P042 și glicemia la 6 luni a pacientului P134) nu se pot deduce în niciun fel: doar documentul sursă le poate completa (70 de ani, respectiv 135 mg/dL). Când sursa nu este disponibilă, valoarea se lasă lipsă și se notează în jurnal; o valoare inventată este mai periculoasă decât una lipsă, chiar și atunci când ipoteza pare evidentă.

8. După corecții, verificați rezultatul. Fila "brut_lucru" trebuie să aibă 160 de rânduri de date, iar statisticile trebuie să coincidă cu cele din fișierul curat: =AVERAGE(M2:M161) dă 131,13; =COUNT(S2:S161) dă 160; =MIN(I2:I161) dă 145 cm; =MAX(M2:M161) dă 169 mmHg; =COUNTBLANK(D2:D161) dă 0; =SORT(UNIQUE(C2:C161)) afișează doar F și M. Dacă toate verificările trec, setul vostru curățat este identic cu mg2026_curat.csv.

Observați cât de mult ar fi schimbat erorile rezultatele dacă nu ar fi fost găsite. Pe fișierul brut necurățat (161 de rânduri, cu valoarea 1540 și cu duplicatul), =AVERAGE(M2:M162) dă 139,70 mmHg în loc de 131,13, iar =STDEV(M2:M162) dă 111,89 mmHg în loc de 13,88: o singură cifră tastată greșit a mutat media grupului la limita hipertensiunii și a înmulțit abaterea standard de aproape opt ori. Mediana, =MEDIAN(M2:M162), rămâne 131 mmHg, exact ca în fișierul curat, ceea ce ilustrează robustețea medianei discutată în capitolul 2. Când media și mediana unei variabile simetrice diferă atât de mult, iar abaterea standard este de mărimea mediei, primul suspect este o eroare de introducere.

L1.4 Explorarea interfeței JASP

JASP este programul în care veți face majoritatea analizelor din semestru. Interfața are trei zone: datele (în stânga, ca o foaie de calcul), panoul de opțiuni al analizei curente (la mijloc) și rezultatele (în dreapta), care se actualizează imediat ce schimbați o opțiune.

1. Deschideți JASP. Din meniul principal (butonul cu trei linii din stânga sus) alegeți Open > Computer > Browse și selectați mg2026_curat.csv. Datele apar în stânga.
2. Verificați tipul fiecărei variabile. Lângă numele fiecărei coloane apare o pictogramă: rigla pentru variabile numerice (Scale), trei bare în trepte pentru variabile ordinale (Ordinal) și trei cercuri pentru variabile nominale (Nominal). JASP ghicește tipul la import, dar nu întotdeauna corect: educatie, activitate și satisfactie sunt importate ca numere, deși sunt ordinale; diabet_ogtt și scadere5 sunt importate ca numere, deși sunt nominale dihotomice. Faceți clic pe pictogramă și schimbați tipul. Această etapă nu este cosmetică: JASP oferă analize diferite în funcție de tipul declarat al variabilei.
3. Alegeți Descriptives > Descriptive Statistics. Mutați în caseta Variables variabilele varsta, imc0, tas0 și glicemie0. În secțiunea Statistics bifați, pe lângă opțiunile implicite (Valid, Missing, Mean, Std. deviation, Minimum, Maximum), și Median, Quartiles, Skewness, Kurtosis și S.E. mean.
4. Comparați rezultatele cu cele din Google Sheets și din tabelul 2.12 din capitolul 2. Trebuie să obțineți: varsta 56,29 ± 8,77 ani, mediana 56; imc_0 $29{,}04 \pm \frac{3{,}91\ \text{kg}}{m^2}$, mediana 28,25; tas_0 $131{,}13 \pm 13{,}88$ mmHg, mediana 131, eroarea standard 1,097; glicemie_0 $105{,}44 \pm 15{,}76$ mg/dL, mediana 104,5. JASP afișează rezultatele cu punct zecimal; în rapoarte le transcrieți cu virgulă.
5. Mutați variabila sex în caseta Split. Rezultatele se afișează separat pentru femei și bărbați; pentru TAS trebuie să vedeți 131,37 ± 14,47 mmHg la femei (n = 86) și 130,85 ± 13,25 mmHg la bărbați (n = 74).

6. Pentru o variabilă nominală, de exemplu fumat, bifați opțiunea Frequency tables (din secțiunea Tables sau direct sub lista de variabile, în funcție de versiune). Tabelul de frecvențe trebuie să arate 74 nefumători, 33 foști fumători și 53 fumători.
7. Salvați analiza ca fișier JASP (File > Save As, LP1_Nume_Prenume.jasp). Fișierul JASP păstrează împreună datele, opțiunile și rezultatele, ceea ce face analiza reproductibilă. Rezultatele se pot exporta și ca PDF sau HTML (File > Export Results).

L1.5 Explorarea interfeței MedCalc

MedCalc are o interfață de tip foaie de calcul, cu meniuri pentru analize și grafice. Este programul preferat pentru curbele ROC și pentru analiza testelor diagnostice, pe care le veți folosi în lucrările 11-12.

1. Deschideți MedCalc și importați fișierul mg2026_curat.csv prin File > Open, alegând tipul de fișier text/CSV. Verificați în fereastra de import că separatorul este virgula și că simbolul zecimal este punctul.
2. Alegeți Statistics > Summary statistics, selectați variabila tas0 și rulați analiza. MedCalc afișează numărul de observații, media, intervalul de încredere de 95% pentru medie (128,96-133,30 mmHg, despre care veți învăța în capitolul 4), abaterea standard, eroarea standard, mediana, percentilele și, opțional, testele de normalitate.
3. Alegeți Graphs > Box-and-whisker plot pentru tas0, eventual cu sex ca variabilă de grupare. Comparați graficul cu figura 2.3D din capitolul 2.
4. Dacă pe calculatorul vostru nu este disponibilă o licență MedCalc, deschideți calculatoarele online gratuite de pe site-ul producătorului. Încercați calculatorul pentru intervalul de încredere al unei proporții: introduceți 29 de pacienți cu diabet din 160 și notați rezultatul; îl veți interpreta în capitolul 4 (metoda Wilson dă 12,9-24,8%).

L1.6 Epi Info, pe scurt

Epi Info 7 (ultima versiune, 7.2.7) a fost dezvoltat de CDC pentru epidemiologia de teren și funcționează complet fără conexiune la internet. CDC a oprit dezvoltarea și suportul programului în septembrie 2025; programul rămâne utilizabil, dar nu mai primește actualizări. Are trei componente principale: Form Designer (construirea formularelor de colectare a datelor), Enter (introducerea datelor) și Analysis (analiza, în două variante, Classic Analysis, bazată pe comenzi, și Visual Dashboard, bazată pe meniuri).

1. Deschideți Epi Info și alegeți Classic Analysis. Cu comanda READ (butonul Read din lista de comenzi), importați fișierul mg2026_curat.csv, alegând tipul de sursă text/CSV.
2. Rulați comanda FREQ pentru variabila fumat și comanda MEANS pentru variabila tas0. Verificați că obțineți aceleași frecvențe (74, 33, 53) și aceeași medie (131,13 mmHg) ca în celelalte programe.
3. Observați că Epi Info afișează și programul generat de comenzi, care poate fi salvat și rulat din nou. Veți reveni la Epi Info în lucrările despre analiza riscului, unde tabelele 2x2 cu risc relativ și raport de șanse se obțin direct.

L1.7 Încărcarea rezultatelor pe Moodle

1. Descărcați foaia Google Sheets ca fișier Excel: Fișier > Descărcați > Microsoft Excel (.xlsx). Fișierul trebuie să conțină filele: date (setul demonstrativ), dictionar, brut_original, brut_lucru și jurnal.

2. Pregătiți un document scurt (o pagină, PDF) cu: tabelul celor 12 probleme găsite și corectate, statisticile de verificare (n, media și abaterea standard a TAS înainte și după curățare) și două-trei fraze despre ce ați învățat din exercițiul de curățare.
3. Încărcați pe Moodle, la activitatea LP1, fișierele LP1_Nume_Prenume.xlsx, LP1_Nume_Prenume.jasp și LP1_Nume_Prenume.pdf, înainte de termenul afișat pe platformă [52].

Exerciții suplimentare

1. Pe fișierul BD_Studenti de pe Moodle (baza de date a studenților folosită în anii anteriori la disciplină): stabiliți volumul eșantionului (N), clasificați fiecare variabilă după tip, scrieți un dicționar de date complet (nume, descriere, tip, unitate sau coduri, interval plauzibil) și verificați cu funcțiile MIN, MAX, COUNT, COUNTBLANK și UNIQUE dacă fișierul conține valori lipsă, valori în afara intervalelor plauzibile sau codificări inconsecvente. Adăugați o coloană calculată cu IMC-ul fiecărui student. Nu se dau rezultate numerice; verificarea se face la lucrarea următoare.
2. Problemă cu date în enunț. Un coleg a introdus datele a opt pacienți dintr-un studiu de hipertensiune în tabelul de mai jos. Găsiți erorile, propuneți modul de verificare pentru fiecare și calculați media TAS înainte și după corecție, presupunând că fișa sursă confirmă pentru P4 valoarea 152 mmHg și că valoarea lipsă a pacientului P7 nu poate fi recuperată.

Tabelul L1.2. Date introduse pentru opt pacienți (exercițiul 2).

id	sex	varsta	tas	fumat
P1	F	58	138	nefumator
P2	M	61	145	fumator
P3	m	49	132	nefumator
P4	F	66	1520	fost fumator
P5	F	55	128	Nefumator
P6	M	72	160	fumator
P7	F	63		nefumator
P2	M	61	145	fumator

Rezolvare. Erorile sunt: (1) P3, sex "m" în loc de "M" (codificare inconsecventă; se detectează cu UNIQUE; se uniformizează la M după confirmare); (2) P4, TAS 1520 mmHg, imposibilă biologic (se detectează cu MAX; se corectează la 152 conform fișei sursă); (3) P5, fumat "Nefumator" cu majusculă (UNIQUE; se uniformizează la nefumator); (4) P7, TAS lipsă (COUNTBLANK; rămâne lipsă și se notează în jurnal); (5) rândul P2 duplicat (COUNTIF pe id; se șterge al doilea rând). Înainte de corecție, media TAS calculată de =AVERAGE pe cele 7 valori numerice (inclusiv duplicatul și valoarea 1520) este (138 + 145 + 132 + 1520 + 128 + 160 + 145)/7 = 2 368/7 = 338,3 mmHg, o valoare absurdă. După corecție rămân 7 pacienți distincți, dintre care 6 cu TAS cunoscută: (138 + 145 + 132 + 152 + 128 + 160)/6 = 855/6 = 142,5 mmHg. Funcția AVERAGE ignoră celula goală, dar numărul de pacienți analizați (n = 6 din 7) trebuie raportat explicit.

3. Pe fișierul mg2026_curat.csv, creați o coloană nouă, "obez", cu formula =IF(K2>=30;1;0), și numărați pacienții obezi cu =SUM(...). Trebuie să obțineți 54. Creați apoi o coloană "tas_crescuta" cu =IF(M2>=140;1;0); trebuie să obțineți 41 de pacienți. Discutați ce informație se pierde prin aceste dihotomizări (capitolul 2, secțiunea 2.1.4).

Ce se predă și barem orientativ

Se predau, pe Moodle, cele trei fișiere descrise la L1.7. Baremul orientativ, din 10 puncte: setul demonstrativ cu structură corectă (un rând pe pacient, nume de variabile fără spații) și IMC calculat cu formulă, 1,5 puncte; dicționarul de date complet, 1,5 puncte; regulile de validare funcționale, 1 punct; importul corect al fișierului curat, cu statisticile de verificare, 1 punct; identificarea celor 11 erori și a duplicatului, 2 puncte (câte un punct pentru fiecare jumătate a listei); jurnalul modificărilor, cu sursa fiecărei corecții, 1 punct; fișierul JASP cu tipurile de variabile setate corect și statisticile descriptive, 1 punct; documentul PDF cu interpretarea, 1 punct. Un punct se pierde dacă statisticile finale nu coincid cu cele ale fișierului curat (de exemplu, media TAS diferită de 131,13 mmHg).

Varianta fără rețea

Dacă rețeaua nu funcționează, lucrarea se poate face integral pe calculatorul local. Setul demonstrativ, dicționarul, validarea și curățarea se realizează în LibreOffice Calc sau Microsoft Excel, cu aceleași funcții (în versiunea în limba română a programului Excel, numele funcțiilor sunt traduse, de exemplu MEDIE în loc de AVERAGE, iar separatorul este tot ";"); validarea datelor se găsește în meniul Date > Validare date. JASP și Epi Info funcționează fără conexiune la internet, odată instalate. Fișierele mg2026 se copiază de pe stick-ul cadrului didactic sau de pe serverul local al laboratorului. Încărcarea pe Moodle se face ulterior, în termenul afișat.

Lucrarea practică 2. Descrierea și organizarea datelor: chestionarul medical, tabele de frecvențe și grafice

Obiectiv

În această lucrare parcurgeți drumul complet al unei variabile, de la întrebarea pusă pacientului până la tabelul și graficul dintr-un raport. Construiți mai întâi un chestionar medical scurt în Google Forms și vedeți cum tipul fiecărei întrebări determină tipul variabilei pe care o veți analiza. Apoi, pe setul mg2026, construiți tabele de frecvențe, calculați indicatorii descriptivi și desenați graficele potrivite, mai întâi în Google Sheets, apoi în JASP. Toate rezultatele așteptate sunt date în text, ca să vă puteți verifica singuri.

La final, studentul trebuie să poată:

- să proiecteze un chestionar scurt în Google Forms, alegând pentru fiecare întrebare tipul care produce variabila dorită și regulile de validare potrivite;
- să lege formularul de o foaie Google Sheets și să recodifice răspunsurile pentru analiză;
- să construiască tabele de frecvențe absolute, relative și cumulate cu COUNTIF, COUNTIFS, FREQUENCY și cu tabele pivot;
- să calculeze în Google Sheets și în JASP media, mediana, modul, abaterea standard, cuartilele, coeficientul de variație, asimetria și excesul;
- să construiască histograma, box-plotul, diagrama cu bare și diagrama circulară și să interpreteze forma distribuției.

Instrumente

Programe: Google Forms și Google Sheets (cont Google al universității) [43-46], JASP 0.98.1 [47]. Fișiere: mg2026_curat.csv (sau foaia curățată la lucrarea practică 1), BD_Studenti pentru exercițiile suplimentare. Timp orientativ: 100 de minute (aproximativ 30 de minute pentru chestionar și 70 pentru analiza descriptivă).

Coloanele setului mg2026 sunt cele din lucrarea practică 1: C sex, D varsta, E mediu, F educatie, G fumat, H activitate, K imc0, M tas0, R trigliceride, S glicemie0, AB satisfactie; datele ocupă rândurile 2-161.

Desfășurare

L2.1 Proiectarea unui chestionar medical scurt în Google Forms

Scenariul: un cabinet de medicină de familie vrea să afle cât de des își măsoară pacienții cu hipertensiune tensiunea acasă și cât de mulțumiți sunt de controlul bolii. Chestionarul este anonim, se completează în sala de așteptare pe tableta cabinetului și are cel mult 10 întrebări.

1. Deschideți Google Forms din contul universitar și creați un formular nou, numit "LP2 Automonitorizarea tensiunii arteriale". În descriere scrieți scopul chestionarului, faptul că este anonim și voluntar și cât durează completarea.
2. Prima întrebare este întrebarea de consimțământ: "Sunt de acord să completez acest chestionar anonim, iar răspunsurile mele pot fi folosite în scopuri statistice", de tip alegere multiplă, cu variantele "Da" și "Nu", obligatorie. Folosiți ramificarea pe secțiuni (Mergi la secțiunea în funcție de răspuns) astfel încât răspunsul "Nu" să trimită direct la finalul formularului [148]. Fără consimțământ nu se colectează nimic.

3. Adăugați întrebările de mai jos, alegând tipul indicat. Tabelul L2.1 arată legătura dintre tipul întrebării din Google Forms și tipul variabilei care va rezulta. Este legătura cea mai importantă din această lucrare: tipul variabilei se decide când scrieți întrebarea, nu când faceți analiza.

Tabelul L2.1. Tipurile de întrebări din Google Forms și variabilele care rezultă.

Întrebarea	Tipul în Google Forms	Validare	Variabila rezultată
Vârsta dumneavoastră (ani împliniți)	Răspuns scurt	Număr, între 18 și 100	numerică discretă (vârstă în ani)
Sexul	Alegere multiplă: feminin, masculin	obligatorie	nominală dihotomică
Mediul în care locuiți	Alegere multiplă: urban, rural	obligatorie	nominală dihotomică
De câte ori v-ați măsurat tensiunea acasă în ultima săptămână?	Răspuns scurt	Număr întreg, între 0 și 50	numerică discretă
Ultima valoare a tensiunii sistolice măsurată acasă (mmHg), dacă o știți	Răspuns scurt	Număr, între 70 și 260; neobligatorie	numerică continuă, cu valori lipsă permise
Ce aparat folosiți?	Alegere multiplă: de braț, de încheietură, nu am aparat	obligatorie	nominală
Ce vă împiedică să vă măsurați mai des?	Casete de selectare (mai multe răspunsuri)	-	mai multe variabile dihotomice (câte una pentru fiecare variantă)
Cât de mulțumit sunteți de controlul tensiunii?	Scară liniară 1-5 (1 foarte nemulțumit, 5 foarte mulțumit)	obligatorie	ordinală
Nivelul de educație	Listă derulantă: gimnaziu, liceu sau școală profesională, universitar, postuniversitar	obligatorie	ordinală

4. Setați regulile de validare. Pentru întrebările numerice, faceți clic pe cele trei puncte din colțul întrebării, alegeți Validarea răspunsului și setați "Număr", "Între", cu limitele din tabel, plus un mesaj de eroare clar ("Vă rugăm să introduceți un număr între 70 și 260") [45]. Validarea împiedică exact erorile pe care le-ați curățat la lucrarea practică 1: textul "87 mg/dl" în loc de 87, o tensiune de 1540, o greutate de 7,26 kg, o vârstă de 5 ani.
5. Nu cereți date de identificare: nici nume, nici adresa de e-mail (în Setări, colectarea adreselor de e-mail trebuie să fie dezactivată), nici data nașterii completă. Vârsta în ani este suficientă. Este aplicarea directă a principiului minimizării datelor (capitolul 1, secțiunea 1.8.3).
6. Testați formularul: completați-l de 5-6 ori, ca și cum ați fi pacienți diferiți, cu răspunsuri variate, inclusiv unul care refuză consimțământul și unul care nu își știe tensiunea. Rugați și colegul de lângă voi să-l completeze de 2-3 ori de pe telefon.

L2.2 Legarea formularului de Google Sheets și recodificarea răspunsurilor

1. În formular, deschideți fila Răspunsuri și alegeți Asociați cu Foi de calcul (Link to Sheets), cu opțiunea de creare a unei foi noi [46]. Google creează o foaie în care fiecare răspuns este un rând, iar fiecare întrebare o coloană; prima coloană conține marcajul temporal al trimiterii. Structura este exact cea recomandată în capitolul 2: un rând pe respondent, o coloană pe variabilă [147].

2. Observați cum apar diferitele tipuri de întrebări. Răspunsurile la alegerea multiplă apar ca text ("feminin"), scara liniară ca număr (1-5), casetele de selectare ca un text cu toate variantele bifate, separate prin virgulă, în aceeaşi celulă. Acesta din urmă nu este un format bun pentru analiză: fiecare variantă trebuie transformată într-o coloană separată, cu 1 (bifat) şi 0 (nebifat). Pentru varianta "nu am timp", de exemplu, folosiți formula =IF(ISNUMBER(SEARCH("nu am timp";H2));1;0).
3. Recodificați variabilele ordinale introduse ca text. Pentru educație, într-o coloană nouă: =IFS(J2="gimnaziu";1;J2="liceu sau şcoală profesională";2;J2="universitar";3;J2="postuniversitar";4). Analiza ordinală are nevoie de coduri numerice în ordinea corectă; textul ar fi sortat alfabetic, iar "gimnaziu" ar ajunge după "universitar" doar din întâmplare.
4. Eliminați rândurile respondenților care au refuzat consimțământul (dacă ramificarea a funcționat, celelalte coloane sunt goale) şi redenumiți coloanele cu nume scurte, fără spații: varsta, sex, mediu, nr_masurari, tas_casa, aparat, satisfactie, educatie. Adăugați o filă "dictionar".

L2.3 Tabele de frecvențe în Google Sheets pe setul mg2026

Pentru analiza descriptivă lucrați pe setul mg2026, care are 160 de pacienți şi permite rezultate cu sens. Deschideți foaia curățată la lucrarea practică 1 sau importați din nou mg2026_curat.csv (cu procedura de la L1.2) şi lucrați într-o filă nouă, "frecvente".

1. Variabilă nominală: statutul de fumător. Scrieți în A2:A4 categoriile nefumator, fost fumator, fumator. În B2 scrieți =COUNTIF(date!$G$2:$G$161;A2) şi copiați în jos. În C2, frecvența relativă: =B2/SUM($B$2:$B$4)*100. Trebuie să obțineți 74 (46,3%), 33 (20,6%) şi 53 (33,1%). Numele filei cu date ("date") se înlocuieşte cu numele filei voastre.
2. Variabilă ordinală: satisfacția (coloana AB). În A7:A11 scrieți valorile 1-5. Frecvența absolută: =COUNTIF(date!$AB$2:$AB$161;A7). Frecvența relativă: =B7/160*100. Frecvența cumulată absolută: în D7 =B7, în D8 =D7+B8, copiat în jos. Frecvența cumulată relativă: =D7/160*100. Rezultatul trebuie să fie tabelul 2.4 din capitolul 2: frecvențele 4, 19, 54, 55, 28; frecvențele relative 2,5; 11,9; 33,8; 34,4; 17,5%; frecvențele cumulate 2,5; 14,4; 48,1; 82,5; 100,0%. Frecvența cumulată trece de 50% la valoarea 4, deci mediana este 4.
3. Acelaşi lucru pentru educație (coloana F): 20, 57, 64, 19 pacienți (12,5; 35,6; 40,0; 11,9%), cu frecvențele cumulate 12,5; 48,1; 88,1; 100,0% şi mediana 3. Pentru activitatea fizică (coloana H): 21, 42, 60, 34, 3 pacienți, cu mediana 3.
4. Variabilă continuă grupată în clase: TAS (coloana M). Calculați mai întâi numărul de clase după Sturges, =1+3,322*LOG10(160), care dă 8,32, şi lățimea =(MAX(date!M2:M161)-MIN(date!M2:M161))/8, care dă 8,4; rotunjiți lățimea la 10 mmHg. În A14:A20 scrieți limitele inferioare 100, 110,..., 160. Frecvența fiecărei clase cu COUNTIFS: =COUNTIFS(date!$M$2:$M$161;">="&A14;date!$M$2:$M$161;"<"&A14+10). Trebuie să obțineți 11, 21, 35, 52, 25, 13, 3, adică tabelul 2.5 din capitolul 2.
5. Aceleași frecvențe se obțin cu funcția FREQUENCY, care primeşte datele şi limitele superioare ale claselor. Scrieți în F14:F19 limitele superioare 109, 119, 129, 139, 149, 159, apoi într-o singură celulă =FREQUENCY(date!M2:M161;F14:F19). Funcția întoarce 7 valori (ultima pentru valorile peste 159) şi le scrie automat în celulele de dedesubt. Verificați că suma frecvențelor este 160.

6. Clasele de IMC folosite în clinică au lățimi inegale: sub 25 (normoponderal), 25-29,9 (supraponderal), 30-34,9 (obezitate gradul I), 35-39,9 (gradul II), ≥ 40 (gradul III). Construiți acest tabel cu COUNTIFS pe coloana K. Rezultatul așteptat: 18, 88, 43, 8 și 3 pacienți; obezi în total 43 + 8 + 3 = 54 (33,8%).

L2.4 Tabele pivot și tabele de contingență

Un tabel pivot rezumă automat datele după una sau mai multe variabile categoriale, fără formule [44].

1. Selectați datele (A1:AB161) și alegeți Inserare > Tabel pivot (Insert > Pivot table), cu crearea unei file noi.
2. În editorul tabelului pivot adăugați la Rânduri variabila fumat, la Coloane variabila sex, iar la Valori variabila id, rezumată prin COUNTA. Trebuie să obțineți tabelul de contingență sex x fumat: femei, 16 foști fumători, 30 fumători, 40 nefumători; bărbați, 17, 23 și 34; totaluri 33, 53, 74 și 160.
3. Schimbați afișarea valorilor în procente din totalul coloanei (opțiunea "Afișați ca", Show as > % of column). Comparați proporția fumătorilor activi la femei (30/86 = 34,9%) și la bărbați (23/74 = 31,1%).
4. Construiți al doilea tabel pivot, brat x satisfactie. Rezultatul așteptat: brațul A, satisfacția 1-5: 1, 13, 28, 9, 2; brațul B: 3, 3, 14, 24, 9; brațul C: 0, 3, 12, 22, 17. Chiar înainte de orice test statistic, tabelul arată că pacienții din brațul C sunt mai mulțumiți: 39 din 54 (72,2%) au ales 4 sau 5, față de 11 din 53 (20,8%) în brațul A. Testul potrivit pentru această comparație, Mann-Whitney sau Kruskal-Wallis, va fi aplicat în lucrarea practică 7.
5. Tabelul pivot poate calcula și medii: puneți brat la Rânduri și tas0 la Valori, rezumat prin AVERAGE. Trebuie să obțineți 131,53 (A), 132,79 (B) și 129,11 (C) mmHg.

L2.5 Statistici descriptive în Google Sheets

Într-o filă nouă, "descriptive", construiți un tabel cu variabilele pe coloane și indicatorii pe rânduri. Formulele pentru TAS (coloana M) sunt:

Tabelul L2.2. Formule Google Sheets pentru statisticile descriptive ale TAS și rezultatele așteptate.

Indicator	Formula (configurare România)	Rezultat așteptat
n	=COUNT(date!M2:M161)	160
Media	=AVERAGE(date!M2:M161)	131,13
Mediana	=MEDIAN(date!M2:M161)	131
Modul	=MODE(date!M2:M161)	138
Abaterea standard	=STDEV(date!M2:M161)	13,88
Varianța	=VAR(date!M2:M161)	192,66
Eroarea standard	=STDEV(date!M2:M161)/SQRT(COUNT(date!M2:M161))	1,097
Coeficientul de variație (%)	=STDEV(date!M2:M161)/AVERAGE(date!M2:M161)*100	10,6
Minim, maxim	=MIN(date!M2:M161), =MAX(date!M2:M161)	102, 169
Q1, Q3	=QUARTILE(date!M2:M161;1), =QUARTILE(date!M2:M161;3)	122, 140
Percentilele 10 și 90	=PERCENTILE(date!M2:M161;0,1), =PERCENTILE(date!M2:M161;0,9)	112; 148,2
Asimetria	=SKEW(date!M2:M161)	0,01
Excesul	=KURT(date!M2:M161)	-0,10

Observați argumentul 0,1 din funcția PERCENTILE: în configurarea România, zecimala se scrie cu virgulă, iar argumentele se separă cu punct și virgulă. Copiați formulele și pentru celelalte variabile, schimbând litera coloanei. Rezultatele așteptate:

Tabelul L2.3. Rezultate așteptate pentru alte variabile din mg2026.

Variabila (coloana)	Media	AS	Mediana	Q1	Q3	Asimetria	Excesul
varsta (D)	56,29	8,77	56	50,75	62	0,01	-0,09
imc0 (K)	29,04	3,91	28,25	26,2	31,4	1,42	3,68
glicemie0 (S)	105,44	15,76	104,5	93	116	0,37	-0,23
trigliceride (R)	161,18	76,30	151,5	107	203,25	0,95	1,12

Calculați și media geometrică a trigliceridelor, =GEOMEAN(date!R2:R161), care trebuie să dea 144,0 mg/dL, și comparați-o cu media aritmetică (161,18) și cu mediana (151,5). Calculați limitele pentru valorile extreme ale IMC: =QUARTILE(date!K2:K161;3)+1,5*(QUARTILE(date!K2:K161;3)-QUARTILE(date!K2:K161;1)) dă 39,2 kg/m², iar =COUNTIF(date!K2:K161;">"&39,2) dă 3 pacienți.

L2.6 Grafice în Google Sheets

1. Diagrama cu bare pentru fumat. Selectați tabelul de frecvențe de la L2.3, pasul 1 (categoriile și frecvențele absolute), apoi Inserare > Diagramă. În editorul diagramei alegeți tipul "Diagramă cu coloane". În fila Personalizare adăugați titlul ("Statutul de fumător, mg2026, n = 160"), titlurile axelor și etichetele de date. Verificați că axa verticală pornește de la 0.
2. Diagrama circulară pentru educație. Selectați tabelul de frecvențe al educației (cu etichetele categoriilor în text, nu codurile 1-4) și alegeți tipul "Diagramă circulară". Afișați procentele pe sectoare: 12,5; 35,6; 40,0; 11,9%. Nu folosiți varianta 3D.
3. Histograma glicemiei. Selectați coloana glicemie0 (S1:S161) și alegeți tipul "Histogramă". În Personalizare > Histogramă setați dimensiunea intervalului (bucket size) la 10. Forma așteptată: un vârf larg între 90 și 120 mg/dL (34 de pacienți în fiecare dintre clasele 90-99, 100-109 și 110-119), cu o coadă scurtă spre dreapta, până la 154 mg/dL. Încercați și un interval de 5, apoi de 20, și observați cum se schimbă aspectul.
4. Box-plotul nu există ca tip standard de diagramă în Google Sheets. Îl veți construi în JASP, la pasul următor. Puteți, în schimb, să afișați rezumatul în cinci numere al TAS într-un mic tabel: 102, 122, 131, 140, 169.

L2.7 Analiza descriptivă și grafice în JASP

1. Deschideți mg2026_curat.csv în JASP și verificați tipurile variabilelor, ca la lucrarea practică 1: educatie, activitate și satisfactie ordinale; sex, mediu, fumat, brat, diabet_ogtt nominale.
2. Descriptives > Descriptive Statistics, cu variabilele varsta, imc0, tas0, glicemie0, trigliceride. Bifați Mean, Std. deviation, Median, Quartiles, Minimum, Maximum, Skewness, Kurtosis, Shapiro-Wilk test (îl veți interpreta în capitolul 3). Comparați cu tabelul L2.3; valorile trebuie să coincidă.
3. Pentru variabilele fumat, educatie și satisfactie, bifați Frequency tables. JASP afișează frecvențele, procentele, procentele valide (care exclud valorile lipsă) și procentele cumulate, exact ca tabelul 2.4.

4. În secțiunea Basic plots (sau Plots) bifați Distribution plots pentru imc0 și trigliceride. Comparați histogramele cu figura 2.2 din capitolul 2: ambele distribuții au coada spre dreapta.
5. Pentru box-plot, puneți tas0 la Variables și sex la Split, apoi bifați Boxplots (eventual și opțiunile Boxplot element, Violin element și Jitter element, care arată și punctele individuale). Rezultatul trebuie să semene cu figura 2.3D: mediana 131 mmHg la ambele sexe, cutiile aproape suprapuse. În tabel, femeile au $131{,}37 \pm 14{,}47$ mmHg$(n = 86)$, iar bărbații $130{,}85 \pm 13{,}25$ mmHg$(n = 74)$.
6. Repetați box-plotul pentru imc0, fără împărțire pe sexe. Identificați pe grafic cele trei valori extreme de peste 39,2 kg/m² (40,0; 43,4 și 48,3).

L2.8 Interpretarea distribuției

Scrieți, pentru raportul predat, câte un paragraf de interpretare pentru TAS, IMC și trigliceride. Modelul pentru TAS: "Tensiunea sistolică la includere a avut o distribuție aproximativ simetrică (asimetrie 0,01, exces -0,10), cu media 131,13 ± 13,88 mmHg, practic egală cu mediana (131 mmHg; Q1-Q3 122-140), și valori între 102 și 169 mmHg. Pentru această variabilă, media și abaterea standard sunt indicatorii potriviți." Pentru IMC și trigliceride, folosiți asimetria, relația dintre medie și mediană și box-plotul pentru a justifica raportarea prin mediană (Q1-Q3).

Exerciții suplimentare

1. Pe fișierul BD_Studenti de pe Moodle: realizați tabelele de frecvențe și graficele potrivite pentru gen, culoarea ochilor, orele de somn și tipul de locuință; calculați indicatorii tendinței centrale și ai dispersiei pentru vârstă, IMC (calculat cu formula din greutate și înălțime), orele de somn și orele de studiu; comparați, cu un grafic potrivit, culoarea ochilor la fete și la băieți și orele de somn la fete și la băieți. Justificați alegerea fiecărui indicator și a fiecărui grafic după tipul variabilei. Nu se dau rezultate numerice; ele se discută la lucrarea următoare.
2. Tabel cu valori lipsă: HbA1c la pacienți cu diabet. Un medic a grupat valorile HbA1c ale celor 40 de pacienți cu diabet de tip 2 de pe lista sa. Completați valorile A, B, C, D, E și F și răspundeți: ce proporție dintre pacienți au atins ținta HbA1c < 7%? Ce proporție au HbA1c ≥ 8%?

Tabelul L2.4. HbA1c la 40 de pacienți cu diabet de tip 2.

HbA1c (%)	**Frecvența absolută**	**Frecvența relativă (%)**	**Frecvența cumulată (%)**
< 6,5	5	12,5	12,5
6,5-6,9	9	A	35,0
7,0-7,9	B	30,0	C
8,0-8,9	8	D	85,0
≥ 9,0	E	F	100,0
Total	40	100,0	

Rezolvare. $A = \frac{9}{40} \times 100 = 22{,}5\%$ (verificare: 12,5 + 22,5 = 35,0). B = 0,30 x 40 = 12 pacienți. C = 35,0 + 30,0 = 65,0%. $D = \frac{8}{40} \times 100 = 20{,}0\%$ (verificare: 65,0 + 20,0 = 85,0). F = 100,0 - 85,0 = 15,0%, iar E = 0,15 x 40 = 6 pacienți. Verificarea sumei: 5 + 9 + 12 + 8 + 6 = 40. Ținta HbA1c < 7% a fost atinsă de 5 + 9 = 14 pacienți, adică 35,0% (frecvența cumulată a clasei 6,5-6,9). HbA1c ≥ 8% au 8 + 6 = 14 pacienți, adică 20,0 + 15,0 = 35,0%. Clasa mediană este 7,0-7,9: a 20-a și a 21-a valoare ordonată se află în această clasă, pentru că frecvența cumulată trece de la 35,0% (14 pacienți) la 65,0% (26 de pacienți). Interpretare clinică: doar o treime dintre

pacienți sunt la țintă, iar o altă treime au un control glicemic prost, care cere intensificarea tratamentului.

3. Tabel cu valori lipsă: satisfacția față de serviciile unui ambulatoriu. La un sondaj de satisfacție au răspuns 120 de pacienți, pe o scală de la 1 (foarte nemulțumit) la 5 (foarte mulțumit). Completați valorile lipsă, determinați mediana și modul și calculați proporția pacienților mulțumiți sau foarte mulțumiți.

Tabelul L2.5. Satisfacția a 120 de pacienți ai unui ambulatoriu.

Satisfacția	**Frecvența absolută**	**Frecvența relativă (%)**	**Frecvența cumulată absolută**	**Frecvența cumulată (%)**
1	6	5,0	6	5,0
2	A	12,5	21	B
3	33	C	54	45,0
4	D	35,0	E	80,0
5	F	20,0	120	100,0
Total	120	100,0		

Rezolvare. A = 0,125 x 120 = 15 pacienți (verificare: 6 + 15 = 21). $B = \frac{21}{120} \times 100 = 17{,}5\%$ (sau 5,0 + 12,5). $C = \frac{33}{120} \times 100 = 27{,}5\%$ (verificare: 17,5 + 27,5 = 45,0). D = 0,35 x 120 = 42 de pacienți. E = 54 + 42 = 96 (verificare: 96/120 = 80,0%). F = 0,20 x 120 = 24 de pacienți (verificare: 96 + 24 = 120). Mediana: n = 120, deci mediana este media valorilor din pozițiile 60 și 61; frecvența cumulată este 54 la valoarea 3 și 96 la valoarea 4, deci ambele poziții au valoarea 4, iar mediana este 4. Modul este tot 4 (42 de pacienți). Mulțumiți sau foarte mulțumiți: 35,0 + 20,0 = 55,0% (66 de pacienți). Raportarea corectă pentru această variabilă ordinală este "mediana 4 (Q1-Q3 3-4)"; cuartilele se găsesc la fel ca mediana, în pozițiile 30-31 (valoarea 3) și 90-91 (valoarea 4).

4. Construiți în Google Forms un formular de colectare a datelor pentru 5 variabile din setul mg2026 (sex, vârstă, TAS, glicemie, satisfacție), cu validările potrivite, și verificați că nu puteți introduce niciuna dintre erorile din mg2026_brut.csv (de exemplu, TAS 1540, glicemie "87 mg/dl", sex "Feminin").

Ce se predă și barem orientativ

Se încarcă pe Moodle: linkul formularului Google Forms (sau o captură de ecran a întrebărilor), foaia Google Sheets descărcată ca.xlsx (filele răspunsuri, dictionar, frecvente, pivot, descriptive), fișierul JASP și un raport PDF de 1-2 pagini cu tabelele de frecvențe, statisticile descriptive, cel puțin patru grafice (bare, circulară, histogramă, box-plot) și paragrafele de interpretare de la L2.8. Baremul orientativ, din 10 puncte: chestionarul, cu tipuri de întrebări corect alese, validări și consimțământ, 2 puncte; legarea de Sheets și recodificarea, 1 punct; tabelele de frecvențe corecte (fumat, satisfacție, TAS pe clase), 2 puncte; tabelele pivot, 1 punct; statisticile descriptive corecte, cu indicatorul potrivit tipului de variabilă, 2 puncte; graficele, cu titluri, axe și unități, 1 punct; interpretarea distribuției, 1 punct.

Varianta fără rețea

Fără conexiune la internet, chestionarul se proiectează pe hârtie (cu aceleași tipuri de întrebări și aceleași reguli de validare scrise ca instrucțiuni), iar răspunsurile se introduc într-o foaie LibreOffice Calc sau Microsoft Excel, cu validarea datelor setată din meniul Date. Tabelele de frecvențe se construiesc cu aceleași funcții (COUNTIF, COUNTIFS, FREQUENCY; în Excel în limba română, CONTARDACĂ, FRECVENȚĂ), tabelele pivot din meniul Inserare > PivotTable, iar statisticile descriptive și graficele (inclusiv box-plotul) în JASP, care funcționează fără rețea. Formularul Google se construiește ulterior, pentru tema de casă.

Lucrarea practică 3. Distribuții și testarea normalității

Obiectiv

Lucrarea transformă în pași de lucru capitolul 3. Pornim de la cinci valori fiziologice măsurate la includerea pacienților în studiul mg2026 (vârsta, tensiunea sistolică, IMC, glicemia a jeun și trigliceridele) și ne întrebăm, pentru fiecare, dacă distribuția ei poate fi considerată normală. Răspunsul nu îl dă un singur număr: îl construim din grafice, din indicatorii de formă și din testele de normalitate, apoi îl folosim ca să decidem ce indicatori descriptivi raportăm și ce tip de test vom putea aplica în lucrările următoare. Lucrăm în paralel în JASP, MedCalc și Google Sheets, pentru că la examenul practic veți avea de ales unul dintre ele.

La final, studentul trebuie să poată:

- să obțină în JASP și MedCalc asimetria, excesul, erorile lor standard, testul Shapiro-Wilk, histograma cu curba normală și graficul Q-Q pentru o variabilă continuă;
- să calculeze în Google Sheets asimetria (SKEW), excesul (KURT), erorile lor standard, scorurile z (STANDARDIZE) și proporțiile teoretice sub curba normală (NORM.DIST, NORM.INV);
- să construiască de mână, în Google Sheets, un grafic Q-Q;
- să aplice o transformare logaritmică, să verifice efectul ei și să retransforme media în medie geometrică;
- să identifice valorile extreme prin scorul z și prin regula 1,5 x IQR;
- să redacteze o decizie argumentată de tipul "variabila X poate/nu poate fi tratată ca normală", în formatul cerut la temele disciplinei.

Instrumente

Programe: JASP 0.98.1 sau mai nou [47]; MedCalc 24.0 [50] (sau calculatoarele lui online); Google Sheets cu setările regionale România (separator zecimal virgula, separator de argumente ";") [43]; opțional Epi Info 7.2.7 [51]. Fișiere: mg2026_curat.csv (160 de rânduri), de pe Moodle, platforma Emediqual [52]; pentru exercițiile suplimentare, fișierele clasice ale disciplinei BD_Studenti și Probleme BPOC. Timp orientativ: 100 de minute.

După importul în Google Sheets, coloanele folosite azi sunt: D = varsta, K = imc0, M = tas0, R = trigliceride, S = glicemie0, V = hba1c0. Datele ocupă rândurile 2-161. Valorile așteptate din această lucrare provin din fișierul de rezultate al setului (calculate cu Python) și coincid cu cele afișate de JASP și MedCalc; diferențe în ultima zecimală pot apărea din rotunjire.

Desfășurare

L3.1 Pregătirea fișierului

1. Deschideți mg2026_curat.csv în JASP (File > Open > Computer). În vizualizarea datelor, verificați că varsta, imc0, tas0, glicemie0, trigliceride și hba1c0 sunt marcate ca variabile continue (iconița cu riglă). Dacă o variabilă apare ca nominală (iconița cu trei cercuri), înseamnă de obicei că o celulă conține text; reveniți la capitolul 2 și la mg2026_brut.csv ca să vedeți de unde poate veni problema.
2. În JASP, creați o variabilă nouă cu logaritmul natural al trigliceridelor: apăsați semnul "+" din dreapta ultimei coloane, dați-i numele ln_trig, alegeți "R" ca mod de calcul și scrieți formula log(trigliceride) (în R, funcția log este logaritmul natural). Prima valoare trebuie să fie 5,1417, pentru că primul pacient are trigliceride 171 mg/dL.

3. Importați același fișier în Google Sheets (File > Import > Upload; separatorul virgulă). Verificați că zecimalele IMC au fost citite ca numere (aliniate la dreapta), nu ca text. În coloana AC scrieți antetul ln_trig, iar în AC2 formula =LN(R2), copiată până la rândul 161.

L3.2 Forma distribuției în JASP

4. Meniul Descriptives > Descriptive Statistics. Mutați în caseta Variables: varsta, tas0, imc0, glicemie0, trigliceride și ln_trig.
5. În secțiunea Statistics bifați: Mean, Median, Std. deviation, Minimum, Maximum, Quartiles; la Distribution bifați Skewness, Kurtosis și Shapiro-Wilk test. JASP va afișa automat și erorile standard ale asimetriei și excesului.
6. În secțiunea Plots (sau Basic plots, în funcție de versiune) bifați Distribution plots (cu opțiunea de afișare a densității), Q-Q plots și Boxplots.
7. Comparați tabelul JASP cu tabelul L3.1. Toate valorile trebuie să coincidă până la rotunjire. Eroarea standard a asimetriei este aceeași pentru toate variabilele (0,192), pentru că depinde doar de n; la fel eroarea standard a excesului (0,381).

Tabelul L3.1. Valori de autoverificare pe mg2026 (n = 160).

Variabila	Media	AS	Mediana	Asimetria	Excesul	Shapiro-Wilk W	p
varsta	56,29	8,77	56,0	0,01	-0,09	0,993	0,687
tas0	131,13	13,88	131,0	0,01	-0,10	0,990	0,303
imc0	29,04	3,91	28,25	1,42	3,68	0,904	< 0,001
glicemie0	105,44	15,76	104,5	0,37	-0,23	0,982	0,037
trigliceride	161,18	76,30	151,5	0,95	1,12	0,945	< 0,001
ln_trig	4,97	0,49	5,02	-0,28	-0,28	0,988	0,200

8. Pentru fiecare variabilă, calculați rapoartele asimetrie / 0,192 și exces / 0,381 și notați-le în caiet. Trebuie să obțineți, de exemplu, 7,42 și 9,64 pentru IMC, 4,97 și 2,94 pentru trigliceride, 1,92 și -0,60 pentru glicemie. Marcați rapoartele care depășesc 2 în valoare absolută.
9. Priviți graficele Q-Q. Pentru tas0 și varsta, punctele trebuie să urmeze diagonala. Pentru imc0, ultimele 3-4 puncte din dreapta urcă mult deasupra liniei (pacienții cu IMC 40,0; 43,4 și 48,3). Pentru trigliceride, punctele formează o curbă concavă în sus, tipică asimetriei pozitive. Pentru ln_trig, curba se îndreaptă. Desenați în caiet, schematic, forma Q-Q pentru IMC și pentru trigliceride.
10. Întrebarea de control: cum interpretați glicemia, pentru care Shapiro-Wilk dă p = 0,037, dar asimetria raportată la eroarea ei standard este sub 2 și media diferă de mediană cu mai puțin de 1 mg/dL? Răspunsul așteptat, pe care îl găsiți argumentat în secțiunea 3.8.5 a cursului, este că abaterea de la normalitate este mică, explicabilă prin pacienții cu diabet nediagnosticat, și nu împiedică folosirea mediei și a testelor parametrice la n = 160; se raportează însă transparent.
11. Opțional, verificați normalitatea pe grupuri: în aceeași fereastră, mutați variabila sex în caseta Split. Pentru glicemie veți observa că abaterea vine mai ales din grupul femeilor (asimetrie 0,57, Shapiro-Wilk p = 0,022), în timp ce la bărbați distribuția este compatibilă cu normalitatea (p = 0,736). Pentru tas0 pe brațe (Split: brat), toate cele trei teste sunt nesemnificative (p = 0,782; 0,373; 0,818). Rețineți ideea: pentru compararea grupurilor contează normalitatea în fiecare grup.

L3.3 Aceleași analize în MedCalc

12. În MedCalc, File > Open și alegeți mg2026_curat.csv (la import, confirmați că prima linie conține numele variabilelor și că separatorul este virgula).
13. Meniul Statistics > Summary statistics. La Variable alegeți tas0. În fereastra de opțiuni bifați testul pentru distribuția normală (Shapiro-Wilk, alternativ Kolmogorov-Smirnov) și, dacă este disponibilă, opțiunea pentru statisticile de formă. Rezultatul afișează media, AS, eroarea standard, IC 95% pentru medie, mediana cu IC, asimetria și excesul (fiecare cu o valoare p a unui test propriu) și concluzia testului de normalitate ("accept Normality" sau "reject Normality").
14. Comparați valorile cu tabelul L3.1. Atenție la două detalii. Primul: formularea "accept Normality" din MedCalc este o prescurtare; în raport scrieți "nu se respinge ipoteza de normalitate". Al doilea: pentru testul Kolmogorov-Smirnov, verificați în documentația programului dacă se aplică corecția Lilliefors. Pe IMC, testul fără corecție dă $p \approx 0{,}08$, iar cu corecția Lilliefors $p \approx 0{,}002$; Shapiro-Wilk dă $p < 0{,}001$. Diferența arată de ce testul clasic, fără corecție, este prea îngăduitor.
15. Graficele: Graphs > Histogram (bifați suprapunerea curbei normale) și Graphs > Normal plot (graficul Q-Q al MedCalc). Pentru trigliceride, histograma cu curba normală suprapusă arată clar că vârful real este la stânga vârfului curbei, iar coada din dreapta depășește curba.
16. Repetați pentru imc0, glicemie0 și trigliceride. Pentru trigliceride, în fereastra Summary statistics bifați și opțiunea de transformare logaritmică (dacă versiunea voastră o oferă) sau creați în foaia de date MedCalc o variabilă nouă, definită prin formula LN(trigliceride). Media geometrică afișată trebuie să fie 144,0 mg/dL.

L3.4 Indicatorii de formă și graficul Q-Q în Google Sheets

17. Într-o foaie nouă, numită "Forma", construiți un tabel cu variabilele pe rânduri și indicatorii pe coloane. Pentru TAS, formulele sunt:

- n: =COUNT(Date!M2:M161), rezultat 160;
- media: =AVERAGE(Date!M2:M161), rezultat 131,13;
- mediana: =MEDIAN(Date!M2:M161), rezultat 131;
- AS: =STDEV(Date!M2:M161), rezultat 13,88;
- asimetria: =SKEW(Date!M2:M161), rezultat 0,01;
- excesul: =KURT(Date!M2:M161), rezultat -0,10;
- Q1 și Q3: =QUARTILE(Date!M2:M161; 1) și =QUARTILE(Date!M2:M161; 3), rezultate 122 și 140.

Copiați rândul pentru celelalte variabile, schimbând coloana. Funcțiile SKEW și KURT din Google Sheets dau aceleași valori corectate pentru eșantion ca JASP și MedCalc.

18. Erorile standard ale asimetriei și excesului, cu formulele exacte, pentru n în celula B2:

- ES asimetrie: =SQRT(6*B2*(B2-1)/((B2-2)*(B2+1)*(B2+3))), rezultat 0,192;
- ES exces: =2*C2*SQRT((B2^2-1)/((B2-3)*(B2+5))), unde C2 conține ES asimetriei, rezultat 0,381.

Calculați rapoartele și formatați condiționat (Format > Conditional formatting) celulele cu valoare absolută mai mare de 2, cu fundal roșu. Google Sheets nu are un test Shapiro-Wilk încorporat; pentru valoarea p folosiți JASP sau MedCalc.

19. Construiți graficul Q-Q pentru TAS. Într-o foaie nouă, "QQ_TAS":

- în A2:A161 copiați valorile TAS și sortați-le crescător (Data > Sort range);
- în B2 scrieți rangul, =ROW()-1, și copiați în jos (1 până la 160);
- în C2 calculați proporția cumulată, =(B2-0,5)/160;
- în D2 calculați cuantila normală teoretică, =NORM.S.INV(C2); prima valoare este -2,73, ultima 2,73;
- selectați D2:D161 și A2:A161 (D pe axa X), Insert > Chart > Scatter chart; adăugați o linie de tendință liniară.

Punctele trebuie să stea foarte aproape de linie; corelația dintre cele două coloane, =CORREL(D2:D161; A2:A161), este 0,996. Repetați pentru trigliceride: corelația scade la 0,972, iar punctele se curbează la capătul din dreapta. Statistica W a testului Shapiro-Wilk se bazează pe exact această idee: cât de bine se aliniază valorile ordonate cu cuantilele normale.

20. Histograma: selectați coloana R (trigliceride), Insert > Chart, Chart type: Histogram chart. În Customize > Histogram, încercați lățimi ale claselor (Bucket size) de 20 și apoi de 50 mg/dL și observați cum se schimbă aspectul. Este motivul pentru care nu judecăm normalitatea doar după histogramă.

L3.5 Scoruri z și proporții sub curba normală

21. În foaia de date, coloana AD, antet z_tas. În AD2: =STANDARDIZE(M2; AVERAGE($M$2:$M$161); STDEV($M$2:$M$161)), copiată până la rândul 161. Primul pacient are TAS 139 mmHg și z = 0,57. Verificați că media coloanei este 0 și AS este 1 (=AVERAGE(AD2:AD161), =STDEV(AD2:AD161)).
22. Numărați pacienții cu scor z mai mare de 2 în valoare absolută: =COUNTIF(AD2:AD161; ">2") + COUNTIF(AD2:AD161; "<-2"). Rezultat: 10 pacienți (6,25%), aproape de cei 4,6% așteptați teoretic pentru o normală. Niciun pacient nu are |z| > 3; scorul z maxim este 2,73, iar cel minim -2,10.
23. Scorul z și percentila unui pacient cu TAS 160 mmHg: =STANDARDIZE(160; 131,13; 13,88) dă 2,08; =NORM.S.DIST(2,08; TRUE) dă 0,9812. Pacientul se află la percentila 98. Verificați același rezultat în tabelul E.1 din anexa E (rândul 2,0, coloana 0,08).
24. Proporția teoretică și cea observată pentru TAS ≥ 140 mmHg:

- teoretic: =1-NORM.DIST(140; AVERAGE(M2:M161); STDEV(M2:M161); TRUE), rezultat 0,261 (26,1%);
- observat: =COUNTIF(M2:M161; ">=140")/COUNT(M2:M161), rezultat 0,256 (25,6%; 41 de pacienți).

25. Intervalul care ar trebui să conțină 95% dintre pacienți: =NORM.INV(0,025; AVERAGE(M2:M161); STDEV(M2:M161)) și =NORM.INV(0,975;...) dau 103,9 și 158,3 mmHg. Numărați câți pacienți se află efectiv între cele două limite cu =COUNTIFS(M2:M161; ">=103,9"; M2:M161; "<=158,3"): rezultatul este 150, adică 93,8%.
26. Repetați pașii 24-25 pentru IMC, cu pragul de 30 kg/m². Modelul normal prezice 40,3% pacienți obezi, dar în eșantion sunt 33,8%. Notați în caiet de ce modelul normal greșește aici (asimetria pozitivă trage media în sus și "lățește" curba).

L3.6 Transformarea logaritmică a trigliceridelor

27. În foaia "Forma", adăugați rândul pentru ln_trig (coloana AC). Asimetria trebuie să scadă de la 0,95 la -0,28, iar excesul de la 1,12 la -0,28. În JASP, Shapiro-Wilk pentru ln_trig dă p = 0,200, față de p < 0,001 pentru trigliceride.

28. Încercați și transformarea radical: coloana AE, =SQRT(R2). Asimetria trebuie să fie 0,32. Comparați cu logaritmul și argumentați alegerea (vezi tabelul 3.7 din curs). Încercați și inversul, =1/R2: asimetria crește la 1,48, adică transformarea este prea puternică.
29. Media geometrică, pe două căi: =EXP(AVERAGE(AC2:AC161)) și =GEOMEAN(R2:R161). Ambele trebuie să dea 144,0 mg/dL. Comparați cu media aritmetică (161,18) și cu mediana (151,5).
30. Intervalul de încredere de 95% pentru media geometrică se calculează pe scala logaritmică și se retransformă:

- limita inferioară: =EXP(AVERAGE(AC2:AC161) - CONFIDENCE.T(0,05; STDEV(AC2:AC161); 160));
- limita superioară: =EXP(AVERAGE(AC2:AC161) + CONFIDENCE.T(0,05; STDEV(AC2:AC161); 160)).

Sheets afișează 133,44 și 155,42 mg/dL, valorile din cursul 3. Ca experiment, rotunjiți limitele de pe scala logaritmică la două zecimale înainte de exponențiere și observați cât se deplasează fiecare limită; concluzia, pe care o notați în caiet: pe scala logaritmică se rotunjește doar la final.

31. Intervalul de referință pentru trigliceride pe scala logaritmică: =EXP(AVERAGE(AC2:AC161) - 1,96*STDEV(AC2:AC161)) și =EXP(AVERAGE(AC2:AC161) + 1,96*STDEV(AC2:AC161)) dau 55,3 și 374,9 mg/dL. Comparați cu intervalul naiv, calculat pe valorile netransformate, 11,6-310,7 mg/dL, și cu percentilele observate, =PERCENTILE(R2:R161; 0,025) și =PERCENTILE(R2:R161; 0,975), care dau 53,8 și 364,0 mg/dL.

L3.7 Valorile extreme

32. Pentru IMC, calculați gardurile Tukey: Q1 = 26,2, Q3 = 31,4, IQR = 5,2; gardul inferior =QUARTILE(K2:K161;1)-1,5*(QUARTILE(K2:K161;3)-QUARTILE(K2:K161;1)) = 18,4; gardul superior = 39,2. Numărați valorile peste gardul superior cu COUNTIF: rezultatul este 3 (40,0; 43,4; 48,3).
33. Repetați pentru trigliceride: gardurile sunt -37,4 și 347,6 mg/dL, iar 5 valori depășesc gardul superior (364, 364, 373, 405, 442). Pentru ln_trig, nicio valoare nu iese din garduri. Concluzia, pe care o scrieți în caiet: "valorile extreme" ale trigliceridelor sunt coada normală a unei distribuții log-normale, nu erori.
34. Scorul z al pacientului cu IMC 48,3 este 4,92. Verificați-l cu STANDARDIZE. Decideți, în scris, ce faceți cu acest pacient (răspuns așteptat: se verifică în fișa sursă, se păstrează, deoarece valoarea este plauzibilă, și se folosesc mediana și IQR sau teste neparametrice pentru IMC; nu se șterge).

L3.8 Redactarea concluziei în formatul a-g

35. Redactați, pentru IMC, rezolvarea în formatul temelor disciplinei. Modelul așteptat:

(a) Ipoteza de cercetare: IMC al pacienților cu risc cardiometabolic din populația studiată are o distribuție normală.

(b) Ipotezele statistice: H0, IMC provine dintr-o populație cu distribuție normală; H1, distribuția IMC în populație nu este normală.

(c) Metoda: testul Shapiro-Wilk, recomandat pentru eșantioane mici și medii, completat de asimetrie și exces raportate la erorile lor standard și de graficul Q-Q.

(d) Rezolvarea: JASP, Descriptives > Descriptive Statistics, cu opțiunile de la pașii 5-6.

(e) Rezultatul: asimetria 1,42 (ES 0,192), excesul 3,68 (ES 0,381), W = 0,904, $p < 0,001$; graficul Q-Q arată abateri mari în coada dreaptă.

(f) Concluzia statistică: $p < 0,05$, se respinge H0.

(g) Concluzia de cercetare: distribuția IMC se abate semnificativ de la normalitate, cu asimetrie pozitivă datorată unui grup mic de pacienți cu obezitate severă. IMC se va descrie prin mediană și IQR (28,25; 26,2-31,4 kg/m²), iar în comparații se vor folosi teste neparametrice sau se va raporta și rezultatul acestora.

36. Faceți același lucru pentru TAS (concluzia așteptată: nu se respinge normalitatea, p = 0,303; se raportează media $\pm$AS, $131{,}13 \pm 13{,}88$ mmHg).

Exerciții suplimentare

Exercițiile 1-3 se rezolvă pe fișierele clasice ale disciplinei, de pe Moodle; rezultatele numerice nu sunt date aici, iar cadrul didactic le verifică la laborator.

1. Fișierul BD_Studenti. Pentru variabilele Varsta, IMC (calculat ca greutate / înălțime², cu înălțimea în metri), Ore_somn și Inaltime, obțineți în JASP asimetria, excesul, erorile lor standard și testul Shapiro-Wilk, plus graficele Q-Q. Repetați analiza separat pe sexe (Split: Gen). Pentru fiecare variabilă, scrieți o frază de decizie de tipul celor de la pasul 35. Atenție la variabilele cu puține valori distincte (orele de somn raportate în ore întregi): ce efect are acest lucru asupra testului Shapiro-Wilk?
2. Fișierul Probleme BPOC. Analizați distribuția capacității vitale inițiale (CVV1), a frecvenței cardiace inițiale (FC1), a tensiunii arteriale maxime inițiale (TMAX1) și a debitului expirator maxim inițial (DEM1). Pentru fiecare, indicați dacă ați folosi la comparația pe sexe testul t nepereche sau testul Mann-Whitney și de ce. Pentru FC1 și FC2, verificați normalitatea diferenței FC1 - FC2 (creați o variabilă nouă în JASP), pentru că aceasta, nu fiecare valoare în parte, contează la testul t pereche.
3. Pe oricare dintre cele două fișiere, alegeți variabila cea mai asimetrică, aplicați transformarea logaritmică și raportați asimetria înainte și după, media geometrică și intervalul ei de încredere de 95%.
4. Problemă cu date în enunț (cu rezolvare). Într-o populație de referință de nou-născuți la termen, greutatea la naștere are media 3400 g și abaterea standard 450 g, cu distribuție aproximativ normală (date fictive). (a) Ce proporție de nou-născuți cântăresc sub 2500 g? (b) Ce greutate separă cei 10% nou-născuți cei mai grei? (c) Un nou-născut are 4500 g: calculați scorul z și percentila.

Rezolvare. (a) z = (2500 - 3400) / 450 = -2,00; $\Phi(-2{,}00) = 1 - 0{,}9772 = 0{,}0228$. Aproximativ 2,3% dintre nou-născuți ar avea sub 2500 g. În Sheets: =NORM.DIST(2500; 3400; 450; TRUE). (b) $\Phi(z) = 0{,}90$ pentru z = 1,28; x = 3400 + 1,28 x 450 = 3976 g (=NORM.INV(0,9; 3400; 450) dă 3977 g). (c) z = (4500 - 3400) / 450 = 2,44; $\Phi(2{,}44) = 0{,}9927$; nou-născutul se află la percentila 99,3. Observație: în realitate, distribuția greutății la naștere are o coadă stângă mai groasă decât normala (prematuri, restricție de creștere), așa că proporția reală sub 2500 g este de obicei mai mare decât prezice modelul normal; exercițiul ilustrează din nou că modelul trebuie verificat pe date.

5. Problemă cu date în enunț (cu rezolvare). O variabilă are, pe 50 de pacienți, asimetria 0,62 și excesul 0,40. Erorile standard pentru n = 50 sunt 0,337 (asimetrie) și 0,662 (exces). Este distribuția compatibilă cu normalitatea?

Rezolvare. Rapoartele sunt 0,62 / 0,337 = 1,84 și 0,40 / 0,662 = 0,60, ambele sub 2. Indicatorii nu semnalează o abatere semnificativă, dar asimetria de 0,62 nu este neglijabilă ca mărime: cu 50 de pacienți, puterea de a o detecta este limitată. Decizia se ia privind și histograma, graficul Q-Q și diferența dintre medie și mediană; dacă variabila este o concentrație serică sau o durată, este prudent să se raporteze și mediana (IQR) și să se verifice rezultatul cu un test neparametric.

Ce se predă și barem orientativ

Se încarcă pe Moodle, cu numele LP3_Nume_Prenume, un singur fișier (PDF sau document Google exportat) care conține:

- tabelul de formă pentru cele cinci variabile și ln_trig, cu media, mediana, AS, asimetria, excesul, rapoartele la erorile standard și Shapiro-Wilk p (2 puncte);
- graficele Q-Q pentru TAS și trigliceride, unul din JASP sau MedCalc și unul construit de mână în Google Sheets (2 puncte);
- calculele de la L3.5: scorul z și percentila pentru 160 mmHg, proporțiile teoretică și observată pentru TAS ≥ 140 și pentru IMC ≥ 30, cu un comentariu de două fraze (2 puncte);
- rezultatele transformării logaritmice: asimetria înainte și după, media geometrică și IC 95% (1,5 puncte);
- analiza valorilor extreme pentru IMC și trigliceride, cu decizia argumentată (1 punct);
- rezolvarea în formatul a-g pentru IMC și TAS (1,5 puncte).

Greșelile care scad nota cel mai des: "testul a confirmat normalitatea" în loc de "nu se respinge ipoteza de normalitate"; folosirea argumentului FALSE în NORM.DIST; ștergerea pacientului cu IMC 48,3; raportarea mediei ± AS pentru IMC fără niciun comentariu despre asimetrie.

Varianta fără rețea

Dacă nu există acces la internet, Google Sheets nu poate fi folosit, dar JASP și MedCalc rulează local. Toți pașii din L3.2 și L3.3 se pot face integral offline. Pașii din L3.4-L3.7 se pot reface în orice program de calcul tabelar instalat local (Microsoft Excel sau LibreOffice Calc), cu aceleași nume de funcții (SKEW, KURT, STANDARDIZE, NORM.DIST, NORM.INV, NORM.S.INV, GEOMEAN, QUARTILE, PERCENTILE); în versiunile românești ale Excel, separatorul de argumente este tot ";". Epi Info 7.2.7 poate calcula, în modulul Classic Analysis, media, AS și cuartilele (comanda MEANS), dar nu oferă teste de normalitate. Ca ultimă soluție, scorurile z și proporțiile sub curbă se calculează de mână, cu tabelul E.1 din anexa E, pe valorile din tabelul L3.1.

Lucrarea practică 4. Estimare și intervale de încredere: efectul dimensiunii eșantionului

Obiectiv

Lucrarea pune în practică capitolul 4. Calculăm intervale de încredere pentru medii și pentru proporții pe setul mg2026, mai întâi cu programele, apoi de mână în Google Sheets, ca să vedem că nu există nicio "cutie neagră": fiecare număr afișat de JASP sau MedCalc se poate reconstrui din medie, abatere standard și n. Construim pe aceleași date intervalele clasice de 68,3%, 95,4% și 99,7% pentru valori și pentru medie, simulăm în Google Sheets cum se schimbă intervalul când luăm eșantioane de 10, 40 și 160 de pacienți și încheiem cu calculul dimensiunii eșantionului în MedCalc, G*Power și Epi Info. Este una dintre puținele lucrări în care veți "vedea" cu ochii voștri hazardul eșantionării.

La final, studentul trebuie să poată:

- să obțină IC 95% și IC 99% pentru o medie în JASP, MedCalc și Google Sheets și să le verifice prin calcul de mână;
- să construiască și să interpreteze, pe aceleași date, intervalele pentru valori individuale și pentru medie;
- să calculeze IC 95% pentru o proporție prin metodele Wald, Wilson și exactă;
- să organizeze o simulare simplă în Google Sheets și să descrie cum depind lățimea și poziția intervalului de mărimea eșantionului;
- să calculeze dimensiunea eșantionului pentru o medie, o proporție și compararea a două medii, cu cel puțin două programe;
- să raporteze corect o estimare: valoare punctuală, IC 95%, unitate de măsură, metodă.

Instrumente

Programe: JASP 0.98.1 sau mai nou [47]; MedCalc 24.0 [50]; G*Power [77]; Epi Info 7.2.7, modulul StatCalc [51] (CDC a oprit dezvoltarea și asistența tehnică pentru Epi Info în septembrie 2025; programul rămâne utilizabil, iar fiecare pas are și o variantă în MedCalc sau G*Power); Google Sheets cu setările regionale România [43]. Fișiere: mg2026_curat.csv de pe Moodle [52]; pentru exercițiile suplimentare, BD_Studenti și Probleme BPOC. Timp orientativ: 100 de minute.

Coloanele din Google Sheets folosite azi: M = tas0, P = colesterol, Q = hdl, S = glicemie0, $X = \text{diabet}_{\text{ogtt}}\left(\frac{0}{1}\right)$, D = varsta. Datele ocupă rândurile 2-161.

Desfășurare

L4.1 Intervalul de încredere pentru medie în JASP

1. Deschideți mg2026_curat.csv în JASP. Meniul T-Tests > One Sample T-Test. Mutați tas0 în caseta Variables. La Test value lăsați 0: nu ne interesează testul, ci intervalul, iar cu valoarea de test 0 diferența afișată este chiar media.
2. Bifați "Location parameter" (sau "Mean difference", în funcție de versiune) și "Confidence interval", cu nivelul 95%. Bifați și "Descriptives". Rezultatul așteptat: media 131,13 mmHg, AS 13,88, ES 1,097, IC 95% 128,96-133,30 mmHg. Valoarea t și p afișate (testul față de 0) nu au sens clinic și se ignoră.

3. Schimbați nivelul de încredere la 99%: intervalul devine 128,27-133,99 mmHg. Apoi la 90%: 129,32-132,95 mmHg. Notați cele trei intervale și lățimile lor (3,6; 4,3; 5,7 mmHg).
4. Repetați pentru glicemie0 (IC 95% așteptat 102,98-107,90 mg/dL) și pentru colesterol (203,63-215,15 mg/dL).
5. Opțional: în Descriptives > Descriptive Statistics, versiunile recente ale JASP oferă direct un interval de încredere pentru medie, iar opțiunea de bootstrap permite intervale pentru alte statistici. Dacă le aveți, verificați că dau același IC 95% pentru TAS.

L4.2 Același interval în MedCalc

6. În MedCalc, Statistics > Summary statistics, variabila tas0. Rezultatul conține rândurile "Arithmetic mean", "95% CI for the mean", "Standard deviation", "Standard error of the mean" și, mai jos, "Median" cu "95% CI for the median". Comparați: IC pentru medie 128,96-133,30 mmHg; intervalul pentru mediană trebuie să fie în jur de 129-134 mmHg (programele pot folosi variante ușor diferite ale metodei rangurilor).
7. Pentru trigliceride, același meniu afișează un IC pentru medie de 149,26-173,09 mg/dL. Dacă activați opțiunea de transformare logaritmică (acolo unde este disponibilă), MedCalc raportează media geometrică 144,0 mg/dL cu IC 95% 133,44-155,42 mg/dL. Notați de ce cele două intervale nu se contrazic (estimează parametri diferiți, vezi secțiunea 4.9.1).

L4.3 Calculul de mână în Google Sheets

8. Într-o foaie nouă, "IC_medie", construiți tabelul pas cu pas pentru TAS:

- B2, n: =COUNT(Date!M2:M161), rezultat 160;
- B3, media: =AVERAGE(Date!M2:M161), rezultat 131,13;
- B4, AS: =STDEV(Date!M2:M161), rezultat 13,88;
- B5, ES: =B4/SQRT(B2), rezultat 1,097;
- B6, t critic: =T.INV.2T(0,05; B2-1), rezultat 1,975;
- B7, marja: =B6*B5, rezultat 2,167;
- B8, limita inferioară: =B3-B7, rezultat 128,96;
- B9, limita superioară: =B3+B7, rezultat 133,30.

9. Verificați marja cu funcția dedicată: =CONFIDENCE.T(0,05; B4; B2) trebuie să dea tot 2,167. Calculați și =CONFIDENCE.NORM(0,05; B4; B2), care folosește z = 1,96 și dă 2,151. Diferența este mică la n = 160; în pasul 16 o veți vedea crescând.
10. Pentru IC 99%, schimbați 0,05 în 0,01 în B6: t = 2,607, marja 2,861, intervalul 128,27-133,99 mmHg.

L4.4 Intervale pentru valori și intervale pentru medie

11. În aceeași foaie, construiți tabelul L4.1 cu formule (k în coloana A: 1, 2, 3):

- limita inferioară pentru valori: =$B$3-A12*$B$4; superioară: =$B$3+A12*$B$4;
- limita inferioară pentru medie: =$B$3-A12*$B$5; superioară: =$B$3+A12*$B$5.

Tabelul L4.1. Valori de autoverificare pentru TAS (mg2026).

Încredere	**Valori individuale** $(m \pm k \times s)$	**Media** $(m \pm k \times \text{ES})$
68,3% $(k = 1)$	117,3-145,0 mmHg	130,03-132,23 mmHg
95,4% $(k = 2)$	103,4-158,9 mmHg	128,94-133,33 mmHg
99,7% $(k = 3)$	89,5-172,8 mmHg	127,84-134,42 mmHg

12. Numărați câți pacienți se află efectiv în fiecare interval pentru valori, cu =COUNTIFS(Date!M2:M161; ">="&C12; Date!M2:M161; "<="&D12). Rezultate așteptate: 112 pacienți (70,0%), 150 (93,8%) și 160 (100%). Numărați apoi câți pacienți se află în intervalul de 95,4% pentru medie: doar 25 de pacienți. Scrieți în caiet, cu cuvintele voastre, de ce al doilea număr este atât de mic și de ce asta nu este o problemă.
13. Repetați tabelul pentru vârstă (m = 56,29; s = 8,77): intervalul de 95,4% pentru valori este 38,75-73,83 ani, iar pentru medie 54,90-57,68 ani. Este exact tipul de problemă de la examen: "estimați cu o încredere de 95,4% valorile vârstei" față de "estimați cu o încredere de 95,4% media vârstei în populație".

L4.5 Intervalul de încredere pentru o proporție

14. Proporția pacienților cu diabet confirmat prin TTGO. În foaia "IC_prop":

- B2, k: =SUM(Date!X2:X161), rezultat 29;
- B3, n: =COUNT(Date!X2:X161), rezultat 160;
- B4, p: =B2/B3, rezultat 0,181;
- B5, z: =NORM.S.INV(0,975), rezultat 1,96;
- B6, ES Wald: =SQRT(B4*(1-B4)/B3), rezultat 0,0305;
- B7 și B8, limitele Wald: =B4-B5*B6 și =B4+B5*B6, rezultate 0,122 și 0,241.

15. Intervalul Wilson, în trei celule:

- B10, centrul: =(B4+B5^2/(2*B3))/(1+B5^2/B3), rezultat 0,189;
- B11, marja: =B5/(1+B5^2/B3)*SQRT(B4*(1-B4)/B3+B5^2/(4*B3^2)), rezultat 0,059;
- B12 și B13, limitele: =B10-B11 și =B10+B11, rezultate 0,129 și 0,248.

IC 95% Wilson: 12,9%-24,8%, față de Wald 12,2%-24,1%.

16. Intervalul exact (Clopper-Pearson): în JASP, Frequencies > Binomial Test, variabila diabet_ogtt (setată ca nominală), bifați "Confidence interval"; pentru nivelul 1 rezultatul este 12,5%-25,0%. În Epi Info, Classic Analysis, comanda FREQ diabet_ogtt afișează aceleași limite exacte. În MedCalc, calculatorul pentru o proporție (Tests > One proportion sau calculatoarele online) dă tot intervalul exact.
17. Test de sensibilitate: în B2 și B3 scrieți manual 2 și 20 (două complicații la 20 de pacienți). Intervalul Wald devine -3,1% până la 23,1%, cu o limită negativă imposibilă; intervalul Wilson este 2,8%-30,1%. Apoi scrieți 0 și 50: Wald dă 0-0%, Wilson 0-7,1%, iar regula lui trei dă 3/50 = 6%. Readuceți formulele inițiale la final.

L4.6 Simularea: cum depinde intervalul de mărimea eșantionului

În această parte tratăm cei 160 de pacienți din mg2026 ca pe o "populație" cu media cunoscută, 131,13 mmHg, și extragem din ea eșantioane aleatoare, exact ca în figura 4.2 din curs.

18. Într-o foaie nouă, "Simulare", în A1 scrieți antetul "esantion". În A2 scrieți formula =INDEX(Date!$M$2:$M$161; RANDBETWEEN(1; 160)) și copiați-o până la A161. Fiecare celulă alege la întâmplare un pacient din cei 160 (cu întoarcere, deci același pacient poate apărea de mai multe ori). La fiecare modificare a foii, valorile se schimbă; pentru a forța o nouă extragere, apăsați Delete pe o celulă goală sau folosiți File > Settings > Calculation > Recalculation: "On change and every minute".
19. Calculați intervalele pentru primele 10, primele 40 și toate cele 160 de valori ale coloanei A:

- n = 10, în D2: =AVERAGE(A2:A11); E2: =CONFIDENCE.T(0,05; STDEV(A2:A11); 10); limitele =D2-E2 și =D2+E2; lățimea =2*E2;
- n = 40: aceleași formule pe A2:A41, cu 40;
- n = 160: pe A2:A161, cu 160.

În coloana următoare scrieți =IF(AND(131,13>=F2; 131,13<=G2); "DA"; "NU"), care spune dacă intervalul conține media "populației".

20. Copiați rezultatele (Edit > Paste special > Values only) într-un tabel de repetări, apoi provocați o nouă extragere și repetați. Faceți cel puțin 20 de repetări (în echipă de câte doi studenți, 40). Pentru fiecare n, calculați la final media lățimilor și numărul de intervale care conțin 131,13.
21. Ce trebuie să observați. Pe 10.000 de repetări făcute cu calculatorul, lățimea medie a IC 95% a fost 19,3 mmHg pentru n = 10 (cu variații mari, de la aproximativ 12 la 27 mmHg între eșantioane), 8,8 mmHg pentru n = 40 (7,2-10,4) și 4,3 mmHg pentru n = 160 (3,9-4,7). Proporția intervalelor care au conținut media populației a fost 95,3%, 94,9% și, respectiv, 94,9%, adică foarte aproape de 95% la toate cele trei mărimi. Cu 20 de repetări, veți avea probabil 18-20 de "DA" pentru fiecare n; dacă apare un "NU", nu este o eroare, ci exact cele 5% promise. Comparați și cu eșantioanele fixe din curs (tabelul 4.3): n = 10, IC 116,38-143,62 mmHg (lățime 27,2); n = 40, 123,08-133,67 (10,6); n = 160, 128,96-133,30 (4,3).
22. Concluzia, pe care o scrieți în caiet în 3-4 fraze: lățimea intervalului scade aproximativ cu $\sqrt{n}$ (de la 10 la 40 de pacienți se înjumătățește, la fel de la 40 la 160); nivelul de încredere de 95% nu depinde de n, ci se referă la proporția intervalelor care conțin parametrul; în eșantioanele mici nu doar lățimea, ci și poziția intervalului variază mult de la o repetare la alta.
23. Varianta cu populație normală, opțional. Înlocuiți formula din A2 cu =NORM.INV(RAND(); 131; 14), care generează valori dintr-o populație normală cu media 131 și AS 14 mmHg, și repetați pașii 19-20, verificând acoperirea față de 131. Este exact simularea din figura 4.2.

L4.7 Calculul dimensiunii eșantionului

24. Pentru o medie, de mână în Sheets: =ROUNDUP((1,96*14/3)^2; 0) dă 84 de pacienți (marja ±3 mmHg, σ = 14 mmHg). Încercați marjele 2 și 5 mmHg: 189 și 31. Unele programe folosesc distribuția t în locul lui z și dau ceva mai mult (pentru ±3 mmHg, 87); diferența nu are importanță practică.
25. Pentru o proporție: =ROUNDUP(1,96^2*0,2*0,8/0,05^2; 0) dă 246 (prevalență anticipată 20%, marja ±5 puncte procentuale). Cu p = 0,5, 385.
26. Epi Info, StatCalc > Population survey. Introduceți: Population size 1000, Expected frequency 20%, Acceptable margin of error 5%, Design effect 1, Clusters 1. La nivelul de încredere 95%, programul afișează 198 de subiecți: este eșantionul de 246 corectat pentru o populație finită de 1000 (246 / (1 + 245 / 1000) = 197,6). Schimbați Population size la 999999: rezultatul devine 246. Faceți același calcul cu Population size 500 (rezultat așteptat: 166).
27. G*Power, compararea a două medii: Test family "t tests", Statistical test "Means: Difference between two independent means (two groups)", Type of power analysis "A priori: Compute required sample size". Introduceți Tail(s): Two, Effect size d: 0,5, α err prob: 0,05, Power (1-β err prob): 0,80, Allocation ratio N2/N1: 1. Rezultat: 64 pe grup, 128 în total. Repetați cu d = 0,3 (176 pe grup) și cu d = 0,8 (26 pe grup). Butonul

"Determine" vă permite să calculați d din medii și abatere standard: pentru o diferență de 5 mmHg și AS 14 mmHg, d = 0,357, iar eșantionul necesar este 125 pe grup.

28. MedCalc, meniul Sampling: alegeți calculul pentru compararea a două medii, introduceți α = 0,05, β = 0,20 (putere 80%), diferența 5 și AS 14 și comparați cu G*Power. Folosiți apoi calculul pentru o singură proporție (prevalență 20%, precizie ±5%).
29. Ajustarea pentru pierderi: dacă anticipați 20% pierderi, =ROUNDUP(64/(1-0,2); 0) dă 80 pe grup. Verificați de ce formula =64*1,2 (rezultat 77) este greșită: calculați câți pacienți rămân din 77 după pierderi de 20%.
30. Puterea unui eșantion existent: în G*Power, Type of power analysis "Post hoc: Compute achieved power", $d = 0{,}357\left(\frac{5\text{ mmHg}}{14\text{ mmHg}}\right)$, α = 0,05, Sample size group 1: 86, group 2: 74. Rezultatul este în jur de 61%: comparația TAS între sexe din mg2026 avea o putere modestă pentru o diferență de 5 mmHg. Rețineți: acest calcul se face pentru o diferență relevantă clinic stabilită dinainte, nu pentru diferența observată.

L4.8 Raportarea unei estimări

31. Scrieți, pentru TAS, glicemie și prevalența diabetului, câte o frază de raportare de tip articol. Model: "Media TAS la includere a fost 131,13 mmHg (AS 13,88; IC 95% pentru medie 128,96-133,30 mmHg; n = 160). Prevalența diabetului confirmat prin TTGO a fost 18,1% (29/160; IC 95% Wilson 12,9-24,8%)." Verificați că fiecare frază conține: estimarea, unitatea de măsură, intervalul cu nivelul lui, metoda (pentru proporții) și n.

Exerciții suplimentare

Exercițiile 1-3 se rezolvă pe fișierele clasice ale disciplinei; rezultatele numerice nu sunt date aici.

1. Fișierul BD_Studenti. (a) Estimați, cu o încredere de 68,3%, valorile înălțimii studenților. (b) Estimați, cu o încredere de 95,4%, media înălțimii ideale la nivelul populației studenților. (c) Estimați, cu o încredere de 99,7%, valorile IMC ale studenților și, separat, media IMC în populație. Comentați diferența dintre ultimele două intervale. (d) Calculați IC 95% cu distribuția t pentru media orelor de somn, separat pentru fete și băieți, și spuneți dacă intervalele se suprapun.
2. Fișierul Probleme BPOC. Calculați IC 95% pentru media frecvenței cardiace inițiale (FC1) și pentru media diferenței FC1 - FC2. Calculați IC 95% Wilson pentru proporția pacienților cu dispnee la includere (DISPN = 1), pe tot lotul și separat sub/peste 50 de ani.
3. Pe BD_Studenti, extrageți în Google Sheets eșantioane aleatoare de 10 și de 30 de studenți (metoda de la pasul 18) și comparați lățimea IC 95% pentru media înălțimii cu cea obținută pe tot fișierul.
4. Problemă cu date în enunț (cu rezolvare). Un studiu-pilot pe 15 pacienți hipertensivi găsește o scădere medie a TAS de 9 mmHg, cu AS 12 mmHg. (a) Calculați IC 95% pentru scăderea medie. (b) Câți pacienți ar fi necesari pentru a estima scăderea medie cu o marjă de ±2 mmHg?

Rezolvare. (a) ES $= \frac{12}{\sqrt{15}} = \frac{12}{3{,}873} = 3{,}098$ mmHg; gl = 14, t = 2,145; marja = 2,145 x 3,098 = 6,65 mmHg; IC 95%: 9 ± 6,65, adică 2,35-15,65 mmHg. Intervalul exclude 0, dar este foarte larg: scăderea reală ar putea fi minoră sau importantă. $(b)\ n = \left(\frac{1{,}96\times 12}{2}\right)^2 = (11{,}76)^2 = 138{,}3$, deci 139 de pacienți (cu distribuția t, un program va da 1-2 pacienți în plus). Observați că AS din pilot este ea însăși imprecisă pe 15 pacienți; un calcul prudent ar folosi o valoare ceva mai mare.

5. Problemă cu date în enunț (cu rezolvare). Într-un sondaj în cabinet, 45 din 150 de pacienți vaccinați antigripal declară o reacție locală. Calculați IC 95% Wald și Wilson.

Rezolvare. $p = \frac{45}{150} = 0{,}30$; ES $= \sqrt{\frac{0{,}30 \times 0{,}70}{150}} = 0{,}0374$; Wald: 0,30 ± 0,073, adică 22,7%-37,3%. Wilson: $\frac{z^2}{n} = 0{,}0256$; centrul = (0,30 + 0,0128) / 1,0256 = 0,305; marja = $\left(\frac{1{,}96}{1{,}0256}\right) \times \sqrt{0{,}0014 + 0{,}0000427} = 1{,}911 \times 0{,}0380 = 0{,}0726$; IC 23,2%-37,8%. Pentru p = 30% și n = 150, cele două metode dau rezultate foarte apropiate; diferența devine importantă doar pentru proporții extreme sau n mic.

Ce se predă și barem orientativ

Se încarcă pe Moodle, cu numele LP4_Nume_Prenume, un fișier care conține:

- intervalele de încredere de 90%, 95% și 99% pentru media TAS, cu captura din JASP sau MedCalc și tabelul de calcul din Sheets (2 puncte);
- tabelul L4.1 completat pentru TAS și pentru vârstă, cu numărul de pacienți din fiecare interval și comentariul de la pasul 12 (2 puncte);
- IC 95% pentru proporția diabetului prin metodele Wald, Wilson și exactă, plus testul de sensibilitate de la pasul 17 (1,5 puncte);
- tabelul simulării (minimum 20 de repetări pentru n = 10, 40, 160), cu media lățimilor, numărul de intervale care conțin 131,13 și concluzia de la pasul 22 (2,5 puncte);
- calculele de dimensiune a eșantionului de la L4.7, cu capturile din G*Power sau MedCalc și Epi Info (1,5 puncte);
- frazele de raportare de la pasul 31 (0,5 puncte).

Greșelile frecvente: folosirea abaterii standard în locul erorii standard pentru intervalul mediei (sau invers); interpretarea IC 95% al mediei ca interval în care se află 95% dintre pacienți; rotunjirea în jos a dimensiunii eșantionului; ajustarea pentru pierderi prin înmulțire cu 1,2.

Varianta fără rețea

JASP, MedCalc, G*Power și Epi Info rulează local, deci pașii L4.1, L4.2, L4.5 (variantele JASP și Epi Info) și L4.7 se fac fără internet. Pentru L4.3, L4.4 și simulare se folosește un program de calcul tabelar instalat local (Excel sau LibreOffice Calc), cu aceleași funcții: AVERAGE, STDEV, T.INV.2T, CONFIDENCE.T, CONFIDENCE.NORM, NORM.S.INV, NORM.INV, RAND, RANDBETWEEN, INDEX, COUNTIFS. În Excel, recalcularea extragerilor aleatoare se forțează cu tasta F9. Dacă nu este disponibil niciun calculator, intervalele se calculează de mână, cu valorile critice din tabelele E.2 și E.3 ale anexei E, pe statisticile din tabelul L4.1.

Lucrarea practică 5. Testarea ipotezelor statistice: formularea ipotezelor, erori și decizie

Obiectiv

Lucrarea transformă în gesturi practice raționamentul din capitolul 5: de la o întrebare clinică la ipotezele H0 și H1, de la date la statistica testului și la valoarea p, de la valoarea p la o decizie formulată corect. Lucrăm pe testul cel mai simplu, testul t pentru o singură medie, pentru că acolo fiecare pas se poate verifica de mână. Partea a doua a lucrării este o simulare: veți genera în Google Sheets sute de "studii" în care ipoteza nulă este adevărată și veți vedea cu ochii voștri că aproximativ 5% dintre ele ies "semnificative". Simularea vă va ajuta să înțelegeți, mai bine decât orice definiție, ce înseamnă α și de ce un rezultat nesemnificativ nu dovedește absența efectului.

La final, studentul trebuie să poată:

- să formuleze ipotezele statistice H0 și H1 (în cuvinte și matematic) pentru o întrebare clinică dată și să justifice alegerea unui test bilateral sau unilateral;
- să aplice testul t pentru o singură medie în JASP, MedCalc și Google Sheets și să citească în rezultate statistica t, gradele de libertate, valoarea p, intervalul de încredere și mărimea efectului;
- să calculeze de mână statistica t și să obțină valoarea p din tabel și cu funcțiile T.DIST.2T și T.INV.2T;
- să simuleze eroarea de tip I și puterea testului în Google Sheets și să interpreteze proporția rezultatelor semnificative;
- să calculeze, cu o formulă simplă și cu un program, numărul de pacienți necesar pentru o putere dată;
- să redacteze rezolvarea unei probleme în formatul a-g al temelor disciplinei.

Instrumente

Programe: JASP 0.98.1 [47]; MedCalc 24.0 [50]; Google Sheets, cu setările regionale România (separatorul de argumente este ";", iar separatorul zecimal este virgula) [43]; opțional G*Power pentru calculul puterii [77]. Fișiere: mg2026_curat.csv (160 de pacienți) de pe platforma Moodle a disciplinei [52]; pentru exercițiile suplimentare, fișierele clasice Date_HbA1c, BD_Studenti și BPOC. Timp orientativ: 100 de minute (20 de minute ipoteze, 35 de minute testul t în cele trei programe, 25 de minute simularea, 20 de minute puterea și tema).

Coloanele din mg2026_curat.csv, importat în Google Sheets cu antetul pe rândul 1 și datele pe rândurile 2-161, sunt: A id, B brat, C sex, D varsta, E mediu, F educatie, G fumat, H activitate, I inaltime, J greutate0, K imc0, L talie, M tas0, N tad0, O fc0, P colesterol, Q hdl, R trigliceride, S glicemie0, T glicemie3, U glicemie6, V hba1c0, W findrisc, X diabet_ogtt, Y greutate6, Z tas6, AA scadere5, AB satisfactie. Verificați ordinea după import; formulele de mai jos presupun această așezare.

Desfășurare

L5.1 Formularea ipotezelor

Înainte de a deschide orice program, scrieți pe hârtie ipotezele. Este regula care previne cele mai multe greșeli de la examenul practic: testul se alege după întrebare, nu după ce "iese" în program.

1. Citiți întrebarea și identificați populația, variabila și comparația. Exemplu: "Diferă media TAS a pacienților cu risc cardiometabolic de 125 mmHg?" Populația: adulții cu risc cardiometabolic din

cabinetele de medicină de familie. Variabila: TAS la includere (numerică, mmHg). Comparația: media populației față de o valoare fixă, 125 mmHg.

2. Scrieți ipoteza de cercetare în limbaj clinic: "media TAS a pacienților cu risc cardiometabolic diferă de 125 mmHg".
3. Scrieți H0 în cuvinte și matematic: "media TAS în populație este 125 mmHg", $\mu = 125$. Scrieți H1: "media TAS în populație diferă de 125 mmHg", $\mu \neq 125$.
4. Decideți bilateral sau unilateral. Implicit, bilateral; unilateralul cere o justificare scrisă înainte de analiză (de exemplu, un studiu de non-inferioritate).
5. Fixați $\alpha = 0,05$.

Exercițiu în clasă (10 minute, pe perechi). Scrieți ipotezele H0 și H1, în cuvinte și matematic, pentru întrebările de mai jos; precizați parametrul (μ, μD, π, ρ) și tipul testului (bilateral sau unilateral).

- (a) Media glicemiei a jeun în mg2026 diferă de 100 mg/dL?
- (b) Media HbA1c diferă de 5,7%?
- (c) Glicemia a jeun se modifică între includere și 6 luni la aceiași pacienți?
- (d) TAS diferă între femei și bărbați?
- (e) Scăderea glicemiei diferă între cele trei brațe ale studiului?
- (f) Proporția pacienților cu diabet confirmat prin TTGO diferă între obezi și neobezi?
- (g) IMC și TAS sunt corelate?
- (h) Un nou glucometru nu subestimează glicemia cu mai mult de 5 mg/dL față de laborator.

Verificare. Pentru (a): H0: $\mu = 100$; H1: $\mu \neq 100$. Pentru (c): H0: $\mu D = 0$, unde D = glicemie0 - glicemie6; H1: $\mu D \neq 0$. Pentru (e): H0: $\mu A = \mu B = \mu C$; H1: cel puțin o medie diferă. Pentru (g): H0: $\rho = 0$; H1: $\rho \neq 0$. Pentru (h): ipoteze de non-inferioritate, unilaterale, cu marja de 5 mg/dL fixată dinainte. Ipotezele se scriu întotdeauna cu parametrii populației, niciodată cu media eșantionului.

L5.2 Testul t pentru o singură medie în JASP

1. Deschideți JASP și încărcați fișierul: meniul principal (cele trei linii din stânga sus) > Open > Computer > Browse > mg2026_curat.csv. Verificați în vizualizarea datelor că tas0 și glicemie0 apar ca variabile continue (pictograma cu riglă); dacă nu, faceți clic pe pictogramă și alegeți Scale.
2. Din bara de module alegeți T-Tests > One Sample T-Test (în secțiunea Classical).
3. Mutați tas0 în caseta Variables. În caseta Test value scrieți 125.
4. Bifați la Tests: Student și Wilcoxon signed-rank. La Alt. Hypothesis lăsați "≠ Test value" (bilateral).
5. La Additional Statistics bifați: Location parameter (diferența medie față de valoarea testată), Confidence interval (95%), Effect size (d Cohen), Descriptives și, dacă vreți un grafic, Descriptives plots.

6. La Assumption Checks bifați Normality (Shapiro-Wilk) și, opțional, Q-Q plot.

Ce trebuie să obțineți pentru tas0 față de 125: media 131,13 mmHg, AS 13,88, n = 160; testul Student t = 5,59, gl = 159, p < 0,001; diferența medie (location parameter) 6,13 mmHg, cu IC 95% 3,96-8,30 mmHg; d Cohen = 0,44; Shapiro-Wilk p = 0,303; testul Wilcoxon p < 0,001. Observați că JASP raportează intervalul de încredere pentru diferența față de 125, nu pentru medie; adunând 125 obțineți intervalul mediei, 128,96-133,30 mmHg, cunoscut din capitolul 4.

7. Schimbați Test value la 130 și citiți din nou rezultatul. Trebuie să obțineți t = 1,03, p = 0,304, IC 95% al diferenței de la -1,04 la 3,30 mmHg (care conține zero). Scrieți decizia corectă: "nu se respinge H0", nu "se acceptă H0".
8. Repetați analiza pentru glicemie0, cu Test value 100. Trebuie să obțineți: media 105,44 mg/dL, AS 15,76; t = 4,36, gl = 159, p < 0,001; diferența medie 5,44 mg/dL (IC 95% 2,98-7,90); d = 0,35; Shapiro-Wilk p = 0,037 (abatere ușoară de la normalitate); Wilcoxon p < 0,001. Discutați cu colegul: este necesar testul Wilcoxon aici? Cu n = 160 și o asimetrie mică (0,37), testul t este robust, iar cele două teste duc la aceeași concluzie.
9. Repetați pentru hba1c0 cu Test value 5,7. Rezultatul așteptat: t = -3,48, gl = 159, p < 0,001; media este sub 5,7%.

L5.3 Același test în MedCalc

1. Deschideți MedCalc și importați fișierul: File > Import > CSV (sau File > Open, cu tipul de fișier CSV). Verificați separatorul zecimal din opțiunile de import, deoarece fișierul folosește punctul zecimal.
2. Din meniul Statistics alegeți T-tests > One sample t-test.
3. La Variable selectați tas0, iar la Test mean (valoarea de comparat) scrieți 125. Confirmați cu OK.
4. Citiți în fereastra de rezultate: media, IC 95% pentru medie, diferența față de valoarea testată, statistica t, gradele de libertate și valoarea p. Rezultatele trebuie să coincidă cu cele din JASP (t = 5,59; p < 0,0001).
5. MedCalc permite și calculul direct din date rezumative, fără fișier: în meniul Tests căutați testul pentru o medie (Test for one mean), introduceți media 131,13, AS 13,88, n = 160 și valoarea 125. Rezultatul trebuie să fie același. Este o metodă utilă pentru verificarea cifrelor dintr-un articol.

L5.4 Același test în Google Sheets, pas cu pas

Google Sheets nu are o funcție dedicată testului t pentru o medie, dar îl puteți construi din funcții simple; tocmai de aceea este cel mai bun exercițiu de înțelegere. Lucrați într-o foaie nouă (de exemplu, "LP5") și scrieți etichetele în coloana A și formulele în coloana B.

1. B1, media: =AVERAGE(Date!M2:M161) (dacă datele sunt în foaia "Date"). Rezultat: 131,13.
2. B2, abaterea standard: =STDEV(Date!M2:M161). Rezultat: 13,88.
3. B3, n: =COUNT(Date!M2:M161). Rezultat: 160.
4. B4, valoarea testată μ0: 125.
5. B5, eroarea standard: =B2/SQRT(B3). Rezultat: 1,097.
6. B6, statistica t: =(B1-B4)/B5. Rezultat: 5,59.
7. B7, gradele de libertate: =B3-1. Rezultat: 159.
8. B8, valoarea p bilaterală: =T.DIST.2T(ABS(B6);B7). Rezultat: aproximativ 0,0000001, adică p < 0,001. Funcția cere o valoare pozitivă, de aceea folosim ABS.
9. B9, valoarea critică pentru α = 0,05 bilateral: =T.INV.2T(0,05;B7). Rezultat: 1,975.
10. B10 și B11, limitele IC 95% ale mediei: =B1-B9*B5 și =B1+B9*B5. Rezultat: 128,96 și 133,30.
11. B12, decizia automată: =IF(B8<0,05;"se respinge H0";"nu se respinge H0").
12. B13, mărimea efectului d: =(B1-B4)/B2. Rezultat: 0,44.

O verificare elegantă cu funcția T.TEST. Testul t pentru o medie este echivalent cu un test t pereche între date și o coloană constantă egală cu μ0. Scrieți 125 în celulele Date!AD2:AD161 (scrieți 125 în AD2 și trageți în jos) și calculați =T.TEST(Date!M2:M161;Date!AD2:AD161;2;1). Rezultatul trebuie să fie identic cu cel din B8. Al treilea argument (2) cere testul bilateral, iar al patrulea (1) cere testul pereche.

Schimbați acum valoarea din B4 la 130, apoi la 140, și urmăriți cum se modifică simultan t, p, decizia și poziția lui μ0 față de intervalul de încredere. Completați tabelul de mai jos; el reproduce, în foaia voastră, echivalența celor trei căi de decizie din capitolul 5.

μ0	t	p	μ0 în IC 95%?	Decizia
125	5,59	< 0,001	nu	se respinge H0
130				
132				
140				

L5.5 Valoarea p calculată de mână din statistica t

La examenul scris nu aveți calculator statistic, dar trebuie să puteți încadra valoarea p între praguri folosind tabelul valorilor critice (tabelul 5.6 din capitol).

1. Calculați t și gradele de libertate.
2. Căutați în tabel rândul cu gradele de libertate (sau cel mai apropiat rând mai mic, ca să fiți prudenți).
3. Comparați |t| cu valorile critice pentru 0,05, 0,01 și 0,001. Dacă |t| este mai mic decât valoarea critică pentru 0,05, atunci p > 0,05; dacă se află între valorile pentru 0,05 și 0,01, atunci 0,01 < p < 0,05; și așa mai departe.
4. Verificați apoi în Google Sheets cu =T.DIST.2T(t;gl).

Exerciții de încadrare (rezolvați întâi cu tabelul, apoi verificați în Sheets): t = 2,10 cu gl = 20 (rezultatul exact trebuie să fie 0,049); t = 0,83 cu gl = 9 (0,43); t = 3,10 cu gl = 9 (0,013); t = 1,80 cu gl = 25 (bilateral 0,084; unilateral =T.DIST.RT(1,8;25) = 0,042); t = 4,36 cu gl = 159 (< 0,001). Pentru penultimul exemplu, discutați: de ce este periculos să decideți după ce ați văzut datele că testul "ar trebui" să fie unilateral?

Exercițiul 5 al lucrării: calculați de mână t pentru un eșantion mic. Luați primii 10 pacienți din fișier (rândurile 2-11) și testați dacă media TAS diferă de 125 mmHg. Calculați media, AS, ES, t, gl și încadrați p cu tabelul. Comparați cu rezultatul pe 160 de pacienți și explicați, în două fraze, de ce concluzia poate fi diferită. (Rezultatul depinde de pacienții aleși; în capitol, un eșantion aleatoriu de 10 pacienți a dat t = 0,83, p = 0,43.)

L5.6 Simularea erorii de tip I în Google Sheets

Scopul este să vedeți ce înseamnă α = 0,05. Vom genera 200 de "studii" în care două grupuri de câte 20 de pacienți provin din aceeași populație (TAS cu media 130 mmHg și AS 14 mmHg). Prin construcție, H0 este adevărată în fiecare studiu. Orice rezultat semnificativ este, deci, o eroare de tip I.

1. Creați o foaie nouă, "Simulare". În celula A1 scrieți "studiul", în B1:U1 etichetele grupului 1 (g1_1... g1_20), în V1:AO1 etichetele grupului 2 (g2_1... g2_20), în AP1 "p".
2. În A2 scrieți 1, în A3 =A2+1 și trageți până la A201 (200 de studii).
3. În B2 scrieți =NORM.INV(RAND();130;14) și copiați formula în tot blocul B2:AO201. Fiecare celulă este acum o valoare TAS aleatoare dintr-o populație normală cu media 130 și AS 14.
4. În AP2 scrieți =T.TEST(B2:U2;V2:AO2;2;2) și trageți până la AP201. Fiecare rând primește valoarea p a unui test t pentru eșantioane independente între cele două grupuri ale aceluiași studiu.

5. În AR1 scrieți "proporția p < 0,05", iar în AR2 =COUNTIF(AP2:AP201;"<0,05")/200. Formatați ca procent.
6. Modificați orice celulă goală (de exemplu, scrieți ceva în AT1 și ștergeți) pentru a forța recalcularea funcției RAND. Notați de 5 ori proporția obținută.

Ce trebuie să obțineți: proporții în jurul a 5%, de obicei între 2% și 8% (cu 200 de studii, variația întâmplătoare este de aproximativ ±3 puncte procentuale). În simularea din capitol (1000 de studii), proporția a fost 4,9%. Construiți și o histogramă a valorilor p din coloana AP (Insert > Chart > Histogram): ar trebui să fie aproximativ plată, cu valori p răspândite uniform între 0 și 1. Concluzia de scris în caiet: când H0 este adevărată, testul declară "semnificativ" aproximativ α din studii; α este rata de alarme false pe care o acceptăm.

7. Simularea puterii. Schimbați formula din grupul 2 (V2:AO201) în =NORM.INV(RAND();137;14), adică o diferență reală de 7 mmHg între grupuri (d = 0,5). Acum H0 este falsă, iar proporția rezultatelor semnificative estimează puterea. Ce trebuie să obțineți: în jur de 34% (valoarea teoretică pentru d = 0,5 și 20 de pacienți pe grup este 33,8%). Cu alte cuvinte, două din trei astfel de studii ar rata un efect real de 7 mmHg. Aceasta este eroarea de tip II.
8. Încercați și o diferență mare, media grupului 2 egală cu 144 mmHg (d = 1): puterea ar trebui să crească la aproximativ 87%.
9. Comparații multiple. Readuceți media grupului 2 la 130 (H0 adevărată). Gândiți-vă la fiecare grup de 10 rânduri ca la un singur "articol" cu 10 teste. Câte dintre cele 20 de "articole" au cel puțin un p < 0,05? Teoretic, 1 - 0,95^10 = 40% dintre ele. Verificați, numărând sau cu =COUNTIF pe blocuri de câte 10 rânduri.

L5.7 Calculul puterii și al dimensiunii eșantionului

1. Formula aproximativă pentru numărul de pacienți pe grup la compararea a două medii, în Google Sheets: =2*(NORM.S.INV(0,975)+NORM.S.INV(0,8))^2*σ^2/Δ^2. Pentru σ = 14 mmHg și Δ = 7 mmHg (d = 0,5): rezultatul este 62,8, adică aproximativ 63 pe grup; calculul exact, cu distribuția t, dă 64 pe grup (valoarea folosită în capitolul 5). Pentru Δ = 5 mmHg: aproximativ 123 pe grup. Pentru Δ = 4,2 mmHg (d = 0,3): aproximativ 175 pe grup (exact 176).
2. Formula aproximativă pentru putere, dat fiind n pe grup: =NORM.S.DIST(d*SQRT(n/2)-NORM.S.INV(0,975);TRUE). Pentru d = 0,5 și n = 20: aproximativ 0,35, foarte aproape de simulare. Pentru d = 0,5 și n = 64: aproximativ 0,81.
3. În JASP, calculul puterii se face din modulul Power (dacă nu apare în bara de module, se activează de la butonul "+" din dreapta barei): alegeți testul t pentru eșantioane independente, introduceți mărimea efectului, α și puterea dorită și citiți n.
4. În MedCalc, meniul Sampling oferă calculul dimensiunii eșantionului pentru compararea a două medii (introduceți diferența, AS, α și β). Alternativ, G*Power: t tests > Means: Difference between two independent means (two groups) > A priori.
5. Completați tabelul și comentați cum se schimbă n când crește puterea sau scade α.

Δ (mmHg)	σ	d	α	Putere	n pe grup (formula)
7	14	0,5	0,05	80%	aproximativ 63
7	14	0,5	0,05	90%	
7	14	0,5	0,01	80%	
5	14	0,36	0,05	80%	
10	14	0,71	0,05	80%	

Valori de autoverificare (formula aproximativă, rotunjită în sus): 85 pe grup pentru putere 90%; 94 pe grup pentru α = 0,01; 123-124 pe grup pentru Δ = 5 mmHg (123,07 înainte de rotunjire); 31 pe grup pentru Δ = 10 mmHg. Calculul exact cu distribuția t dă valori cu 1-2 pacienți mai mari (de exemplu, 86 și 96 pentru primele două rânduri).

L5.8 Tema în formatul a-g: un model complet

Problemă. Este media glicemiei a jeun la includere în mg2026 diferită de pragul de 100 mg/dL?

(a) Ipoteza de cercetare: media glicemiei a jeun în populația adulților cu risc cardiometabolic din cabinetele de medicină de familie diferă de pragul de 100 mg/dL al glicemiei a jeun modificate.

(b) Ipotezele statistice. H0: media glicemiei a jeun în populație este 100 mg/dL; μ = 100. H1: media diferă de 100 mg/dL; μ ≠ 100. Test bilateral, α = 0,05.

(c) Metoda: o variabilă numerică, o medie comparată cu o valoare fixă, σ necunoscut; testul t pentru o singură medie. Condiții: observații independente; normalitate verificată cu Shapiro-Wilk (p = 0,037, abatere ușoară, acceptabilă la n = 160 datorită teoremei limită centrale); verificare suplimentară cu testul Wilcoxon.

(d) Modul de rezolvare: JASP, T-Tests > One Sample T-Test, variabila glicemie0, Test value 100, cu Student și Wilcoxon, Location parameter, Effect size, Descriptives, Normality; verificare în Google Sheets cu formulele din L5.4.

(e) Rezultatul: $m = 105{,}44$ mg/dL, s = 15,76, n = 160; t(159) = 4,36; p < 0,001; diferența medie 5,44 mg/dL (IC 95% 2,98-7,90); d = 0,35; Wilcoxon p < 0,001. Graficul: histograma glicemiei cu o linie verticală la 100 mg/dL sau graficul descriptiv din JASP cu media și IC 95%.

(f) Concluzia statistică: p < 0,001 < 0,05, se respinge H0; diferența este extrem de semnificativă.

(g) Concluzia de cercetare: media glicemiei a jeun a pacienților cu risc cardiometabolic este mai mare decât pragul de 100 mg/dL, cu aproximativ 3-8 mg/dL; pacientul tipic din studiu se află în zona glicemiei a jeun modificate. Efectul este mic spre mediu (d = 0,35), iar rezultatul descrie populația studiată, nu populația generală.

Exerciții suplimentare

Rezolvați fiecare exercițiu în formatul a-g, în JASP și în Google Sheets. Pentru fișierele de pe Moodle nu se dau rezultatele; verificați-vă reciproc cu colegii și comparați valorile obținute în cele două programe.

1. Fișierul Date_HbA1c. Este media HbA1c din primul an al datelor diferită de ținta terapeutică de 7%? Formulați ipotezele, verificați normalitatea și aplicați testul potrivit. Repetați pentru ultimul an. Comentați diferența dintre semnificația statistică și atingerea țintei la nivel individual.
2. Fișierul BD_Studenti. Diferă înălțimea medie a studentelor de o valoare de referință aleasă de cadrul didactic (de exemplu, media națională pentru femeile tinere, comunicată la lucrare)? Dar a studenților? Care dintre cele două teste are mai multă putere și de ce?
3. Fișierul BPOC. Este frecvența cardiacă inițială medie a pacienților diferită de 80 bătăi/min? Verificați normalitatea; dacă este cazul, raportați și testul Wilcoxon.
4. Cu date în enunț (cu rezolvare completă). La 16 pacienți cu hipotiroidism tratați, TSH are media 3,1 mUI/L și AS 1,2 mUI/L. Ținta laboratorului este 2,5 mUI/L. (a) H0: μ = 2,5; H1: μ ≠ 2,5. (b) ES $= \frac{1{,}2}{\sqrt{16}} = 0{,}30$. (c) $t = \frac{3{,}1-2{,}5}{0{,}30} = 2{,}00$, gl = 15. (d) t critic (15; 0,05) = 2,131; 2,00 < 2,131, deci p > 0,05 (exact, T.DIST.2T(2;15) = 0,064). (e) IC 95% pentru medie: $3{,}1 \pm 2{,}131 \times 0{,}30 = 2{,}46 - \frac{3{,}74 m_{\text{UI}}}{L}$, interval care conține 2,5. (f) Nu se respinge H0. (g) Datele nu permit concluzia că media TSH diferă de țintă; intervalul este însă larg,

deci rezultatul este nedeterminat, iar cu un eșantion mai mare aceeași diferență medie de 0,6 mUI/L ar putea deveni semnificativă.

5. Cu date în enunț (cu rezolvare completă). Un studiu planificat vrea să detecteze o diferență a glicemiei a jeun de 8 mg/dL între două grupuri, cu $\sigma = 16$ mg/dL, $\alpha = 0{,}05$ bilateral și putere 80%. $d = \frac{8}{16} = 0{,}5$; $n = \frac{2\times(1{,}96+0{,}84)^2\times16^2}{8^2} = \frac{2\times7{,}84\times256}{64} = 62{,}7$, adică 63 de pacienți pe grup (64 cu calculul exact). Dacă se așteaptă 15% pierderi din urmărire, se recrutează 64 / 0,85 = 75,3, adică 76 de pacienți pe grup.

Ce se predă și barem orientativ

Se predă, pe Moodle, un singur fișier (PDF sau document) care conține: tabelul cu ipotezele pentru cele 8 întrebări din L5.1 (1 punct); rezultatul testului t pentru tas0 față de 125 și față de 130 din JASP, cu captura tabelului și decizia corect formulată (2 puncte); foaia Google Sheets din L5.4, cu formulele vizibile și tabelul completat pentru cele patru valori μ0 (2 puncte); rezultatul simulării din L5.6 (cele 5 proporții, histograma valorilor p, puterea estimată pentru diferența de 7 mmHg) cu o interpretare de 3-4 fraze (2 puncte); tabelul de putere din L5.7 completat (1 punct); o temă în formatul a-g, pe una dintre problemele suplimentare 1-3 (2 puncte). Total: 10 puncte. Se depunctează: ipoteze scrise cu media eșantionului, expresia "se acceptă H0", p raportat ca 0,000, lipsa intervalului de încredere și lipsa concluziei de cercetare în cuvinte.

Varianta fără rețea

Dacă rețeaua sau platforma nu funcționează, lucrarea se poate face integral cu calculatorul de buzunar și cu tabelul valorilor critice t din capitolul 5, plus JASP instalat local (programul nu are nevoie de internet după instalare). Fișierul mg2026_curat.csv se copiază de pe stick-ul cadrului didactic. Pentru Google Sheets se folosește orice program de calcul tabelar instalat local (LibreOffice Calc sau Excel), cu aceleași funcții: T.DIST.2T, T.INV.2T, NORM.INV, RAND, COUNTIF (în LibreOffice în română, numele funcțiilor pot apărea traduse, iar separatorul de argumente este tot ";"). Simularea din L5.6 se poate face și "manual", în grup: fiecare student aruncă de 20 de ori două zaruri, notează suma și compară media lui cu a colegului de bancă printr-un test t calculat de mână; la nivelul grupei, proporția comparațiilor "semnificative" va fi mică, iar discuția despre eroarea de tip I rămâne aceeași.

Lucrarea practică 6. Teste parametrice: testul t și ANOVA

Obiectiv

În această lucrare aplicați pe setul mg2026 testele parametrice din capitolul 6: testul t pentru eșantioane pereche, testul t pentru eșantioane independente (Student și Welch), ANOVA cu un factor cu testele post hoc și ANOVA cu măsurători repetate. Lucrăm în paralel în JASP, MedCalc și Google Sheets. Scopul nu este să apăsați butoane, ci să recunoașteți designul, să verificați condițiile de aplicare înainte de a citi valoarea p și să puteți reface în Google Sheets, cu formule simple, un rezultat obținut într-un program specializat. Fiecare pas are cifrele pe care trebuie să le obțineți, ca să vă puteți verifica singuri.

La final, studentul trebuie să poată:

- să aleagă între testul t pereche, testul t independent și ANOVA pe baza designului și să formuleze ipotezele în cuvinte și matematic;
- să ruleze testul t pereche și testul t independent în JASP și MedCalc, cu verificarea normalității și a egalității variațelor, și să interpreteze diferența dintre variantele Student și Welch;
- să ruleze ANOVA cu un factor în JASP, cu testul Levene, graficul Q-Q al reziduurilor, testele post hoc Tukey și Games-Howell și graficul mediilor;
- să construiască în Google Sheets tabelul ANOVA cu funcțiile AVERAGEIF, DEVSQ, FILTER și F.DIST.RT și să folosească T.TEST cu tipurile 1, 2 și 3;
- să ruleze și să interpreteze ANOVA cu măsurători repetate, inclusiv testul Mauchly și corecția Greenhouse-Geisser;
- să redacteze rezultatele în formatul a-g, cu mărimea efectului și intervalul de încredere.

Instrumente

Programe: JASP 0.98.1 [47]; MedCalc 24.0 [50]; Google Sheets cu setările regionale România (separatorul de argumente ";", virgula zecimală) [43]. Fișiere de pe Moodle [52]: mg2026_curat.csv; pentru exercițiile suplimentare BD_lp3 (glicemia la 0, 3 și 6 luni, statusul de asigurat), BPOC (frecvența cardiacă, VEMS, indicele de permeabilitate bronșică înainte și după tratament, medicamentul primit) și BD_Studenti (înălțimea, înălțimea ideală, sexul, grupa). Timp orientativ: 100 de minute (10 minute pregătirea datelor, 25 de minute testul pereche și cel independent, 35 de minute ANOVA, 15 minute măsurători repetate, 15 minute Google Sheets și tema).

Coloanele din mg2026_curat.csv, în ordinea din fișier: A id, B brat, C sex, D varsta, E mediu, F educatie, G fumat, H activitate, I inaltime, J greutate0, K imc0, L talie, M tas0, N tad0, O fc0, P colesterol, Q hdl, R trigliceride, S glicemie0, T glicemie3, U glicemie6, V hba1c0, W findrisc, X diabet_ogtt, Y greutate6, Z tas6, AA scadere5, AB satisfactie.

Desfășurare

L6.1 Pregătirea datelor: variabilele de modificare

Întrebarea principală a studiului mg2026 privește scăderea glicemiei la 6 luni pe brațe. Pentru ANOVA avem nevoie de o variabilă nouă, calculată pentru fiecare pacient.

1. În JASP, în vizualizarea datelor, faceți clic pe semnul "+" de după ultima coloană (Add computed column). Numele: dglic; tipul: Scale; metoda: formula R sau constructorul vizual. Formula:

glicemie0 - glicemie6. Confirmați cu Compute column. Media coloanei noi trebuie să fie 10,19 mg/dL.

2. Creați în același mod dgr = greutate0 - greutate6 (scăderea în greutate, kg). Media: 2,31 kg.
3. În Google Sheets, în foaia cu date, scrieți în AC1 "dglic", în AC2 =S2-U2 și trageți până la AC161; în AD1 "dgr", în AD2 =J2-Y2 și trageți. Verificați: =AVERAGE(AC2:AC161) dă 10,19; =AVERAGE(AD2:AD161) dă 2,31.
4. Notați convenția de semn: valorile pozitive înseamnă scădere. O greșeală frecventă la examen este interpretarea inversă a semnului diferenței.

L6.2 Testul t pentru eșantioane pereche în JASP

Întrebarea: se modifică glicemia a jeun între includere și 6 luni la aceiași pacienți? H0: $\mu D = 0$; H1: $\mu D \neq 0$, unde D = glicemie0 - glicemie6.

1. T-Tests > Paired Samples T-Test. Mutați perechea glicemie0 și glicemie6 în caseta Variable pairs (prima variabilă în stânga, a doua în dreapta).
2. La Tests bifați Student și Wilcoxon signed-rank; ipoteza alternativă "Measure 1 ≠ Measure 2".
3. La Additional Statistics bifați Location parameter, Confidence interval, Effect size, Descriptives și Descriptives plots (sau graficul "raincloud", dacă este disponibil în versiunea voastră).
4. La Assumption Checks bifați Normality (Shapiro-Wilk). JASP testează normalitatea diferențelor, adică exact condiția corectă.

Ce trebuie să obțineți: media glicemiei 105,4 mg/dL la includere și 95,2 mg/dL la 6 luni; t = 11,14, gl = 159, p < 0,001; diferența medie 10,19 mg/dL, IC 95% 8,38-11,99; d Cohen (pentru date pereche, dz) = 0,88; Shapiro-Wilk pe diferențe p = 0,725; testul Wilcoxon p < 0,001.

5. Adăugați perechile glicemie0-glicemie3 și glicemie3-glicemie6 în aceeași analiză. Valori așteptate: 0-3 luni, diferența 6,08 mg/dL (IC 95% 4,61-7,54), t = 8,22, dz = 0,65, Shapiro-Wilk p = 0,133; 3-6 luni, diferența 4,11 mg/dL (IC 95% 2,10-6,13), t = 4,03, p < 0,001, dz = 0,32, Shapiro-Wilk p = 0,868. Atenție: trei teste pereche pe aceleași date sunt comparații multiple; analiza corectă pentru trei momente este ANOVA cu măsurători repetate (L6.5).
6. Perechea tas0-tas6: diferența 4,78 mmHg (IC 95% 3,36-6,19), t = 6,68, gl = 159, p < 0,001, dz = 0,53, Shapiro-Wilk p = 0,129.
7. Exercițiu de control: rulați, greșit intenționat, testul t independent pe aceleași date (glicemia la includere și la 6 luni tratate ca două grupuri). În Google Sheets, =T.TEST(S2:S161;U2:U161;2;2). Comparați valoarea p cu cea a testului pereche și explicați diferența (capitolul 6, secțiunea 6.3.3).

L6.3 Testul t pentru eșantioane independente în JASP

Întrebarea: diferă TAS între femei și bărbați? H0: $\mu F = \mu M$; H1: $\mu F \neq \mu M$.

1. T-Tests > Independent Samples T-Test. Mutați tas0 în Variables și sex în Grouping Variable.
2. La Tests bifați Student, Welch și Mann-Whitney.
3. La Additional Statistics bifați Location parameter, Confidence interval, Effect size (Cohen's d), Descriptives și Descriptives plots.
4. La Assumption Checks bifați Normality și Equality of variances (testul Levene).

Ce trebuie să obțineți: femei $131{,}37 \pm 14{,}47$ mmHg$(n = 86)$, bărbați $130{,}85 \pm 13{,}25$ mmHg$(n = 74)$; Shapiro-Wilk p = 0,426 (F) și 0,648 (M); Levene p ≈ 0,52 (variațele pot fi considerate egale); testul Student t = 0,24, gl = 158, p = 0,81; testul Welch t = 0,24, gl = 157,4, p = 0,81; diferența F - M 0,52 mmHg (IC 95% aproximativ -3,8 la 4,9); d = 0,04; Mann-Whitney p = 0,80. JASP calculează diferența în ordinea alfabetică a grupurilor (F minus M), de aceea semnul este pozitiv; în capitol am raportat M minus F, -0,5 mmHg. Semnul nu schimbă concluzia, dar trebuie precizat în text.

5. Repetați pentru glicemie0 pe sexe. Valori așteptate: F 104,81 ± 16,30, M 106,16 ± 15,19; Levene p ≈ 0,43; Student t = -0,54, gl = 158, p = 0,591; Welch p = 0,589; diferența F - M -1,35 mg/dL (IC 95% -6,27 la 3,57); d = -0,09. Atenție la verificarea normalității: în grupul femeilor, Shapiro-Wilk p = 0,022. Cu 86 de paciente și o abatere moderată, testul t rămâne acceptabil; testul Mann-Whitney (p = 0,44) confirmă concluzia. Scrieți în temă ambele rezultate.
6. Repetați pentru inaltime pe sexe: diferența F - M -13,37 cm (IC 95% -15,28 la -11,45), t = -13,79, gl = 158, p < 0,001, d ≈ -2,19; Levene p > 0,7. Un efect enorm, evident și pe box-plot.
7. Repetați pentru varsta pe sexe: F 55,42 ± 8,95, M 57,31 ± 8,50; Welch t = -1,37, p = 0,17; d = -0,22. Discutați de ce verificarea vârstei pe sexe contează înaintea interpretării altor diferențe între sexe (vârsta este un posibil factor de confuzie).

L6.4 ANOVA cu un factor în JASP

Întrebarea principală a studiului: diferă scăderea glicemiei la 6 luni între brațele A, B și C? H0: μA = μB = μC; H1: cel puțin o medie diferă.

1. ANOVA > ANOVA. Mutați dglic în Dependent Variable și brat în Fixed Factors.
2. La Display bifați Descriptive statistics și Estimates of effect size (η² și ω²).
3. La Assumption Checks bifați Homogeneity tests (Levene) și Q-Q plot of residuals. La Homogeneity corrections lăsați None pentru analiza principală; bifați și Welch pentru comparație.
4. La Post Hoc Tests mutați brat în caseta din dreapta și bifați Tukey; adăugați Games-Howell și Bonferroni pentru comparație; bifați Confidence intervals și Effect size.
5. La Descriptives Plots puneți brat pe axa orizontală și bifați Display error bars, Confidence interval 95%.
6. La Nonparametrics mutați brat pentru testul Kruskal-Wallis.
7. Pentru normalitatea pe grupuri, folosiți Descriptives > Descriptive Statistics, cu dglic la Variables și brat la Split, și bifați Shapiro-Wilk test.

Ce trebuie să obțineți: A 3,0 ± 8,8 mg/dL$(n = 53)$, B 11,1 ± 11,2$(n = 53)$, C 16,3 ± 10,6$(n = 54)$. Tabelul ANOVA: brat SS = 4806,9, gl = 2, MS = 2403,4; Residuals SS = 16459,5, gl = 157, MS = 104,8; F = 22,93, p < 0,001; η² = 0,226; ω² = 0,215. Testul Levene: p = 0,089 sau p = 0,163, în funcție de varianta testului folosită de program (centrată pe medie sau pe mediană); în ambele cazuri peste 0,05. ANOVA Welch: F = 26,05, p < 0,001. Shapiro-Wilk pe brațe: A p = 0,575, B p = 0,052, C p = 0,903. Tukey: B - A 8,1 mg/dL (IC 95% 3,4-12,8; p < 0,001), C - A 13,3 (8,6-18,0; p < 0,001), C - B 5,2 (0,6-9,9; p = 0,024). Games-Howell: p = 0,0002, < 0,001 și 0,038. Kruskal-Wallis: H = 36,63, gl = 2, p < 0,001.

8. Repetați pentru dgr (scăderea în greutate). Valori așteptate: A 0,71 ± 2,48 kg, B 2,84 ± 2,36, C 3,37 ± 2,39; F(2; 157) = 18,25, p < 0,001; η² = 0,189; Levene p = 0,79; Tukey: B - A 2,13 kg (p < 0,001), C - A 2,66 kg (p < 0,001), C - B 0,53 kg (p = 0,49); Kruskal-Wallis H = 27,50, p < 0,001.
9. Repetați pentru tas0 cu fumat ca factor. Valori așteptate: foști fumători 127,55 ± 14,39$(n = 33)$, fumători 137,04 ± 14,63$(n = 53)$, nefumători 128,50 ± 11,76$(n = 74)$; F(2; 157) = 7,85, p < 0,001 (0,0006); η² = 0,091; Levene p ≈ 0,11; Tukey: fumători - foști fumători 9,49 mmHg (p = 0,004), fumători - nefumători 8,54 mmHg (p = 0,001), nefumători - foști fumători 0,95 mmHg (p = 0,94).
10. Interpretați pe hârtie, în 4-5 fraze, rezultatul de la punctul 7: ce brațe diferă, cât de mare este efectul, ce înseamnă intervalul de încredere al diferenței C - B care începe la 0,6 mg/dL.

L6.5 ANOVA cu măsurători repetate în JASP

Întrebarea: se modifică glicemia între 0, 3 și 6 luni? H0: μ0 = μ3 = μ6.

1. ANOVA > Repeated Measures ANOVA. La Repeated Measures Factors redenumiți factorul "timp" și nivelurile "0 luni", "3 luni", "6 luni".

2. Mutați glicemie0, glicemie3 și glicemie6 în Repeated Measures Cells, în ordinea corectă.
3. La Assumption Checks bifați Sphericity tests, iar la Sphericity corrections bifați None și Greenhouse-Geisser.
4. La Post Hoc Tests mutați timp și bifați Bonferroni. La Descriptives Plots puneți timp pe axa orizontală.

Ce trebuie să obțineți: mediile 105,4 / 99,4 / 95,2 mg/dL; $F(2; 318) = 64{,}97$, $p < 0{,}001$; testul Mauchly $W = 0{,}873$, $p < 0{,}001$ (sfericitatea nu este respectată); ε Greenhouse-Geisser = 0,887, gradele de libertate corectate 1,77 și 282,2, $p < 0{,}001$; comparațiile Bonferroni: toate cele trei perechi $p < 0{,}001$. Testul Friedman (ANOVA > Repeated Measures ANOVA > Nonparametrics, sau capitolul 7): $\chi^2 = 74{,}67$, $p < 0{,}001$.

5. Model mixt (opțional, pentru studenții care termină mai repede): adăugați brat în Between Subject Factors. Valori așteptate: efectul timpului $F(2; 314) = 73{,}56$, $p < 0{,}001$; efectul brațului $F(2; 157) = 8{,}50$, $p < 0{,}001$; interacțiunea timp × braț $F(4; 314) = 11{,}51$, $p < 0{,}001$. Mediile la 0 / 3 / 6 luni: A 107,8 / 105,7 / 104,8; B 106,3 / 101,3 / 95,2; C 102,3 / 91,2 / 86,0 mg/dL. Interacțiunea semnificativă înseamnă că evoluția în timp diferă între brațe: aceasta este, de fapt, întrebarea despre eficacitatea intervențiilor. Desenați graficul mediilor cu o linie pentru fiecare braț (Descriptives Plots, brat la Separate lines).

L6.6 Aceleași analize în MedCalc

1. Importați mg2026_curat.csv (File > Import sau File > Open, tip CSV) și verificați că variabilele numerice sunt recunoscute ca numere.
2. Testul pereche: Statistics > T-tests > Paired samples t-test; Variable 1: glicemie0, Variable 2: glicemie6. Rezultatul trebuie să coincidă cu L6.2 (diferența medie, IC 95%, $t = 11{,}14$, $p < 0{,}0001$).
3. Testul independent: Statistics > T-tests > Independent samples t-test. La Sample 1 alegeți variabila tas0 cu criteriul de selecție sex="F", iar la Sample 2 aceeași variabilă cu sex="M". MedCalc raportează și testul F pentru egalitatea variațelor, testul t clasic și varianta corectată pentru variațe inegale.
4. ANOVA: Statistics > ANOVA > One-way analysis of variance. La Data puteți scrie direct expresia glicemie0-glicemie6 (sau o variabilă calculată în prealabil), la Factor alegeți brat. Bifați testul Levene și un test post hoc (Tukey-Kramer sau Bonferroni). Comparați tabelul cu cel din JASP.
5. Din date rezumative, fără fișier: meniul Tests > Comparison of means (t-test) permite introducerea mediei, AS și n pentru două grupuri. Introduceți valorile IMC dintr-un articol publicat, $28{,}15 \pm \frac{6{,}49\ \mathrm{kg}}{m^2}\ (n = 52)$ și $24{,}47 \pm \frac{4{,}60\ \mathrm{kg}}{m^2}\ (n = 84)$ [71], și comparați rezultatul cu valoarea p raportată de autori (p = 0,001). Ar trebui să obțineți un t în jur de 3,6-3,9 (în funcție de varianta testului) și $p < 0{,}001$.

L6.7 Testele t și tabelul ANOVA în Google Sheets

Funcția T.TEST(interval1; interval2; cozi; tip) dă direct valoarea p. Argumentul "cozi" este 1 sau 2 (folosiți 2, bilateral). Argumentul "tip" este 1 pentru eșantioane pereche, 2 pentru eșantioane independente cu varianță egală (Student) și 3 pentru eșantioane independente cu variațe inegale (Welch).

1. Testul pereche: =T.TEST(S2:S161;U2:U161;2;1). Rezultat: aproximativ 1,1 × 10^-21, adică $p < 0{,}001$.
2. Testul independent pe sexe, cu FILTER pentru a separa grupurile: =T.TEST(FILTER(M2:M161;C2:C161="F");FILTER(M2:M161;C2:C161="M");2;2) dă 0,814 (Student), iar cu tipul 3 dă 0,813 (Welch).
3. Statistica t pentru testul pereche, construită de mână: în celule separate calculați media diferențelor =AVERAGE(AC2:AC161) (10,19), abaterea standard =STDEV(AC2:AC161) (11,57), eroarea

standard =STDEV(AC2:AC161)/SQRT(160) (0,915), t = media / ES (11,14), p =T.DIST.2T(t;159) și limitele IC 95% = media ± T.INV.2T(0,05;159) × ES (8,38 și 11,99).

4. Testul F pentru raportul varianțelor: =F.TEST(FILTER(M2:M161;C2:C161="F");FILTER(M2:M161;C2:C161="M")) dă aproximativ 0,44. Testul F este sensibil la abaterea de la normalitate; Levene, din JASP, este preferabil, dar ambele duc aici la aceeași concluzie.
5. Tabelul ANOVA pentru dglic pe brațe, construit cu formule. Într-o foaie nouă scrieți în A2:A4 literele A, B, C. În B2 =COUNTIF(Date!B$2:B$161;A2) (n pe braț: 53, 53, 54). În C2 =AVERAGEIF(Date!B$2:B$161;A2;Date!AC$2:AC$161) (mediile 3,02; 11,09; 16,33). În D2 =DEVSQ(FILTER(Date!AC$2:AC$161;Date!B$2:B$161=A2)) (suma pătratelor abaterilor în fiecare braț: 3987,0; 6532,5; 5940,0). Trageți formulele până la rândul 4.
6. Media generală: =AVERAGE(Date!AC2:AC161) (10,19). SS total: =DEVSQ(Date!AC2:AC161) (21266,4). SS în interior: =SUM(D2:D4) (16459,5). SS între: =SUMPRODUCT(B2:B4;(C2:C4-media generală)^2) sau, mai simplu, SS total - SS în interior (4806,9).
7. MS între = SS între / 2 = 2403,4; MS în interior = SS în interior / 157 = 104,8; F = MS între / MS în interior = 22,93; p =F.DIST.RT(F;2;157), aproximativ 1,8 × 10^-9; valoarea critică =F.INV.RT(0,05;2;157) = 3,05; eta² = SS între / SS total = 0,226. Așezați rezultatele într-un tabel cu coloanele Sursa, SS, gl, MS, F, p, exact ca tabelul 6.5 din capitol.
8. Varianta cu supliment: din Extensions > Add-ons se poate instala un supliment de analiză statistică (de exemplu, XLMiner Analysis ToolPak), care oferă "Anova: Single Factor". Suplimentul cere datele fiecărui grup într-o coloană separată; le puteți obține cu =FILTER(Date!AC2:AC161;Date!B2:B161="A") și analogele pentru B și C. Rezultatul trebuie să coincidă cu tabelul construit de mână.

L6.8 Tema în formatul a-g: modelul pentru ANOVA

(a) Ipoteza de cercetare: scăderea glicemiei a jeun la 6 luni diferă între pacienții care au primit consiliere standard, consiliere cu aplicație și consiliere cu aplicație și teleconsultație.

(b) H0: scăderea medie a glicemiei este aceeași în cele trei populații, $\mu A = \mu B = \mu C$. H1: cel puțin un braț are o scădere medie diferită de celelalte. $\alpha = 0,05$.

(c) Metoda: variabilă numerică (dglic) comparată între trei grupuri independente; ANOVA cu un factor, urmată de testul Tukey; condițiile verificate: independența (randomizare individuală), normalitatea pe brațe (Shapiro-Wilk p între 0,052 și 0,903), omogenitatea varianțelor (Levene p > 0,05, raportul AS 1,27). Alternativa în caz de abatere: ANOVA Welch și Kruskal-Wallis.

(d) Programul: JASP, ANOVA > ANOVA, cu Post Hoc Tests Tukey și Descriptives Plots; verificare în Google Sheets cu tabelul construit în L6.7.

(e) Rezultatul: A 3,0 ± 8,8, B 11,1 ± 11,2, *C* 16,3 ± 10,6 mg/dL; F(2; 157) = 22,93, p < 0,001, $\eta^2 = 0,226$; Tukey B - A 8,1 (IC 95% 3,4-12,8), C - A 13,3 (8,6-18,0), C - B 5,2 (0,6-9,9), toate p < 0,05. Graficul: box-plot pe brațe sau graficul mediilor cu IC 95%.

(f) Concluzia statistică: p < 0,001, se respinge H0; diferențele dintre brațe sunt extrem de semnificative, iar toate cele trei perechi diferă semnificativ.

(g) Concluzia de cercetare: adăugarea aplicației de automonitorizare la consiliere a crescut scăderea glicemiei a jeun la 6 luni cu aproximativ 8 mg/dL, iar teleconsultația lunară a adus un beneficiu suplimentar de aproximativ 5 mg/dL față de aplicație, cu o estimare mai imprecisă (IC 95% 0,6-9,9 mg/dL). Brațul de tratament explică aproape un sfert din variabilitatea scăderii glicemiei. Studiul este randomizat, deci diferențele pot fi atribuite intervențiilor; datele sunt însă simulate în scop didactic.

Exerciții suplimentare

Exercițiile 1-6 se lucrează pe fișierele clasice ale disciplinei de pe Moodle; rezultatele nu se dau, iar fiecare se rezolvă în formatul a-g, cu verificarea condițiilor de aplicare și cu graficul potrivit. Exercițiile 7 și 8 au datele în enunț și rezolvarea completă.

1. Fișierul BD_lp3. Există diferențe semnificative între glicemia la momentul zero și glicemia la 3 luni? Dar între 3 și 6 luni? Rezolvați apoi întrebarea pe toate cele trei momente cu ANOVA cu măsurători repetate și comparați concluziile.
2. Fișierul BD_lp3. Diferă glicemia la 6 luni între pacienții asigurați și cei neasigurați? Diferă vârsta între cele două grupuri? Dacă vârsta diferă, cum ar putea influența interpretarea diferenței de glicemie?
3. Fișierul BD_lp3. Diferă evoluția glicemiei la 0, 3 și 6 luni între asigurați și neasigurați? Folosiți modelul mixt din L6.5 (statusul de asigurat ca factor între subiecți) și interpretați termenul de interacțiune.
4. Fișierul BPOC. A scăzut semnificativ frecvența cardiacă după tratament (FC1 față de FC2)? A crescut VEMS procentual (VEMSP1 față de VEMSP2)? Verificați normalitatea diferențelor și raportați, dacă este cazul, testul Wilcoxon.
5. Fișierul BPOC. Diferă indicele de permeabilitate bronșică inițial (IPB1) între pacienții care au primit medicamentele A, B și C? Aplicați ANOVA cu testul Levene și testul post hoc potrivit; dacă variațele diferă, folosiți ANOVA Welch și Games-Howell.
6. Fișierul BD_Studenti. (a) Diferă înălțimea între sexe? (b) Diferă înălțimea reală de înălțimea ideală declarată (test pereche)? (c) Diferă înălțimea între grupele de studiu? Pentru (c), discutați de ce un rezultat semnificativ ar putea reflecta doar compoziția pe sexe a grupelor.
7. Date în enunț, cu rezolvare. Frecvența cardiacă (bătăi/min) la 6 pacienți cu BPOC înainte și după un beta-blocant cardioselectiv: înainte 96, 88, 102, 90, 94, 100; după 84, 82, 90, 86, 85, 91. Diferențele sunt 12, 6, 12, 4, 9, 9; $m_D = \frac{52}{6} = 8{,}67$; abaterile față de medie: 3,33; -2,67; 3,33; -4,67; 0,33; 0,33; suma pătratelor 11,11 + 7,11 + 11,11 + 21,78 + 0,11 + 0,11 = 51,33; $s_D = \sqrt{\frac{51{,}33}{5}} = 3{,}20$; ESD $= \frac{3{,}20}{\sqrt{6}} = 1{,}31$; $t = \frac{8{,}67}{1{,}31} = 6{,}63$, gl = 5; t critic (5; 0,05) = 2,571, deci p < 0,05 (exact p = 0,0012). IC 95%: 8,67 ± 2,571 × 1,31 = 5,30-12,03 bătăi/min. Concluzie: frecvența cardiacă a scăzut semnificativ, în medie cu aproximativ 9 bătăi/min; cu 6 pacienți și fără grup de control, rezultatul este doar orientativ.
8. Date în enunț, cu rezolvare. Hemoglobina (g/dL) la trei grupuri de câte 4 gravide, în trimestrele I, II și III (grupuri independente): T1 12,5; 13,0; 12,8; 12,9; T2 11,8; 12,0; 11,6; 12,2; T3 11,5; 11,9; 11,2; 11,8. Mediile: 12,80; 11,90; 11,60; media generală 12,10. SS între = 4 × [(0,70)² + (-0,20)² + (-0,50)²] = 4 × 0,78 = 3,12, gl = 2, MS = 1,56. SS în interior: T1 0,0900 + 0,0400 + 0,0000 + 0,0100 = 0,14; T2 0,0100 + 0,0100 + 0,0900 + 0,0900 = 0,20; T3 0,0100 + 0,0900 + 0,1600 + 0,0400 = 0,30; total 0,64, gl = 9, MS = 0,0711. $F = \frac{1{,}56}{0{,}0711} = 21{,}9$; F critic (2; 9; 0,05) = 4,26, deci p < 0,05 (exact p < 0,001). $\eta^2 = \frac{3{,}12}{3{,}76} = 0{,}83$. Testul Tukey: q(3; 9; 0,05) = 3,95; HSD $= 3{,}95 \times \sqrt{\frac{0{,}0711}{4}} = 3{,}95 \times 0{,}133 = 0{,}53$ g/dL; T1 - T2 = 0,90 și T1 - T3 = 1,20 sunt semnificative, T2 - T3 = 0,30 nu. Concluzie: hemoglobina medie este mai mare în trimestrul I decât în trimestrele II și III, în concordanță cu hemodiluția fiziologică a sarcinii; diferența dintre trimestrele II și III nu a putut fi pusă în evidență pe acest eșantion foarte mic.

Ce se predă și barem orientativ

Se predă pe Moodle un singur fișier care conține: rezultatele testului pereche pentru glicemie0-glicemie6 (tabelul JASP, testul de normalitate pe diferențe, graficul) cu interpretarea (2 puncte); rezultatele testului independent pentru tas0 și glicemie0 pe sexe, cu testul Levene și cu justificarea alegerii dintre Student, Welch și Mann-Whitney (2 puncte); ANOVA pentru dglic pe brațe, cu verificarea condițiilor, testul Tukey și graficul mediilor (2 puncte); tabelul ANOVA construit în Google Sheets, cu formulele vizibile (1,5 puncte); ANOVA cu măsurători repetate cu testul Mauchly și corecția Greenhouse-Geisser (1 punct); o temă în formatul a-g pe unul dintre exercițiile suplimentare 1-6 (1,5 puncte). Total: 10 puncte. Se depunctează: testul independent aplicat pe date pereche, normalitatea verificată pe fiecare serie în loc de diferențe la testul pereche, teste t multiple în locul ANOVA, lipsa testului post hoc după un F semnificativ, lipsa mărimii efectului și a intervalului de încredere, formularea "se acceptă H0".

Varianta fără rețea

Fără rețea, JASP și MedCalc funcționează local, după instalare; fișierul mg2026_curat.csv se copiază de pe stick-ul cadrului didactic. Partea de Google Sheets se înlocuiește cu LibreOffice Calc sau Excel: funcțiile T.TEST, T.DIST.2T, T.INV.2T, DEVSQ, AVERAGEIF, COUNTIF, F.DIST.RT și F.INV.RT există și acolo (în Excel, FILTER este disponibil în versiunile recente; altfel, sortați datele după braț și selectați manual intervalele fiecărui grup). În Excel, ANOVA se obține și din Data > Data Analysis > Anova: Single Factor, după activarea suplimentului Analysis ToolPak. Dacă nu este disponibil niciun calculator, exercițiile suplimentare 7 și 8 se rezolvă integral de mână, cu tabelele valorilor critice t și F și cu tabelul q pentru testul Tukey.

Lucrarea practică 7. Teste neparametrice: Mann-Whitney, Wilcoxon, Kruskal-Wallis, Friedman și chi-pătrat

Obiectiv

În această lucrare aplicați pe calculator testele din capitolul 7, pe setul de date mg2026, în trei programe diferite, și verificați că obțineți aceleași rezultate. Accentul nu cade pe apăsarea butoanelor, ci pe trei lucruri care vor conta la examenul practic: alegerea testului potrivit pentru tipul de variabilă și de design, citirea corectă a tabelelor de ieșire (care statistică, ce valoare p, ce mărime a efectului) și formularea concluziei. Fiecare pas vă dă cifrele pe care trebuie să le obțineți, ca să vă puteți autoverifica; dacă rezultatul vostru diferă, căutați greșeala înainte de a merge mai departe.

La final, studentul trebuie să poată:

- să aplice în JASP testele Mann-Whitney, Wilcoxon, Kruskal-Wallis (cu post hoc Dunn) și Friedman și să extragă statistica, valoarea p și mărimea efectului;
- să construiască în JASP un tabel de contingență cu frecvențele așteptate, testul chi-pătrat, corecția Yates, testul Fisher și V Cramér;
- să introducă un tabel 2 x 2 și un tabel 2 x n în Epi Info StatCalc și să interpreteze ieșirea (chi-pătrat necorectat, Yates, Fisher, OR, RR, testul pentru trend);
- să reproducă în Google Sheets frecvențele observate, frecvențele așteptate și valoarea p a testului chi-pătrat;
- să verifice condițiile de aplicare (frecvențe așteptate, date pereche sau independente) și să aleagă varianta corectă a testului;
- să formuleze concluzia statistică și concluzia clinică pentru fiecare test.

Instrumente

- JASP 0.98.1 [47], modulele implicite T-Tests, ANOVA și Frequencies.
- MedCalc 24.0 [50] (pe calculatoarele din laborator).
- Epi Info 7.2.7 [51], componenta StatCalc și, opțional, Classic Analysis.
- Google Sheets, cu setările regionale pentru România (separatorul de argumente este ";", iar separatorul zecimal este virgula).
- Fișierul mg2026_curat.csv (160 de pacienți, 28 de coloane) de pe platforma Moodle a disciplinei.
- Timp orientativ: 100 de minute (JASP 50 de minute, Epi Info 15 minute, Google Sheets 20 de minute, concluzii și încărcare 15 minute).

Coloanele fișierului, în ordine, sunt: A id, B brat, C sex, D varsta, E mediu, F educatie, G fumat, H activitate, I inaltime, J greutate0, K imc0, L talie, M tas0, N tad0, O fc0, P colesterol, Q hdl, R trigliceride, S glicemie0, T glicemie3, U glicemie6, V hba1c0, W findrisc, X diabet_ogtt, Y greutate6, Z tas6, AA scadere5, AB satisfactie. Datele sunt pe rândurile 2-161.

Desfășurare

L7.1 Pregătirea datelor în JASP

1. Deschideți JASP și încărcați fișierul: meniul principal (cele trei linii din stânga sus) > Open > Computer > Browse, apoi selectați mg2026_curat.csv.

2. Verificați tipul fiecărei variabile, afișat prin pictograma din stânga numelui coloanei (riglă pentru continuă, bare în trepte pentru ordinală, trei cercuri pentru nominală). Faceți clic pe pictogramă pentru a o schimba. Setați: brat, sex, mediu, fumat, diabet_ogtt și scadere5 ca **nominale**; educatie, activitate și satisfactie ca **ordinale**; restul ca **continue**.
3. Creați variabila "dglic" (scăderea glicemiei la 6 luni): faceți clic pe semnul "+" de lângă ultima coloană, dați numele dglic, alegeți tipul continuu și metoda de calcul cu formula R, apoi scrieți `glicemie0 - glicemie6` și apăsați Compute column. Verificați: media coloanei dglic trebuie să fie 10,19.
4. În același mod creați variabila "obez", nominală, cu formula R `ifelse(imc0 >= 30, 1, 0)`. Trebuie să obțineți 54 de valori 1 și 106 valori 0.
5. Creați variabila "glic100_0" cu formula `ifelse(glicemie0 >= 100, 1, 0)` și "glic100_6" cu formula `ifelse(glicemie6 >= 100, 1, 0)`. Trebuie să obțineți 97, respectiv 62 de valori 1.

De ce este importantă setarea tipului de variabilă? JASP nu vă lasă să puneți o variabilă continuă în căsuța de "Rows" a unui tabel de contingență și, invers, oferă numai teste potrivite pentru tipul declarat. Un tip greșit este primul motiv pentru care "nu apare testul" la laborator.

L7.2 Testul Mann-Whitney U în JASP: satisfacția în brațul C față de brațul A

Testul Mann-Whitney compară două grupuri, deci trebuie să excludeți temporar brațul B.

1. Faceți clic pe numele coloanei brat; în fereastra de etichete care se deschide, debifați "Filter" pentru valoarea B (sau folosiți butonul de filtru, simbolul pâlnie). Rândurile pacienților din brațul B devin gri: sunt excluse din analize. Verificați în bara de jos: 107 rânduri active.
2. T-Tests > Classical > Independent Samples T-Test.
3. Dependent Variables: satisfactie (dacă JASP nu acceptă variabila ordinală, setați-o temporar ca "Scale"). Grouping Variable: brat.
4. La Tests, debifați Student și bifați **Mann-Whitney**. La Additional Statistics bifați **Effect size** (pentru Mann-Whitney, JASP raportează corelația biserială a rangurilor) și **Descriptives**.
5. Pentru descriere, folosiți Descriptives > Descriptive Statistics, cu satisfactie la Variables și brat la Split, bifând Median și Quartiles (sau Interquartile range).

Ce trebuie să obțineți: mediana 3 (Q1 2, Q3 3) în brațul A și 4 (Q1 3, Q3 5) în brațul C; W (statistica Mann-Whitney raportată de JASP) = 596,5 sau 2265,5, în funcție de grupul luat ca referință (ordinea A, C dă 596,5); p < 0,001; corelația biserială a rangurilor 0,58 în valoare absolută (semnul depinde de ordinea grupurilor). Notați în fișa de lucru: "$U = \frac{2265,5}{596,5}$ sunt aceeași statistică, văzută din cele două grupuri; U1 + U2 = 54 x 53 = 2862".

6. Repetați pentru brațele B și A (filtrați acum brațul C). Trebuie să obțineți p < 0,001 și corelația biserială a rangurilor 0,42. Apoi pentru C și B: p = 0,079, corelația 0,19.
7. La final, reactivați toate brațele (bifați din nou "Filter" pentru B și C). Uitarea filtrului activ este o sursă clasică de rezultate greșite în exercițiile următoare.

Interpretare scrisă (în fișa de lucru): "Satisfacția a fost mai mare în brațul C decât în brațul A (mediana 4 față de 3; Mann-Whitney p < 0,001; corelația biserială a rangurilor 0,58, efect mare)."

L7.3 Testul Wilcoxon pentru date pereche în JASP

Exercițiul A: glicemia la 0 și la 6 luni, toți pacienții.

1. T-Tests > Classical > Paired Samples T-Test.
2. Mutați glicemie0 și glicemie6 în căsuța Variable Pairs (ele formează o pereche).
3. La Tests bifați atât Student, cât și **Wilcoxon signed-rank**. La Additional Statistics bifați Effect size, Descriptives, iar la Assumption Checks, Normality (Shapiro-Wilk pe diferențe).

Ce trebuie să obțineți: testul t pereche t = 11,14, gl 159, p < 0,001; Shapiro-Wilk pe diferențe p = 0,725 (diferențele sunt normale, deci testul t este analiza principală); Wilcoxon p < 0,001. Când ambele teste concordă, raportați testul ales dinainte; aici, testul t pereche, cu diferența medie 10,19 mg/dL (IC 95% 8,38-11,99).

Exercițiul B: greutatea la 0 și la 6 luni, doar brațul A.

1. Filtrați brațele B și C (rămân 53 de rânduri active).
2. Paired Samples T-Test cu perechea greutate0 - greutate6, bifând Student, Wilcoxon signed-rank și Normality.

Ce trebuie să obțineți: Wilcoxon W = 870 (suma rangurilor pozitive pentru diferența greutate0 - greutate6; dacă inversați ordinea variabilelor obțineți 456), p = 0,052; testul t pereche t = 2,07, gl 52, p = 0,043; Shapiro-Wilk pe diferențe p = 0,696. Mediana diferențelor este 1,1 kg, media 0,71 kg.

Discutați în fișa de lucru, în 4-5 rânduri: care test este cel corect aici și de ce (distribuția diferențelor este normală, deci testul t pereche este justificat); de ce Wilcoxon dă o valoare p puțin mai mare (folosește doar rangurile, eficiența lui relativă este de aproximativ 95%); de ce nu avem voie să alegem testul după ce am văzut care valoare p este sub 0,05.

3. Opțional, testul semnului: din coloana diferențelor, numărați câte sunt pozitive (32), negative (19) și zero (2). În Frequencies > Binomial Test, pe o variabilă care codifică semnul diferenței, cu Test value 0,5, trebuie să obțineți p = 0,092.

L7.4 Testul Kruskal-Wallis cu post hoc Dunn în JASP

1. Reactivați toate rândurile. ANOVA > Classical > ANOVA.
2. Dependent Variable: dglic; Fixed Factors: brat.
3. Deschideți secțiunea **Nonparametrics** și mutați brat în căsuța pentru Kruskal-Wallis. Bifați **Dunn's post hoc test** (în unele versiuni, Dunn apare în secțiunea Post Hoc Tests, la Type: Dunn).
4. Pentru comparație, cereți și ANOVA clasică, cu Assumption Checks > Homogeneity tests și Post Hoc Tests > Tukey.

Ce trebuie să obțineți: Kruskal-Wallis H = 36,63, gl 2, p < 0,001; Dunn cu corecție Bonferroni: A față de B p = 0,0003, A față de C p < 0,001, B față de C p = 0,12 (neajustat 0,041). ANOVA: F(2;157) = 22,93, p < 0,001; Levene p = 0,163; Tukey C față de B p = 0,024.

5. Repetați pentru satisfactie (Dependent Variable: satisfactie). Trebuie să obțineți H = 32,16, gl 2, p < 0,001; Dunn Bonferroni: A față de B p = 0,0003, A față de C p < 0,001, B față de C p = 0,29.
6. Calculați de mână mărimea efectului $\varepsilon^2 = \frac{H}{N-1}$: 36,63/159 = 0,23 pentru dglic și 32,16/159 = 0,20 pentru satisfacție.

Întrebare de control (răspundeți în fișă): de ce comparația B-C este semnificativă la Tukey (p = 0,024) dar nu la Dunn cu Bonferroni (p = 0,12)? Care este analiza principală pentru dglic și de ce?

L7.5 Testul Friedman în JASP

1. ANOVA > Classical > Repeated Measures ANOVA.
2. În căsuța Repeated Measures Factors, redenumiți factorul "RM Factor 1" în "timp" și nivelurile în "0 luni", "3 luni", "6 luni".
3. Mutați glicemie0, glicemie3 și glicemie6, în această ordine, în căsuța Repeated Measures Cells.
4. Deschideți secțiunea **Nonparametrics**, mutați factorul timp în căsuța pentru Friedman și bifați Conover's post hoc tests.
5. Pentru descriere: Descriptives > Descriptive Statistics pe cele trei variabile, cu Median și Quartiles.

Ce trebuie să obțineți: medianele 104,5, 99,0 și 92,5 mg/dL; Friedman $\chi^2 = 74{,}67$, gl 2, $p < 0{,}001$; Kendall's W = 0,23 (dacă JASP nu îl afișează, calculați-l: 74,67/(160 x 2)); ANOVA cu măsurători repetate $F = 64{,}97$, $p < 0{,}001$. Comparațiile post hoc Conover trebuie să arate diferențe semnificative între toate cele trei momente; ca verificare independentă, testele Wilcoxon cu corecție Bonferroni din tabelul 7.7 al capitolului 7 dau p ajustat < 0,001 (0-3 luni), 0,0007 (3-6 luni) și < 0,001 (0-6 luni).

6. Filtrați doar brațul A și repetați. Trebuie să obțineți χ^2 Friedman = 6,04, p = 0,049, cu medianele 111, 104 și 105 mg/dL. Discutați: rezultatul global este la limită; rămâne vreo comparație semnificativă după corecție? (Răspuns de verificat: nu, cea mai mică valoare ajustată Bonferroni este 0,075.)

L7.6 Tabele de contingență și testul chi-pătrat în JASP

Exercițiul A: brațul și scăderea în greutate ≥ 5% (tabel 3 x 2).

1. Reactivați toate rândurile. Frequencies > Classical > Contingency Tables.
2. Rows: brat; Columns: scadere5.
3. La Statistics bifați χ^2 și, la Nominal, **Phi and Cramér's V**. La Cells bifați Counts **Expected** și Percentages **Row**.

Ce trebuie să obțineți: frecvențele observate 6/47, 16/37, 16/38; frecvențele așteptate 12,59/40,41 (A), 12,59/40,41 (B), 12,82/41,18 (C); $\chi^2 = 6{,}77$, gl 2, p = 0,034; V Cramér = 0,21; procentele pe rânduri 11,3%, 30,2% și 29,6%. Verificați regula de aplicare: cea mai mică frecvență așteptată este 12,6, deci testul este valid.

4. Graficul: Descriptives > Descriptive Statistics, variabila scadere5 cu Split pe brat, Plots > Distribution plots; sau, mai simplu, construiți graficul cu bare al procentelor în Google Sheets (pasul L7.8). Rețineți observația de la laborator: pentru chi-pătrat, reprezentarea grafică nu se face din meniul Contingency Tables, ci din Descriptives.

Exercițiul B: obezitatea și diabetul (tabel 2 x 2).

1. Contingency Tables cu Rows: obez, Columns: diabet_ogtt.
2. Bifați χ^2, **χ^2 continuity correction**, iar la secțiunea pentru tabele 2 x 2, **Log odds ratio** și **Fisher's exact test**. La Cells: Expected, Row percentages.

Ce trebuie să obțineți: observat 95/11 (neobezi) și 36/18 (obezi); așteptat 86,79/19,21 și 44,21/9,79; $\chi^2 = 12{,}70$, gl 1, $p < 0{,}001$; cu corecție de continuitate 11,20, p = 0,0008; Fisher p = 0,0008; $\varphi = 0{,}28$. Log odds ratio = 1,46, deci $\text{OR} = e^{1{,}46} = 4{,}32$(; IC 95% $1{,}86 - 10{,}03$), indiferent dacă îl calculați ca obezi față de neobezi.

3. Scrieți în fișă de ce, la acest tabel, cele trei variante (Pearson, Yates, Fisher) duc la aceeași concluzie (toate frecvențele așteptate sunt peste 5, cea mai mică 9,79), și în ce situație ar fi dat concluzii diferite.

Exercițiul C: o situație în care chi-pătratul nu este valid.

Filtrați doar pacienții sub 50 de ani (adăugați o variabilă calculată `ifelse(varsta < 50, 1, 0)` și filtrați după ea) și construiți tabelul obez x diabet_ogtt pentru acest subgrup. Notați câte celule au frecvența așteptată sub 5 și ce test alegeți. (Subgrupul are 34 de pacienți și doar 2 cazuri de diabet; testul potrivit este Fisher.)

L7.7 Aceleași teste în MedCalc

Deschideți mg2026_curat.csv în MedCalc (File > Open, tipul de fișier text/CSV; verificați că virgula zecimală este recunoscută corect). Denumirile meniurilor pot diferi puțin între versiuni; căutați testele după nume.

1. Mann-Whitney: meniul Statistics, la testele pe ranguri, Mann-Whitney test (independent samples). Variable: satisfactie; Grouping/Filter: brat = "A" pentru primul eșantion și brat = "C" pentru al doilea. Ce trebuie să obțineți: aceeași valoare p (< 0,001) ca în JASP.
2. Wilcoxon: Wilcoxon test (paired samples), cu greutate0 și greutate6, Filter: brat = "A". Ce trebuie să obțineți: p = 0,052.
3. Kruskal-Wallis: la analiza variației, Kruskal-Wallis test, Variable: dglic (creați-o înainte cu Variable editor sau scrieți direct expresia glicemie0-glicemie6), Factor: brat. Ce trebuie să obțineți: H = 36,63, p < 0,001, plus comparațiile post hoc oferite de program.
4. Friedman: Friedman test, cu cele trei coloane de glicemie. Ce trebuie să obțineți: χ^2 = 74,67 (sau o valoare foarte apropiată, în funcție de corecția pentru egalități), p < 0,001.
5. Chi-pătrat: la tabelele de contingență, Chi-squared test, cu brat și scadere5. Ce trebuie să obțineți: χ^2 = 6,77, gl 2, p = 0,034. MedCalc are și testul pentru trend (Chi-squared test for trend), util aici pentru că brațele au intensitate crescătoare: ce trebuie să obțineți este o valoare χ^2 pentru trend în jur de 4,9, cu p = 0,027.
6. McNemar: McNemar test, cu glic100_0 și glic100_6 (create ca variabile 0/1). Ce trebuie să obțineți: perechile discordante 36 (≥ 100 doar la includere) și 1 (≥ 100 doar la 6 luni), p < 0,001. Valoarea statisticii depinde de varianta afișată: 33,11 fără corecție sau 31,24 cu corecție.

L7.8 Tabele 2 x 2 și 2 x n în Epi Info StatCalc

Epi Info StatCalc lucrează cu tabele deja sumarizate: nu are nevoie de fișierul de date, ci doar de frecvențe.

Tabel 2 x 2: obezitate x diabet.

1. Deschideți Epi Info 7 > StatCalc > Tables (2 x 2, 2 x n).
2. Introduceți: expuși (obezi) cu boala (diabet) 18, expuși fără boală 36, neexpuși cu boală 11, neexpuși fără boală 95. Atenție la orientarea tabelului: în StatCalc, rândurile sunt expunerea, coloanele sunt boala, iar celula din stânga sus este "expus și bolnav".
3. Citiți ieșirea. Ce trebuie să obțineți: Chi-square uncorrected 12,70 (p = 0,0004); Chi-square Yates corrected 11,20 (p = 0,0008); Fisher exact, bilateral, aproximativ 0,0008; Odds ratio 4,32 (IC 95% 1,86-10,03); Risk ratio 3,21 (IC 95% 1,64-6,31). StatCalc afișează și o variantă Mantel-Haenszel a chi-pătratului, puțin mai mică decât cea necorectată (12,62), și intervale mid-p; nu le confundați cu valorile din JASP.
4. Interpretați, în fișă: diabetul este de 3,2 ori mai frecvent la obezi (raportul prevalențelor, studiu transversal), iar asocierea este semnificativă statistic.

Tabel 2 x n: brațul și scăderea ≥ 5%.

5. În același modul, adăugați coloane pentru a obține trei niveluri ale expunerii (A, B, C) și introduceți cazurile (scădere ≥ 5%: 6, 16, 16) și non-cazurile (47, 37, 38). Ce trebuie să obțineți: chi-pătrat aproximativ 6,77, cu 2 grade de libertate, p = 0,034.
6. StatCalc > Chi square for trend: introduceți scorurile de expunere 0, 1, 2 pentru brațele A, B, C, cu cazurile și non-cazurile de mai sus. Ce trebuie să obțineți: chi-pătrat pentru trend aproximativ 4,9, p = 0,027 (Epi Info folosește varianta Mantel a testului, care poate diferi la a doua zecimală). Explicați în fișă de ce testul pentru trend are 1 grad de libertate, iar testul global 2.

Opțional, în Classic Analysis: READ (fișierul CSV), apoi TABLES brat scadere5 pentru un tabel 3 x 2 cu chi-pătratul afișat automat, și MEANS dglic brat, care afișează ANOVA, testul Bartlett și testul Kruskal-Wallis în aceeași ieșire.

L7.9 Chi-pătrat pas cu pas în Google Sheets

Importați mg2026_curat.csv într-o foaie nouă (File > Import > Upload, Separator type: comma). Lucrați într-o a doua foaie, numită "chi".

1. Tabelul observat (brațul pe rânduri, scadere5 pe coloane). În celulele A2:A4 scrieți A, B, C; în B1 scrieți "da", în C1 "nu". În B2 scrieți formula:

   =COUNTIFS(mg2026!$B$2:$B$161;$A2;mg2026!$AA$2:$AA$161;1)

   iar în C2:

   =COUNTIFS(mg2026!$B$2:$B$161;$A2;mg2026!$AA$2:$AA$161;0)

   Copiați formulele în jos până la rândul 4. Trebuie să obțineți 6/47, 16/37, 16/38.
2. Totalurile: în D2 =SUM(B2:C2) (copiați în jos), în B5 =SUM(B2:B4) (copiați la dreapta), în D5 =SUM(D2:D4). Trebuie să obțineți 53, 53, 54 pe rânduri, 38 și 122 pe coloane, 160 total.
3. Tabelul așteptat, în F2:G4. În F2 scrieți:

   =$D2*B$5/$D$5

   și copiați formula în tot blocul F2:G4. Observați simbolurile $: rândul totalului de coloană și coloana totalului de rând rămân fixe. Trebuie să obțineți 12,59/40,41, 12,59/40,41, 12,82/41,18.
4. Valoarea p direct: =CHISQ.TEST(B2:C4;F2:G4). Trebuie să obțineți 0,034. Atenție: funcția primește tabelul observat și tabelul așteptat, nu datele brute, și nu aplică nicio corecție de continuitate.
5. Statistica χ^2: =SUMPRODUCT((B2:C4-F2:G4)^2/F2:G4). Trebuie să obțineți 6,77. Verificați p: =CHISQ.DIST.RT(6,77;2) dă 0,034, iar valoarea critică =CHISQ.INV.RT(0,05;2) dă 5,99.
6. V Cramér: =SQRT(6,77/(160*1)), adică 0,21.
7. Graficul: selectați coloana cu procentele pe rânduri (calculați în H2 =B2/D2, formatat ca procent) și inserați un grafic cu bare (Insert > Chart, Column chart). Trebuie să arate ca figura 7.4 din capitolul 7.
8. Rangurile pentru un Mann-Whitney manual (opțional): pentru exemplul din capitolul 7 (scorurile 2, 3, 3, 5, 6 și 4, 6, 7, 7, 8, 9 în A2:A12), în B2 scrieți =RANK.AVG(A2;$A$2:$A$12;1) și copiați în jos. Suma rangurilor primelor cinci valori, =SUM(B2:B6), trebuie să fie 17,5, iar U1 = 17,5 - 5 x 6/2 = 2,5.

Repetați pașii 1-6 pentru tabelul obezitate x diabet (folosind o coloană auxiliară cu =IF(K2>=30;1;0) în foaia de date). Trebuie să obțineți χ^2 = 12,70 și p = 0,0004 (fără corecție Yates).

Exerciții suplimentare

1. În fișierul BPOC de pe Moodle, alegeți o variabilă ordinală sau numerică asimetrică și comparați-o între sexe cu testul Mann-Whitney; raportați medianele, U, p și corelația biserială a rangurilor. Justificați de ce nu ați folosit testul t.
2. În același fișier, comparați o variabilă măsurată la începutul și la sfârșitul tratamentului (de exemplu o valoare funcțională respiratorie de la momentul 1 și momentul 2) cu testul Wilcoxon și cu testul t pereche. Verificați normalitatea diferențelor și decideți care test este cel potrivit.
3. În fișierul BD_Studenti, investigați asocierea dintre sex și opinia despre biostatistică (tabel de contingență). Verificați frecvențele așteptate; dacă una dintre categorii are prea puțini studenți, propuneți o comasare cu sens și refaceți testul.

4. În fișierul Date_HbA1c, comparați proporția pacienților care ating ținta terapeutică (HbA1c < 7,5%) între primul și ultimul an. Atenție: dacă sunt aceiași pacienți, testul potrivit este McNemar, nu chi-pătratul obișnuit.
5. Problemă cu date în enunț: într-un studiu de cohortă la pensionari, 30 din 80 de persoane expuse la pulberi și 30 din 240 de persoane neexpuse au prezentat scăderea capacității pulmonare. Construiți tabelul, calculați frecvențele așteptate și χ^2 de mână, apoi verificați în Epi Info StatCalc.

   Rezolvare: totalul cu scădere este 60 din 320 (18,75%). Frecvențele așteptate: expuși $\frac{80\times60}{320} = 15$ cu scădere și 65 fără; neexpuși $\frac{240\times60}{320} = 45$ cu scădere și 195 fără. Observat: 30/50 și 30/210. |O - E| = 15 în fiecare celulă. $\chi^2 = \frac{225}{15} + \frac{225}{65} + \frac{225}{45} + \frac{225}{195} = 15 + 3{,}46 + 5 + 1{,}15 = 24{,}62$, gl 1, p < 0,001. Riscul la expuși 30/80 = 37,5%, la neexpuși 30/240 = 12,5%, RR = 3,0. Expunerea la pulberi este un factor de risc semnificativ pentru scăderea capacității pulmonare în acest eșantion.

Ce se predă și barem orientativ

Se încarcă pe Moodle un singur fișier PDF (sau documentul Google exportat ca PDF) care conține:

- capturile de ecran ale tabelelor de ieșire din JASP pentru pașii L7.2-L7.6 (2 puncte);
- tabelul de verificare completat, cu valorile obținute de voi alături de cele din acest text, și explicația oricărei diferențe (2 puncte);
- ieșirea Epi Info pentru tabelul 2 x 2 și pentru testul pentru trend, cu interpretarea OR și RR (1,5 puncte);
- foaia Google Sheets (link sau captură) cu tabelele observat și așteptat și formulele vizibile (1,5 puncte);
- răspunsurile la întrebările de control din L7.3, L7.4 și L7.6 (2 puncte);
- concluzia statistică și clinică, în 2-3 fraze, pentru fiecare dintre cele cinci teste (1 punct).

Total: 10 puncte. Se depunctează: test aplicat pe date filtrate greșit (filtru uitat activ), media și abaterea standard raportate pentru satisfacție, chi-pătrat raportat fără verificarea frecvențelor așteptate, concluzia "H0 se acceptă".

Varianta fără rețea

Dacă rețeaua sau platforma Moodle nu sunt disponibile, fișierul mg2026_curat.csv se copiază de pe stick-ul cadrului didactic sau din folderul partajat al laboratorului. JASP, MedCalc și Epi Info funcționează local, fără conexiune. Pasul L7.9 (Google Sheets) se înlocuiește cu același calcul într-o foaie de calcul instalată local (LibreOffice Calc sau Excel), unde funcțiile se numesc la fel (CHISQ.TEST, CHISQ.DIST.RT, RANK.AVG, COUNTIFS) și separatorul de argumente este tot ";" în configurarea românească. Dacă niciun program nu este disponibil, se lucrează de mână pe tabelul obezitate x diabet (frecvențe așteptate, χ^2, gl, comparația cu 3,84) și pe exemplul Mann-Whitney din figura 7.1, cu tabelele de valori critice din anexa volumului. Fișierul cu rezultatele se predă la următoarea lucrare.

Lucrarea practică 8. Selecția testului statistic și raportarea standardizată a rezultatelor

Obiectiv

Lucrările practice anterioare v-au învățat câte un test pe rând. În această lucrare nu vi se mai spune ce test să aplicați. Primiți 15 întrebări de cercetare formulate pe setul mg2026, iar pentru fiecare trebuie să alegeți singuri testul, să îl justificați, să îl aplicați în program și să raportați rezultatul în formatul standardizat al disciplinei (a-g). În ultima parte a lucrării evaluați raportul unui coleg după o grilă comună. Este, practic, o repetiție generală pentru examenul practic, iar greșelile făcute aici sunt cele mai ieftine greșeli pe care le veți face vreodată.

La final, studentul trebuie să poată:

- să identifice, pentru o întrebare de cercetare dată, scopul analizei, tipul variabilei rezultat și designul (independent sau pereche, numărul de grupuri);
- să aleagă testul potrivit și alternativa lui, folosind algoritmul din capitolul 8, și să verifice condițiile de aplicare în program;
- să aplice testul în JASP (sau MedCalc, Epi Info) și să extragă statistica, gradele de libertate, valoarea p și o mărime a efectului;
- să redacteze rezolvarea completă în formatul a-g, cu grafic potrivit;
- să evalueze critic, după o grilă, raportul unui coleg și să formuleze observații constructive;
- să recunoască întrebările la care testarea statistică nu este răspunsul potrivit (de exemplu comparațiile de la includere într-un studiu randomizat).

Instrumente

- JASP 0.98.1 [47] (programul principal); MedCalc 24.0 [50] și Epi Info 7.2.7 [51] pentru verificări; Google Sheets pentru graficele simple și pentru documentul raportului.
- Fișierul mg2026_curat.csv de pe Moodle, cu variabilele calculate în lucrarea practică 7 (dglic, obez, glic100_0, glic100_6). Pentru această lucrare mai creați: dgr (`greutate0 - greutate6`), dtas (`tas0 - tas6`) și tas140_0, tas140_6 (`ifelse(tas0 >= 140, 1, 0)`, respectiv pentru tas6).
- Șablonul de raport (secțiunea L8.2), disponibil pe Moodle ca document Google.
- Figura 8.1 (algoritmul de alegere a testului) și tabelul 8.1 (sinteza testelor) din capitolul 8, deschise alături.
- Timp orientativ: 100 de minute (recapitulare 10 minute, rezolvarea problemelor 60 de minute, evaluare colegială 20 de minute, discuție finală și încărcare 10 minute).

Desfășurare

L8.1 Recapitularea algoritmului (10 minute)

Înainte de calculator, lucrați pe hârtie. Pentru fiecare întrebare din lista de la L8.3, completați doar primele patru coloane ale tabelului de mai jos, fără să deschideți programul. Abia după ce ați decis testul treceți la JASP. Exercițiul vă obligă să alegeți testul din design, nu din ce "iese".

Tabelul L8.1. Fișa de decizie (se completează înainte de analiză).

Nr.	Scopul (diferență / asociere / acord / predicție)	Variabila rezultat și tipul ei	Designul (grupuri, independent sau pereche)	Testul ales	Alternativa (dacă nu sunt îndeplinite condițiile)
1					
...					
15					

Cadrul didactic verifică fișa de decizie a fiecărui student înainte de începerea calculelor (pentru grupele mari, prin sondaj). O alegere greșită se corectează acum, nu după 40 de minute de lucru.

L8.2 Șablonul de raport standardizat

Fiecare problemă se redactează după șablonul de mai jos. Textul total pentru o problemă nu trebuie să depășească o jumătate de pagină, plus graficul. Concizia face parte din cerință: un raport clinic bun spune tot ce trebuie, nimic în plus.

Tabelul L8.2. Șablonul de raport a-g.

Componenta	Ce se scrie	Exemplu de formulare
(a) Ipoteza de cercetare	o propoziție, în limbaj clinic	"Femeile au HDL-colesterolul mai mare decât bărbații."
(b) Ipotezele statistice	H0 și H1 în cuvinte și matematic; bilateral	"H0: μF = μM, nu există diferență între mediile HDL; H1: μF ≠ μM."
(c) Metoda și justificarea	tipul variabilelor; designul; verificarea condițiilor cu valorile p ale testelor de verificare; testul ales și alternativa	"HDL numeric, sexul nominal cu 2 categorii, grupuri independente; SW p = 0,41 și 0,48; Levene p = 0,034, deci testul t Welch."
(d) Rezolvarea și programul	programul, meniul, opțiunile bifate	"JASP, T-Tests > Independent Samples T-Test: Welch, Descriptives, Effect size, Assumption checks."
(e) Rezultatul, cu grafic	statistici descriptive pe grupuri; statistica testului, gl, p; mărimea efectului cu IC 95%; graficul	"F 53,92 ± 11,70; M 44,41 ± 9,66; t = 5,63, gl 157,7, p < 0,001; diferența 9,51 (IC 95% 6,18-12,85); d = 0,88; box-plot."
(f) Concluzia statistică	decizia față de H0	"p < 0,001, respingem H0: diferența este semnificativă statistic."
(g) Concluzia de cercetare	ce înseamnă clinic, cu mărimea efectului și cu limitele	"Femeile au HDL în medie cu 9,5 mg/dL mai mare; efect mare; analiză transversală pe un eșantion de pacienți cu risc cardiometabolic."

Reguli de redactare, verificate la evaluare:

- valorile p se scriu cu trei zecimale (p = 0,034), iar sub 0,001 ca "p < 0,001"; niciodată "p = 0,000";
- variabilele numerice normale se descriu prin medie ± AS, cele asimetrice și cele ordinale prin mediană (IQR), cele nominale prin n (%);
- formularea deciziei este "respingem H0" sau "nu respingem H0", niciodată "acceptăm H0";
- graficul trebuie să corespundă tipului de date: box-plot sau grafic de medii cu IC pentru variabile numerice, grafic cu bare (procente) pentru variabile nominale și ordinale, diagramă de dispersie pentru asocierea a două variabile numerice;
- zecimalele se scriu cu virgulă.

Un exemplu de raport complet, pe o întrebare care nu face parte din lista de mai jos, arată cât de lung trebuie să fie, de fapt, un răspuns bun.

Întrebarea: diferă glicemia a jeun la includere între pacienții cu și fără diabet confirmat prin TTGO?

(a) Pacienții la care TTGO a confirmat diabetul au glicemia a jeun mai mare decât cei fără diabet. (b) H0: media glicemiei a jeun este aceeași în cele două grupuri (μdiabet = μfără diabet); H1: mediile diferă (μdiabet ≠ μfără diabet), test bilateral. (c) Glicemia este numerică, diabetul este nominal cu două categorii, grupurile sunt independente și inegale (29 și 131). Shapiro-Wilk pe grupuri: p = 0,198 (diabet) și p = 0,061 (fără diabet); Levene p = 0,457. Testul ales: t nepereche; alternativa: Mann-Whitney. (d) JASP, T-Tests > Independent Samples T-Test; Dependent Variables: glicemie0; Grouping Variable: diabet_ogtt; bifate Student, Welch, Mann-Whitney, Descriptives, Effect size, Assumption checks; graficul din Descriptives (box-plot pe grupuri). (e) Fără diabet $101{,}85 \pm 13{,}97$ mg/dL(n 131), cu diabet $121{,}62 \pm 13{,}19$ mg/dL(n 29); t = -6,96, gl 158, p < 0,001 (Welch t = -7,22, gl 43,1, p < 0,001); diferența 19,77 mg/dL (IC 95% 14,25-25,28); d Cohen = 1,43; Mann-Whitney p < 0,001; box-plot atașat. (f) p < 0,001 < 0,05, deci respingem H0: diferența este semnificativă statistic. (g) Pacienții cu diabet confirmat au avut glicemia a jeun în medie cu aproximativ 20 mg/dL mai mare, un efect foarte mare. Rezultatul este de așteptat, pentru că glicemia a jeun face parte din criteriile de diagnostic; dacă am vrea să știm cât de bine identifică glicemia a jeun diabetul confirmat prin TTGO, întrebarea ar fi de predicție, iar metoda potrivită ar fi curba ROC (AUC 0,844 în mg2026, capitolul 12).

L8.3 Cele 15 probleme

Rezolvați problemele în ordine. Problemele 1-10 sunt obligatorii pentru toți studenții; problemele 11-15 se împart pe grupe (fiecare student rezolvă cel puțin două, după indicația cadrului didactic). Pragul de semnificație este α = 0,05, testele sunt bilaterale. Răspunsurile așteptate, pentru cadrul didactic, se găsesc în fișierul de răspunsuri al capitolului 8.

Problema L1. Diferă circumferința taliei între femei și bărbați? Atenție la rezultatul testului de normalitate pe fiecare grup și la mărimea grupurilor: decideți și argumentați dacă folosiți testul t sau testul Mann-Whitney.

Problema L2. Diferă trigliceridele între pacienții cu și fără diabet confirmat prin TTGO? Țineți cont de distribuția trigliceridelor (capitolul 3) și de mărimea foarte inegală a grupurilor.

Problema L3. Diferă HbA1c la includere între cele trei brațe? Rezolvați tehnic, apoi discutați în (g) dacă această întrebare are sens într-un studiu randomizat.

Problema L4. Diferă colesterolul total la includere între cele trei brațe? (Aceeași discuție ca la L3; comparați cele două rezultate.)

Problema L5. Diferă scăderea în greutate la 6 luni (dgr) între cele trei brațe? Dacă testul global este semnificativ, precizați care brațe diferă între ele.

Problema L6. Diferă IMC-ul între nefumători, foști fumători și fumători? Verificați distribuția IMC în fiecare grup și omogenitatea variațelor înainte de a alege testul.

Problema L7. Diferă satisfacția la 6 luni între femei și bărbați?

Problema L8. Diferă vârsta între pacienții din mediul urban și cei din mediul rural?

Problema L9. Diferă proporția pacienților cu scădere în greutate de cel puțin 5% între femei și bărbați? Calculați și raportați frecvențele așteptate.

Problema L10. Este statutul de fumător asociat cu mediul de rezidență?

Problema L11. Este statutul de fumător distribuit la fel în cele trei brațe? Calculați testul, apoi explicați în (g) cum interpretați o diferență de la includere într-un studiu randomizat.

Problema L12. S-a schimbat proporția pacienților cu glicemia a jeun ≥ 100 mg/dL între includere și 6 luni, separat în brațul A și în brațul B? Alegeți testul pentru date pereche potrivit și comentați diferența dintre cele două brațe.

Problema L13. Este glicemia a jeun la includere asociată cu HbA1c la includere? Construiți diagrama de dispersie și alegeți coeficientul de corelație potrivit.

Problema L14. S-a modificat glicemia a jeun la 0, 3 și 6 luni în brațul A? Verificați condițiile ANOVA cu măsurători repetate (normalitate, sfericitate) și decideți care test îl stabiliți ca principal. Comparați apoi rezultatul cu cel al alternativei și discutați.

Problema L15. Este vârsta asociată cu scorul FINDRISC? Alegeți coeficientul de corelație și discutați în (g) dacă asocierea poate fi parțial "construită" de felul în care se calculează scorul FINDRISC.

L8.4 Evaluarea colegială (20 de minute)

După ce ați terminat problemele 1-10, schimbați raportul cu colegul de lângă voi (prin partajarea documentului Google, cu drept de comentare). Fiecare evaluează raportul celuilalt pentru trei probleme alese de cadrul didactic, după grila din tabelul L8.3. Comentariile se scriu direct în document, pe marginea textului. Scopul nu este să "prindeți" greșeli, ci să le găsiți acum, în locul examinatorului.

Tabelul L8.3. Grila de evaluare colegială (pentru o problemă; 0 = absent sau greșit, 1 = parțial, 2 = complet și corect).

Criteriul	Ce verificați	Punctaj
1. Ipotezele	H0 și H1 sunt scrise în cuvinte și matematic, cu parametrul corect (μ, Me, p)	0-2
2. Alegerea testului	testul corespunde tipului variabilei și designului; este menționată alternativa	0-2
3. Verificarea condițiilor	normalitatea pe grupuri sau pe diferențe, Levene, frecvențele așteptate, după caz, cu valori	0-2
4. Rezultatul	descrierea potrivită (medie ± AS sau mediană și IQR sau n și %); statistica, gl, p; mărimea efectului	0-2
5. Graficul	tipul graficului se potrivește datelor; axele au titlu și unitate	0-2
6. Concluziile	decizia statistică corectă ("nu respingem", nu "acceptăm"); concluzia clinică ține cont de mărimea efectului și de design	0-2

Pe lângă punctaj, evaluatorul scrie o observație pozitivă ("ce ai făcut foarte bine") și o sugestie concretă de îmbunătățire. După evaluare, fiecare student are 5 minute să își corecteze raportul. În raportul final se păstrează comentariile colegului și se marchează ce s-a corectat.

Întrebări care merită puse la evaluare, din experiența laboratorului:

- Colegul a verificat normalitatea pe tot eșantionul sau pe fiecare grup?
- Pentru o problemă cu date pereche (L12, L14), a folosit un test pentru date pereche?
- A raportat media pentru satisfacție (L7)? Dacă da, descrierea nu se potrivește cu testul.
- A scris "nu există diferență" după un rezultat nesemnificativ, în loc de "nu am găsit o diferență semnificativă"?
- La L3, L4 și L11, a observat că testarea diferențelor de la includere într-un studiu randomizat este discutabilă?
- La L14, a ales testul înainte de a vedea valoarea p, sau l-a ales pe cel care "iese"?

L8.5 Discuția finală și încărcarea

În ultimele 10 minute, cadrul didactic discută cu toată grupa trei probleme în care alegerea a fost cel mai des greșită și una în care testele au dat rezultate discordante (de regulă L1, L12 sau L14). Fiecare student încarcă pe Moodle raportul final (PDF), cu fișa de decizie, cele 10-12 probleme rezolvate și comentariile colegului.

Exerciții suplimentare

1. Pe fișierul BPOC de pe Moodle, formulați voi trei întrebări de cercetare: una care cere un test pentru două grupuri independente, una pentru date pereche și una pentru două variabile calitative. Rezolvați-le în formatul a-g, fără a vă uita la rezolvările din alte lucrări.
2. Pe fișierul Date_HbA1c, rezolvați problema "Variază semnificativ media HbA1c între anii studiați?" în două variante: presupunând măsurători independente (alți pacienți în fiecare an) și presupunând măsurători repetate pe aceiași pacienți. Explicați de ce testul se schimbă.
3. Problemă cu date în enunț, cu rezolvarea completă. Un studiu a comparat durata durerii postoperatorii (zile) după două tehnici de hernioplastie: tehnica 1 (n = 6): 2, 3, 3, 4, 5, 9; tehnica 2 (n = 6): 4, 5, 6, 6, 8, 10. Alegeți și aplicați testul.

   Rezolvare. (a) Tehnica 1 produce o durată mai scurtă a durerii. (b) H0: distribuțiile duratei sunt aceleași (Me1 = Me2); H1: distribuțiile diferă. (c) Variabila este numerică, dar o durată în zile, pe grupuri de câte 6, cu o valoare izolată (9 și 10 zile): Mann-Whitney U; grupuri independente. (d) Calcul de mână. Seria comună ordonată: 2, 3, 3, 4, 4, 5, 5, 6, 6, 8, 9, 10. Rangurile: 2 → 1; 3 și 3 → 2,5; 4 și 4 → 4,5; 5 și 5 → 6,5; 6 și 6 → 8,5; 8 → 10; 9 → 11; 10 → 12. R1 = 1 + 2,5 + 2,5 + 4,5 + 6,5 + 11 = 28; R2 = 4,5 + 6,5 + 8,5 + 8,5 + 10 + 12 = 50; verificare $28 + 50 = 78 = \frac{12 \times 13}{2}$. U1 = 28 - 21 = 7; U2 = 50 - 21 = 29; U = 7. (e) Valoarea critică pentru n1 = n2 = 6 (tabelul 7.15) este 5; medianele 3,5 și 6 zile; $r_{\text{rb}} = 1 - \frac{2 \times 7}{36} = 0{,}61$. (f) U = 7 > 5, nu respingem H0 (p exact aproximativ 0,09). (g) Durerea a durat mai puțin după tehnica 1 (mediana 3,5 față de 6 zile), cu un efect estimat mare, dar diferența nu este semnificativă statistic pe 12 pacienți; este nevoie de un studiu mai mare.

Ce se predă și barem orientativ

Se încarcă pe Moodle un singur fișier PDF cu:

- fișa de decizie completată înainte de analiză (tabelul L8.1), pentru toate cele 15 probleme (1 punct);
- problemele L1-L10 rezolvate în formatul a-g, fiecare cu graficul ei (5 puncte, câte 0,5 puncte pe problemă);
- cel puțin două probleme din L11-L15, în același format (1,5 puncte);
- grila de evaluare completată pentru raportul colegului, cu observația pozitivă și sugestia de îmbunătățire (1,5 puncte);
- lista corecturilor făcute după evaluarea colegială (1 punct).

Total: 10 puncte. Erori care anulează punctajul unei probleme: test pentru grupuri independente pe date pereche; chi-pătrat raportat cu frecvențe așteptate sub 5 fără nicio discuție; test ales după rezultat (de exemplu raportarea alternativei doar pentru că dă p < 0,05); concluzia "acceptăm H0".

Varianta fără rețea

Fără acces la Moodle, fișierul de date se copiază din folderul partajat al laboratorului sau de pe stick, iar raportul se redactează într-un editor de text local și se predă la lucrarea următoare. Evaluarea colegială se face pe foi tipărite: studenții își schimbă rapoartele și completează grila din tabelul L8.3 de mână. Dacă nici calculatoarele nu sunt disponibile, problemele se rezolvă pe hârtie pornind de la statisticile descriptive date de cadrul didactic (medii, abateri standard, mediane, tabele de frecvențe), iar cerința se limitează la componentele (a), (b), (c), (f) și (g) ale formatului, plus calculul de mână al unui test chi-pătrat sau Mann-Whitney ales de cadrul didactic.

Lucrarea practică 9. Analiza corelației și regresiei: Pearson, Spearman și regresia liniară simplă

Obiectiv

Lucrarea transformă în paşi de lucru capitolul 9: pornind de la două variabile măsurate pe aceiaşi pacienți, studentul desenează diagrama de corelație, alege coeficientul potrivit, îl testează, estimează dreapta de regresie, verifică reziduurile și formulează concluzia în formatul temelor disciplinei. Lucrăm pe setul mg2026, cu patru programe diferite, ca să vedeți că rezultatele coincid indiferent de instrument și ca să vă puteți descurca cu oricare dintre ele la examenul practic.

La final, studentul trebuie să poată:

- să construiască o diagramă de corelație cu dreapta de regresie, ecuația și R^2 afișate;
- să calculeze coeficienții Pearson, Spearman și Kendall cu valoarea p și intervalul de încredere, în JASP și MedCalc;
- să obțină în Google Sheets r, R^2, panta, interceptul, eroarea standard a estimării și o predicție, folosind funcțiile CORREL, PEARSON, RSQ, SLOPE, INTERCEPT, FORECAST, STEYX și LINEST;
- să ruleze o regresie liniară simplă în JASP, MedCalc și Epi Info și să citească tabelul coeficienților;
- să verifice condițiile regresiei pe graficele reziduurilor;
- să redacteze rezolvarea unei probleme de corelație și regresie în formatul a-g.

Instrumente

Programe: JASP 0.98.1 [47]; MedCalc 24.0 [50]; Epi Info 7.2.7 [51]; Google Sheets, cu setările regionale România (separatorul zecimal este virgula, iar separatorul de argumente din formule este ";") [43]. Fișiere: mg2026_curat.csv (160 de rânduri, de pe Moodle, platforma Emediqual [52]); pentru exercițiile suplimentare, fișierele clasice ale disciplinei BD_Studenti și Probleme BPOC. Timp orientativ: 100 de minute.

În mg2026_curat.csv, coloanele folosite azi sunt (după importul în Google Sheets): D = varsta, H = activitate, I = inaltime, J = greutate0, K = imc0, M = tas0, Q = hdl, R = trigliceride, S = glicemie0, V = hba1c0. Datele ocupă rândurile 2-161.

Desfășurare

L9.1 Primul pas: diagrama de corelație

1. Deschideți fișierul în JASP (File > Open > Computer > mg2026_curat.csv). Verificați în vizualizarea datelor că imc0, tas0, glicemie0 și hba1c0 sunt marcate ca variabile continue (iconița cu riglă), iar activitate și educatie ca ordinale. Dacă JASP a citit o variabilă greșit, faceți clic pe iconița din antetul coloanei și schimbați tipul.
2. În JASP, meniul Descriptives > Descriptive Statistics: puneți tas0 și imc0 în caseta Variables; la Plots bifați "Scatter plots" (în unele versiuni opțiunea se numește "Correlation plots"). Veți vedea norul de puncte cu IMC pe o axă și TAS pe cealaltă. Puneți-vă cele patru întrebări din capitolul 9: formă, sens, intensitate, valori izolate. Răspunsul așteptat: nor alungit, crescător, destul de împrăștiat, cu câțiva pacienți cu IMC peste 40 care trag privirea spre dreapta.

3. În Google Sheets: importați fișierul (File > Import > Upload, cu separatorul virgulă; verificați că zecimalele au fost citite corect). Selectați coloanele K și M (IMC și TAS), apoi Insert > Chart. În Chart editor, la Chart type alegeți Scatter chart. Asigurați-vă că pe axa X este IMC (prima coloană selectată). În fila Customize > Series bifați Trendline, la Type alegeți Linear, la Label alegeți "Use equation" și bifați "Show R²". Graficul trebuie să afișeze o ecuație de forma y = 1,43x + 89,5 și $R^2 = 0{,}164$.
4. Repetați graficul pentru glicemie0 (S) și hba1c0 (V). Norul este vizibil mai strâns ($R^2 = 0{,}545$). Comparați cele două grafice: așa arată diferența dintre o corelație slabă (r = 0,405) și una moderată spre puternică (r = 0,738).

L9.2 Coeficienții de corelație în JASP

5. Meniul Regression > Correlation (în versiunile recente, modulul se numește Regression și conține "Correlation"). Mutați în caseta Variables: imc0, tas0, varsta, glicemie0, hba1c0, hdl, trigliceride.
6. În secțiunea Sample Correlation Coefficient bifați Pearson's r, Spearman's rho și Kendall's tau-b. În Additional Options bifați "Report significance", "Flag significant correlations" și "Confidence intervals" (95%). La Plots bifați "Scatter plots" și, pentru imaginea de ansamblu, "Heatmap".
7. JASP afișează o matrice. Pentru perechile de interes trebuie să regăsiți valorile din tabelul L9.1.

Tabelul L9.1. Valori de autoverificare pe mg2026 (n = 160).

Pereche	Pearson r (IC 95%)	p	Spearman rho	Kendall tau-b	Colton
IMC - TAS	0,405 (0,266-0,527)	< 0,001	0,372	0,258	slabă
vârstă - TAS	0,256 (0,106-0,396)	0,001	0,174	0,122	slabă
glicemie - HbA1c	0,738 (0,659-0,802)	< 0,001	0,728	0,557	moderată
HDL - trigliceride	0,076 (-0,080 la 0,229)	0,34	0,075	0,050	fără corelație
înălțime - greutate	0,652 (0,552-0,733)	< 0,001	0,693	0,505	moderată

8. Pentru variabila ordinală activitate, rulați separat Spearman cu imc0: rho = -0,012, p = 0,88. Pentru educatie și satisfactie: rho = 0,125, p = 0,115. Observați că pentru variabilele ordinale nu se raportează Pearson; dacă îl bifați, JASP îl calculează oricum, dar interpretarea lui nu este justificată.
9. Ce să notați în caiet: pentru IMC-TAS, de ce raportăm ambele coeficiente (IMC nu este normal distribuit: Shapiro-Wilk p < 0,001, asimetrie 1,42) și faptul că cele două valori (0,405 și 0,372) sunt apropiate, deci valorile extreme nu distorsionează mult rezultatul.

L9.3 Regresia liniară simplă în JASP

10. Meniul Regression > Linear Regression. Puneți tas0 la Dependent Variable și imc0 la Covariates. Metoda: Enter.
11. La Statistics bifați: Estimates, Confidence intervals (95%), Model fit, Descriptives; la Residuals bifați Durbin-Watson și Casewise diagnostics (cu pragul implicit de 3 abateri standard).
12. La Plots bifați: "Residuals vs. predicted", "Residuals histogram" (cu "Standardized residuals") și "Q-Q plot standardized residuals".

13. Tabelul Model Summary trebuie să arate R = 0,405, $R^2 = 0{,}164$, R^2 ajustat $= 0{,}158$ și RMSE = 12,73. RMSE (engl. root mean square error) este, în JASP, eroarea standard a estimării (SEE) din curs.
14. Tabelul ANOVA: suma pătratelor pentru regresie 5015,2, pentru reziduuri 25 617,0, total 30 632,2; F(1; 158) = 30,93, p < 0,001.
15. Tabelul Coefficients: interceptul 89,47 (ES 7,56; t = 11,84; IC 74,54-104,40), panta pentru imc0 1,435 (ES 0,258; t = 5,56; p < 0,001; IC 0,925-1,944). Verificați că t pentru pantă este același cu cel obținut de mână pentru r în capitolul 9 și că $F = t^2 (5{,}56^2 = 30{,}93)$.
16. Graficele: reziduurile față de valorile prezise trebuie să formeze o bandă orizontală fără curbură și fără formă de pâlnie (comparați cu figura 9.5 din capitol); histograma și graficul Q-Q trebuie să arate o distribuție aproximativ normală. Dacă vreți un test formal, salvați reziduurile (opțiunea de adăugare a reziduurilor ca variabilă nouă în date, acolo unde versiunea o permite) și aplicați Shapiro-Wilk din Descriptives: p = 0,878.
17. Repetați regresia cu varsta ca predictor: TAS = 108,28 + 0,406 x vârsta, IC pentru pantă 0,166-0,646, $R^2 = 0{,}066$. Și cu glicemie0 ca predictor pentru hba1c0: HbA1c = 3,379 + 0,0209 x glicemia, $R^2 = 0{,}545$, RMSE = 0,301.

L9.4 Google Sheets: aceleași rezultate prin funcții

18. Într-o foaie nouă (sau în coloanele libere din dreapta datelor) scrieți etichetele și formulele de mai jos. Toate folosesc intervalele de pe rândurile 2-161. Atenție la ordinea argumentelor: la SLOPE, INTERCEPT, RSQ, STEYX, FORECAST și LINEST se dă întâi Y (variabila dependentă), apoi X; la CORREL și PEARSON ordinea nu contează.

Tabelul L9.2. Funcțiile Google Sheets pentru corelație și regresie (TAS în funcție de IMC).

Ce calculăm	Formula	Rezultat așteptat
coeficientul Pearson	=CORREL(K2:K161;M2:M161)	0,405
același, altă funcție	=PEARSON(K2:K161;M2:M161)	0,405
coeficientul de determinare	=RSQ(M2:M161;K2:K161)	0,164
panta b	=SLOPE(M2:M161;K2:K161)	1,435
interceptul a	=INTERCEPT(M2:M161;K2:K161)	89,47
eroarea standard a estimării	=STEYX(M2:M161;K2:K161)	12,73
predicția pentru IMC 30	=FORECAST(30;M2:M161;K2:K161)	132,5
predicția pentru IMC 35	=FORECAST(35;M2:M161;K2:K161)	139,7
statistica t pentru r (r în celula X2)	=X2*SQRT(158)/SQRT(1-X2^2)	5,56
valoarea p bilaterală (t în celula X3)	=T.DIST.2T(X3;158)	aproximativ 0,0000001

19. Funcția LINEST întoarce toate statisticile regresiei într-un bloc de 5 rânduri și 2 coloane. Selectați o celulă goală care are dedesubt și în dreapta spațiu liber și scrieți =LINEST(M2:M161;K2:K161;TRUE;TRUE). Blocul obținut se citește astfel: rândul 1, panta 1,435 și interceptul 89,47; rândul 2, erorile lor standard 0,258 și 7,56; rândul 3, R² = 0,164 și SEE = 12,73; rândul 4, F = 30,93 și gradele de libertate reziduale 158; rândul 5, suma pătratelor regresiei 5015,2 și suma pătratelor reziduale 25 617,0. LINEST nu afișează etichete, așa că scrieți-le voi alături.
20. Intervalul de încredere pentru pantă: dacă panta este în celula Y1 și eroarea ei standard în Y2, scrieți =Y1-T.INV.2T(0,05;158)*Y2 și =Y1+T.INV.2T(0,05;158)*Y2. Trebuie să obțineți 0,925 și 1,944.
21. Spearman în Sheets nu are o funcție dedicată. Se procedează ca în definiție: într-o coloană nouă calculați rangul fiecărui IMC cu =RANK.AVG(K2;K$2:K$161;1), trageți formula până la rândul 161; faceți la fel pentru TAS. Apoi =CORREL pe cele două coloane de

ranguri dă rho = 0,372. RANK.AVG atribuie valorilor egale media rangurilor, exact ca în capitolul 9; funcția RANK simplă nu face acest lucru și dă un rezultat ușor diferit.

22. Pentru activitate și IMC, aplicați aceeași procedură și verificați rho = -0,012. Cu multe valori egale (activitatea are doar 5 categorii), este important să folosiți RANK.AVG.

L9.5 MedCalc

23. Importați fișierul (File > Import, sau deschideți direct fișierul CSV, cu virgula ca separator de câmpuri și punctul ca separator zecimal; verificați în foaia de date că valorile sunt numerice).
24. Statistics > Correlation > Correlation coefficient: Variable Y = tas0, Variable X = imc0. MedCalc raportează mărimea eșantionului, coeficientul r = 0,405, nivelul de semnificație P < 0,0001 și IC 95% pentru r 0,266-0,527.
25. Statistics > Correlation > Rank correlation: aceleași variabile. Bifați atât Spearman's rho, cât și Kendall's tau, dacă versiunea oferă ambele opțiuni. Valori așteptate: rho = 0,372 (IC 95% aproximativ 0,23-0,50) și tau = 0,258.
26. Statistics > Regression > Regression: Variable Y = tas0, Variable X = imc0, model "y = a + b x". În opțiuni bifați reprezentarea reziduurilor. Rezultatul include coeficientul de determinare $R^2 = 0{,}164$, R² ajustat 0,158, abaterea standard reziduală (Residual standard deviation) 12,73, ecuația y = 89,47 + 1,435 x, tabelul coeficienților cu eroarea standard, IC 95%, t și P și tabelul ANOVA. MedCalc afișează automat și diagrama de dispersie cu dreapta de regresie; dacă nu, folosiți Graphs > Scatter diagram, cu opțiunea de afișare a dreptei de regresie și a benzii de încredere.
27. Statistics > Regression > Regression, cu Y = hba1c0 și X = glicemie0: verificați HbA1c = 3,379 + 0,0209 x glicemia.

L9.6 Epi Info 7

28. Din meniul principal, Analyze Data (Classic Analysis). Comanda Read (Import): alegeți tipul de fișier text/CSV și încărcați mg2026_curat.csv. Verificați cu List sau cu Frequencies că variabilele au fost citite.
29. Statistics > Linear Regression: Outcome Variable = tas0, Other Variables = imc0. Rezultatul afișează coeficientul de corelație sub forma "r^2" = 0,16, tabelul coeficienților (pentru imc0: 1,435, eroarea standard 0,258; pentru constantă: 89,47, eroarea standard 7,56) și un tabel ANOVA. În loc de statistica t, Epi Info afișează pentru fiecare coeficient un test F, care este t²: pentru imc0, F = 30,93, iar pentru constantă F = 140,1. Valorile p sunt aceleași.
30. Alternativ, în Visual Dashboard: Add Analysis Gadget > Advanced Statistics > Linear Regression, cu aceleași variabile. Rezultatele sunt identice.
31. Epi Info nu are un meniu separat pentru Spearman în Classic Analysis; pentru date ordinale folosiți JASP, MedCalc sau procedura cu ranguri din Sheets.

L9.7 Redactarea rezultatului în formatul a-g

32. Redactați rezolvarea problemei "Există o asociere între IMC și TAS la includere?" pe modelul de mai jos și comparați cu varianta voastră.

(a) Ipoteza de cercetare: pacienții cu IMC mai mare au, în medie, tensiunea sistolică mai mare.

(b) Ipotezele statistice: H0: ρ = 0 (IMC și TAS nu sunt corelate liniar în populație); H1: ρ ≠ 0.

(c) Metoda: două variabile numerice pe aceiași pacienți, deci corelație și regresie liniară simplă; X = IMC (predictor), Y = TAS (răspuns). TAS este normală (Shapiro-Wilk p = 0,303), IMC asimetric (p < 0,001), de aceea raportăm Pearson și Spearman.

(d) Programul: JASP, Regression > Correlation și Regression > Linear Regression; verificare în Google Sheets.

(e) Rezultatul: r = 0,405 (IC 95% 0,266-0,527), p < 0,001; rho = 0,372, p < 0,001; TAS = 89,47 + 1,435 x IMC (IC 95% pentru pantă 0,925-1,944); $R^2 = 0{,}164$; SEE = 12,73 mmHg; graficul de dispersie cu dreapta de regresie (figura 9.4 din capitol).

(f) Concluzia statistică: p < 0,001, H0 se respinge; corelația este pozitivă și slabă după regula lui Colton.

(g) Concluzia de cercetare: la pacienții cu risc cardiometabolic din studiu, fiecare kg/m² în plus se asociază cu o TAS mai mare în medie cu aproximativ 1,4 mmHg. IMC explică doar 16,4% din variația TAS; 83,6% ține de alți factori. Predicția pentru IMC 30 este 132,5 mmHg, pentru IMC 35 este 139,7 mmHg; modelul nu trebuie folosit pentru IMC în afara intervalului 23,5-48,3 kg/m².

33. Repetați redactarea pentru perechea înălțime-greutate și răspundeți la cele două întrebări clasice ale lucrării: ce proporție din variația greutății ține de alți factori decât înălțimea ($1 - R^2 = 0{,}576$, adică 57,6%) și ce greutate se prezice pentru 1,78 m (93,1 kg) și pentru 2,35 m (formula dă 155,0 kg, dar este o extrapolare în afara domeniului 145-192 cm, deci predicția nu este credibilă).

Exerciții suplimentare

Exercițiile 1-4 se rezolvă pe fișierele clasice ale disciplinei de pe Moodle. Nu dăm rezultatele numerice: le veți compara cu colegii și cu cadrul didactic la lucrare.

1. Fișierul BD_Studenti. Investigați asocierea dintre greutate și înălțime: (1) ce tip de investigație statistică este și care sunt caracteristicile ei; (2) ipotezele statistice în cuvinte și matematic; (3) reprezentarea grafică potrivită; (4) prelucrarea statistică; (5) ce proporție din variația greutății este cauzată de alți factori decât variația înălțimii; (6) interpretarea în raport cu graficul și ipotezele; (7) ce greutate ar trebui să aibă un student de 1,78 m și unul de 2,35 m, cu comentariul despre extrapolare.
2. Fișierul BD_Studenti. Calculați IMC = greutate (kg) / înălțime² (m²) pentru fiecare student și IMC dorit (din greutatea dorită și aceeași înălțime). Investigați asocierea dintre IMC actual și IMC dorit. Alegeți coeficientul după verificarea normalității celor două variabile. Discutați separat cum ați arăta, pe aceleași date, dacă studenții își doresc în medie un IMC diferit de cel actual (indiciu: aceasta este o întrebare de comparație pereche, cap. 6, nu de corelație; o corelație mare nu înseamnă că valorile sunt egale).
3. Fișierul Probleme BPOC. În ce măsură vârsta pacientului (ani) poate prezice debitul expirator maxim inițial (DEM1)? Scrieți ecuația dreptei de regresie, interpretați panta și R², verificați reziduurile și precizați intervalul de vârstă în care ecuația poate fi folosită.
4. Fișierul Probleme BPOC. Există o corelație între tensiunea arterială sistolică inițială (TMAX1) și frecvența cardiacă inițială (FC1)? Alegeți între Pearson și Spearman după verificarea condițiilor și justificați alegerea.
5. Problemă cu date în enunț, cu rezolvare completă. La cinci pacienți s-au măsurat glicemia a jeun (mg/dL) și HbA1c (%): (90; 5,3), (100; 5,2), (110; 5,8), (125; 5,7), (140; 6,4). Calculați r, testați-l, scrieți dreapta de regresie a HbA1c în funcție de glicemie și prezicеți HbA1c pentru o glicemie de 120 mg/dL.

Rezolvare. Mediile: $m_x = \frac{565}{5} = 113$ mg/dL, $m_y = \frac{28,4}{5} = 5,68\%$. Abaterile lui X: -23, -13, -3, 12, 27; ale lui Y: -0,38; -0,48; 0,12; 0,02; 0,72. Produsele: 8,74; 6,24; -0,36; 0,24; 19,44, deci Sxy = 34,30. Sxx = 529 + 169 + 9 + 144 + 729 = 1580. Syy = 0,1444 + 0,2304 + 0,0144 + 0,0004 + 0,5184 = 0,908. $r = \frac{34,30}{\sqrt{1580 \times 0,908}} = \frac{34,30}{37,88} = 0,906$, corelație puternică după Colton. Test: $t = \frac{0,906 \times \sqrt{3}}{\sqrt{1-0,820}} = \frac{1,569}{0,424} = 3,70$; valoarea critică pentru 3 grade de libertate este 3,182; cum 3,70 > 3,182, H0 se respinge (p = 0,034). Panta $b = \frac{34,30}{1580} = 0,0217$ puncte procentuale pe mg/dL; interceptul a = 5,68 - 0,0217 x 113 = 3,23. Ecuația: HbA1c = 3,23 + 0,0217 x glicemia. Pentru 120 mg/dL: 3,23 + 0,0217 x 120 = 5,83%. $R^2 = 0,82$. Comentariu: cu doar cinci pacienți, corelația este semnificativă abia la limită, deși r este 0,9; pe mg2026 (n = 160) corelația este mai mică (0,738), dar estimată mult mai precis. Verificați rezultatul în Sheets cu =CORREL, =SLOPE, =INTERCEPT și =FORECAST.

Ce se predă și barem orientativ

Se predă pe Moodle, la activitatea lucrării, un singur document (PDF sau link Google Docs) care conține: graficul de dispersie IMC-TAS cu dreapta de regresie, ecuația și R² (din Sheets sau JASP); tabelul L9.1 completat cu valorile obținute de voi; captura tabelului Coefficients din JASP și graficul reziduurilor; blocul LINEST din Sheets, cu etichete; redactarea a-g pentru IMC-TAS și pentru înălțime-greutate; rezolvarea exercițiului suplimentar 1 sau 3.

Barem orientativ (10 puncte): graficul corect, cu axele potrivite (X predictor, Y răspuns) și dreapta afișată, 1 punct; coeficienții Pearson și Spearman cu p și IC, alegerea justificată, 2 puncte; dreapta de regresie cu interpretarea pantei și a interceptului, 2 puncte; verificarea reziduurilor, 1 punct; R² și 1 - R² interpretate corect, 1 punct; predicția și comentariul despre extrapolare, 1 punct; redactarea a-g completă și corectă, 2 puncte. Se depunctează formularea "se acceptă H0", interpretarea corelației drept cauzalitate și lipsa unităților de măsură la pantă.

Varianta fără rețea

Dacă nu aveți acces la internet sau la programe, lucrarea se poate face integral cu calculatorul de buzunar și cu un extras tipărit din mg2026 (primii 20 de pacienți, coloanele imc0 și tas0, primit de la cadrul didactic). Pașii sunt cei din problema rezolvată 9.1: tabelul cu abaterile, produsele și pătratele, Sxy, Sxx, Syy, apoi r, t (comparat cu valoarea critică din tabelul distribuției t primit la lucrare), b, a, o predicție și R². Diagrama de corelație se desenează pe hârtie milimetrică, cu dreapta trasă prin punctul mediilor (mx, my) și printr-un al doilea punct calculat din ecuație. Coeficientul Spearman se calculează cu ranguri, după tabelul 9.4. Rezultatele pe 20 de pacienți vor diferi de cele pe 160: comparați-le și explicați de ce intervalul de încredere pentru r este mai larg.

Lucrarea practică 10. Analize epidemiologice: riscul relativ, raportul de șanse și interpretarea studiilor observaționale

Obiectiv

Lucrarea aplică metodele din capitolul 10 pe problemele clasice ale disciplinei și pe setul mg2026. Accentul cade pe trei deprinderi care se verifică la examenul practic: transformarea unui enunț în tabelul 2 x 2 corect, alegerea măsurii de asociere potrivite designului (RR, raportul prevalențelor sau OR) și interpretarea intervalului de încredere. Calculele se fac de mână o dată, apoi se verifică în Epi Info, MedCalc, JASP și Google Sheets.

La final, studentul trebuie să poată:

- să recunoască tipul de studiu dintr-un enunț și să construiască tabelul 2 x 2 cu notația a, b, c, d;
- să calculeze RR, OR, intervalele lor de încredere 95% și testul chi-pătrat sau Fisher, de mână și în Epi Info StatCalc;
- să folosească calculatoarele de risc relativ și raport de șanse din MedCalc și tabelele de contingență din JASP;
- să construiască în Google Sheets o foaie de calcul reutilizabilă pentru orice tabel 2 x 2;
- să estimeze impactul la nivelul populației: riscul atribuibil, fracția atribuibilă și NNT sau NNH;
- să formuleze concluzia statistică și pe cea clinică pentru un studiu observațional, cu limitele lui.

Instrumente

Programe: Epi Info 7.2.7 [51], modulele StatCalc și Classic Analysis; MedCalc 24.0 [50]; JASP 0.98.1 [47]; Google Sheets cu setări regionale România (virgula zecimală, ";" ca separator de argumente) [43]. Fișiere: mg2026_curat.csv de pe Moodle [52]; pentru exercițiile suplimentare, fișierul Probleme BPOC. Timp orientativ: 100 de minute.

Desfășurare

L10.1 De la enunț la tabelul 2 x 2

1. Pentru fiecare problemă, înainte de orice calcul, răspundeți în scris la patru întrebări: Care este expunerea (factorul de risc)? Care este rezultatul (boala)? De unde pornește studiul: de la expunere, de la boală sau de la o "fotografie" a populației? Ce totaluri sunt fixate de design?
2. Construiți tabelul cu expunerea pe rânduri (expuși sus) și boala pe coloane (bolnavi în stânga). Completați întâi totalurile date în enunț, apoi celulele date direct, apoi celulele obținute prin scădere. Verificați la final că a + b + c + d = N.
3. Exersați pe cele patru enunțuri clasice. Fumat-BPOC: 1200 de persoane sănătoase urmărite 5 ani, 300 fumători, 50 de fumători și 70 de nefumători cu BPOC; cohortă, deci a = 50, b = 250, c = 70, d = 830. Monitor-acuitate: 1000 de persoane, 200 expuse cel puțin 8 ore pe zi, dintre care 90 cu acuitate scăzută, și 150 cu acuitate scăzută printre cei 800 neexpuși; transversal, a = 90, b = 110, c = 150, d = 650. Contraceptive orale-artrită reumatoidă: 100 de cazuri și 200 de martore, expuse 40 de cazuri și 120 de martore; caz-control, a = 40, b = 120, c = 60, d = 80. Pulberi-capacitate pulmonară: 320 de pensionari, 80 expuși, câte 30 cu capacitatea scăzută în fiecare grup; a = 30, b = 50, c = 30, d = 210.

4. Greșelile de verificat: în problema pulberilor, "30 în ambele grupuri" nu înseamnă risc egal (30 din 80 față de 30 din 240); în problema fumat-BPOC, numărul nefumătorilor este 900, nu 1200; în problema caz-control, totalurile fixate sunt pe coloane (100 de cazuri și 200 de martore), nu pe rânduri.

L10.2 Epi Info StatCalc: Tables (2 x 2 x n)

5. Din meniul principal Epi Info, alegeți StatCalc, apoi Tables (2 x 2 x n). Apare un tabel cu rândurile Exposed + / Exposed - și coloanele Outcome + / Outcome - (denumirile pot fi ușor diferite în funcție de versiune). Introduceți pentru fumat-BPOC: 50, 250, 70, 830, în ordinea a, b, c, d.
6. Rezultatele apar imediat, în două blocuri. Blocul măsurilor de asociere conține raportul de șanse (Odds ratio) 2,37 cu IC 95% 1,61-3,50 (metoda seriilor Taylor, echivalentă cu metoda logaritmică din curs), o estimare de maximă verosimilitate a OR (MLE Odds Ratio, aproximativ 2,37, cu un interval exact puțin mai larg), riscul relativ (Risk Ratio) 2,14 cu IC 95% 1,53-3,01, diferența de risc (Risk Difference) 8,89% cu IC 95% aproximativ 4,3-13,5% și fracțiile etiologice (la expuși 53,3%, în populație 22,2%). Blocul testelor conține chi-pătrat necorectat 19,75, chi-pătrat Mantel-Haenszel 19,74 și chi-pătrat corectat Yates 18,78, fiecare cu valoarea p (toate sub 0,0001), plus testele exacte (mid-p și Fisher).
7. Chi-pătrat Mantel-Haenszel dintr-un singur tabel este chi-pătratul obișnuit înmulțit cu (N - 1) / N: $\frac{19{,}753 \times 1199}{1200} = 19{,}737$. Diferența contează doar la eșantioane mici.
8. Introduceți varianta mică (5, 25, 7, 83). Notați: RR 2,14 (IC 95% 0,73-6,25), OR 2,37 (IC 95% 0,69-8,13), chi-pătrat necorectat 1,98 (p = 0,16), Fisher exact bilateral p = 0,17. StatCalc semnalează că o frecvență așteptată este sub 5 (frecvența așteptată pentru celula a este 3) și recomandă testul exact. Concluzie: aceleași măsuri de asociere, dar rezultat nesemnificativ.
9. Introduceți pe rând celelalte probleme clasice și completați tabelul L10.1 din secțiunea L10.8. Pentru problema ECG-CHD (8, 39, 10, 134), observați rezultatul la limită: chi-pătrat necorectat p = 0,040, Yates p = 0,077, Fisher p = 0,0485, IC 95% pentru RR 1,03-5,85.
10. Analiza stratificată. StatCalc permite introducerea mai multor straturi (de aici "x n" din denumire): după primul tabel, treceți la stratul următor. Introduceți asocierea obezitate-diabet din mg2026 pe sexe: femei 10, 19, 5, 52; bărbați 8, 17, 6, 43. Rezultatul sumar afișează OR brut (Crude) 4,32 și OR ajustat Mantel-Haenszel 4,31, cu IC 95% aproximativ 1,86-10,0. Cum OR ajustat este practic egal cu OR brut, sexul nu este un factor de confuzie. Valorile OR pe straturi (5,47 la femei, 3,37 la bărbați) diferă, dar testul de omogenitate (Breslow-Day, acolo unde este afișat) nu arată o diferență semnificativă.

L10.3 Epi Info Classic Analysis pe datele brute mg2026

11. Analyze Data > Classic. Comanda Read: importați mg2026_curat.csv.
12. Creați variabila de expunere. Comanda Define: obez (tip numeric). Apoi Assign (sau Recode): obez = 1 dacă imc0 ≥ 30, altfel 0. În Classic Analysis se poate scrie, de exemplu, IF imc0 >= 30 THEN obez = 1 ELSE obez = 0 END. Verificați cu Frequencies obez: 54 de valori 1 și 106 valori 0.
13. Statistics > Tables: Exposure Variable = obez, Outcome Variable = diabet_ogtt. Epi Info afișează tabelul 2 x 2 și aceleași statistici ca StatCalc. Verificați ordinea rândurilor și a coloanelor: celula din stânga sus trebuie să fie "obez = 1 și diabet = 1", adică 18. Dacă

programul a pus categoriile 0 primele, OR afișat va fi inversul (1 / 4,32 = 0,23); în acest caz recodificați variabilele ca Yes/No sau verificați opțiunile de afișare.

14. Valorile așteptate: OR 4,32 (IC 95% 1,86-10,03), RR (aici raportul prevalențelor, fiind o analiză transversală) 3,21 (IC 95% 1,64-6,31), chi-pătrat necorectat 12,70 (p = 0,0004), Yates 11,20 (p = 0,0008), Fisher exact p = 0,0008.

L10.4 MedCalc: calculatoarele de risc relativ și raport de șanse

15. În MedCalc, meniul Tests conține calculatoare care lucrează direct cu frecvențele, fără fișier de date. Alegeți Relative risk (în unele versiuni "Relative risk & Number needed to treat"). Introduceți pentru grupul expus numărul de evenimente (50) și totalul grupului (300), iar pentru grupul de comparație 70 și 900.
16. MedCalc raportează riscul relativ 2,14, IC 95% 1,53-3,01, statistica z = 4,41 și nivelul de semnificație P < 0,0001. Statistica z este $\frac{\ln(\text{RR})}{\text{ES}[\ln(\text{RR})]} = \frac{0{,}762}{0{,}1727} = 4{,}41$, exact testul construit pe scara logaritmică din curs. Dacă se cere și NNT, MedCalc îl afișează ca număr necesar pentru a produce un efect nefavorabil (NNT harm sau NNH): 11,25, cu IC 95% aproximativ 7,4-23,1.
17. Tests > Odds ratio: introduceți pentru fumat-BPOC numărul de expuși și neexpuși printre cazuri și printre necazuri (a = 50, c = 70 printre bolnavi; b = 250, d = 830 printre sănătoși, în câmpurile corespunzătoare ale calculatorului). Rezultat: OR 2,37, IC 95% 1,61-3,50, z = 4,35, P < 0,0001. Citiți cu atenție etichetele câmpurilor: calculatorul de OR din MedCalc este gândit pentru studii caz-control și cere frecvențele pe coloane (cazuri, martori), nu pe rânduri.
18. Repetați pentru contraceptive-artrită (cazuri: 40 expuse, 60 neexpuse; martore: 120 expuse, 80 neexpuse): OR 0,44, IC 95% 0,27-0,73, P = 0,001. Nu folosiți calculatorul de RR pentru această problemă: designul caz-control nu permite RR.
19. Pentru studiul randomizat mg2026, calculatorul de risc relativ cu 16 evenimente din 54 (brațul C) și 6 din 53 (brațul A) dă RR 2,62 (IC 95% 1,11-6,17), P = 0,028, iar NNT (benefit) 5,5, cu IC 95% aproximativ 3,0-29,1. Comparați cu valoarea din curs. Observați că P din calculatorul MedCalc (testul z pe scara logaritmică) diferă puțin de p al testului chi-pătrat (0,019); ambele sunt sub 0,05.

L10.5 JASP: tabele de contingență cu raportul de șanse

20. Deschideți mg2026_curat.csv în JASP. Creați o variabilă calculată: clic pe semnul "+" de lângă ultima coloană, numele obez, tip nominal, definită prin formula ifelse(imc0 >= 30, 1, 0) (editorul R sau constructorul drag-and-drop). Verificați frecvențele: 54 și 106.
21. Meniul Frequencies > Contingency Tables. Rows: obez; Columns: diabet_ogtt. La Statistics bifați χ^2, "χ^2 continuity correction", "Log odds ratio (2 x 2 only)" cu intervalul de încredere 95% și, dacă este disponibil, testul exact Fisher. La Cells bifați Expected counts și Row percentages.
22. Citiți rezultatele: frecvențele așteptate 9,8; 44,2; 19,2; 86,8 (toate peste 5, deci chi-pătrat este valid); χ^2 = 12,70, p < 0,001; cu corecție 11,20, p = 0,0008. JASP raportează logaritmul raportului de șanse: ln(OR) = 1,463, cu IC 95% 0,620-2,305. Transformarea în OR se face prin exponențiere: $e^{1{,}463} = 4{,}32$; $e^{0{,}620} = 1{,}86$; $e^{2{,}305} = 10{,}03$.
23. Atenție la ordinea categoriilor. JASP ordonează categoriile alfabetic sau numeric (0 înaintea lui 1), deci în tabelul afișat celula "obez = 0, diabet = 0" poate apărea în stânga sus. Raportul de șanse rămâne același dacă sunt inversate și rândurile, și coloanele, dar devine inversul (0,23, cu ln = -1,463) dacă este inversată doar una dintre ele. Verificați

întotdeauna semnul logaritmului cu tabelul observat. În JASP puteți schimba ordinea nivelurilor din fereastra variabilei (clic pe antetul coloanei).

24. JASP lucrează cu date individuale. Pentru o problemă dată doar prin frecvențe (de exemplu, monitor-acuitate), creați un fișier mic cu patru rânduri și trei coloane: expunere, rezultat și frecvența (90, 110, 150, 650), apoi în Contingency Tables puneți frecvența în caseta Counts.

L10.6 Google Sheets: o foaie de calcul pentru orice tabel 2 x 2

25. Într-o foaie nouă, puneți a în B2, b în C2, c în B3, d în C3 (de exemplu 50, 250, 70, 830). Scrieți apoi formulele din tabelul L10.2. După ce verificați rezultatele pe fumat-BPOC, schimbați doar cele patru celule și foaia recalculează totul pentru orice problemă.

Tabelul L10.2. Formulele Google Sheets pentru analiza unui tabel 2 x 2 (a în B2, b în C2, c în B3, d în C3).

Celula	Ce calculează	Formula	Fumat-BPOC
F2	riscul la expuși R1	=B2/(B2+C2)	0,1667
F3	riscul la neexpuși R0	=B3/(B3+C3)	0,0778
F4	RR	=F2/F3	2,14
F5	ES[ln(RR)]	=SQRT(1/B2-1/(B2+C2)+1/B3-1/(B3+C3))	0,1727
F6	limita inferioară IC RR	=EXP(LN(F4)-1,96*F5)	1,53
F7	limita superioară IC RR	=EXP(LN(F4)+1,96*F5)	3,01
F8	OR	=(B2*C3)/(C2*B3)	2,37
F9	ES[ln(OR)]	=SQRT(1/B2+1/C2+1/B3+1/C3)	0,1987
F10	limita inferioară IC OR	=EXP(LN(F8)-1,96*F9)	1,61
F11	limita superioară IC OR	=EXP(LN(F8)+1,96*F9)	3,50
F12	diferența de risc (RA)	=F2-F3	0,0889
F13	NNT sau NNH	=ROUNDUP(1/ABS(F12);0)	12
F14	fracția atribuibilă la expuși	=(F4-1)/F4	0,533
F15	proporția expușilor pe	=(B2+C2)/SUM(B2:C3)	0,25
F16	fracția atribuibilă populațională	=F15*(F4-1)/(1+F15*(F4-1))	0,222

26. Testul chi-pătrat în Sheets cere tabelul frecvențelor așteptate. În H2:I3 scrieți frecvențele așteptate: în H2 =(B2+C2)*(B2+B3)/SUM($B$2:$C$3), apoi formulele analoage pentru celelalte trei celule (total rând x total coloană / N). Verificați: 30, 270, 90, 810. Apoi =CHISQ.TEST(B2:C3;H2:I3) întoarce direct valoarea p (aproximativ 0,0000088 pentru fumat-BPOC). Valoarea statisticii se obține cu =CHISQ.INV.RT(p;1) sau cu formula Σ (O - E)² / E: 19,75. CHISQ.TEST nu aplică corecția Yates.

27. Pentru caz-control, foaia calculează și RR, dar rezultatul nu are sens. Adăugați o celulă de avertizare (de exemplu, o listă derulantă "design" cu valorile cohortă, transversal, caz-control și o formulă =IF(K1="caz-control";"RR nu se interpretează";F4)).

L10.7 Estimarea riscului la nivelul populației

28. Pe fumat-BPOC, completați frazele de interpretare cu valorile din foaia de calcul. Riscul atribuibil: "fumatul adaugă aproximativ 8,9 cazuri de BPOC la 100 de fumători în 5 ani". NNH: "la fiecare 12 fumători apare un caz în plus". Fracția atribuibilă la expuși: "aproximativ 53% dintre cazurile de BPOC de la fumători sunt atribuibile fumatului".

Fracția atribuibilă populațională: "aproximativ 22% dintre toate cazurile din această populație ar fi evitabile fără fumat". Toate frazele presupun că relația este cauzală și neconfundată.

29. Pe mg2026, asocierea obezitate-diabet: proporția obezilor pe = 54 / 160 = 0,3375; RP = 3,21; fracția atribuibilă populațională = 0,43. Prevalența diabetului în tot eșantionul (18,1%) minus prevalența la neobezi (10,4%) dă riscul atribuibil populațional, 7,7 puncte procentuale; împărțit la 18,1%, dă din nou 0,43. Formulați fraza cu "dacă", pentru că datele sunt transversale.
30. Pe studiul randomizat mg2026 (brațul C față de A, scădere ≥ 5%): diferența absolută 18,3 puncte procentuale (IC 95% 3,4-33,2), NNT 5,5, rotunjit la 6 (IC 95% 3,0-29,1). Interpretare: "tratând 6 pacienți cu pachetul complet în loc de consilierea standard, obținem în medie încă un pacient cu scădere ponderală relevantă". Aici interpretarea cauzală este justificată de randomizare.

L10.8 Problemele clasice: tabel de autoverificare

31. Completați singuri coloanele, apoi comparați cu tabelul L10.1. Toate intervalele sunt calculate prin metoda logaritmică.

Tabelul L10.1. Rezultate de autoverificare pentru problemele clasice și pentru mg2026.

Problema (a, b, c, d)	Design	Măsura principală (IC 95%)	OR (IC 95%)	Test	Concluzie
Fumat-BPOC (50, 250, 70, 830)	cohortă	RR 2,14 (1,53-3,01)	2,37 (1,61-3,50)	χ^2 19,75, p < 0,001	factor de risc
Fumat-BPOC mic (5, 25, 7, 83)	cohortă	RR 2,14 (0,73-6,25)	2,37 (0,69-8,13)	Fisher p = 0,17	nesemnificativ
Monitor-acuitate (90, 110, 150, 650)	transversal	RP 2,40 (1,94-2,96)	3,55 (2,55-4,93)	χ^2 60,44, p < 0,001	factor de risc
Contraceptive-AR (40, 120, 60, 80)	caz-control	doar OR	0,44 (0,27-0,73)	χ^2 10,71, p = 0,001	factor de protecție
ECG-deces CHD (8, 39, 10, 134)	cohortă	RR 2,45 (1,03-5,85)	2,75 (1,02-7,44)	Fisher p = 0,0485	la limită
Pulberi (30, 50, 30, 210)	transversal	RP 3,00 (1,94-4,65)	4,20 (2,32-7,60)	χ^2 24,62, p < 0,001	factor de risc
Vârstă, neexpuși (30, 210, 12, 188)	transversal	RP 2,08 (1,10-3,96)	2,24 (1,11-4,50)	χ^2 5,34, p = 0,021	factor de risc
Obezitate-diabet mg2026 (18, 36, 11, 95)	transversal	RP 3,21 (1,64-6,31)	4,32 (1,86-10,03)	χ^2 12,70, p < 0,001	factor de risc
Brațul C față de A mg2026 (16, 38, 6, 47)	randomizat	RR 2,62 (1,11-6,17)	3,30	χ^2 5,49, p = 0,019	intervenție eficace

32. Pentru problema monitor-acuitate, răspundeți și la întrebarea clasică "care este șansa (odds) de acuitate scăzută la cei care lucrează mai puțin de 8 ore pe zi": 150 / 650 = 0,231. Pentru problema contraceptivelor, explicați în două fraze de ce nu se calculează RR.

L10.9 Interpretarea unui studiu observațional

33. Alegeți unul dintre articolele colegilor discutate în capitolul 10 (de exemplu, cohorta de 110 paciente cu RR = 2,75, IC 95% 1,34-5,64 [114] sau studiul caz-control potrivit 1:2 pe

435 de pacienți [116]) și completați o fișă de lectură cu următoarele rubrici: designul și argumentul din text care îl identifică; expunerea și rezultatul; măsura de asociere și dacă este potrivită designului; intervalul de încredere și interpretarea lui; factorii de confuzie luați în calcul (dacă rezumatul îi menționează); ce criterii Bradford Hill sunt susținute; ce ar mai trebui raportat conform STROBE [25].

34. Discutați în grupă: un OR ajustat de 7,20 (IC 95% 4,26-12,17) obținut într-un studiu transversal în care rezultatul are o prevalență de 24,1% [4] poate fi tradus prin "risc de 7 ori mai mare"? Răspunsul se găsește în tabelul 10.4 din capitol.

Exerciții suplimentare

1. Fișierul Probleme BPOC (fără rezultate numerice date). Pacienții cu vârsta de peste 50 de ani au un risc mai mare de a prezenta dispnee la începutul studiului (DISPN = 1), comparativ cu cei sub 50 de ani? Creați variabila grupei de vârstă, construiți tabelul 2 x 2, precizați designul (datele de la includere sunt transversale), calculați RP și OR cu IC 95% în Epi Info și în Sheets, aplicați testul potrivit și formulați concluziile în formatul a-g.
2. Fișierul Probleme BPOC (fără rezultate numerice date). Este prezența ralurilor asociată cu grupa de vârstă? Calculați OR și verificați dacă intervalul conține 1. Comparați rezultatul testului chi-pătrat cu cel al testului Fisher și explicați eventualele diferențe.
3. Problemă cu date în enunț, cu rezolvare completă. Într-un sezon gripal, un cabinet de medicină de familie urmărește 1000 de pacienți cu boli cronice: 400 vaccinați antigripal, dintre care 20 fac gripă confirmată, și 600 nevaccinați, dintre care 72 fac gripă. Analizați datele.

Rezolvare. Design: cohortă (observațională, pacienții au ales dacă se vaccinează). Tabelul: a = 20 (vaccinați bolnavi), b = 380, c = 72 (nevaccinați bolnavi), d = 528. Riscurile: $R_1 = \frac{20}{400} = 5{,}0\%$; $R_0 = \frac{72}{600} = 12{,}0\%$. $\text{RR} = \frac{5{,}0}{12{,}0} = 0{,}42$; $\text{ES}[\ln(\text{RR})] = \sqrt{\frac{1}{20} - \frac{1}{400} + \frac{1}{72} - \frac{1}{600}} = \sqrt{0{,}0597} = 0{,}244$; $\text{IC}\,95\% = \exp(-0{,}875 \pm 0{,}479) = 0{,}26 - 0{,}67$. $\text{OR} = \frac{20\times528}{380\times72} = \frac{10\,560}{27\,360} = 0{,}39$(; IC 95% 0,23 − 0,64). χ^2 = 14,08, p < 0,001. Vaccinarea apare ca factor de protecție: riscul de gripă este cu 58% mai mic la vaccinați (eficacitatea vaccinală estimată = 1 - RR = 58%). Diferența absolută de risc este 7,0 puncte procentuale (IC 95% 3,6-10,4), iar $\text{NNT} = \frac{1}{0{,}07} = 14{,}3$, rotunjit la 15 (IC 95% aproximativ 10-28): trebuie vaccinați 15 pacienți pentru a preveni un caz de gripă într-un sezon. Limita principală: studiul este observațional, iar pacienții care aleg să se vaccineze pot avea și alte comportamente protectoare (efectul "vaccinatului sănătos"), deci estimarea poate fi optimistă.

Ce se predă și barem orientativ

Se predă pe Moodle un document care conține: tabelul L10.1 completat cu valorile proprii (din Epi Info sau Sheets) pentru cel puțin cinci probleme; captura ecranului StatCalc pentru fumat-BPOC și pentru varianta mică; foaia Google Sheets din L10.6 (link cu drept de vizualizare); rezultatul JASP pentru obezitate-diabet, cu transformarea log OR în OR; frazele de interpretare din L10.7; fișa de lectură din L10.9 pentru un articol; rezolvarea exercițiului suplimentar 1.

Barem orientativ (10 puncte): tabelele 2 x 2 construite corect și designul identificat, 2 puncte; RR și OR calculate și alese corect în funcție de design, 2 puncte; intervalele de încredere și testul potrivit (chi-pătrat sau Fisher), 2 puncte; măsurile de impact (RA, NNT/NNH, fracția atribuibilă) cu interpretare, 1 punct; foaia de calcul funcțională, 1 punct; fișa de lectură critică, 1 punct; concluzia clinică formulată prudent, cu limitele designului, 1 punct. Se depunctează calculul RR într-un studiu caz-control, folosirea totalului general ca numitor pentru riscul la neexpuși și concluzia "factorul nu are efect" dintr-un rezultat nesemnificativ.

Varianta fără rețea

Toate problemele clasice se pot rezolva cu calculatorul de buzunar, dacă acesta are funcțiile ln și e^x. Pașii sunt: tabelul 2 x 2 și verificarea totalurilor; R1, R0, RR și OR; ES pe scara logaritmică (formulele din capitolul 10, secțiunea 10.7); intervalul pe scara logaritmică și exponențierea capetelor; frecvențele așteptate și $\chi^2 = \frac{\sum(O-E)^2}{E}$, comparat cu valoarea critică 3,84 pentru 1 grad de libertate și α = 0,05; RA, NNT sau NNH și fracțiile atribuibile. Când o frecvență așteptată este sub 5, testul exact Fisher nu se poate face practic de mână; notați că ar fi testul potrivit și bazați concluzia pe intervalul de încredere, menționând că este aproximativ. Comparați la următoarea ședință rezultatele de mână cu cele din Epi Info.

Lucrarea practică 11. Analiza bayesiană: raționamentul probabilistic în diagnostic și screening

Obiectiv

Lucrarea transformă teoria din capitolul 11 în instrumente pe care le puteți folosi și după examen: un calculator bayesian reutilizabil în Google Sheets, un arbore de frecvențe naturale, un tabel de diagnostic diferențial și primele analize bayesiene în JASP. Toate calculele se fac întâi de mână, pe hârtie, apoi în program, ca să vă puteți verifica.

La final, studentul trebuie să poată:

- să calculeze probabilitatea post-test din probabilitatea pre-test, sensibilitate și specificitate, atât prin teorema lui Bayes, cât și prin forma cu șanse;
- să construiască în Google Sheets un calculator bayesian și un arbore de frecvențe naturale care se actualizează automat la schimbarea prevalenței;
- să folosească nomograma Fagan pentru o estimare grafică rapidă;
- să rezolve un tabel de diagnostic diferențial cu trei boli și trei semne clinice;
- să obțină din setul mg2026 tabelul 2 x 2 al scorului FINDRISC și să calculeze probabilitățile post-test pentru diferite contexte de screening;
- să ruleze în JASP un test t bayesian și un tabel de contingență bayesian și să interpreteze factorul BF10.

Instrumente

Google Sheets (configurare regională România: virgula zecimală și separatorul de argumente ";"), JASP 0.98.1 [47], opțional MedCalc 24.0 [50]. Fișierul mg2026_curat.csv de pe Moodle (160 de rânduri). Nomograma Fagan tipărită (figura 11.3 din capitol) și o riglă. Timp orientativ: 100 de minute.

În fișierul mg2026_curat.csv, importat în Google Sheets cu rândul 1 drept antet, coloanele folosite în această lucrare sunt: C (sex), K (imc0), M (tas0), S (glicemie0), W (findrisc) și X (diabet_ogtt, 1 = diabet confirmat prin TTGO, 0 = fără diabet). Datele ocupă rândurile 2-161.

Desfășurare

L11.1 Calculatorul bayesian în Google Sheets (20 de minute)

Deschideți o foaie nouă și numiți-o "Bayes". Construiți calculatorul în coloanele A (etichete) și B (valori), exact ca în tabelul de mai jos. Celulele B1-B3 sunt date de intrare; restul sunt formule.

Tabelul L11.1. Structura calculatorului bayesian.

Celula	Eticheta (coloana A)	Conținutul coloanei B
B1	Probabilitate pre-test	0,181
B2	Sensibilitate	0,759
B3	Specificitate	0,748
B4	RV+	=B2/(1-B3)
B5	RV-	=(1-B2)/B3
B6	Șansa pre-test	=B1/(1-B1)
B7	Șansa post-test (test pozitiv)	=B6*B4
B8	Probabilitate post-test (test pozitiv)	=B7/(1+B7)
B9	Șansa post-test (test negativ)	=B6*B5

Celula	Eticheta (coloana A)	Conținutul coloanei B
B10	Probabilitate post-test (test negativ)	=B9/(1+B9)
B11	VPP prin formula Bayes (verificare)	=B2*B1/(B2*B1+(1-B3)*(1-B1))
B12	VPN prin formula Bayes (verificare)	=B3*(1-B1)/(B3*(1-B1)+(1-B2)*B1)

Formatați B8, B10, B11 și B12 ca procent cu o zecimală (Format > Number > Percent). Cu valorile pentru FINDRISC ≥ 12 din setul mg2026 trebuie să obțineți: RV+ 3,01; RV- 0,32; șansa pre-test 0,221; probabilitatea post-test 40,0% după un scor pozitiv și 6,7% după unul negativ. B11 trebuie să fie egal cu B8 (40,0%), iar B12 cu 1 - B10 (93,3%). Dacă cele două căi nu dau același rezultat, aveți o greșeală într-o formulă.

Acum folosiți calculatorul ca pe un instrument. Schimbați doar B1 și completați tabelul L11.2 (rezultatele corecte sunt cele din tabelul 11.7 al capitolului).

Tabelul L11.2. Probabilități post-test pentru FINDRISC ≥ 12, de completat.

Probabilitate pre-test (B1)	Post-test pozitiv (B8)	Post-test negativ (B10)
0,02	5,8%	0,7%
0,05		
0,10		
0,30		
0,50		

Pentru o tabelare automată, în D1 scrieți "pre-test", în E1 "post +", în F1 "post -". În D2:D7 introduceți 0,02; 0,05; 0,10; 0,181; 0,30; 0,50. În E2 scrieți =D2/(1-D2)*$B$4/(1+D2/(1-D2)*$B$4), iar în F2 =D2/(1-D2)*$B$5/(1+D2/(1-D2)*$B$5), apoi trageți formulele în jos până la rândul 7. Semnul $ fixează referința la RV, ca să nu se deplaseze la copiere. Selectați D1:F7 și inserați un grafic de tip linie (Insert > Chart, Line chart): ați refăcut o porțiune din figura 11.2.

Verificare de mână (obligatorie, pe caiet): pentru pre-test 0,10, șansa este 0,111; 0,111 x 3,01 = 0,334; 0,334/1,334 = 25,1%.

L11.2 Arborele frecvențelor naturale (15 minute)

În aceeași foaie, începând cu celula A15, construiți un arbore de frecvențe naturale pentru o populație de 1000 de persoane, care folosește B1-B3.

Tabelul L11.3. Arborele frecvențelor naturale în Google Sheets.

Celula	Eticheta	Formula
B15	Populație	1000
B16	Bolnavi	=B15*B1
B17	Nebolnavi	=B15-B16
B18	Bolnavi cu test pozitiv (AP)	=B16*B2
B19	Bolnavi cu test negativ (FN)	=B16-B18
B20	Nebolnavi cu test pozitiv (FP)	=B17*(1-B3)
B21	Nebolnavi cu test negativ (AN)	=B17-B20
B22	Total teste pozitive	=B18+B20
B23	VPP din arbore	=B18/B22
B24	VPN din arbore	=B21/(B19+B21)

Pasul 1. Setați B1 = 0,10, B2 = 0,80, B3 = 0,90. Trebuie să obțineți 100 de bolnavi, 80 AP, 20 FN, 900 de nebolnavi, 90 FP, 810 AN, 170 de teste pozitive, VPP 47,1% și VPN 97,6%, adică exact figura 11.1 din capitol.

Pasul 2. Păstrați Se și Sp și schimbați prevalența la 0,01 și apoi la 0,50. Notați VPP și VPN. La 1% ar trebui să obțineți 8 AP și 99 FP (VPP 7,5%); la 50%, 400 AP și 50 FP (VPP 88,9%).

Pasul 3. Scrieți pe caiet, în două fraze, cum i-ați explica unui pacient rezultatul pozitiv în fiecare dintre cele trei situații, folosind numere întregi ("din 170 de persoane cu test pozitiv, 80 au boala"). Exercițiul nu are un singur răspuns corect, dar va fi discutat la final.

L11.3 Nomograma Fagan (10 minute)

Folosiți nomograma tipărită (figura 11.3 din capitol) și o riglă. Pentru fiecare situație, trageți o linie din punctul pre-test, prin valoarea RV, citiți probabilitatea post-test și apoi verificați-o cu calculatorul din L11.1 (introduceți în B4 sau B5 valoarea RV direct, în locul formulei, sau calculați separat).

Tabelul L11.4. Exerciții cu nomograma Fagan.

Situație	Pre-test	RV	Post-test (nomogramă)	Post-test (calcul)
FINDRISC ≥ 12, eșantion mg2026	18,1%	3,01	aproximativ 40%	40,0%
FINDRISC < 12, eșantion mg2026	18,1%	0,32		6,7%
Test de confirmare foarte specific	10%	10		52,6%
Test de excludere, suspiciune mare	50%	0,1		9,1%
Elastografie peste pragul de confirmare a cirozei [3]	28,1%	19,7		88,5%

Diferențele dintre citire și calcul ar trebui să fie de câteva procente. Observați pe nomogramă de ce același RV mută mult probabilitatea în zona de mijloc și puțin la extreme.

L11.4 Tabelul de diagnostic diferențial cu trei boli (20 de minute)

Scenariu (date fictive, plauzibile). O pacientă de 34 de ani are la hemoleucogramă anemie microcitară ușoară. Medicul de familie ia în considerare trei diagnostice: anemia feriprivă (probabilitate a priori 0,60), talasemia minoră (0,15) și anemia din bolile cronice (0,25). Trei rezultate sunt disponibile: feritina sub 30 ng/mL (F), lărgimea distribuției eritrocitare (RDW) crescută (R) și indicele Mentzer (VEM/număr de eritrocite) sub 13 (M). Frecvența fiecărui rezultat în fiecare boală este dată mai jos.

Tabelul L11.5. Probabilitățile a priori și frecvențele rezultatelor (date fictive).

Diagnostic	p(D)	$p(F/D)$	$p(R/D)$	$p(M/D)$
Anemie feriprivă	0,60	0,90	0,80	0,15
Talasemie minoră	0,15	0,05	0,20	0,80
Anemia bolilor cronice	0,25	0,10	0,40	0,20

Pacienta are feritina normală (F-), RDW normal (R-) și indicele Mentzer sub 13 (M+).

Pasul 1. Pe caiet, calculați verosimilitatea combinației observate în fiecare boală (pentru un rezultat absent folosiți 1 - p): feriprivă 0,10 x 0,20 x 0,15 = 0,003; talasemie 0,95 x 0,80 x 0,80 = 0,608; boli cronice 0,90 x 0,60 x 0,20 = 0,108.

Pasul 2. În Google Sheets, într-o foaie nouă "Diferential", introduceți tabelul L11.5 în A1:E4 (antet pe rândul 1, bolile pe rândurile 2-4). În G1, H1, I1 scrieți rezultatele observate codificate 1 (prezent) sau 0 (absent): G1 = 0 (F), H1 = 0 (R), I1 = 1 (M).

Pasul 3. În F2 scrieți formula verosimilității combinate:

=IF($G$1=1;C2;1-C2)*IF($H$1=1;D2;1-D2)*IF($I$1=1;E2;1-E2)

În G2 produsul cu probabilitatea a priori: =B2*F2. În H2 probabilitatea posterioară: =G2/SUM($G$2:$G$4). Copiați F2:H2 în jos până la rândul 4.

Rezultate de verificare: produse 0,0018; 0,0912; 0,0270; suma 0,1200; probabilități posterioare: anemie feriprivă 1,5%, talasemie minoră 76,0%, anemia bolilor cronice 22,5%.

Pasul 4. Schimbați codificarea în G1:I1 pentru o pacientă cu feritină scăzută, RDW crescut și indice Mentzer peste 13 (1; 1; 0). Ce diagnostic devine cel mai probabil? Notați rezultatul. (Ar trebui să obțineți aproximativ 97,8% pentru anemia feriprivă.)

Pasul 5. Discuție: ce s-ar întâmpla dacă pacienta ar avea, de fapt, o anemie sideroblastică, care nu este în tabel? Ce presupunere a modelului ar fi încălcată dacă feritina și RDW ar fi corelate în interiorul aceleiași boli?

Tabelul obținut este un model "Bayes naiv" funcțional. Îl puteți reutiliza pentru orice problemă de diagnostic diferențial, schimbând numele bolilor, probabilitățile a priori și frecvențele semnelor; problema rezolvată 11.5 (durerea toracică) și exercițiul 7 din capitolul 11 (cefaleea) se pot verifica imediat.

L11.5 Screening cu FINDRISC pe setul mg2026 (15 minute)

Importați mg2026_curat.csv într-o foaie nouă "Date" (File > Import > Upload, cu "Replace current sheet" sau "Insert new sheet"). Verificați că W conține scorul FINDRISC și X diagnosticul TTGO.

Pasul 1. Obțineți tabelul 2 x 2 pentru pragul ≥ 12, cu funcția COUNTIFS, într-o foaie "Screening":

- AP: =COUNTIFS(Date!X2:X161;1;Date!W2:W161;">=12") trebuie să dea 22
- FN: =COUNTIFS(Date!X2:X161;1;Date!W2:W161;"<12") trebuie să dea 7
- FP: =COUNTIFS(Date!X2:X161;0;Date!W2:W161;">=12") trebuie să dea 33
- AN: =COUNTIFS(Date!X2:X161;0;Date!W2:W161;"<12") trebuie să dea 98

Pasul 2. Calculați prevalența (29/160 = 18,1%), Se (75,9%), Sp (74,8%), VPP (40,0%) și VPN (93,3%), apoi copiați Se și Sp în calculatorul din L11.1.

Pasul 3. Faceți pragul variabil: scrieți pragul într-o celulă (de exemplu B1 = 12) și înlocuiți ">=12" cu ">="&B1 și "<12" cu "<"&B1. Încercați pragurile 10, 14 și 16. Trebuie să regăsiți valorile din tabelul 12.5 al capitolului 12 (de exemplu la ≥ 14: AP 17, FP 16, FN 12, AN 115).

Pasul 4. Pentru pragul ≥ 12, calculați probabilitatea post-test pentru trei pacienți fictivi: o femeie de 30 de ani fără factori de risc (pre-test 2%), un bărbat de 55 de ani supraponderal (10%) și o pacientă obeză cu prediabet cunoscut (30%). Formulați recomandarea pentru fiecare: liniștire, glicemie a jeun sau TTGO.

Pasul 5 (testare secvențială). Pentru un pacient cu FINDRISC ≥ 12 și glicemie a jeun ≥ 110 mg/dL, înmulțiți succesiv rapoartele de verosimilitate (FINDRISC RV+ 3,01; glicemie RV+ 2,78). Obțineți o probabilitate post-test de aproximativ 64,9%. Apoi numărați direct în date: =COUNTIFS(Date!X2:X161;1;Date!W2:W161;">=12";Date!S2:S161;">=110") dă 20, iar aceeași formulă cu X = 0 dă 19. Probabilitatea observată este 20/39 = 51,3%. Explicați în două fraze diferența (indicație: secțiunea 11.5.2).

L11.6 Analize bayesiene în JASP (20 de minute)

Deschideți mg2026_curat.csv în JASP. Verificați tipul variabilelor: sex nominal, tas0 și imc0 continue (scale), diabet_ogtt nominal.

Pasul 1. Testul t bayesian pentru eșantioane independente. Din meniul T-Tests, la secțiunea Bayesian, alegeți Independent Samples T-Test. Mutați tas0 la Variables și sex la Grouping Variable. Lăsați ipoteza alternativă bilaterală și priorul implicit (Cauchy, lățimea 0,707). Bifați Prior and posterior și Bayes factor robustness check la Plots. Rezultat așteptat: BF10 aproximativ

0,18 (dacă alegeți afișarea BF01, aproximativ 5,7). Interpretare: dovezi moderate pentru absența unei diferențe a TAS între sexe. Comparați cu testul t frecventist din lucrarea practică despre testele t (p = 0,81): valoarea p nu permitea afirmarea absenței unei diferențe, factorul Bayes o susține moderat.

Pasul 2. Graficul de robustețe arată cum se schimbă BF10 când lățimea priorului variază. Dacă concluzia (dovezi pentru H0) rămâne aceeași pentru toate lățimile rezonabile, rezultatul este robust. Notați cum se modifică BF10 pentru priorul "wide" și "ultrawide".

Pasul 3. Tabel de contingență bayesian. Creați întâi variabila obez: în JASP, folosiți butonul "+" de lângă numele coloanelor pentru a adăuga o coloană calculată (Computed column), cu formula imc0 >= 30, de tip nominal. (Alternativ, creați coloana în Google Sheets cu =IF(K2>=30;1;0) și reimportați fișierul.) Apoi, din meniul Frequencies, secțiunea Bayesian, alegeți Contingency Tables. Puneți obez la Rows și diabet_ogtt la Columns. La Sampling alegeți Joint multinomial, cu concentrația priorului implicită (1). Rezultat așteptat: BF10 aproximativ 85, adică dovezi foarte puternice pentru asocierea dintre obezitate și diabet. Dacă alegeți Independent multinomial, rows fixed, BF10 scade la aproximativ 64; concluzia rămâne aceeași. Notați că tipul de eșantionare este o alegere de design care trebuie justificată și raportată.

Pasul 4 (opțional). Din meniul Frequencies, secțiunea Bayesian, alegeți Binomial Test pentru variabila diabet_ogtt, cu priorul implicit Beta(1; 1). Bifați graficul Prior and posterior. Intervalul credibil central de 95% pentru proporția de diabet trebuie să fie aproximativ 12,9-24,8%, ca în figura 11.4 a capitolului. Schimbați priorul în Beta(10; 40) și observați cât de puțin se modifică intervalul (aproximativ 13,6-24,1%).

Pasul 5. Raportați rezultatele în formatul temelor disciplinei: (a) ipoteza de cercetare; (b) ipotezele statistice; (c) metoda și justificarea (de ce un test bayesian); (d) programul și setările (priorul, tipul de eșantionare); (e) rezultatul cu graficul; (f) concluzia statistică, folosind scala din tabelul 11.8 al capitolului; (g) concluzia de cercetare în cuvinte.

Exerciții suplimentare

1. Pe fișierul BPOC de pe Moodle, considerați dispneea (DISPN) drept "boală" și ralurile (RAL) drept "simptom". Construiți tabelul boală-simptom cu COUNTIFS și calculați p(B+), p(S+ / B+), p(S+ / B-) și p(B+ / S+). Verificați p(B+ / S+) cu teorema lui Bayes. Comentați dacă ralurile ar fi un "test" util pentru dispnee în acest eșantion.
2. Pe fișierul Date_HbA1c de pe Moodle, alegeți un prag clinic pentru HbA1c (de exemplu 6,5%) și o variabilă care poate servi drept referință, dacă există una în fișier. Construiți tabelul 2 x 2, calculați Se, Sp, RV+ și RV- și folosiți calculatorul din L11.1 pentru trei probabilități pre-test diferite. Discutați limitele referinței alese.
3. Problemă cu date (rezolvare completă). Un test rapid pentru antigenul streptococic are Se 86% și Sp 95%. La un copil cu faringită, medicul estimează o probabilitate pre-test de 40%. Calculați probabilitatea post-test după un test pozitiv și după unul negativ.

Rezolvare. $RV^{+} = \frac{0,86}{0,05} = 17,2$; $RV^{-} = \frac{0,14}{0,95} = 0,147$. Șansa pre-test = 0,40/0,60 = 0,667. Test pozitiv: 0,667 x 17,2 = 11,47, probabilitatea 11,47/12,47 = 92,0%. Test negativ: 0,667 x 0,147 = 0,098, probabilitatea 0,098/1,098 = 8,9%. Cu un prag de tratament de 25% (problema rezolvată 11.6), testul pozitiv justifică tratamentul, iar cel negativ coboară probabilitatea sub prag.

4. Problemă cu date (rezolvare completă). Folosind tabelul de diagnostic diferențial din L11.4, o pacientă are feritina scăzută (F+), RDW normal (R-) și indicele Mentzer peste 13 (M-). Calculați probabilitățile posterioare.

Rezolvare. Verosimilități: feriprivă 0,90 x 0,20 x 0,85 = 0,153; talasemie 0,05 x 0,80 x 0,20 = 0,008; boli cronice 0,10 x 0,60 x 0,80 = 0,048. Produse cu a priori: 0,60 x 0,153 = 0,0918; 0,15 x

0,008 = 0,0012; 0,25 x 0,048 = 0,0120. Suma 0,1050. Probabilități posterioare: feriprivă 87,4%, talasemie 1,1%, boli cronice 11,4%.

Ce se predă și barem orientativ

Se predă pe Moodle un singur fișier Google Sheets (link cu drept de vizualizare) sau exportat ca.xlsx, cu foile "Bayes", "Diferential" și "Screening", plus un document scurt (maximum o pagină) cu rezultatele din JASP (capturi ale tabelelor cu BF10) și interpretarea în formatul a-g.

Barem orientativ (10 puncte): calculatorul bayesian corect, cu verificare încrucișată B8 = B11 (2 puncte); arborele de frecvențe naturale funcțional pentru trei prevalențe (1,5 puncte); exercițiile cu nomograma, cu citirea și calculul (1 punct); tabelul de diagnostic diferențial cu formulele corecte și cele două scenarii (2 puncte); tabelul 2 x 2 al FINDRISC obținut cu COUNTIFS și prag variabil, plus explicația testării secvențiale (1,5 puncte); analizele JASP raportate corect, cu interpretarea BF10 pe scala din curs (2 puncte).

Varianta fără rețea

Dacă nu aveți acces la Google Sheets sau la JASP, toate calculele din L11.1-L11.5 se pot face cu un calculator de buzunar și pe hârtie, folosind tabelele din această lucrare ca șabloane: transformați probabilitatea în șansă, înmulțiți cu RV, transformați înapoi. Arborele de frecvențe naturale se desenează de mână pentru 1000 de persoane, iar nomograma Fagan tipărită înlocuiește calculatorul. Pentru L11.5, tabelul 2 x 2 al FINDRISC este dat în capitolul 12 (tabelul 12.3), iar pentru L11.6 folosiți rezultatele de referință date mai sus (BF10 0,18 pentru TAS și sex; aproximativ 85 pentru obezitate și diabet) și scrieți doar interpretarea. Aceeași logică se poate aplica în LibreOffice Calc sau Microsoft Excel, cu aceleași formule (în configurarea cu punct zecimal, separatorul de argumente devine ",").

Lucrarea practică 12. Performanța testelor diagnostice: sensibilitate, specificitate, valori predictive și curba ROC

Obiectiv

În această lucrare evaluați un test diagnostic de la cap la coadă: construiți tabelul 2 x 2, calculați toți indicatorii (întâi de mână, apoi în Google Sheets și MedCalc), alegeți un prag, construiți curba ROC și comparați două teste. Testul principal este scorul FINDRISC din setul mg2026, cu TTGO ca standard de referință; testul de laborator analizat în final este glicemia a jeun.

La final, studentul trebuie să poată:

- să calculeze dintr-un tabel 2 x 2 prevalența, Se, Sp, VPP, VPN, acuratețea, rata erorii (b + c)/N, indicele Youden, RV+, RV- și DOR, cu intervale de încredere;
- să folosească procedura de evaluare a unui test diagnostic din MedCalc și să recalculeze valorile predictive pentru o altă prevalență;
- să construiască în Google Sheets un tabel de praguri și curba ROC corespunzătoare și să estimeze AUC prin metoda trapezelor;
- să ruleze în MedCalc analiza ROC (AUC cu IC 95%, tabelul criteriilor, indicele Youden) și compararea a două curbe ROC;
- să obțină aceleași rezultate, acolo unde versiunea permite, în JASP și să citească un tabel 2 x 2 în Epi Info;
- să redacteze evaluarea performanței unui test de laborator în formatul a-g al temelor disciplinei.

Instrumente

Google Sheets (configurare regională România: virgula zecimală, separatorul ";"), MedCalc 24.0 [50], JASP 0.98.1 [47], Epi Info 7.2.7 [51]. Fișierul mg2026_curat.csv de pe Moodle. Timp orientativ: 100 de minute.

Coloanele din mg2026_curat.csv folosite aici: S (glicemie0, mg/dL), V (hba1c0, %), W (findrisc, 0-26), X (diabet_ogtt: 1 diabet confirmat prin TTGO, 0 fără diabet); datele ocupă rândurile 2-161. În setul mg2026 sunt 29 de persoane cu diabet și 131 fără diabet.

Desfășurare

L12.1 Indicatorii unui tabel 2 x 2 în Google Sheets (15 minute)

Pornim de la problema clasică a cursului: un test aplicat la 440 de pacienți, cu a = 150, b = 30, c = 4, d = 256 (bolnavii pe primul rând, testul pozitiv pe prima coloană).

Pasul 1. Într-o foaie nouă "2x2", introduceți tabelul: B1 "T+", C1 "T-", A2 "B+", A3 "B-", apoi B2 = 150, C2 = 30, B3 = 4, C3 = 256. În D2 scrieți =B2+C2, în D3 =B3+C3, în B4 =B2+B3, în C4 =C2+C3, în D4 =D2+D3.

Pasul 2. Începând cu A6, construiți lista indicatorilor (eticheta în coloana A, formula în coloana B):

Tabelul L12.1. Formulele indicatorilor în Google Sheets.

Indicator	Formula în coloana B	Rezultat așteptat
Prevalența	=D2/D4	40,9%
Sensibilitatea	=B2/D2	83,3%
Specificitatea	=C3/D3	98,5%
VPP	=B2/B4	97,4%
VPN	=C3/C4	89,5%
Acuratețea	=(B2+C3)/D4	92,3%
Rata erorii	=(C2+B3)/D4	7,7%
Indicele Youden	=B2/D2+C3/D3-1	0,818
RV+	=(B2/D2)/(1-C3/D3)	54,2
RV-	=(1-B2/D2)/(C3/D3)	0,17
DOR	=(B2*C3)/(C2*B3)	320

Atenție la rata erorii: formula conține celulele C2 (fals negativi, b) și B3 (fals pozitivi, c). O formulă care folosește C3 (adevărat negativi) este greșită; verificați că acuratețea plus rata erorii dau 1.

Pasul 3. Intervalul de încredere Wilson pentru sensibilitate. Dacă numărătorul este în B2 și numitorul în D2, limita inferioară este:

=(B2/D2+1,96^2/(2*D2)-1,96*SQRT(B2/D2*(1-B2/D2)/D2+1,96^2/(4*D2^2)))/(1+1,96^2/D2)

iar limita superioară se obține înlocuind "-1,96*SQRT" cu "+1,96*SQRT". Pentru Se = 150/180 trebuie să obțineți aproximativ 77,2-88,1%. Adaptați formula pentru specificitate (C3 și D3): aproximativ 96,1-99,4%.

Pasul 4. Valorile predictive la altă prevalență. Într-o celulă (de exemplu F2) scrieți o prevalență nouă, 0,10, și calculați VPP cu teorema lui Bayes: =(B2/D2)*F2/((B2/D2)*F2+(1-C3/D3)*(1-F2)), iar VPN cu =(C3/D3)*(1-F2)/((C3/D3)*(1-F2)+(1-B2/D2)*F2). La prevalența de 10% trebuie să obțineți VPP 85,8% și VPN 98,2%. Comparați cu valorile de la prevalența de 40,9%.

Pasul 5. Refaceți pe caiet, fără program, cele două probleme de reconstituire din capitolul 12 (problemele rezolvate 12.2 și 12.3), apoi verificați-le introducând tabelele obținute în foaia "2x2". Toate formulele se recalculează automat.

L12.2 Evaluarea testului în MedCalc (10 minute)

MedCalc are o procedură dedicată pentru tabelul 2 x 2 al unui test diagnostic, în meniul Tests, opțiunea Diagnostic test (2 x 2 table). Dacă versiunea instalată organizează altfel meniurile, căutați procedura după nume; aceeași procedură există gratuit online, pe site-ul MedCalc, sub numele "Diagnostic test evaluation calculator".

Pasul 1. Atenție la orientarea tabelului: în MedCalc boala este pe coloane (Disease present / Disease absent), iar testul pe rânduri (Test positive / Test negative). Introduceți: Test positive și Disease present = 150; Test positive și Disease absent = 4; Test negative și Disease present = 30; Test negative și Disease absent = 256. Dacă introduceți 30 în locul lui 4, toți indicatorii vor fi greșiți, deși programul nu vă va avertiza.

Pasul 2. Citiți rezultatele. Programul afișează sensibilitatea, specificitatea, RV+, RV-, prevalența, VPP, VPN și acuratețea, fiecare cu IC 95%. Pentru Se și Sp, MedCalc folosește intervale binomiale exacte (Clopper-Pearson), puțin mai largi decât cele Wilson: Se 83,3% (IC

95% 77,1-88,5), Sp 98,5% (IC 95% 96,1-99,6), acuratețea 92,3% (IC 95% 89,4-94,6), RV+ 54,2 (IC 95% aproximativ 20,4-143,6), RV- 0,17 (IC 95% aproximativ 0,12-0,23).

Pasul 3. În câmpul pentru prevalență (Disease prevalence), introduceți 10%. VPP și VPN se recalculează pentru noua prevalență; trebuie să regăsiți valorile din L12.1, pasul 4. Notați cât de larg este intervalul de încredere al RV+: cu doar 4 pozitive false, estimarea este foarte imprecisă.

L12.3 Tabelul pragurilor și curba ROC pentru FINDRISC în Google Sheets (20 de minute)

Importați mg2026_curat.csv într-o foaie "Date". Într-o foaie nouă "ROC", construiți tabelul pragurilor.

Pasul 1. În A1:I1 scrieți antetul: Prag, AP, FP, FN, AN, Se, Sp, 1-Sp, J. În A2:A8 scrieți pragurile 16, 15, 14, 12, 11, 10, 8 (ordonate de la cel mai strict la cel mai permisiv).

Pasul 2. În B2 (AP): =COUNTIFS(Date!$X$2:$X$161;1;Date!$W$2:$W$161;">="&A2). În C2 (FP): =COUNTIFS(Date!$X$2:$X$161;0;Date!$W$2:$W$161;">="&A2). În D2 (FN): =29-B2. În E2 (AN): =131-C2. În F2: =B2/29. În G2: =E2/131. În H2: =1-G2. În I2: =F2+G2-1. Copiați rândul 2 până la rândul 8.

Pasul 3. Verificați rezultatele cu tabelul 12.5 din capitol. La pragul ≥ 12 trebuie să obțineți AP 22, FP 33, FN 7, AN 98, Se 0,759, Sp 0,748; la ≥ 11, J maxim = 0,595.

Pasul 4. Adăugați deasupra pragurilor punctul (0; 0) și dedesubt punctul (1; 1) într-o zonă separată cu două coloane (1-Sp și Se), de exemplu K1:L9, în ordinea 1-Sp crescătoare. Selectați zona și inserați un grafic de tip Scatter chart; în Chart editor > Customize > Series bifați "Line" ca să unească punctele. Adăugați seria diagonalei, cu punctele (0; 0) și (1; 1). Titlul axelor: "1 - specificitate" și "sensibilitate". Graficul trebuie să semene cu figura 12.2 din capitol.

Pasul 5. Calculați AUC prin metoda trapezelor. Dacă 1-Sp este în K2:K10 și Se în L2:L10 (cu (0; 0) pe primul rând și (1; 1) pe ultimul), în M3 scrieți =(K3-K2)*(L3+L2)/2, copiați până la M10, apoi =SUM(M3:M10). Cu cele șapte praguri de mai sus veți obține o arie de aproximativ 0,845; cu toate pragurile posibile ale scorului (0-25), aria devine 0,852. Explicați pe caiet de ce aria cu puține praguri este mai mică.

Pasul 6 (opțional). Repetați pentru glicemia a jeun (coloana S), cu pragurile 126, 120, 115, 110, 105, 100. Rezultate de control: la ≥ 110, AP 24, FP 39 (Se 82,8%, Sp 70,2%); aria trapezelor cu aceste șase praguri este 0,836 (problema rezolvată 12.7).

L12.4 Analiza ROC și compararea curbelor în MedCalc (20 de minute)

Deschideți mg2026_curat.csv în MedCalc (File > Open, sau importați fișierul CSV și verificați că separatorul zecimal a fost recunoscut corect).

Pasul 1. Din meniul Statistics, alegeți ROC curves > ROC curve analysis. La Variable selectați findrisc, la Classification variable selectați diabet_ogtt (codificare 1 = boală prezentă). La Methodology puteți alege metoda DeLong sau Hanley și McNeil pentru eroarea standard. Bifați opțiunea de afișare a valorilor criteriului (Criterion values and coordinates of the ROC curve) și graficul ROC.

Pasul 2. Rezultate așteptate: AUC 0,852; testul AUC = 0,5 cu p < 0,001. Intervalul de încredere depinde de metoda aleasă: 0,762-0,943 cu formula Hanley-McNeil, aproximativ 0,782-0,923 cu eroarea standard DeLong; MedCalc poate afișa implicit un interval "binomial exact", ușor diferit. Notați în raport ce metodă a folosit programul.

Pasul 3. Indicele Youden. MedCalc afișează criteriul asociat indicelui Youden maxim: J = 0,595, cu criteriul "> 10". Pentru un scor cu valori întregi, "> 10" înseamnă exact "≥ 11", pragul

din capitol (Se 93,1%, Sp 66,4%). Această notație îi încurcă pe mulți studenți la examen: citiți întotdeauna semnul din fața criteriului.

Pasul 4. În tabelul criteriilor, găsiți rândul "> 11" (adică ≥ 12) și verificați Se 75,86% și Sp 74,81%, cu +LR 3,01 și -LR 0,32. Tabelul criteriilor este, de fapt, tabelul pragurilor pe care l-ați construit manual în L12.3.

Pasul 5. Compararea a două curbe ROC. Din meniul Statistics, alegeți ROC curves > Comparison of ROC curves. Selectați variabilele findrisc și glicemie0, cu clasificarea diabet_ogtt. Rezultate așteptate: AUC FINDRISC 0,852, AUC glicemie 0,844; diferența 0,008 (IC 95% aproximativ -0,060 la 0,077), z aproximativ 0,24, p aproximativ 0,81 (metoda DeLong pentru curbe corelate, pentru că ambele teste au fost aplicate acelorași pacienți). Concluzie: nu s-a găsit o diferență semnificativă între capacitățile de discriminare ale celor două teste. Graficul arată două curbe care se intersectează.

Pasul 6 (opțional). Adăugați în comparație hba1c0. AUC pentru HbA1c este 0,764; comparați vizual și numeric cu celelalte două teste.

L12.5 Curba ROC în JASP (10 minute)

JASP nu are un meniu separat numit "ROC" în modulele de bază, dar curba ROC se poate obține prin regresia logistică, iar în versiunile recente și prin modulul Machine Learning. Opțiunile disponibile diferă între versiuni; descrierea de mai jos corespunde organizării obișnuite, iar dacă nu găsiți o opțiune, folosiți MedCalc (L12.4) sau calculul din Sheets (L12.3).

Pasul 1. Deschideți mg2026_curat.csv. Setați diabet_ogtt ca variabilă nominală.

Pasul 2. Din meniul Regression, alegeți Logistic Regression. Puneți diabet_ogtt la Dependent Variable și findrisc la Covariates.

Pasul 3. În secțiunea de statistici sau de performanță a modelului (Statistics / Performance Diagnostics), bifați Confusion matrix și, la Performance metrics, AUC, Sensitivity, Specificity, Precision și F-measure. În secțiunea de grafice de performanță, bifați ROC plot.

Pasul 4. Interpretare. AUC raportată pentru modelul cu un singur predictor este 0,852, aceeași cu AUC a scorului FINDRISC, pentru că modelul logistic transformă scorul monoton într-o probabilitate, fără să schimbe ordinea pacienților. Matricea de confuzie și sensibilitatea afișate de JASP folosesc însă, implicit, pragul de 0,5 pentru probabilitatea prezisă, nu un prag al scorului; cu o prevalență de 18,1%, foarte puțini pacienți depășesc o probabilitate prezisă de 0,5, deci sensibilitatea afișată va fi mică și specificitatea mare. Acesta este un exemplu bun de "aceeași curbă ROC, alt punct de operare". Dacă versiunea permite modificarea pragului (cutoff), încercați 0,18 și comparați cu tabelul pragurilor.

Pasul 5. Alternativa din modulul Machine Learning (dacă este activat, din butonul "+" al barei de module): o procedură de clasificare raportează și matricea de confuzie, AUC și curba ROC, de obicei pe un set de testare separat. Rezultatele vor diferi de cele de mai sus pentru că evaluarea se face pe o parte din date; este, de fapt, o validare internă simplă.

L12.6 Tabelul 2 x 2 în Epi Info (5 minute)

Epi Info nu are o procedură dedicată testelor diagnostice în forma pe care o folosim, dar tabelul 2 x 2 se obține ușor, iar procentele pe rânduri și pe coloane conțin indicatorii căutați.

Pasul 1. În Classic Analysis sau în Visual Dashboard, importați fișierul și creați variabila findrisc12 (1 dacă findrisc ≥ 12, altfel 0), de exemplu cu comanda DEFINE și RECODE sau cu o variabilă calculată.

Pasul 2. Rulați un tabel (Tables) cu findrisc12 ca expunere (pe rânduri) și diabet_ogtt ca rezultat (pe coloane). Epi Info afișează frecvențele, procentele pe rând și pe coloană, testul chi-pătrat și măsurile de asociere (OR, RR).

Pasul 3. Citiți: pe rândul findrisc12 = 1, procentul pe rând al celor cu diabet este VPP $\left(\frac{22}{55} = 40{,}0\%\right)$; pe rândul findrisc12 = 0, procentul pe rând al celor fără diabet este VPN $\left(\frac{98}{105} = 93{,}3\%\right)$. Pe coloana diabet_ogtt = 1, procentul pe coloană al celor cu findrisc12 = 1 este sensibilitatea (22/29 = 75,9%); pe coloana diabet_ogtt = 0, procentul celor cu findrisc12 = 0 este specificitatea (98/131 = 74,8%). OR afișat (9,33) este raportul șanselor de diagnostic.

Pasul 4. Notați că semnificația testului chi-pătrat arată doar că testul și boala sunt asociate, nu cât de bun este testul. Un test cu AUC 0,6 poate avea p < 0,001 într-un eșantion mare.

L12.7 Analiza performanței unui test de laborator: glicemia a jeun pentru diabetul confirmat prin TTGO (15 minute)

Folosiți rezultatele obținute până acum (sau rulați aceleași proceduri pentru glicemie0) și redactați o analiză completă, în formatul temelor disciplinei.

(a) Ipoteza de cercetare: glicemia a jeun discriminează între adulții cu și fără diabet confirmat prin TTGO, în populația cu risc cardiometabolic din medicina de familie.

(b) Ipotezele statistice: H0: AUC = 0,5 (testul nu discriminează); H1: AUC ≠ 0,5.

(c) Metoda: analiza ROC, cu TTGO ca standard de referință; IC 95% pentru AUC; alegerea pragului prin indicele Youden și compararea cu pragul clinic de 126 mg/dL.

(d) Programul: MedCalc (ROC curve analysis), verificat în Google Sheets.

(e) Rezultatele așteptate: AUC 0,844 (IC 95% 0,752-0,937, Hanley-McNeil), p < 0,001. Pragul Youden ≥ 110 mg/dL (MedCalc afișează "> 109" sau valoarea imediat inferioară prezentă în date): Se 82,8%, Sp 70,2%, VPP 38,1%, VPN 94,8%, J = 0,530. La pragul clinic ≥ 126 mg/dL: Se 37,9%, Sp 94,7%, VPP 61,1%, VPN 87,3%, RV+ aproximativ 7,1, RV- aproximativ 0,66. Includeți graficul ROC.

(f) Concluzia statistică: se respinge H0; glicemia a jeun are o capacitate de discriminare foarte bună pe scala din curs (0,81-0,90), dar limita inferioară a intervalului de încredere (0,752) corespunde unei performanțe "bune".

(g) Concluzia de cercetare: în această populație, glicemia a jeun este un test util de triaj la pragul de aproximativ 110 mg/dL (ratează puțini diabetici, dar trimite la TTGO aproximativ 4 din 10 pacienți fără diabet). Pragul de 126 mg/dL confirmă diabetul cu specificitate mare, dar ratează aproape două treimi dintre cazurile identificate prin TTGO, deci un rezultat sub 126 mg/dL nu exclude diabetul. Performanța glicemiei nu diferă semnificativ de cea a scorului FINDRISC (p = 0,81).

Adăugați un paragraf de limite: eșantion mic de bolnavi (29), deci intervale largi pentru sensibilitate; prag ales pe aceleași date (supraestimare optimistă); date simulate; STARD 2015 cere și diagrama de flux a participanților [140].

Exerciții suplimentare

1. Pe fișierul Date_HbA1c de pe Moodle, alegeți o variabilă numerică de laborator și o variabilă dihotomică ce poate servi drept referință (dacă există). Construiți în Google Sheets tabelul pragurilor, curba ROC și AUC prin trapeze; verificați în MedCalc. Raportați în formatul a-g.

2. Pe fișierul BPOC de pe Moodle, evaluați cât de bine "prezice" un indice spirometric la începutul studiului (de exemplu IPB1) prezența ralurilor (RAL). Rulați analiza ROC în MedCalc, raportați AUC cu IC 95% și comentați dacă o astfel de analiză are sens clinic.
3. Problemă cu date (rezolvare completă). Un test are Sp = 0,90, VPP = 0,60, AN = 270 și FN = 10. Reconstituiți tabelul și calculați Se, VPN și rata erorii.

Rezolvare. d = 270 și $\text{Sp} = \frac{d}{c+d} = 0{,}90$, deci c + d = 300 și c = 30. Din $\text{VPP} = \frac{a}{a+30} = 0{,}60$: a = 0,60a + 18, deci 0,40a = 18 și a = 45. b = 10. Tabel: a = 45, b = 10, c = 30, d = 270; N = 355. $\text{Se} = \frac{45}{55} = 81{,}8\%$; $\text{VPN} = \frac{270}{280} = 96{,}4\%$; $\text{RE} = \frac{10+30}{355} = 11{,}3\%$; $\text{AC} = \frac{315}{355} = 88{,}7\%$.

4. Problemă cu date (rezolvare completă). Pentru IL-6 în sepsisul neonatal s-au raportat Se 80,1% și Sp 85,7% [126]. Folosind calculatorul din L12.1 (pasul 4), calculați VPP și VPN la prevalențele de 5% și 30%.

Rezolvare. $\text{RV}^{+} = \frac{0{,}801}{0{,}143} = 5{,}60$; $\text{RV}^{-} = \frac{0{,}199}{0{,}857} = 0{,}232$. La 5%: șansa 0,0526; pozitiv 0,0526 x 5,60 = 0,295, VPP 22,8%; negativ 0,0526 x 0,232 = 0,0122, probabilitate post-test 1,2%, VPN 98,8%. La 30%: șansa 0,429; pozitiv 0,429 x 5,60 = 2,40, VPP 70,6%; negativ 0,429 x 0,232 = 0,0995, probabilitate post-test 9,1%, VPN 90,9%.

Ce se predă și barem orientativ

Se predă pe Moodle fișierul Google Sheets (foile "2x2", "ROC", eventual "Glicemie"), capturile din MedCalc (analiza ROC pentru FINDRISC și compararea curbelor) și documentul de o pagină cu analiza glicemiei a jeun în formatul a-g.

Barem orientativ (10 puncte): foaia "2x2" cu toți indicatorii corecți, inclusiv rata erorii (b + c)/N și IC Wilson (2 puncte); evaluarea în MedCalc cu orientarea corectă a tabelului și recalcularea pentru prevalența de 10% (1 punct); tabelul pragurilor, curba ROC și AUC prin trapeze în Sheets (2 puncte); analiza ROC și compararea curbelor în MedCalc, cu interpretarea criteriului "> 10" (2 puncte); JASP sau Epi Info, cu explicarea diferenței dintre punctul de operare implicit și pragurile scorului (1 punct); analiza glicemiei în formatul a-g, cu limite (2 puncte).

Varianta fără rețea

Fără acces la programe, lucrarea se poate face integral pe hârtie. Indicatorii din L12.1 se calculează cu un calculator de buzunar; tabelul pragurilor FINDRISC și al glicemiei sunt date în capitolul 12 (tabelul 12.5 și problema rezolvată 12.7); curba ROC se desenează pe hârtie milimetrică, iar AUC se estimează prin metoda trapezelor. Pentru analiza ROC și compararea curbelor folosiți rezultatele de referință din L12.4 (AUC 0,852 și 0,844, diferență nesemnificativă, p = 0,81) și concentrați-vă pe interpretare și pe redactarea în formatul a-g. MedCalc oferă și varianta online gratuită a calculatorului pentru tabelul 2 x 2, utilizabilă de pe telefon când rețeaua revine.

Lucrarea practică 13. Proiectarea unui mini-studiu clinic și analiza integrată a unui set complet de date

Obiectiv

Această lucrare practică leagă cele două capete ale unui studiu: cel în care datele se nasc și cel în care se transformă în concluzii. În prima parte veți proiecta un mini-studiu observațional, veți construi formularul de culegere a datelor în Google Forms și veți simula culegerea și centralizarea datelor în Google Sheets. În a doua parte veți analiza complet setul mg2026, de la descriere până la curba ROC și analiza riscului, și veți redacta un raport final în format academic. Lucrarea pregătește direct examenul practic din lucrarea practică 14 și aplică noțiunile din capitolul 13.

La final, studentul trebuie să poată:

- să transforme o întrebare clinică într-o fișă de proiect (PICO, design, variabile, dimensiunea eșantionului);
- să construiască un formular Google Forms cu pagină de consimțământ, secțiuni, ramificare și reguli de validare, legat de o foaie Google Sheets;
- să verifice, să curețe și să documenteze un set de date brut, cu jurnal de modificări;
- să aleagă și să aplice testele potrivite pe setul mg2026 în JASP și MedCalc și să-și verifice rezultatele cu cifrele de control din fișă;
- să redacteze un raport IMRaD de 2-3 pagini, cu tabel, grafic, rezultate raportate corect și limite.

Instrumente

- Google Forms și Google Sheets (cont instituțional), cu setările regionale pentru România: separatorul zecimal este virgula, iar separatorul de argumente în formule este ";" [43,45,46,148].
- JASP 0.98.1 [47], MedCalc 24.0 [50], opțional Epi Info 7.2.7 (StatCalc) [51]; ghidurile de lucru în JASP și culegerea de aplicații a disciplinei [14,48,49].
- Fișierele mg2026_curat.csv și mg2026_brut.csv, de pe platforma Moodle a disciplinei [52].
- Capitolul 13 al acestui volum, pentru dicționarul de date și pentru planul de analiză.
- Timp orientativ: 100 de minute (aproximativ 40 de minute pentru partea de proiectare și formular, 50 pentru analiză, restul pentru raport, care se finalizează acasă).

Desfășurare

L13.1 Fișa de proiect a mini-studiului

Lucrați în echipe de câte 3-4 studenți. Fiecare echipă proiectează un mini-studiu observațional transversal într-un cabinet de medicină de familie. Tema propusă, pe care o puteți adapta: "Activitatea fizică, indicele de masă corporală și tensiunea arterială la adulții care se prezintă în cabinetul de medicină de familie". Completați, într-un document Google Docs comun, o fișă de proiect cu următoarele rubrici.

1. Întrebarea de cercetare în format PICO [24]. Exemplu: P, adulți de 18-80 de ani prezentați la cabinet într-o lună; I (expunerea), activitate fizică redusă (sub 150 de minute pe săptămână); C, activitate fizică cel puțin 150 de minute pe săptămână; O, hipertensiunea arterială (TAS ≥ 140 mmHg sau TAD ≥ 90 mmHg) și valoarea TAS.
2. Designul: transversal, un singur moment de măsurare. Notați explicit ce nu permite acest design (succesiunea temporală, deci concluzia cauzală).

3. Ipotezele statistice pentru rezultatul principal, în cuvinte și matematic. Exemplu: H0, media TAS este aceeași la cei cu activitate redusă și la cei activi (μ1 = μ2); H1, mediile diferă (μ1 ≠ μ2).
4. Variabilele, cu tipul fiecăreia (numerică, ordinală, nominală) și testul pe care îl veți folosi. Minimum 10 variabile: cod, vârstă, sex, mediu, educație, fumat (cu număr de țigări pe zi pentru fumători), activitate fizică în minute pe săptămână, înălțime, greutate, TAS, TAD, un item Likert ("Cât de mulțumit sunteți de starea dumneavoastră de sănătate?", 1-5).
5. Dimensiunea eșantionului. Dacă vreți să estimați prevalența HTA, presupusă 30%, cu o marjă de ±10 puncte procentuale, $n = \frac{1{,}96^2 \times 0{,}30 \times 0{,}70}{0{,}10^2} = 80{,}7$, deci 81 de pacienți. Calculați și pentru marja de ±5 puncte procentuale și comentați diferența (rezultatul corect este 323).
6. Aspectele etice: informarea și consimțământul, anonimizarea (cod de studiu în loc de nume), minimizarea datelor (vârsta în ani, nu data nașterii), avizul comisiei de etică pentru un studiu real [32,35].

Rezultat așteptat: fișa de proiect de o pagină, cu dicționarul de date sub formă de tabel (numele variabilei, descrierea, tipul, unitatea, valorile permise), după modelul tabelului 13.2.

L13.2 Construirea formularului în Google Forms

Deschideți Google Forms din contul instituțional și creați un formular nou, cu titlul studiului. Construiți-l pe secțiuni, în ordinea de mai jos. Butonul pentru o întrebare nouă este semnul "+" din bara laterală; butonul pentru o secțiune nouă este simbolul cu două dreptunghiuri din aceeași bară.

Secțiunea 1, Informare și consimțământ. Adăugați un bloc de text (titlu și descriere) care explică scopul studiului, ce date se culeg, că participarea este voluntară, că datele sunt anonime și că persoana se poate retrage oricând. Adăugați apoi o întrebare de tip "Variante multiple", marcată "Obligatoriu": "Sunteți de acord să participați la acest studiu?", cu variantele "Da, sunt de acord" și "Nu sunt de acord". Din meniul cu trei puncte al întrebării alegeți "Accesați secțiunea în funcție de răspuns" (în interfața în limba engleză, "Go to section based on answer"). Pentru "Da" alegeți "Continuați la secțiunea următoare", pentru "Nu" alegeți "Trimiteți formularul" [148]. Așa, cine refuză nu vede nicio întrebare despre sănătatea lui.

Secțiunea 2, Date generale. Codul participantului ("Răspuns scurt", obligatoriu, cu validarea "Expresie regulată", "Corespunde", modelul P[0-9][0-9][0-9], mesajul de eroare "Codul are forma P001"); vârsta ("Răspuns scurt", validarea "Număr", "Între", 18 și 80, mesajul "Introduceți vârsta în ani împliniți, între 18 și 80"); sexul ("Variante multiple": F, M); mediul ("Variante multiple": urban, rural); educația ("Variante multiple", în ordine: gimnaziu, liceu sau școală profesională, universitar, postuniversitar). Regulile de validare se activează din meniul cu trei puncte al întrebării, opțiunea "Validarea răspunsurilor" ("Response validation") [45].

Secțiunea 3, Măsurători. Înălțimea în cm (validare "Număr", "Între" 120 și 220, mesaj "Înălțimea în centimetri, de exemplu 168"); greutatea în kg (între 30 și 250); TAS (între 70 și 250 mmHg) și TAD (între 40 și 150 mmHg). Nu adăugați o întrebare pentru IMC: el se calculează în foaia de calcul din greutate și înălțime. Observați cum fiecare regulă ar fi oprit una dintre erorile din fișierul mg2026_brut.csv (înălțimea 1,68, greutatea 8,6, TAS 1450).

Secțiunea 4, Stil de viață. Întrebarea "Fumați în prezent?" ("Variante multiple": "Da", "Nu, am renunțat", "Nu, nu am fumat niciodată"), cu ramificare: "Da" duce la secțiunea 4a, celelalte variante la secțiunea 5. Secțiunea 4a conține o singură întrebare: "Câte țigări fumați, în medie, pe zi?" (număr între 1 și 80). Urmează minutele de activitate fizică pe săptămână ("Răspuns scurt", număr între 0 și 3000) și itemul Likert ("Scară liniară", de la 1 la 5, cu etichetele "deloc mulțumit" și "foarte mulțumit").

Secțiunea 5, Final. Un text de mulțumire. În setări (pictograma roată dințată), dezactivați colectarea adreselor de e-mail: pentru un studiu anonim, adresa de e-mail ar fi o dată de identificare inutilă.

Verificați formularul cu butonul "Previzualizare" (pictograma ochi): încercați să introduceți o vârstă de 15 ani, o înălțime de 1,75 și răspunsul "Nu sunt de acord" la consimțământ. Formularul trebuie să refuze primele două și să se încheie imediat la al treilea. Rezultat așteptat: un formular funcțional, cu 5 secțiuni (plus 4a), cel puțin 3 reguli de validare și 2 ramificări.

L13.3 Simularea culegerii și centralizarea datelor în Google Sheets

În fila "Răspunsuri" a formularului, alegeți "Conectați-vă la Foi de calcul" ("Link to Sheets") și creați o foaie nouă [46]. Fiecare membru al echipei completează formularul de 5 ori, cu date fictive, dar plauzibile, ca și cum ar fi pacienți diferiți; cel puțin unul dintre ei refuză participarea, iar cel puțin doi sunt fumători. Echipa obține astfel 15-20 de înregistrări.

Deschideți foaia de răspunsuri și faceți următoarele verificări, într-o filă nouă numită "verificare".

1. Redenumiți coloanele cu nume scurte, fără spații și fără diacritice (cod, varsta, sex, mediu, educatie, inaltime, greutate, tas, tad, fumat, tigari, activitate, satisfactie). Prima coloană, "Marcaj de timp", se păstrează pentru trasabilitate.
2. Numărați înregistrările complete: =COUNTA(B2:B21). Numărați refuzurile: =COUNTIF(B2:B21;"Nu sunt de acord").
3. Verificați duplicatele de cod: =COUNTIF(B$2:B$21;B2) într-o coloană auxiliară; orice valoare mai mare de 1 semnalează un duplicat.
4. Calculați IMC-ul într-o coloană nouă: =ROUND(G2/(F2/100)^2;1), presupunând greutatea în coloana G și înălțimea în F.
5. Calculați descriptivele pentru TAS: =AVERAGE(H2:H21), =STDEV(H2:H21), =MEDIAN(H2:H21), =MIN(H2:H21), =MAX(H2:H21).
6. Construiți un tabel pivot ("Inserare" > "Tabel pivot") cu sexul pe rânduri și fumatul pe coloane, cu valorile numărate [44].
7. Creați o filă "modificari", cu coloanele: data, cine, rândul, variabila, valoarea veche, valoarea nouă, motivul, sursa. Introduceți intenționat o eroare (de exemplu, schimbați manual o TAS în 1500), apoi corectați-o și notați corecția în jurnal.

Discutați în echipă: care dintre regulile de validare au funcționat, ce erori au trecut totuși (de exemplu, o TAS de 95 cu o TAD de 110, ambele în intervalul permis, dar incompatibile clinic) și ce verificare logică ar trebui adăugată. Rezultat așteptat: foaia de calcul cu filele "raspunsuri", "verificare" și "modificari", partajată cu cadrul didactic doar pentru citire.

L13.4 Curățarea fișierului mg2026_brut.csv

Importați fișierul mg2026_brut.csv într-o foaie Google Sheets nouă ("Fișier" > "Importați" > "Încărcați"). Fără să vă uitați în capitolul 13, căutați erorile folosind doar instrumentele de mai jos, apoi comparați lista voastră cu tabelul 13.3.

- Numărul de rânduri cu date: trebuie să obțineți 161 (un rând este duplicat).
- Duplicatele: sortați după coloana id sau folosiți =COUNTIF(A$2:A$162;A2); pacientul P091 apare de două ori.
- Variabilele categoriale: pentru sex, mediu și fumat, folosiți "Date" > "Filtru" și priviți lista de valori distincte. Trebuie să găsiți: la sex, "f" și "Feminin" pe lângă F și M; la mediu, "Urban"; la fumat, "Fumator " (cu spațiu).

- Variabilele numerice: =MIN() și =MAX() pe fiecare coloană. Trebuie să găsiți TAS0 = 1540, înălțimea 1,72 și greutatea 7,26. Folosiți =COUNT() comparat cu =COUNTA(): diferența arată valorile stocate ca text (glicemia "87 mg/dl", HbA1c "5,2").
- Valorile lipsă: =COUNTBLANK() pe fiecare coloană; lipsesc vârsta la P042 și glicemia la 6 luni la P134.

Notați în jurnalul de modificări cum ați trata fiecare eroare. Reamintim regula: erorile de formă se corectează direct, valorile imposibile se corectează doar din documentul sursă. În acest exercițiu, "documentul sursă" este fișierul mg2026_curat.csv: verificați acolo valorile pacienților P034, P057 și P102 și notați, pentru fiecare, ce corectură ați fi ghicit și ce arată sursa. Pentru înălțime și greutate folosiți și IMC-ul păstrat în tabel ca verificare independentă. Rezultat așteptat: 12 probleme identificate, fiecare cu metoda de detectare și decizia de corectare.

L13.5 Analiza completă a setului mg2026

Lucrați pe mg2026_curat.csv. Pentru fiecare analiză, scrieți în raport elementele (a)-(g) din formatul temelor disciplinei: ipoteza de cercetare, ipotezele statistice, metoda și justificarea, programul și modul de rezolvare, rezultatul cu grafic, concluzia statistică și concluzia de cercetare. Cifrele de control de mai jos vă permit să vă verificați.

1. Descrierea. În JASP: "Descriptives" > "Descriptive Statistics", variabilele varsta, imc0, tas0, glicemie0, trigliceride; în "Statistics" bifați media, abaterea standard, mediana, quartilele, asimetria, excesul și "Shapiro-Wilk test"; în "Plots", "Distribution plots" și "Boxplots". Cifre de control: vârsta 56,29 ± 8,77 ani (Shapiro-Wilk p = 0,687); IMC mediana 28,25, IQR 26,2-31,4, asimetrie 1,42, p < 0,001; TAS $131{,}13 \pm 13{,}88$ mmHg$(p = 0{,}303)$; trigliceride mediana 151,5 mg/dL, asimetrie 0,95. Pentru variabilele categoriale ("Frequencies" > "Frequency tables"): F 86, M 74; A 53, B 53, C 54; fumători 53, foști 33, nefumători 74.
2. Tabelul 1 pe brațe. Repetați descrierea cu "Split" după brat. Comparați rezultatele cu tabelul 13.4 și explicați de ce nu adăugați valori p.
3. Intervalul de încredere. În JASP, "T-Tests" > "One Sample T-Test" pe tas0, cu "Location parameter" și "Confidence interval" bifate: IC 95% 128,96-133,30 mmHg. În Google Sheets: =AVERAGE(M2:M161) - T.INV.2T(0,05;159)*STDEV(M2:M161)/SQRT(160), presupunând TAS0 în coloana M.
4. Criteriul principal. Creați variabila dglic = glicemie0 - glicemie6 (în JASP, butonul "+" din antetul tabelului de date, "Computed column"). "ANOVA" > "ANOVA", dependent dglic, factor brat; în "Assumption Checks" bifați "Homogeneity tests"; în "Post Hoc Tests" mutați brat și bifați "Tukey"; în "Additional Options", "Estimates of effect size" (eta²). Cifre de control: medii 3,0 / 11,1 / 16,3 mg/dL; Levene p = 0,163; F(2; 157) = 22,93, p < 0,001; $\eta^2 = 0{,}226$; Tukey B-A 8,1 (3,4-12,8), C-A 13,3 (8,6-18,0), C-B 5,2 (0,6-9,9). În secțiunea "Nonparametrics" a aceluiași meniu, Kruskal-Wallis: H = 36,63, p < 0,001.
5. Date pereche și măsurători repetate. "T-Tests" > "Paired Samples T-Test", glicemie0 și glicemie6, cu "Normality" bifat: diferența medie 10,19 mg/dL (IC 95% 8,38-11,99), t = 11,14, gl 159, p < 0,001, Shapiro-Wilk pe diferențe p = 0,725. "ANOVA" > "Repeated Measures ANOVA" cu glicemie0, glicemie3, glicemie6 ca niveluri ale factorului timp: F = 64,97, p < 0,001; în "Nonparametrics", Friedman $\chi^2 = 74{,}67$, p < 0,001.
6. Proporții și risc. "Frequencies" > "Contingency Tables", rânduri brat, coloane scadere5; în "Statistics" bifați χ^2, în "Cells" frecvențele așteptate și procentele pe rând: A 6/53, B 16/53, C 16/54; $\chi^2 = 6{,}77$, gl 2, p = 0,034; cea mai mică frecvență așteptată 12,6. Pentru C față de A, în MedCalc: "Tests" > "Relative risk", cu 16 evenimente din 54 și 6 din 53: RR 2,62 (IC 95% 1,11-6,17), NNT 5,5 (adică 6). Aceeași analiză în Epi Info, "StatCalc" > "2 x 2 tables".
7. Ordinale. "ANOVA" > "ANOVA" > "Nonparametrics", satisfactie după brat: Kruskal-Wallis H = 32,16, p < 0,001. "T-Tests" > "Independent Samples T-Test" cu "Mann-Whitney" bifat, doar pentru brațele A și C (filtrați datele): U = 2265,5, p < 0,001, corelația biserială a rangurilor 0,58.

8. Corelație și regresie. "Regression" > "Correlation", imc0 și tas0, cu Pearson și Spearman: r = 0,405, rho = 0,372, ambele p < 0,001; diagrama de dispersie din "Plots". "Regression" > "Linear Regression", dependent tas0, covariabilă imc0: TAS = 89,47 + 1,435 x IMC, $R^2 = 0{,}164$. Adăugați varsta, sex și fumat: $R^2 = 0{,}308$, coeficientul IMC 1,389.
9. Asocierea obezitate-diabet. Creați variabila obez (imc0 ≥ 30). "Contingency Tables" obez x diabet_ogtt, cu "Log odds ratio" bifat: 18/54 față de 11/106; χ^2 = 12,70, p < 0,001; OR 4,32 (IC 95% 1,86-10,03); raportul prevalențelor, calculat în MedCalc, 3,21 (IC 95% 1,64-6,31). "Regression" > "Logistic Regression", dependent diabet_ogtt, covariabile obez, varsta, sex, cu "Odds ratios" bifat: OR obezitate 6,01, vârstă 1,11 pe an.
10. Testul diagnostic. În MedCalc: "Statistics" > "ROC curves" > "ROC curve analysis", variabila findrisc, clasificarea diabet_ogtt; bifați "Criterion values and coordinates of the ROC curve". Cifre de control: AUC 0,852 (IC 95% 0,762-0,943), pragul Youden > 10, adică ≥ 11 (Se 93,1%, Sp 66,4%). La pragul ≥ 12 construiți de mână tabelul 2 x 2 (AP 22, FP 33, FN 7, AN 98) și calculați Se 75,9%, Sp 74,8%, VPP 40,0%, VPN 93,3%, RV+ 3,01, RV- 0,32.

Atenție la ultima analiză: MedCalc raportează pragul sub forma "> 10", ceea ce, pentru un scor cu valori întregi, înseamnă "≥ 11". Mulți studenți scriu în raport pragul greșit tocmai din acest motiv.

L13.6 Raportul final în format IMRaD

Raportul individual, de 2-3 pagini, urmează structura unui articol științific: Introducere, Metode, Rezultate și Discuție (IMRaD). El se bazează pe analiza din L13.5, nu pe datele fictive ale echipei.

Introducerea (un paragraf) prezintă problema (riscul cardiometabolic în medicina de familie), ce se știe și întrebarea studiului, în format PICO. Metodele descriu designul (studiu randomizat pragmatic simulat, 3 brațe, 12 cabinete), participanții și criteriile, intervențiile, criteriul principal și cele secundare, analiza statistică (testele, pragul $\alpha = 0{,}05$, programele cu versiunea) și precizează că datele sunt simulate. Rezultatele încep cu fluxul participanților (figura 13.3) și cu Tabelul 1 (un tabel cu 8-10 rânduri este suficient), continuă cu criteriul principal (descriere, test, mărimea efectului, IC) și cu două criterii secundare la alegere, apoi cu o analiză observațională (obezitate-diabet sau FINDRISC). Includeți un singur grafic, ales pentru criteriul principal (box-plot pe brațe sau diagrama diferențelor cu IC). Discuția rezumă rezultatul principal într-o frază, îl interpretează clinic (de exemplu, NNT), discută limitele (lipsa orbirii, criteriu surogat, urmărire scurtă, date simulate fără pierderi) și formulează concluzia.

Respectați regulile de raportare din secțiunea 14.4: valori p exacte, mărimea efectului cu IC 95%, unitățile de măsură, "nu se respinge H0" în loc de "se acceptă H0", fără concluzii cauzale din analizele observaționale. Un model de frază pentru rezultatul principal: "La 6 luni, scăderea medie a glicemiei a jeun a fost 3,0 mg/dL în brațul A, 11,1 mg/dL în brațul B și 16,3 mg/dL în brațul C (ANOVA, $F(2;\ 157) = 22{,}93,\ p < 0{,}001;\ \eta^2 = 0{,}23$)."

Exerciții suplimentare

1. Pe fișierul BD_Studenti de pe Moodle, proiectați formularul Google Forms care ar fi produs acest fișier: pentru fiecare variabilă, alegeți tipul întrebării și regula de validare. Identificați variabilele care au fost culese, probabil, în clase și propuneți întrebarea care ar fi permis culegerea valorii exacte.
2. Pe fișierul BPOC (studiu încrucișat cu măsurători la începutul studiului și după medicație), scrieți planul de analiză pentru întrebarea "se modifică frecvența cardiacă după tratament?" și pentru "diferă capacitatea vitală după tratament în funcție de medicație?". Precizați, pentru fiecare, de ce testul ales este pentru date pereche sau pentru grupuri independente. Aplicați testele și interpretați.
3. Pe fișierul Date_HbA1c, scrieți Tabelul 1 pe două grupuri la alegere, alegând corect între medie ± AS și mediană (IQR) pentru fiecare variabilă, după testul de normalitate.

4. O echipă vrea să estimeze prevalența fumatului printre pacienții unui cabinet, presupusă 25%, cu o marjă de ±5 puncte procentuale și încredere de 95%. Calculați n. Rezolvare: $n = \frac{1{,}96^2 \times 0{,}25 \times 0{,}75}{0{,}05^2} = \frac{3{,}8416 \times 0{,}1875}{0{,}0025} = 288{,}1$, deci 289 de pacienți. Cu o rată de răspuns estimată de 80%, trebuie invitați 289/0,8 = 361,3, adică 362 de pacienți.
5. Într-un formular, întrebarea "Cât de des faceți sport? rar / uneori / des" a produs răspunsuri greu de interpretat. Rescrieți-o astfel încât să producă o variabilă numerică și explicați ce câștigați la analiză. Rezolvare model: "În ultimele 7 zile, câte minute în total ați făcut activitate fizică moderată sau intensă? ___ minute"; variabila numerică permite media, mediana, corelația cu IMC și dihotomizarea ulterioară la un prag (150 de minute), în timp ce variantele "rar / uneori / des" sunt interpretate diferit de fiecare respondent.

Ce se predă și barem orientativ

Se predau pe Moodle, în secțiunea lucrării 13: fișa de proiect (cu dicționarul de date), linkul formularului (cu drept de vizualizare), foaia de calcul cu filele "raspunsuri", "verificare" și "modificari", lista erorilor din mg2026_brut.csv și raportul IMRaD individual, în format PDF.

Tabelul L13.1. Barem orientativ pentru lucrarea practică 13 (100 de puncte).

Componenta	Criterii	Puncte
Fișa de proiect	PICO corect; design numit și limitele lui; ipoteze statistice în cuvinte și matematic; dimensiunea eșantionului calculată corect; aspecte etice	15
Dicționarul de date	toate variabilele, cu tip, unitate, valori permise și coduri	5
Formularul Google Forms	pagină de consimțământ cu ramificare de ieșire; cel puțin 3 reguli de validare cu mesaje clare; ramificare pentru fumători; tipuri de întrebări potrivite; fără e-mail colectat	15
Centralizarea în Sheets	coloane redenumite; verificări (duplicate, IMC, descriptive); tabel pivot; jurnal de modificări	10
Curățarea mg2026_brut	erori identificate (minimum 10 din 12), cu metoda de detectare și decizia corectă	10
Analiza mg2026	teste alese și justificate corect; presupuneri verificate; rezultate concordante cu cifrele de control; formatul (a)-(g)	25
Raportul IMRaD	structură; Tabelul 1; un grafic potrivit; rezultate cu mărimea efectului și IC; limite; concluzie fără exagerări cauzale	20
Total		100

Erori care scad punctajul indiferent de componentă: formularea "se acceptă H0"; test pentru grupuri independente aplicat pe date pereche sau invers; medie ± AS pentru variabile vizibil asimetrice fără discuție; OR prezentat ca risc relativ; concluzii cauzale din analiza obezitate-diabet; pragul FINDRISC raportat greșit ("> 10" transcris ca "≥ 10").

Varianta fără rețea

Dacă nu aveți acces la internet, partea de proiectare se face pe hârtie: desenați formularul pe secțiuni, cu săgeți pentru ramificări și cu regula de validare scrisă lângă fiecare câmp, iar

"culegerea" se simulează prin completarea a 5 formulare tipărite de fiecare student. Centralizarea se face într-un fișier Excel sau LibreOffice Calc, cu aceleași formule (separatorul de argumente rămâne ";" la setările regionale românești). Curățarea fișierului mg2026_brut.csv și analiza mg2026 se fac în JASP și MedCalc, instalate local pe calculatoarele din laborator; Epi Info poate înlocui MedCalc pentru tabelele 2 x 2. Pentru curba ROC fără MedCalc, construiți de mână tabelul cu pragurile 8, 10, 12, 14, 15 și 16 al scorului FINDRISC (tabelul din capitolul 12) și desenați punctele (1 - Sp, Se) pe hârtie milimetrică; aria se poate aproxima prin regula trapezelor.

Lucrarea practică 14. Examenul practic: model de subiect, barem și plan de pregătire

Obiectiv

Ultima lucrare practică este dedicată examenului practic. Veți afla cum este organizat, veți lucra integral două subiecte model, cu rezolvările complete și cifrele așteptate, veți vedea cum se punctează fiecare cerință și care sunt greșelile care costă cele mai multe puncte. Lucrarea se încheie cu un plan de pregătire în două săptămâni și cu o listă de verificare a deprinderilor de lucru în programe. Un student care poate rezolva singur, în timp, ambele subiecte model este pregătit pentru examen.

La final, studentul trebuie să poată:

- să descrie structura și ponderea examenului practic în nota finală;
- să rezolve, în timpul alocat, un subiect cu 6-8 cerințe care acoperă descrierea, estimarea, testarea, corelația și regresia, riscul și testul diagnostic;
- să redacteze fiecare rezolvare în formatul (a)-(g) al disciplinei;
- să-și verifice rezultatele și să evite greșelile frecvente din barem;
- să-și organizeze pregătirea pe baza listei de deprinderi.

Instrumente

- JASP 0.98.1 [47], MedCalc 24.0 [50], Epi Info 7.2.7 [51], Google Sheets cu setările regionale pentru România (separator de argumente ";") [43].
- Fișierul mg2026_curat.csv și fișierele clasice ale disciplinei de pe Moodle [52]; culegerile de aplicații practice [14,149] și ghidul de lucru în JASP [49].
- Capitolele 2-12 și 14 ale acestui volum; tabelul-sinteză al formulelor (tabelul 14.2).
- Timp orientativ: 100 de minute, din care 60 pentru rezolvarea unui subiect model în condiții de examen și 40 pentru corectarea lui pe baza baremului.

Desfășurare

L14.1 Cum se desfășoară examenul practic

Nota finală la Biostatistică se compune din trei părți, conform fișei disciplinei: examenul grilă de 50 de întrebări, corectat în sistem rezidențiat (50%), activitatea și temele de pe parcursul semestrului (20%) și examenul practic în laborator (30%). Examenul practic verifică exact ce ați exersat la lucrări: să deschideți un fișier de date, să alegeți și să aplicați metoda potrivită într-un program, să interpretați rezultatul și să-l scrieți corect.

Examenul se susține individual, la calculator, în sala de lucrări practice. Primiți un subiect cu 6-8 cerințe, pe un set de date cunoscut (de tipul mg2026 sau al fișierelor de la lucrări) sau pe un set mic, nou, ale cărui date sunt date în enunț. Aveți la dispoziție programele folosite în semestru. Răspunsurile se scriu într-un document (sau pe foaia de examen), în formatul cerut la teme: (a) ipoteza de cercetare; (b) ipotezele statistice în cuvinte și matematic; (c) metoda aleasă și justificarea; (d) modul de rezolvare și programul; (e) rezultatul, cu grafic acolo unde se cere; (f) concluzia statistică; (g) concluzia de cercetare în cuvinte. Nu toate cerințele au nevoie de toate cele șapte elemente: o cerință de statistică descriptivă nu are ipoteze statistice, iar una de estimare are interval de încredere, nu test.

Ce se punctează: alegerea corectă a metodei și justificarea ei (de obicei cea mai mare parte din punctaj), corectitudinea cifrelor, interpretarea statistică și interpretarea medicală. O cifră corectă cu o metodă greșită primește puțin; o metodă corectă cu o mică greșeală de calcul, bine interpretată, primește aproape tot. La examen, cea mai frecventă pierdere de puncte nu vine din calcul, ci din concluzii formulate greșit ("se acceptă H0", "corelația dovedește că...").

L14.2 Subiectul model 1 (setul mg2026)

Deschideți fișierul mg2026_curat.csv. Considerați eșantionul reprezentativ pentru adulții cu risc cardiometabolic din cabinetele de medicină de familie din vestul țării. Rezolvați cerințele, cu α = 0,05.

1. Descrieți variabilele HDL-colesterol (hdl) și circumferința taliei (talie). Alegeți indicatorii potriviți, justificați alegerea prin forma distribuției și realizați câte un grafic potrivit.
2. Estimați media HDL în populație, cu încredere de 95%. În ce interval se găsesc aproximativ 95% dintre valorile individuale ale HDL?
3. Diferă HDL între femei și bărbați?
4. Studiați relația dintre circumferința taliei și glicemia a jeun la includere (glicemie0). Realizați diagrama de dispersie cu dreapta de regresie și estimați glicemia unui pacient cu talia de 100 cm.
5. Diferă tensiunea arterială sistolică la includere (tas0) între fumători, foști fumători și nefumători?
6. Este fumatul actual asociat cu hipertensiunea arterială (tas0 ≥ 140 mmHg) la includere? Calculați măsura potrivită a asocierii.
7. În brațul C, s-a modificat TAS între includere și 6 luni (tas0 față de tas6)?
8. Evaluați glicemia a jeun ≥ 110 mg/dL ca test pentru diabetul confirmat prin TTGO (diabet_ogtt): construiți tabelul 2 x 2 și calculați Se, Sp, VPP, VPN, acuratețea, rata de eroare și rapoartele de verosimilitate.

L14.3 Rezolvarea subiectului 1

Cerința 1. HDL: n = 160, media 49,52 mg/dL, AS 11,78, eroarea standard 0,931, mediana 49, IQR 41-59, minim 20, maxim 78, asimetrie 0,04; Shapiro-Wilk W = 0,991, p = 0,371. Nu respingem normalitatea, media și mediana sunt practic egale, deci descriem prin medie ± AS: 49,52 ± 11,78 mg/dL. Graficul potrivit este histograma (cu curba normală suprapusă). Talia: media 88,54 cm, AS 12,11, mediana 86, IQR 80,75-95, minim 65, maxim 140, asimetrie 0,86; Shapiro-Wilk W = 0,961, p < 0,001 (p = 0,0002). Distribuția este asimetrică spre dreapta, media este mai mare decât mediana, deci descriem prin mediană și IQR: 86 (80,75-95) cm. Grafic: box-plot sau histogramă. În JASP: "Descriptives" > "Descriptive Statistics", cu Shapiro-Wilk bifat.

Cerința 2. (c) Estimare prin interval de încredere pentru medie, cu distribuția t (n = 160, gl 159, t = 1,975). (d) ES $= \frac{11{,}78}{\sqrt{160}} = 0{,}931$; IC 95% $= 49{,}52 \pm 1{,}975 \times 0{,}931 = 49{,}52 \pm 1{,}84$. (e) IC 95% pentru medie: 47,68-51,36 mg/dL. Intervalul pentru valorile individuale este $m \pm 1{,}96 \times s = 49{,}52 \pm 23{,}09$, adică 26,4-72,6 mg/dL. (g) Cu 95% încredere, media HDL a populației se află între 47,7 și 51,4 mg/dL; aproximativ 95% dintre pacienți au HDL între 26 și 73 mg/dL. Cele două intervale răspund la întrebări diferite, iar confuzia lor este o greșeală de barem.

Cerința 3. (a) HDL diferă între sexe. (b) H0: μF = μM (nu există diferență între mediile HDL la femei și bărbați); H1: μF ≠ μM. (c) Variabilă numerică, două grupuri independente; normalitatea nu este respinsă în niciun grup (Shapiro-Wilk p = 0,413 la femei, p = 0,477 la bărbați), dar testul Levene respinge egalitatea varianțelor (p = 0,034), deci folosim testul t Welch. (d) JASP: "T-Tests" > "Independent Samples T-Test", cu "Welch" și "Equality of variances" bifate. (e) Femei $53{,}92 \pm 11{,}70$ mg/dL$(n\,86)$, bărbați $44{,}41 \pm 9{,}66$ mg/dL$(n\,74)$; t = 5,63, gl 157,7, p < 0,001; diferența 9,51 mg/dL (IC 95% 6,18-12,85); d Cohen 0,88. (f) Respingem H0. (g)

Femeile au HDL în medie cu aproximativ 9,5 mg/dL mai mare decât bărbații, o diferență mare (d 0,88), concordantă cu fiziologia cunoscută. Varianta Student ar fi dat t = 5,55, cu aceeași concluzie; alegerea Welch se justifică prin rezultatul testului Levene.

Cerința 4. (a) Talia mai mare se asociază cu glicemia mai mare. (b) H0: ρ = 0 (variabilele nu sunt corelate); H1: ρ ≠ 0. (c) Ambele variabile sunt numerice, dar niciuna nu are distribuție normală (talie p = 0,0002; glicemie0 p = 0,037), deci corelația principală este Spearman; raportăm și Pearson și dreapta de regresie, cerută explicit. (d) JASP: "Regression" > "Correlation" (Pearson și Spearman), apoi "Regression" > "Linear Regression", dependent glicemie0, covariabilă talie. (e) rho = 0,430, p < 0,001; r = 0,488, p < 0,001; $R^2 = 0{,}238$. Dreapta: glicemie0 = 49,17 + 0,636 x talie (IC 95% pentru pantă 0,457-0,814); eroarea standard a estimării 13,80 mg/dL. La talia de 100 cm: 49,17 + 63,55 = 112,7 mg/dL. (f) Respingem H0. (g) Există o corelație pozitivă moderată între talie și glicemie: fiecare centimetru în plus se asociază cu o glicemie mai mare cu aproximativ 0,6 mg/dL; talia explică aproximativ 24% din variabilitatea glicemiei. Predicția individuală este imprecisă (eroarea standard a estimării aproape 14 mg/dL), iar asocierea nu dovedește cauzalitatea.

Cerința 5. (a) TAS diferă după statutul de fumător. (b) H0: μfumători = μfoști = μnefumători; H1: cel puțin o medie diferă. (c) Variabilă numerică, trei grupuri independente; Shapiro-Wilk în fiecare grup: p = 0,37 (foști), p = 0,86 (fumători), p = 0,23 (nefumători); Levene p = 0,114: ANOVA unifactorială cu post-hoc Tukey. (d) JASP: "ANOVA" > "ANOVA", dependent tas0, factor fumat, cu "Homogeneity tests", "Tukey" și eta². (e) Foști fumători $127{,}5 \pm 14{,}4$ mmHg$(n\,33)$, fumători $137{,}0 \pm 14{,}6(n\,53)$, nefumători $128{,}5 \pm 11{,}8(n\,74)$; F(2; 157) = 7,85, p < 0,001 (p = 0,0006); $\eta^2 = 0{,}091$. Tukey: fumători față de foști fumători +9,5 mmHg (IC 95% 2,5-16,5; p = 0,005); fumători față de nefumători +8,5 mmHg (IC 95% 2,9-14,2; p = 0,001); foști fumători față de nefumători 1,0 mmHg (IC 95% -5,6 la 7,6; p = 0,94). Kruskal-Wallis, ca verificare, H = 12,77, p = 0,002. (f) Respingem H0. (g) Fumătorii actuali au TAS mai mare cu 8-10 mmHg decât foștii fumători și nefumătorii, care nu diferă între ei. Diferența este relevantă clinic. Fiind date transversale, nu putem afirma că fumatul produce această diferență; regresia multiplă din capitolul 9 arată însă că ea se păstrează la același IMC, vârstă și sex (+6,7 mmHg față de foștii fumători).

Cerința 6. (a) Fumatul este asociat cu HTA. (b) H0: proporția HTA este aceeași la fumători și la nefumători (sau RP = 1); H1: proporțiile diferă. (c) Două variabile dihotomice (fumător da/nu, HTA da/nu), design transversal: chi-pătrat; măsura asocierii este raportul prevalențelor (RP), cu OR raportat suplimentar. (d) Creați variabilele fum (fumat = "fumator") și hta (tas0 ≥ 140) în JASP ("Computed column"), apoi "Frequencies" > "Contingency Tables", cu χ², frecvențele așteptate și "Log odds ratio"; RP în MedCalc, "Tests" > "Relative risk". (e) HTA: 21/53 (39,6%) la fumători, 20/107 (18,7%) la ceilalți; cea mai mică frecvență așteptată 13,6; χ² = 8,15, gl 1, p = 0,004; RP = 2,12 (IC 95% 1,26-3,55); OR = 2,85 (IC 95% 1,37-5,95). (f) Respingem H0. (g) Prevalența HTA este de aproximativ 2 ori mai mare la fumătorii actuali; OR-ul supraestimează această relație, pentru că HTA este frecventă (25,6% în eșantion).

Cerința 7. (a) TAS s-a modificat în brațul C. (b) H0: μd = 0 (media diferențelor tas0 - tas6 este 0); H1: μd ≠ 0. (c) Date pereche (aceiași pacienți, două momente), n = 54; normalitatea diferențelor nu este respinsă (Shapiro-Wilk p = 0,196): testul t pereche. (d) Filtrați brațul C în JASP (pictograma pâlnie, brat = "C"), apoi "T-Tests" > "Paired Samples T-Test", tas0 și tas6, cu "Normality". (e) TAS $129{,}1 \pm 14{,}0$ mmHg la includere și 122,1 ± 16,8 mmHg la 6 luni; diferența medie 6,96 mmHg (AS 8,18; IC 95% 4,73-9,20); t = 6,26, gl 53, p < 0,001; dz = 0,85; Wilcoxon, ca verificare, p < 0,001. (f) Respingem H0. (g) În brațul C, TAS a scăzut în medie cu aproximativ 7 mmHg în 6 luni. Atenție la interpretare: comparația înainte-după într-un singur braț nu arată

efectul intervenției față de control; pentru aceasta se compară scăderile între brațe (în mg2026, ANOVA pe scăderea TAS dă p = 0,056).

Cerința 8. (c) Test diagnostic față de standardul de referință TTGO, prag fixat. (d) Creați variabila test = glicemie0 ≥ 110 și faceți tabelul test x diabet_ogtt. (e) Tabelul:

Tabelul L14.1. Glicemia a jeun ≥ 110 mg/dL față de diabetul confirmat prin TTGO (mg2026, n = 160).

	Diabet TTGO da	Diabet TTGO nu	Total
Glicemie ≥ 110	24	39	63
Glicemie < 110	5	92	97
Total	29	131	160

$\text{Se} = \frac{24}{29} = 82{,}8\%;\ \text{Sp} = \frac{92}{131} = 70{,}2\%;\ \text{VPP} = \frac{24}{63} = 38{,}1\%;\ \text{VPN} = \frac{92}{97} = 94{,}8\%;$ acuratețea = (24 + 92)/160 = 72,5%; rata de eroare = (39 + 5)/160 = 27,5%; $\text{RV}^{+} = \frac{0{,}828}{0{,}298} = 2{,}78;\ \text{RV}^{-} = \frac{0{,}172}{0{,}702} = 0{,}25$. (g) La acest prag, glicemia a jeun ratează aproximativ unul din șase pacienți cu diabet și dă un rezultat fals pozitiv la aproape 30% dintre cei fără diabet. Un rezultat negativ face diabetul puțin probabil (VPN 94,8%), un rezultat pozitiv ridică probabilitatea doar până la 38%. Performanța globală a glicemiei (AUC 0,844) este apropiată de cea a scorului FINDRISC (AUC 0,852).

L14.4 Subiectul model 2 (un set mic, nou)

Un medic de familie testează un program de 8 săptămâni de reducere a sodiului din alimentație la 12 pacienți hipertensivi. Măsoară TAS înainte și după program și notează IMC-ul la începutul programului.

Tabelul L14.2. Datele subiectului 2 (TAS în mmHg, IMC în kg/m²).

Pacient	TAS înainte	TAS după	IMC
1	152	140	29,4
2	148	141	30,8
3	160	150	33,5
4	145	140	26,1
5	158	145	28,9
6	150	142	31,6
7	165	150	32,2
8	142	138	27,5
9	155	146	34,0
10	149	143	26,8
11	162	151	30,1
12	147	145	29,3

Cerințe (α = 0,05):

1. Descrieți TAS înainte de program (medie, AS, mediană, minim, maxim), cu calculul de mână al mediei și al abaterii standard.
2. Estimați media TAS înainte de program în populația pacienților hipertensivi asemănători, cu încredere de 95% (t(11; 0,975) = 2,201).
3. A scăzut TAS după program? Aplicați testul potrivit, cu calculul de mână al statisticii testului.
4. Este IMC-ul corelat cu TAS înainte de program? Calculați r de mână, testați-l (t(10; 0,975) = 2,228) și scrieți dreapta de regresie.

5. Într-o cohortă de 400 de pacienți hipertensivi urmăriți 5 ani, 150 sunt fumători. AVC au făcut 18 fumători și 10 nefumători. Calculați RR cu IC 95%, OR, diferența de risc, NNH și fracția atribuibilă la expuși; testați asocierea.
6. Un test rapid pentru hipertrofia ventriculară stângă are Se 90% și Sp 80%. Se aplică la 1000 de pacienți hipertensivi, cu o prevalență a hipertrofiei de 15%. Reconstituiți tabelul 2 x 2, calculați VPP, VPN, acuratețea, rata de eroare, RV+ și RV-. Cât ar fi VPP la o prevalență de 2%?
7. Ce test alegeți pentru: a) compararea TAS după program între bărbați și femei (6 și 6 pacienți, distribuții asimetrice); b) compararea proporției pacienților cu TAS < 140 mmHg înainte și după program la aceiași pacienți; c) compararea satisfacției (1-5) între trei programe diferite; d) evoluția TAS la 0, 4 și 8 săptămâni la aceiași pacienți, cu distribuții normale?

L14.5 Rezolvarea subiectului 2

Cerința 1. Suma valorilor TAS înainte este 1833, deci $m = \frac{1833}{12} = 152{,}75$ mmHg. Suma pătratelor abaterilor de la medie este 574,25, deci $s^2 = \frac{574{,}25}{11} = 52{,}20$ și s = 7,23 mmHg. Valorile ordonate: 142, 145, 147, 148, 149, 150, 152, 155, 158, 160, 162, 165; mediana este media valorilor de pe pozițiile 6 și 7: (150 + 152)/2 = 151 mmHg; minim 142, maxim 165. Shapiro-Wilk p = 0,805: nu respingem normalitatea, deci media ± AS este o descriere potrivită: 152,75 ± 7,23 mmHg.

Cerința 2. $\text{ES} = \frac{7{,}23}{\sqrt{12}} = 2{,}086$; $\text{IC}\,95\% = 152{,}75 \pm 2{,}201 \times 2{,}086 = 152{,}75 \pm 4{,}59$, adică 148,16-157,34 mmHg. Cu un eșantion de 12 pacienți folosim distribuția t cu 11 grade de libertate, nu valoarea 1,96.

Cerința 3. (a) Programul scade TAS. (b) H0: μd = 0; H1: μd ≠ 0 (sau unilateral μd > 0, dacă s-a stabilit dinainte). (c) Date pereche, n = 12; diferențele (înainte - după) sunt 12, 7, 10, 5, 13, 8, 15, 4, 9, 6, 11, 2; Shapiro-Wilk pe diferențe p > 0,99: testul t pereche. (d) Media diferențelor $\text{md} = \frac{102}{12} = 8{,}5$ mmHg; suma pătratelor abaterilor diferențelor este 167, deci $\text{sd} = \sqrt{\frac{167}{11}} = \sqrt{15{,}18} = 3{,}90$; $\text{ES} = \frac{3{,}90}{\sqrt{12}} = 1{,}125$; $t = \frac{8{,}5}{1{,}125} = 7{,}56$, gl 11. Valoarea critică bilaterală este 2,201; 7,56 este mult mai mare, deci p < 0,001 (programul dă p = 0,00001). IC 95% pentru md: $8{,}5 \pm 2{,}201 \times 1{,}125 = 8{,}5 \pm 2{,}48$, adică 6,0-11,0 mmHg; $\text{dz} = \frac{8{,}5}{3{,}90} = 2{,}18$. Testul Wilcoxon dă, ca verificare, p < 0,001 (toate cele 12 diferențe sunt pozitive). (f) Respingem H0. (g) TAS a scăzut în medie cu 8,5 mmHg (IC 95% 6,0-11,0). Fără grup de control, scăderea nu poate fi atribuită cu certitudine programului: regresia spre medie și efectul de a fi urmărit sunt explicații alternative.

Cerința 4. Media IMC este 30,02 kg/m². Sumele de lucru: $S_{\text{xy}} = \sum(x - m_x)\left(y - m_y\right) = 121{,}85$; $S_{\text{xx}} = \sum(x - m_x)^2 = 70{,}06$; Syy = 574,25.

$$r = \frac{S_{\text{xy}}}{\sqrt{S_{\text{xx}} \times S_{\text{yy}}}} = \frac{121{,}85}{\sqrt{70{,}06 \times 574{,}25}} = \frac{121{,}85}{200{,}57} = 0{,}608 t = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}608 \times 3{,}162}{\sqrt{0{,}630}} = \frac{1{,}922}{0{,}794} = 2{,}42$$

Cum 2,42 > 2,228, respingem H0: ρ = 0 (p = 0,036). $R^2 = 0{,}369$. Panta $b = \frac{S_{\text{xy}}}{S_{\text{xx}}} = \frac{121{,}85}{70{,}06} = 1{,}739$; interceptul a = 152,75 - 1,739 x 30,02 = 100,54. Dreapta: TAS = 100,54 + 1,739 x IMC; la IMC 30, TAS prezisă 152,7 mmHg. Corelația Spearman (rho = 0,629, p = 0,028) duce la aceeași concluzie. Interpretare: corelație pozitivă moderată spre bună; IMC explică aproximativ 37% din variabilitatea TAS în acest grup mic. Cu 12 pacienți, intervalul de încredere pentru pantă este foarte larg (0,14-3,34 mmHg pe kg/m²), deci estimarea este imprecisă.

Cerința 5. Tabelul: fumători 18 cu AVC, 132 fără; nefumători 10 cu AVC, 240 fără. Riscuri: 18/150 = 0,12 și 10/250 = 0,04.

$\mathrm{RR} = \frac{0{,}12}{0{,}04} = 3{,}0;\ \mathrm{ES}(\ln \mathrm{RR}) = \sqrt{\frac{1}{18} - \frac{1}{150} + \frac{1}{10} - \frac{1}{250}} = \sqrt{0{,}0556 - 0{,}0067 + 0{,}1000 - 0{,}0040} = \sqrt{0{,}1449} = 0{,}381;\ \mathrm{IC}\,95\%\,\mathrm{RR} = \exp(1{,}099 \pm 1{,}96 \times 0{,}381) = \exp(0{,}352)$ la $\exp(1{,}845) = 1{,}42$ la $6{,}33;\ \mathrm{OR} = \frac{18\times240}{132\times10} = \frac{4320}{1320} = 3{,}27(;\ \mathrm{IC}\,95\%\,1{,}47 - 7{,}30);\ \mathrm{DR} = 0{,}12 - 0{,}04 = 0{,}08(;\ \mathrm{IC}\,95\%\,0{,}023 - 0{,}137);\ \mathrm{NNH} = \frac{1}{0{,}08} = 12{,}5$, adică 13; FA la expuși = (3 - 1)/3 = 0,667

Frecvențele așteptate (cea mai mică 10,5) permit chi-pătrat: χ^2 = 9,22, gl 1, p = 0,002 (cu corecția Yates 8,03, p = 0,005). Concluzie: în această cohortă, fumătorii hipertensivi au un risc de AVC de 3 ori mai mare; la fiecare 13 fumători apare un AVC în plus în 5 ani față de nefumători; dacă relația este cauzală, două treimi dintre AVC-urile fumătorilor ar fi atribuibile fumatului. Fiind un design de cohortă, RR este măsura potrivită; OR-ul (3,27) este apropiat, pentru că AVC-ul este relativ rar.

Cerința 6. Bolnavi: 150; sănătoși: 850. AP = 0,90 x 150 = 135; FN = 15; AN = 0,80 x 850 = 680; FP = 170.

Tabelul L14.3. Tabelul reconstituit pentru testul rapid (n = 1000, prevalență 15%).

	Hipertrofie da	**Hipertrofie nu**	**Total**
Test pozitiv	135	170	305
Test negativ	15	680	695
Total	150	850	1000

$\mathrm{VPP} = \frac{135}{305} = 44{,}3\%;\ \mathrm{VPN} = \frac{680}{695} = 97{,}8\%$; acuratețea = 815/1000 = 81,5%; rata de eroare = (170 + 15)/1000 = 18,5%; $\mathrm{RV}^{+} = \frac{0{,}90}{0{,}20} = 4{,}5;\ \mathrm{RV}^{-} = \frac{0{,}10}{0{,}80} = 0{,}125$. La o prevalență de 2% (20 de bolnavi la 1000): AP 18, FP 196, deci $\mathrm{VPP} = \frac{18}{214} = 8{,}4\%$, iar VPN crește la 99,7%. Interpretare: un rezultat negativ exclude practic hipertrofia; un rezultat pozitiv trebuie confirmat ecocardiografic, mai ales la prevalențe mici.

Cerința 7. a) Testul Mann-Whitney (două grupuri independente mici, asimetrice). b) Testul McNemar (proporții pe date pereche). c) Testul Kruskal-Wallis (variabilă ordinală, trei grupuri independente). d) ANOVA cu măsurători repetate (trei momente, aceiași pacienți, distribuții normale); alternativa neparametrică ar fi testul Friedman.

L14.6 Baremul și greșelile frecvente

Fiecare subiect se punctează din 100 de puncte, repartizate pe cerințe proporțional cu dificultatea. În interiorul unei cerințe de testare, punctajul orientativ este următorul.

Tabelul L14.4. Repartizarea orientativă a punctajului pentru o cerință de testare statistică (10 puncte).

Element	**Ce se cere**	**Puncte**
(a)-(b) Ipotezele	ipoteza de cercetare; H0 și H1 în cuvinte și matematic	2
(c) Metoda	testul corect, justificat prin tipul variabilei, numărul de grupuri, independența și verificarea presupunerilor	3
(d)-(e) Rezolvarea și rezultatul	programul și meniul; statistica testului, gradele de libertate, p,	3

Element	Ce se cere	Puncte
	mărimea efectului sau IC; graficul cerut	
(f) Concluzia statistică	"se respinge H0" sau "nu se respinge H0", cu pragul	1
(g) Concluzia de cercetare	interpretarea medicală, cu sensul și mărimea efectului, fără exagerări	1

Pentru cerințele de descriere se punctează alegerea indicatorilor după forma distribuției, cifrele și graficul. Pentru cerințele de estimare, formula sau meniul, intervalul corect și interpretarea lui. Pentru cerințele de risc și de test diagnostic, tabelul 2 x 2 corect orientat, măsura potrivită designului și interpretarea.

Greșelile care costă cele mai multe puncte, din experiența sesiunilor anterioare:

- Test pentru grupuri independente aplicat pe date pereche (sau invers). Este greșeala cu cel mai mare cost, pentru că invalidează întreaga cerință.
- Lipsa verificării normalității înainte de un test parametric, sau verificarea ei pe variabila greșită (la t pereche, normalitatea se verifică pe diferențe, nu pe fiecare moment).
- "Se acceptă H0" în loc de "nu se respinge H0"; "rezultatul dovedește că nu există diferență".
- Confuzia dintre intervalul de încredere pentru medie și intervalul pentru valorile individuale.
- OR raportat într-un studiu de cohortă când se cere RR, sau RR calculat într-un studiu caz-control.
- Tabelul 2 x 2 orientat greșit (bolnavii pe rânduri și testul pe coloane, apoi formulele aplicate ca și cum ar fi invers), ceea ce schimbă Se cu VPP.
- Rata de eroare calculată ca (a + d)/N; formula corectă este (b + c)/N.
- Concluzii cauzale dintr-o corelație sau dintr-un studiu transversal.
- Valori p scrise "p = 0,000" (se scrie p < 0,001) și zecimale în exces.
- Pragul scorului din MedCalc transcris greșit ("> 10" înseamnă "≥ 11" pentru valori întregi).
- Graficul nepotrivit: diagramă circulară pentru o variabilă numerică, histogramă pentru o variabilă nominală.

L14.7 Planul de pregătire în două săptămâni

Planul presupune aproximativ o oră și jumătate pe zi, cinci zile pe săptămână, plus o simulare de examen la sfârșitul fiecărei săptămâni.

Săptămâna 1, consolidarea.

- Ziua 1: tipurile de variabile, statistica descriptivă și graficele (capitolul 2, lucrările 1-2). Pe mg2026, descrieți toate variabilele și alegeți indicatorii; refaceți cerința 1 din subiectul 1.
- Ziua 2: distribuția normală, testele de normalitate, scorul z, transformarea logaritmică (capitolul 3). Pe mg2026: TAS, IMC, trigliceridele înainte și după logaritmare.
- Ziua 3: intervalele de încredere și dimensiunea eșantionului (capitolul 4). Calculați de mână și în program IC pentru TAS, HDL, proporția diabetului (Wald și Wilson).
- Ziua 4: testele t, ANOVA și echivalentele neparametrice (capitolele 6-8). Refaceți cerințele 3, 5 și 7 din subiectul 1 și cerința 3 din subiectul 2, fără să vă uitați la rezolvare.
- Ziua 5: chi-pătrat, riscul relativ, raportul cotelor, NNT (capitolele 7 și 10). Problemele cu fumat-BPOC, monitor-acuitate vizuală, contraceptive-artrită reumatoidă de la lucrări; cerința 6 din subiectul 1 și cerința 5 din subiectul 2.
- Sfârșitul săptămânii: subiectul model 1 în 60 de minute, cu cronometru, apoi corectare cu baremul.

Săptămâna 2, integrarea.

- Ziua 6: corelația și regresia (capitolul 9). Calculați de mână r pentru subiectul 2; pe mg2026, IMC-TAS, vârstă-TAS, talie-glicemie; regresia multiplă.
- Ziua 7: raționamentul bayesian și testele diagnostice (capitolele 11-12). Reconstituiți tabele 2 x 2 din Se, Sp și prevalență sau din Se și VPP; curba ROC a scorului FINDRISC în MedCalc.
- Ziua 8: alegerea testului (capitolul 8 și tabelul 14.1). Luați 20 de întrebări din exercițiile capitolelor și, pentru fiecare, scrieți doar testul și justificarea, în 1 minut.
- Ziua 9: redactarea (secțiunea 14.4). Rescrieți concluziile de la toate cerințele rezolvate până acum în formatul (f)-(g), verificând formulările interzise.
- Ziua 10: recapitulare pe fișierele de la lucrări (BD_Studenti, BPOC, Date_HbA1c), câte o întrebare din fiecare tip.
- Sfârșitul săptămânii: subiectul model 2 în 60 de minute, apoi o variantă proprie: schimbați variabilele din subiectul 1 (de exemplu, colesterolul în loc de HDL, mediul în loc de sex) și rezolvați-o.

L14.8 Lista de verificare a deprinderilor de lucru în programe

Bifați fiecare deprindere doar după ce ați executat-o singuri, fără fișa de lucru.

Google Sheets:

- ☐ import CSV, setările regionale România, separatorul ";";
- ☐ =AVERAGE, =MEDIAN, =STDEV, =MIN, =MAX, =QUARTILE, =COUNTIF, =COUNTBLANK;
- ☐ IC pentru medie cu =T.INV.2T și =SQRT;
- ☐ tabel pivot pentru frecvențe și tabele de contingență;
- ☐ histogramă și box-plot din meniul "Inserare" > "Diagramă".

JASP:

- ☐ deschiderea unui CSV și verificarea tipului fiecărei variabile (nominal, ordinal, scale);
- ☐ coloană calculată (diferențe, IMC, dihotomizare la un prag) și filtrarea pe un subgrup;
- ☐ Descriptive Statistics cu Shapiro-Wilk, histograme, box-plot, "Split" pe grupuri;
- ☐ One Sample, Independent Samples (cu Welch și Mann-Whitney) și Paired Samples T-Test (cu Wilcoxon);
- ☐ ANOVA cu Levene, Tukey și eta²; Kruskal-Wallis; Repeated Measures ANOVA și Friedman;
- ☐ Contingency Tables cu χ^2, frecvențe așteptate, procente pe rând și log odds ratio;
- ☐ Correlation (Pearson, Spearman) cu diagrama de dispersie; Linear Regression simplă și multiplă; Logistic Regression cu OR.

MedCalc:

- ☐ Tests > Relative risk și Odds ratio din frecvențe introduse direct;
- ☐ Statistics > ROC curves > ROC curve analysis, cu tabelul pragurilor și indicele Youden;
- ☐ calculatorul de dimensiune a eșantionului pentru o medie și pentru o proporție.

Epi Info:

- ☐ StatCalc > 2 x 2 tables (RR, OR, chi-pătrat, Fisher);
- ☐ StatCalc > Sample size pentru un studiu transversal.

Calcul de mână (cu calculatorul de buzunar):

- ☐ media, AS, ES, IC 95% pentru medie și pentru o proporție;
- ☐ t pereche din diferențe; r din sumele de produse;
- ☐ RR, OR, DR, NNT cu IC; Se, Sp, VPP, VPN, RV, rata de eroare; probabilitatea post-test prin cote.

Exerciții suplimentare

1. Pe fișierul BD_Studenti, alcătuiți un subiect de examen cu 6 cerințe după modelul subiectului 1 (o descriere, un IC, un test pentru două grupuri independente, un test pentru date pereche, o asociere între variabile categoriale, o corelație) și schimbați-l cu un coleg; fiecare rezolvă subiectul celuilalt și îl corectează după baremul din L14.6.
2. Pe fișierul BPOC, rezolvați în formatul (a)-(g): "Frecvența cardiacă se modifică semnificativ după tratament?" și "Vârsta de peste 50 de ani este factor de risc pentru TAS inițială peste 140 mmHg?". Precizați designul fiecărei întrebări și de ce măsura potrivită pentru a doua este raportul prevalențelor sau OR, nu RR.
3. Pe fișierul Date_HbA1c, alegeți o variabilă asimetrică și una simetrică și scrieți descrierea lor completă, cu grafic.
4. Problemă cu date în enunț. Un test are Se 80% și VPP 50% într-un eșantion de 400 de persoane cu prevalența 10%. Reconstituiți tabelul. Rezolvare: bolnavi 40, AP 32, FN 8; din VPP 50%, AP + FP = 64, deci FP 32; AN = 360 - 32 = 328. Sp $= \frac{328}{360} = 91{,}1\%$; VPN $= \frac{328}{336} = 97{,}6\%$; acuratețea (32 + 328)/400 = 90,0%; rata de eroare (32 + 8)/400 = 10,0%; $RV^+ = \frac{0{,}80}{0{,}089} = 9{,}0$.
5. Problemă cu date în enunț. La 10 pacienți, glicemia înainte și după un tratament are diferențele (mg/dL): 15, 8, 22, -3, 10, 12, 5, 18, 9, 14. Calculați media diferențelor, AS, t și concluzia (t(9; 0,975) = 2,262). Rezolvare: md $= \frac{110}{10} = 11{,}0$; suma pătratelor abaterilor de la medie este 442, sd $= \sqrt{\frac{442}{9}} = \sqrt{49{,}11} = 7{,}01$; ES $= \frac{7{,}01}{\sqrt{10}} = 2{,}22$; $t = \frac{11{,}0}{2{,}22} = 4{,}96 > 2{,}262$, deci respingem H0 ($p < 0{,}001$); IC 95% $11{,}0 \pm 2{,}262 \times 2{,}22 = 6{,}0 - 16{,}0$ mg/dL. Normalitatea diferențelor nu este respinsă (Shapiro-Wilk p = 0,97).

Ce se predă și barem orientativ

La simularea din laborator se predă pe Moodle documentul cu rezolvarea subiectului model 1, în formatul (a)-(g), cu graficele inserate. Cadrul didactic îl corectează după baremul din L14.6 și vă indică, pentru fiecare cerință, ce element a lipsit. La examenul propriu-zis, punctajul obținut se convertește în nota examenului practic, care reprezintă 30% din nota finală.

Tabelul L14.5. Repartizarea orientativă a celor 100 de puncte pe cerințele subiectului 1.

Cerința	Conținut	Puncte
1	descriere, alegerea indicatorilor, grafice	12
2	IC pentru medie și intervalul valorilor individuale	10
3	test pentru două grupuri independente	14
4	corelație și regresie, predicție	14
5	ANOVA cu post-hoc	14
6	chi-pătrat și măsura asocierii	12
7	test pentru date pereche	12
8	test diagnostic, tabel 2 x 2	12
Total		100

Varianta fără rețea

Dacă programele nu pot fi folosite, examenul se poate susține în varianta "pe hârtie", cu subiectul de tipul 2: setul mic de date este dat în enunț, iar calculele se fac cu calculatorul de buzunar, pe baza tabelului-sinteză al formulelor (tabelul 14.2) și a tabelelor de valori critice t, χ^2 și z primite odată cu subiectul. Graficele se schițează de mână (histogramă, box-plot, diagramă de dispersie cu dreapta de regresie). Pentru pregătire, rezolvați subiectul 2 integral de mână și verificați apoi cifrele în JASP sau MedCalc, când aveți acces la ele.

Anexa A. Răspunsuri la întrebările de autoevaluare și la exercițiile propuse

Răspunsurile sunt date pe capitole. La întrebările de tip complement multiplu se notează toate variantele corecte; la exercițiile cu calcule se indică rezultatul și metoda, iar rezolvarea completă se reface după modelul problemelor rezolvate din capitol.

Capitolul 1

Întrebări: 1 b; 2 c; 3 c; 4 b; 5 c; 6 a, b, d; 7 b, c, d; 8 a, b, d; 9 a, b, d; 10 a, c, d

Explicații: La întrebarea 3, capcana este varianta d (0,621), care este $P(\text{obez} \mid \text{diabet}) = \frac{18}{29}$, nu $P(\text{diabet} \mid \text{obez}) = \frac{18}{54} = 0{,}333$; condiția indică totalul la care se raportează celula. La întrebarea 8, două evenimente mutual exclusive cu probabilități nenule nu pot fi independente: dacă s-a produs B, probabilitatea lui A devine zero, deci P(A|B) = 0 ≠ P(A); iar P(A) + P(B) = P(A sau B) nu poate depăși 1. La întrebarea 7, un interval de încredere îngust arată o eroare aleatoare mică, dar nu spune nimic despre eroarea sistematică; un studiu foarte mare poate fi foarte precis și totuși deplasat.

Exerciții propuse: 1. $P(D \mid \text{urban}) = \frac{19}{95} = 0{,}200$; $P(D \mid \text{rural}) = \frac{10}{65} = 0{,}154$; $P(\text{urban și } D) = \frac{19}{160} = 0{,}119$; $P(\text{urban}) \times P(D) = 0{,}594 \times 0{,}181 = 0{,}108$. Valorile sunt apropiate, dar nu identice; diferența este mică și poate fi întâmplătoare, deci evenimentele sunt aproximativ independente în eșantion (decizia formală cere un test chi-pătrat). 2. $P(\text{fumător} \mid \text{femeie}) = \frac{30}{86} = 0{,}349$; $P(\text{femeie} \mid \text{fumător}) = \frac{30}{53} = 0{,}566$; $P(\text{femeie sau fumător}) = \frac{86+53-30}{160} = \frac{109}{160} = 0{,}681$. Primele două diferă pentru că se raportează la totaluri diferite (86 de femei, respectiv 53 de fumători). 3. Binomială n = 8, p = 0,25: $(a)\ 0{,}75^8 = 0{,}100$; $(b)\ 8 \times 0{,}25 \times 0{,}75^7 = 0{,}267$; (c) 1 - 0,100 - 0,267 = 0,633. 4. $k = \frac{2\,400}{120} = 20$; dosarele 9, 29, 49, 69, 89,..., ultimul 9 + 20 x 119 = 2 389. 5. Fracția de eșantionare 90/1 500 = 6%; 600 x 0,06 = 36 (18-39 ani), 36 (40-64 ani), 300 x 0,06 = 18 (≥ 65 ani). 6. (a) eroare de informare (de raportare, dezirabilitate socială: consumul este subraportat); (b) confuzie (activitatea fizică este legată și de suplimente, și de riscul de fractură); (c) eroare de selecție prin pierderi din urmărire (cei rămași nu mai reprezintă grupul randomizat); (d) eroare de selecție de tip spectru (pacienți cu boală avansată față de voluntari sănătoși; performanța testului va fi supraestimată). 7. Unitatea statistică: rețeta; n = 1 498; populația țintă: prescripțiile (și, indirect, vârstnicii tratați) din mediul urban din vestul României; studiu transversal; variabile numerice: vârsta, numărul de medicamente pe rețetă; variabile calitative: sexul, prezența unui medicament potențial inadecvat (da/nu); dacă un pacient are mai multe rețete, observațiile nu mai sunt independente, iar pacienții cu multe rețete cântăresc mai mult în rezultate. 8. Poisson $\lambda = 1{,}5$: $P(X = 0) = e^{-1{,}5} = 0{,}223$; $P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0{,}934 = 0{,}066$. 9. Probabilitatea totală: $P(\text{BPOC}) = 0{,}30 \times 0{,}15 + 0{,}70 \times 0{,}03 = 0{,}045 + 0{,}021 = 0{,}066 (6{,}6\%)$; Bayes: $P(\text{fumător} \mid \text{BPOC}) = \frac{0{,}045}{0{,}066} = 0{,}682 (68{,}2\%)$. 10. (a) P: adulți cu risc cardiometabolic din cabinetele de medicină de familie; I: consiliere + aplicație + teleconsultație lunară (brațul C); C: consiliere standard (brațul A); O: modificarea tensiunii sistolice de la includere la 6 luni (tas0 - tas6); T: 6 luni. (b) Datele sunt pseudonimizate (există o listă de corespondență care permite reidentificarea); rămân date cu caracter personal, din categoria specială a datelor privind sănătatea, supuse GDPR și Legii 190/2018.

Capitolul 2

Întrebări: 1 b; 2 c; 3 b; 4 b; 5 b; 6 a, b, c; 7 a, b, d, e; 8 a, b, c, e; 9 a, b, c, e; 10 a, b, d, e

Explicații: La întrebarea 2, abaterea standard (7,0) este mai mare decât jumătate din medie (11,6) la o variabilă care nu poate fi negativă, deci distribuția este asimetrică spre dreapta și mediana descrie mai bine valoarea tipică. La întrebarea 5, n = 8 este par, iar mediana este media valorilor din pozițiile 4 și 5, (4 + 5)/2 = 4,5. La întrebarea 7, varianta c este falsă: abaterea standard descrie variabilitatea pacienților și nu scade sistematic când n crește; doar eroarea standard, $\frac{s}{\sqrt{n}}$, scade. La întrebarea 10, varianta c este falsă: o valoare extremă plauzibilă biologic și verificată nu se elimină; regula 1,5 x IQR semnalează valori de verificat.

Exerciții propuse: 1. Grupa sanguină: nominală; scorul Apgar: ordinală (scor discret 0-10); numărul de sarcini: numerică discretă, scală de raport; temperatura corporală: numerică continuă, scală de interval; clasa NYHA: ordinală; HbA1c: numerică continuă, scală de raport; acordul Likert: ordinală; codul pacientului: identificator, nu se analizează; durata spitalizării: numerică (discretă când se numără zile întregi), scală de raport; fumat da/nu: nominală dihotomică. 2. A = 0,35 x 40 = 14; B = 15,0 + 35,0 = 50,0%; $C = \frac{10}{40} = 25{,}0\%$ (verificare 50,0 + 25,0 = 75,0); D = 0,125 x 40 = 5; E = 40 - 6 - 14 - 10 - 5 = 5 (12,5%); glicemie ≥ 126 mg/dL: 12,5 + 12,5 = 25,0% (10 pacienți). 3. $m = \frac{800}{6} = 133{,}33$ mmHg; mediana (130 + 136)/2 = 133 mmHg; AMA $= \frac{56}{6} =$ 9,33 mmHg; suma pătratelor abaterilor 693,33; $s^2 = \frac{693{,}33}{5} = 138{,}67$ mmHg^2; s = 11,78 mmHg; CV = 8,8%; ES $= \frac{11{,}78}{\sqrt{6}} = 4{,}81$ mmHg. 4. Media 85/10 = 8,5 zile; mediana (5 + 6)/2 = 5,5 zile; două moduri, 4 și 5 (distribuție bimodală în eșantion); Q1 = 4, Q3 = 8 (medianele jumătăților; programele dau 4,25 și 7,75), IQR = 4; limitele -2 și 14, deci 31 de zile este valoare extremă (12 nu este); fără ea media devine 54/9 = 6,0 zile; se raportează mediana 5,5 (Q1-Q3 4-8). 5. k = 1 + 3,322 x log10(250) = 1 + 3,322 x 2,398 = 8,97, adică 9 clase; $h = \frac{412-68}{9} = 38{,}2$ mg/dL, rotunjit la 40 mg/dL (de exemplu, clase 60-99, 100-139,..., 400-439). 6. Dacă 3 este ES, AS = 3 x √64 = 24 mg/dL (CV 21%, plauzibil); dacă 3 ar fi AS, CV ar fi 2,7%, o omogenitate neplauzibilă biologic, deci interpretarea ca ES este cea plauzibilă; ES pentru TAS la femei = 14,47/√86 = 1,56 mmHg. 7. Media ponderată = $\frac{120\times45{,}2+60\times61{,}8+20\times70{,}4}{200} = \frac{10\,540}{200} = 52{,}7$ ani; media simplă a mediilor, 59,13 ani, supraestimează vârsta pentru că dă secției mici (20 de pacienți) aceeași pondere ca secției mari. 8. $\frac{(a)\,26}{53} = 49{,}1\%$; (b) IMC, trigliceridele și scorul FINDRISC sunt raportate ca mediană (Q1-Q3) pentru că sunt asimetrice sau sunt scoruri; (c) după randomizare, diferențele inițiale dintre brațe sunt prin definiție întâmplătoare, deci un test ar verifica o ipoteză nulă despre care știm că este adevărată; se judecă mărimea clinică a dezechilibrelor, nu valoarea p. 9. (a) diagramă cu bare (o diagramă circulară cu 4 categorii este acceptabilă); (b) box-plot alăturate pe brațe; (c) diagramă de dispersie; (d) grafic liniar; (e) histogramă; (f) bare stivuite la 100%, câte una pentru fiecare sex, cu categoriile în ordinea scalei. 10. Media aritmetică = 1 111 000/4 = 277 750 copii/mL; logaritmii zecimali sunt 3, 4, 5, 6, cu media 4,5, deci media geometrică = 10 la puterea 4,5 = 31 623 copii/mL; încărcătura virală variază multiplicativ (distribuție aproximativ log-normală), iar pe scară logaritmică o scădere de "1 log" înseamnă o reducere de 10 ori, ușor de comparat între pacienți.

Capitolul 3

Întrebări: 1 c; 2 b; 3 b; 4 b; 5 c; 6 a, b, c, e; 7 b, c, e; 8 a, b, c, d; 9 b, c, d, e; 10 a, b, d, e

Explicații: La întrebarea 4, m - 2s = 11,6 - 14,0 = -2,4 zile; o variabilă care nu poate fi negativă și are m - 2s sub zero nu poate fi normală, iar duratele de spitalizare au aproape

întotdeauna coadă lungă spre dreapta. La întrebarea 7, un p de 0,30 înseamnă doar că nu se respinge ipoteza de normalitate; cu 12 observații testul are putere mică, deci nu "dovedește" nimic și nu obligă la un test parametric. La întrebarea 10, varianta c este falsă: prin construcție, 5% dintre persoanele sănătoase au valori în afara intervalului, iar 1 - 0,95^20 = 0,64 explică varianta e.

Exerciții propuse:

1. z(125) = (125 - 131,13) / 13,88 = -0,44, Φ = 0,3300, percentila 33; z(150) = 1,36, Φ = 0,9131, percentila 91. În Sheets: =NORM.S.DIST(-0,44; TRUE) și =NORM.S.DIST(1,36; TRUE).
2. z(240) = 0,83, proporția teoretică 1 - 0,7967 = 20,3% (observat 23,8%); $z(190) = -0{,}53$, proporția teoretică 29,8% (observat 30,0%). Potrivire bună; coada dreaptă este puțin mai groasă decât prezice modelul, diferență de mărimea fluctuațiilor de eșantionare.
3. $49{,}52 \pm 1{,}96 \times 11{,}78 = 49{,}52 \pm 23{,}09$, adică 26,4-72,6 mg/dL; teoretic 95% (observat 94,4%). HDL < 40: z = -0,81, Φ = 0,2090, adică 20,9% (observat 19,4%).
4. Binomială n = 6, p = 0,33: media 1,98; varianța 6 x 0,33 x 0,67 = 1,33; $P(0) = 0{,}67^6 = 0{,}0905$; $P(1) = 6 \times 0{,}33 \times 0{,}67^5 = 0{,}2673$; $P(X \geq 2) = 1 - 0{,}0905 - 0{,}2673 = 0{,}6422$.
5. Poisson $\lambda = 0{,}5$: $P(X \geq 1) = 1 - e^{-0{,}5} = 1 - 0{,}6065 = 0{,}3935$; $P(X \geq 3) = 1 - (0{,}6065 + 0{,}3033 + 0{,}0758) = 0{,}0144$.
6. $4{,}3 \pm 1{,}96 \times 0{,}35 = 4{,}3 \pm 0{,}69$, adică 3,61-4,99 mmol/L; $z = \frac{5{,}1-4{,}3}{0{,}35} = 2{,}29$, Φ = 0,9890, percentila 98,9; valoarea este în afara intervalului și se repetă (hemoliza probei poate crește fals potasemia) înainte de orice decizie.
7. m - 2s = 9,1 - 11,6 = -2,5 zile: distribuția este asimetrică spre dreapta; mai potrivite ar fi fost mediana și IQR (eventual minimul și maximul), iar comparația cu un test neparametric sau pe scală logaritmică.
8. 0,39 / 0,192 = 2,03 și 0,74 / 0,381 = 1,94: abatere la limită, confirmată de Shapiro-Wilk p = 0,034; scorul este discret, cu multe valori egale. Se raportează mediana (IQR), eventual și media ± AS; pentru comparații cu n = 160 testul t este robust, dar un test neparametric (Mann-Whitney, Kruskal-Wallis) este o alegere sigură; ideal se raportează că ambele duc la aceeași concluzie.
9. Media geometrică $e^{4{,}97} = 144{,}0$; factorul $e^{0{,}49} = 1{,}63$; intervalul de 95%: $e^{4{,}97-0{,}96} = e^{4{,}01} = 55{,}1$ până la $e^{4{,}97+0{,}96} = e^{5{,}93} = 376{,}3$ (aproximativ 55-376, asimetric față de 144).
10. IQR = 23; gardurile 93 - 34,5 = 58,5 și 116 + 34,5 = 150,5, deci 154 este valoare extremă după regula Tukey; $z = \frac{154-105{,}44}{15{,}76} = 3{,}08$, deci și după criteriul z. Valoarea este plauzibilă într-o populație cu diabet nediagnosticat: se verifică în buletinul de laborator, se păstrează și se face o analiză de sensibilitate.
11.
    (a) puncte pe linie; (b) curbă în formă de "U": puncte deasupra liniei la ambele capete, cu urcare abruptă la dreapta (asimetrie pozitivă); (c) formă de "S" culcat: puncte sub linie la stânga și peste linie la dreapta; (d) puncte pe linie, cu un singur punct mult deasupra liniei în colțul din dreapta sus.

Capitolul 4

Întrebări: 1 b; 2 c; 3 b; 4 d; 5 c; 6 a, c, d, e; 7 a, b, c, e; 8 a, b, c, d; 9 a, b, c, e; 10 a, b, c, d

Explicații: La întrebarea 3, intervalul 0,90-2,18 include valoarea nulă 1, deci p ≥ 0,05; asta nu dovedește absența asocierii (varianta c), ci arată că datele sunt compatibile atât cu o scădere, cât și cu o dublare a șanselor. La întrebarea 4 se aplică regula lui trei: 3 / 60 = 0,05, adică 5%. La întrebarea 9, varianta d este falsă: un prag α mai permisiv (0,10) micșorează valoarea critică și

deci eșantionul necesar, cu prețul mai multor rezultate fals pozitive. La întrebarea 10, varianta e confundă precizia cu mărimea efectului: un interval larg arată o estimare imprecisă.

Exerciții propuse:

1. ES = $\frac{36,91}{\sqrt{160}} = 2,918$ mg/dL; marja 1,975 x 2,918 = 5,76; IC 95% 203,63-215,15 mg/dL.
2. ES = $\frac{11,78}{\sqrt{160}} = 0,931$ mg/dL. Valori individuale: 68,3% 37,74-61,30; 95,4% 25,96-73,08; 99,7% 14,18-84,86 mg/dL. Medie: 68,3% 48,59-50,45; 95,4% 47,66-51,38; 99,7% 46,73-52,31 mg/dL.
3. ES = $\frac{1,5}{\sqrt{25}} = 0,30\%$; t(24; 0,975) = 2,064 (n mic, σ estimat din eșantion); marja 0,62; IC 95% 6,58-7,82%.
4. $p = \frac{41}{160} = 25,6\%$; ES = 0,0345; Wald 18,9-32,4%; Wilson 19,5-32,9%.
5. Regula lui trei: 3 / 120 = 2,5% (metoda exactă dă 3,0%).
6. Diferența 8 mmHg; ES = $\sqrt{\frac{15^2}{40} + \frac{14^2}{45}} = \sqrt{5,625 + 4,356} = 3,16$ mmHg; IC 95% aproximativ 8 ± 6,19, adică 1,8-14,2 mmHg (cu t Welch, aproximativ 80 gl, 1,7-14,3). Intervalul exclude 0, deci diferența este semnificativă la pragul 0,05, dar estimarea este imprecisă.
7. OR = $\frac{20\times40}{30\times10} = 2,67$; ln OR = 0,981; ES = $\sqrt{\frac{1}{20} + \frac{1}{30} + \frac{1}{10} + \frac{1}{40}} = 0,456$; 0,981 ± 0,895 dă 0,086-1,876; IC 95% 1,09-6,52. Exclude 1 (asociere semnificativă), dar limita inferioară este aproape de 1.
8. (a) $n = \left(\frac{1,96\times10}{2}\right)^2 = 96,04$, deci 97; $(b)\ n = \frac{1,96^2\times0,25}{0,04^2} = 600,25$, deci 601.
9. $n = \frac{2\times(1,96+0,84)^2}{0,4^2} = \frac{15,68}{0,16} = 98,0$, deci 99 pe grup (G*Power: 100); cu 20% pierderi, 99 / 0,8 = 123,75, deci 124 pe grup (125 pornind de la 100).
10. Valoarea 1,1 este aproape sigur eroarea standard; AS = $1,1 \times \sqrt{160} = 13,9$ mmHg. OR 1,75 (IC 95% 0,95-3,10): intervalul include 1, deci asocierea nu este semnificativă la pragul 0,05.
11. Lățimea scade cu $\sqrt{n}$: pentru a o înjumătăți, n trebuie mărit de 4 ori, deci aproximativ 160 de pacienți.

Capitolul 5

Întrebări: 1 c; 2 b; 3 b; 4 d; 5 d; 6 a, b, c, e; 7 a, b, d; 8 b, c, e; 9 a, b, d; 10 a, b, d, e

Explicații: La întrebarea 3, valoarea 132 mmHg se află în IC 95% (128,96-133,30), deci testul bilateral la α = 0,05 nu respinge H0; varianta c este greșită pentru că "acceptarea H0" nu este o concluzie permisă, iar media populației nu este cunoscută sigur. La întrebarea 7, trecerea de la un test unilateral justificat la unul bilateral împarte α în două cozi și scade puterea; creșterea abaterii standard mărește zgomotul și scade puterea. La întrebarea 8, varianta d generalizează abuziv: un rezultat nesemnificativ nu demonstrează absența unui efect, iar intervalul de încredere arată doar ce diferențe sunt excluse în populația studiată.

Exerciții propuse: 1. Testul t pentru o medie: ES = $\frac{0,45}{\sqrt{160}} = 0,0356$; $t = \frac{5,58-5,7}{0,0356} = -3,37$ (cu media exactă 5,5775, t = -3,48), gl = 159, p < 0,001; se respinge H0. IC 95% pentru medie 5,51-5,65%. Media HbA1c în populația studiată este semnificativ sub 5,7%. 2. ES = $\frac{4,27}{12,649} = 0,338$; $t = \frac{10,37-12}{0,338} = -4,83$, p < 0,001; IC 95% 9,70-11,04; $d = -\frac{1,63}{4,27} = -0,38$ (efect mic spre mediu). Se respinge H0; media FINDRISC este sub 12. 3. ES = $\frac{9,30}{12,649} = 0,735$; $t = \frac{2,71}{0,735} = 3,68$, p <

0,001 (aproximativ 0,0003); IC 95% 73,26-76,16 bătăi/min; d = 0,29. Rezultat extrem de semnificativ statistic, dar o diferență de aproximativ 3 bătăi/min față de 72 nu are relevanță clinică. 4. Testul z: ES $= \frac{1,2}{\sqrt{36}} = 0,2$ g/dL; $z = \frac{11,2-11,8}{0,2} = -3,0$; |z| > 1,96 (și > 2,576); p = 0,0027. Se respinge H0, diferență foarte semnificativă: hemoglobina medie a copiilor din comunitate este mai mică decât media populației generale. 5. (a) eroare de tip I (fals pozitiv); (b) eroare de tip II (fals negativ), favorizată de eșantionul mic; (c) decizie corectă (H0 adevărată, nerespinsă). 6. $(a)\ 1 - 0,95^8 = 0,337$; $\frac{(b)\ 0,05}{8} = 0,00625$; $(c)\ 1 - 0,95^4 = 0,185$, prag Bonferroni 0,05 / 4 = 0,0125. 7. Fără corecție: 4 (0,002; 0,009; 0,018; 0,035). Bonferroni (prag 0,01): 2. Holm: 0,002 < 0,01; 0,009 < 0,0125; 0,018 > 0,0167, oprire: 2. Benjamini-Hochberg (praguri 0,01; 0,02; 0,03; 0,04; 0,05): 0,035 ≤ 0,04, deci 4. 8. (a) T.DIST.2T(2,5;12) = 0,028 (între valorile critice 2,179 și 3,055); (b) bilateral 0,097, unilateral 0,048; (c) T.DIST.2T(2,9;159) = 0,004 (funcția primește valoarea absolută a lui t). 9. (a) IC 95% conține 0, deci p > 0,05. (b) Efect nedeterminat: intervalul cuprinde atât 0, cât și diferențe peste pragul clinic de 5 mmHg (până la 9 mmHg). (c) Un studiu mai mare, dimensionat pentru o diferență de 5 mmHg (aproximativ 123-125 de pacienți pe grup la σ = 14, putere 80%). 10. (a) ES $= \frac{3,91}{12,649} = 0,309$; $t = \frac{29,04-30}{0,309} = -3,11$, p = 0,002. (b) Distribuția IMC este asimetrică (1,42) și nenormală; se preferă testul Wilcoxon al rangurilor cu semn pentru mediană față de 30 (p < 0,001) sau analiza pe date logaritmate. (c) Mediana 28,25 kg/m² (IQR 26,2-31,4), testul Wilcoxon, p < 0,001: IMC tipic al pacienților este sub pragul de 30 kg/m². 11. $n = \frac{2\times(1,96+0,84)^2\times14^2}{10^2} = \frac{2\times7,84\times196}{100} = 30,7$, adică 31 de pacienți pe grup (calculul exact cu distribuția t dă 32); $d = \frac{10}{14} = 0,71$.

Capitolul 6

Întrebări: 1 b; 2 c; 3 b; 4 c; 5 b; 6 a, b, c, e; 7 a, b, c; 8 a, b, d; 9 a, b, c; 10 a, b, c, e

Explicații: La întrebarea 2, gradele de libertate ale testului pereche se calculează din numărul de perechi (25 - 1 = 24), nu din numărul total de măsurători. La întrebarea 8, varianta c este falsă: ANOVA arată doar că cel puțin o medie diferă; perechea C - B este identificată de testul post hoc. Varianta e este falsă pentru că $\eta^2 = 0,226$ arată un efect mare, care nu depinde de n. La întrebarea 10, alternativa neparametrică pentru măsurători repetate este testul Friedman; Kruskal-Wallis se aplică grupurilor independente.

Exerciții propuse: 1. Diferențele: 0,6; 0,3; 0,1; 0,9; 0,6; 0,3; 0,8; 0,4; mD = 0,50%; suma pătratelor abaterilor 0,52; $s_D = \sqrt{\frac{0,52}{7}} = 0,273$; ESD = 0,096; t = 5,19, gl = 7, p = 0,0013 (t critic 2,365); IC 95% 0,27-0,73%; dz = 1,83. Se respinge H0: HbA1c a scăzut extrem de semnificativ, în medie cu 0,5 puncte procentuale (studiu fără grup de control). 2. $s^2_{\text{comună}} = \frac{24\times144+29\times196}{53} = 172,45$; scomună = 13,13; ES $= 13,13 \times \sqrt{\frac{1}{25}+\frac{1}{30}} = 3,56$; $t = \frac{7}{3,56} = 1,97$, gl = 53, p = 0,054 (t critic 2,006); IC 95% -0,13 la 14,13 mmHg; d = 0,53. Nu se respinge H0 (la limită); intervalul cuprinde atât 0, cât și diferențe mari, deci rezultatul este nedeterminat, iar studiul are putere insuficientă pentru o diferență de 7 mmHg. 3. Student: $s^2_{\text{comună}} = \frac{11\times1,21+39\times10,24}{50} = 8,25$; ES = 0,946; t = -1,37, gl = 50, p = 0,175. Welch: ES $= \sqrt{\frac{1,21}{12}+\frac{10,24}{40}} = 0,597$; t = -2,18, gl = 48,9, p = 0,034. Varianța comună este dominată de grupul mare, care are varianța mult mai mare (raportul AS aproape 3); pentru grupul mic, cu varianța mică, eroarea standard este supraestimată, iar testul Student devine prea conservator. Testul Welch este cel corect; concluzia: diferență semnificativă. 4. M = 126,33; SS între $= 10 \times [(120-126,33)^2 + (128-126,33)^2 + (131-126,33)^2] =$

$646{,}7(\text{gl } 2,\ \text{MS } 323{,}3)$; SS în interior = 9 × (100 + 144 + 121) = 3285 (gl 27, MS 121,7); F = 2,66 < 3,35; p = 0,088; $\eta^2 = 0{,}16$. Nu se respinge H0. 5. gl între = 3, gl în interior = 36, gl total = 39; MS între = 120, MS în interior = 40; F = 3,0 > 2,87; p = 0,043; $\eta^2 = \frac{360}{1800} = 0{,}20$. Se respinge H0 (diferență semnificativă); urmează testele post hoc. 6. Welch: $\text{ES} = \sqrt{\frac{6{,}49^2}{52} + \frac{4{,}60^2}{84}} = \sqrt{0{,}810 + 0{,}252} = 1{,}030$; $t = \frac{3{,}68}{1{,}030} = 3{,}57$, gl ≈ 82,7, p ≈ 0,0006. scomună = 5,40, d ≈ 0,68 (efect mediu spre mare). Varianta Student ar da t ≈ 3,86. Ambele valori sunt compatibile cu p = 0,001 raportat; diferențele pot proveni din testul folosit (eventual neparametric) sau din rotunjiri. 7. Welch: $\text{ES} = \sqrt{\frac{7{,}0^2}{224} + \frac{5{,}8^2}{62}} = \sqrt{0{,}219 + 0{,}543} = 0{,}873$; $t = \frac{2{,}5}{0{,}873} = 2{,}87$, gl ≈ 115, p ≈ 0,005 (calcul didactic, nu rezultat raportat de autori). AS reprezintă aproximativ 60% din medie, iar media minus 2 AS ar fi negativă: distribuția duratei spitalizării este asimetrică spre dreapta. Pe datele individuale se preferă mediana cu IQR și testul Mann-Whitney sau analiza pe date logaritmate. 8. $\text{ESD} = \frac{9{,}05}{12{,}649} = 0{,}716$; $t = \frac{4{,}78}{0{,}716} = 6{,}68$, gl = 159, p < 0,001; IC 95% 3,36-6,19 mmHg; dz = 0,53. Scăderea medie de aproximativ 5 mmHg este relevantă la nivel de populație, dar analiza pe tot lotul nu are grup de control; efectul intervențiilor se evaluează comparând brațele. 9. (a) t pereche / Wilcoxon al rangurilor cu semn; (b) t independent (Student sau Welch) / Mann-Whitney; (c) ANOVA cu măsurători repetate / Friedman; (d) t independent / Mann-Whitney; (e) ANOVA cu un factor / Kruskal-Wallis; (f) t pentru o singură medie / Wilcoxon pentru o mediană. 10. $\text{HSD} = 3{,}51 \times \sqrt{\frac{121{,}7}{10}} = 3{,}51 \times 3{,}49 = 12{,}2$ mmHg. Diferențele: 8, 11 și 3 mmHg, toate sub 12,2; nicio pereche nu diferă, consecvent cu F nesemnificativ. 11. Testul pereche folosește abaterea standard a diferențelor individuale, care depinde de corelația dintre cele două măsurători și nu se poate deduce din AS de la fiecare moment. Ar fi necesare: numărul de pacienți (nedeclarat în rezumat) și diferența medie cu AS a diferențelor sau cu IC 95% (ori mediana diferențelor cu IQR, dacă s-a folosit testul Wilcoxon).

Capitolul 7

Întrebări: 1 c; 2 b; 3 d; 4 a; 5 c; 6 a, c, e; 7 a, c, d; 8 a, b, c, e; 9 a, c, d, e; 10 a, b, d, e Explicații: la întrebarea 2, cele trei valori 7 ocupă pozițiile 3, 4 și 5 și primesc fiecare rangul mediu 4, iar suma rangurilor (21) confirmă calculul. La întrebarea 7, o frecvență observată 0 nu este o problemă dacă frecvențele așteptate sunt mari (b), iar grupurile inegale nu invalidează testul (e); contează frecvențele așteptate, independența observațiilor și folosirea numerelor absolute. La întrebarea 8, satisfacția nu este o măsură a controlului glicemic (d). La întrebarea 10, valoarea p mică nu măsoară tăria asocierii (c); tăria se citește din φ = 0,28 sau din raportul prevalențelor. Exerciții propuse: 1. Rangurile: 4; 6; 6; 1,5; 8; 6; 3; 1,5; suma $36 = \frac{8\times9}{2}$. 2. R1 = 11, R2 = 34; U1 = 11 - 10 = 1, U2 = 34 - 15 = 19; U = 1 ≤ 1, deci se respinge H0 (p exact 0,032); anxietatea este mai mică în grupul cu informare video (mediana 4 față de 8). 3. Diferențele: 8, 2, -2, 12, 6, 0, 7, 9; pacientul 6 are diferența 0 și se elimină, n = 7; rangurile valorilor absolute: 2 și 2 → 1,5; 6 → 3; 7 → 4; 8 → 5; 9 → 6; 12 → 7; W+ = 26,5, W- = 1,5; T = 1,5 ≤ 2, se respinge H0 (p exact 0,047), tensiunea a scăzut (mediana diferențelor 7 mmHg). 4. Suma $40 + 60 + 71 = 171 = \frac{18\times19}{2}$; $H = \frac{\left[\frac{12}{18\times19}\right]\times(40^2+60^2+71^2)}{6} - 57 = 2{,}89 < 5{,}99$, nu se respinge H0 (p = 0,24). 5. $\chi^2 F = \left[\frac{12}{10\times3\times4}\right] \times (14^2 + 20^2 + 26^2) - 120 = 7{,}2 > 5{,}99$, se respinge H0 (p = 0,027); $W = \frac{7{,}2}{10\times2} = 0{,}36$. 6. Vaccinați 10 cu gripă / 190 fără; nevaccinați 24 / 126; așteptate 19,43 / 180,57 / 14,57 / 135,43; χ^2 = 11,83, p < 0,001 (Yates 10,60, p = 0,001); $\text{RR} = \frac{5{,}0\%}{16{,}0\%} = 0{,}31$ (vaccinarea reduce riscul cu

aproximativ 69% în acest eșantion); φ = 0,18. 7. N = 70; totalurile coloanelor 20, 36, 14; așteptate pe rândul 3: 2,57; 4,63; 1,80 (celula rând 1, coloana 3 are exact 5,0); 3 din 9 celule (33%) au E < 5, peste limita de 20%, deci chi-pătratul nu este aplicabil ca atare; soluții: comasarea rândului 3 cu o categorie vecină dacă are sens clinic, test exact (Fisher-Freeman-Halton) sau mai multe date. 8. McNemar (date pereche): $\chi^2 = \frac{(12-4)^2}{16} = 4{,}00$, p = 0,046; cu corecție (8 - 1)²/16 = 3,06, p = 0,080; varianta exactă p = 0,077; cu doar 16 perechi discordante se preferă varianta corectată sau exactă, deci rezultatul este la limită, nesemnificativ. 9. Așteptate 70, 80, 34, 16; χ^2 = 0 + 0,31 + 0,47 + 0,06 = 0,85, gl 3, p = 0,84; distribuția observată este compatibilă cu proporțiile ipotetice. 10. Așteptate 1,5 și 10,5 pe fiecare rând, deci testul exact Fisher; p = 0,217, nu se respinge H0; cu 3 evenimente din 24 de pacienți puterea este foarte mică, iar rezultatul nu dovedește siguranța medicamentului. 11. Mediane 4 (IQR 3-4) în B și 3 (IQR 2-3) în A; U = 2000 (sau 809, dacă programul raportează U-ul grupului A); p < 0,001; $r_{\text{rb}} = 1 - \frac{2\times809}{2809} = 0{,}42$ (efect mediu spre mare).

Capitolul 8

Întrebări: 1 b; 2 c; 3 c; 4 c; 5 b; 6 a, b, d; 7 a, b, d; 8 a, b, c, e; 9 a, b, d; 10 a, b, c, e Explicații: la întrebarea 4, riscul global este 1 - 0,95³ = 0,143, adică aproximativ 14%, nu 15% (3 x 5% este doar o aproximare). La întrebarea 7, un Shapiro-Wilk semnificativ la n = 300 nu obligă la test neparametric (c), iar alegerea testului după valoarea p este o greșeală de procedură (e). La întrebarea 8, schimbările din interiorul brațelor nu dovedesc eficacitatea (d); doar comparația între brațe arată superioritatea unei intervenții. Exerciții propuse: 1. Variabila: număr de vizite (numerică discretă, puternic asimetrică); două grupuri independente; Mann-Whitney U (alternativa: testul t, doar dacă grupurile sunt mari și asimetria moderată; la nivel avansat, regresie Poisson sau binomială negativă); H0: distribuțiile numărului de vizite sunt identice, H1: diferă. 2. SpO2 (numerică), aceiași 25 de pacienți la două momente, date pereche; testul t pereche dacă diferențele sunt normale, altfel Wilcoxon (SpO2 este mărginită superior și frecvent asimetrică); H0: μd = 0. 3. Variabila vaccinat da/nu (nominală), trei grupuri independente; chi-pătrat pe tabel 3 x 2, gl = 2 (frecvențele așteptate sunt mari); H0: p1 = p2 = p3. 4. Scorul 0-21 (numeric sau ordinal), aceiași 30 de rezidenți în trei condiții, măsurători repetate; ANOVA cu măsurători repetate dacă distribuția o permite, altfel Friedman, apoi post hoc cu corecție. 5. Acordul a doi evaluatori pe o variabilă binară: coeficientul kappa Cohen (testul McNemar arată doar dacă un radiolog spune "pneumonie" sistematic mai des decât celălalt, iar chi-pătratul arată asocierea, nu acordul). 6. Scopul este predicția unui rezultat binar în timp: dacă toți pacienții au fost urmăriți complet 10 ani, regresie logistică (OR); dacă durata de urmărire variază sau există pierderi din urmărire, regresie Cox (HR) pe timpul până la diabet. 7. 1 - 0,95³ = 0,143 (14,3%); pragul Bonferroni 0,05/3 = 0,0167; nicio comparație nu rămâne semnificativă (0,03 și 0,04 > 0,0167); corect era ANOVA urmată de Tukey. 8. Cea mai mică frecvență așteptată = $\frac{12\times10}{60} = 2{,}0 < 5$, deci testul exact Fisher. 9. Femei $208{,}08 \pm 35{,}37$ mg/dL$(n\,86)$, bărbați $210{,}91 \pm 38{,}82(n\,74)$; SW p = 0,61 și 0,93; Levene p = 0,65; t Student = -0,48, gl 158, p = 0,63 (Welch p = 0,63); diferența F - M -2,82 mg/dL (IC 95% -14,50 la 8,85); d = -0,08; Mann-Whitney p = 0,76; nu respingem H0, colesterolul nu diferă între sexe. 10. Rural 10/65 (15,4%), urban 19/95 (20,0%); cea mai mică frecvență așteptată 11,8; χ^2 = 0,55, gl 1, p = 0,46 (Yates p = 0,59, Fisher p = 0,53); nu respingem H0, diabetul nu este asociat cu mediul de rezidență în acest eșantion. 11. Răspuns deschis; se verifică coerența dintre design, tipul variabilei și test.

Lucrarea practică 8 - răspunsuri așteptate (pentru cadrul didactic)

L1. Circumferința taliei pe sexe. Femei $86{,}34 \pm 11{,}67$ cm (mediana 84, IQR 78-93), bărbați $91{,}09 \pm 12{,}19$ cm (mediana 90, IQR 83-97). SW: femei p = 0,008, bărbați p = 0,005 (asimetrie ușoară), dar grupurile sunt mari (86 și 74), deci ambele teste sunt acceptabile dacă alegerea este justificată. Levene p = 0,80. Testul t Student t = -2,52, gl 158, p = 0,013 (Welch t = -2,51, p = 0,013); diferența F - M -4,76 cm (IC 95% -8,50 la -1,01); d = -0,40. Mann-Whitney p = 0,0096, r_rb = 0,24. Concluzie: respingem H0; bărbații au talia în medie cu aproximativ 5 cm mai mare (efect mic spre mediu); pragurile clinice ale circumferinței taliei sunt diferite pe sexe.

L2. Trigliceridele după diabet. Fără diabet: mediana 154 mg/dL (IQR 104,5-209,5), SW p < 0,001; cu diabet: mediana 143 (IQR 112-179), SW p = 0,11. Test potrivit: Mann-Whitney (distribuție asimetrică, grupuri inegale 131 și 29); alternativa: testul t pe trigliceridele logaritmate. U = 2033,5 (raportat pentru grupul fără diabet), p = 0,55; r_rb = 0,07; Welch pe valorile brute p = 0,30. Nu respingem H0. În (g) se acceptă comentariul că, în populații reale, trigliceridele sunt de obicei mai mari la diabetici; pe 29 de cazuri și în date simulate diferența nu apare.

L3. HbA1c la includere pe brațe. A 5,61 ± 0,45; B 5,57 ± 0,43; C 5,55 ± 0,46%; Levene p = 0,58; ANOVA F(2;157) = 0,27, p = 0,76, η^2 = 0,003; Kruskal-Wallis p = 0,55. Nu respingem H0. Punctul cheie al problemei este discuția: într-un studiu randomizat, diferențele de la includere sunt întâmplătoare prin definiție; tabelul de la includere se descrie, nu se testează (CONSORT). Ce contează este mărimea clinică a unei eventuale diferențe (aici neglijabilă: 0,06 puncte procentuale).

L4. Colesterolul la includere pe brațe. A 209,32 ± 39,01; B 202,85 ± 37,55; $C\ 215{,}87 \pm 33{,}56$ mg/dL; Levene p = 0,61; ANOVA F(2;157) = 1,68, p = 0,19, η^2 = 0,021; Kruskal-Wallis p = 0,19. Nu respingem H0. Aceeași discuție ca la L3; diferența maximă între brațe (13 mg/dL) nu este clinic importantă pentru rezultatele glicemice.

L5. Scăderea în greutate pe brațe. $A\ 0{,}71 \pm 2{,}48$ kg (mediana 1,1); B 2,84 ± 2,36 (mediana 2,3); C 3,37 ± 2,39 (mediana 3,55); SW pe brațe p = 0,70, 0,26, 0,17; Levene p = 0,79. ANOVA F(2;157) = 18,25, p < 0,001, η^2 = 0,189. Tukey: B - A 2,13 kg (IC 95% 1,02-3,24; p < 0,001); C - A 2,66 kg (IC 95% 1,56-3,77; p < 0,001); C - B 0,53 kg (IC 95% -0,57 la 1,63; p = 0,49). Kruskal-Wallis H = 27,50, p < 0,001. Concluzie: ambele brațe cu aplicație au dus la o scădere mai mare decât consilierea simplă (cu 2,1-2,7 kg), iar teleconsultația nu a adus un beneficiu suplimentar semnificativ față de aplicația singură.

L6. IMC după statutul de fumător. Foști fumători mediana 28,0 (IQR 26,5-30,9); fumători 28,3 (26,3-31,4); nefumători 28,35 (26,1-31,08). SW p < 0,05 în toate grupurile (0,029, < 0,001, 0,003); Levene p = 0,048. Test potrivit: Kruskal-Wallis (asimetrie în toate grupurile, grupul foștilor fumători are doar 33 de pacienți); H = 0,46, gl 2, p = 0,80. ANOVA p = 0,35 și Welch p = 0,38 duc la aceeași concluzie. Nu respingem H0.

L7. Satisfacția pe sexe. Mediana 4 (IQR 3-4) la ambele sexe; Mann-Whitney U = 3221, p = 0,89; r_rb = 0,01. Nu respingem H0. Se depunctează raportarea mediilor (3,56 și 3,49) ca descriere principală și folosirea testului t.

L8. Vârsta după mediul de rezidență. Rural 57,38 ± 9,08 ani (n 65), urban $55{,}55 \pm 8{,}52 (n\ 95)$; SW p = 0,87 și 0,69; Levene p = 0,64; t Student = 1,30, gl 158, p = 0,19 (Welch t = 1,29, p = 0,20); diferența rural - urban 1,84 ani (IC 95% -0,98 la 4,66); d = 0,21; Mann-Whitney p = 0,17. Nu respingem H0.

L9. Scăderea ≥ 5% pe sexe. Femei 23/86 (26,7%), bărbați 15/74 (20,3%). Frecvențele așteptate: femei 20,4 cu scădere și 65,6 fără; bărbați 17,6 și 56,4. χ^2 = 0,92, gl 1, p = 0,34 (Yates p = 0,44, Fisher p = 0,36); φ = 0,08. Nu respingem H0.

L10. Fumatul după mediu. Rural: foşti fumători 14 (21,5%), fumători 16 (24,6%), nefumători 35 (53,8%); urban: 19 (20,0%), 37 (38,9%), 39 (41,1%). Cea mai mică frecvență așteptată 13,4. χ^2 = 3,80, gl 2, p = 0,15; V Cramér = 0,15. Nu respingem H0; tendința (mai mulți fumători în urban) nu atinge pragul de semnificație.

L11. Fumatul pe brațe. A: foşti fumători 5 (9,4%), fumători 18 (34,0%), nefumători 30 (56,6%); B: 11 (20,8%), 18 (34,0%), 24 (45,3%); C: 17 (31,5%), 17 (31,5%), 20 (37,0%). Cea mai mică frecvență așteptată 10,9. χ^2 = 8,60, gl 4, p = 0,072; V Cramér = 0,16. Nu respingem H0. Discuția așteptată: dezechilibrul (de exemplu 9,4% foşti fumători în A față de 31,5% în C) este întâmplător, rezultat al randomizării; nu "invalidează" studiul; dacă fumatul influențează rezultatul principal, el se poate include ca variabilă de ajustare în analiză, decizie care ar trebui luată dinainte.

L12. Glicemia ≥ 100 la 0 și 6 luni, McNemar pe fiecare braț. Brațul A: ≥ 100 la ambele momente 32, doar la includere 3, doar la 6 luni 1, la niciunul 17; proporția 35/53 (66,0%) la includere și 33/53 (62,3%) la 6 luni; χ^2 McNemar = 1,00, p = 0,32 (cu corecție 0,25, p = 0,62; exact p = 0,63); nu respingem H0. Brațul B: la ambele 19, doar la includere 15, doar la 6 luni 0, la niciunul 19; 34/53 (64,2%) față de 19/53 (35,8%); χ^2 = 15,00, p < 0,001 (cu corecție 13,07, p < 0,001; exact p < 0,001); respingem H0. Discuția: rezultatele din interiorul brațelor descriu schimbarea în fiecare braț; comparația eficacității se face între brațe (de exemplu chi-pătrat pe proporția celor care au trecut sub 100 sau analiza scăderii glicemiei pe brațe). Se depunctează chi-pătratul obișnuit pe tabelul "moment x glicemie ≥ 100".

L13. Glicemia și HbA1c la includere. Diagrama de dispersie arată o relație liniară, fără valori extreme importante; HbA1c are distribuție normală (SW p = 0,30), glicemia o abatere ușoară (SW p = 0,037 la n = 160, asimetrie 0,37). Corelația Pearson r = 0,738, p < 0,001 (Spearman rho = 0,727, p < 0,001); $R^2 = 0{,}54$. Respingem H0: ρ = 0. Asociere puternică, pozitivă, așteptată fiziologic (ambele măsoară glicemia, pe termen scurt și pe termen lung); nu este o măsură a acordului dintre cele două, fiind mărimi diferite.

L14. Glicemia la 0, 3 și 6 luni în brațul A. Mediile 107,8 / 105,7 / 104,8 mg/dL; medianele 111 / 104 / 105. SW pe fiecare moment p = 0,20, 0,62, 0,71 (normale); testul Mauchly p = 0,020 (sfericitate încălcată, ε Greenhouse-Geisser 0,88). Analiza principală justificată: ANOVA cu măsurători repetate cu corecția Greenhouse-Geisser: F(2;104) = 2,73, p = 0,078 (fără corecție p = 0,070); nu respingem H0. Alternativa, Friedman: χ^2 = 6,04, gl 2, p = 0,049; comparațiile Wilcoxon cu Bonferroni nu rămân semnificative (cel mai mic p ajustat 0,075). Răspunsul complet recunoaște discordanța, raportează testul stabilit dinainte și concluzionează că în brațul A nu există dovezi convingătoare de scădere a glicemiei în timp. Se depunctează alegerea testului Friedman justificată doar prin faptul că dă p < 0,05.

L15. Vârsta și FINDRISC. Scorul FINDRISC este un scor cu valori întregi și asimetrie ușoară (SW p = 0,034); se acceptă Spearman (rho = 0,456, p < 0,001) sau Pearson (r = 0,442, p < 0,001), cu justificare. Respingem H0. Discuția așteptată: vârsta este una dintre componentele scorului FINDRISC (pacienții primesc puncte pentru grupele de vârstă), deci o parte din corelație este "construită" de definiția scorului; corelația nu arată că vârsta adaugă informație independentă față de celelalte componente ale scorului.

Capitolul 9

Întrebări: 1 b; 2 c; 3 d; 4 b; 5 c; 6 a, b, e; 7 a, b, d; 8 a, b, d; 9 a, b, c, e; 10 a, b, c, e

Explicații: La întrebarea 3, $R^2 = 0{,}60^2 = 0{,}36$, deci proporția datorată altor factori este 1 - 0,36 = 0,64 (64%); varianta "40%" apare când se scade r, nu r², din 1. La întrebarea 6, varianta c este falsă pentru că r măsoară doar relația liniară (o relație în formă de U poate avea r aproape de

zero), iar varianta d este falsă pentru că r nu depinde de unitățile de măsură. La întrebarea 8, $R^2 = 0{,}088^2 = 0{,}0077$, adică sub 1%, nu 8,8%; corelația este semnificativă doar datorită eșantionului mare și nu demonstrează o relație cauzală. La întrebarea 9, normalitatea se cere reziduurilor, nu variabilei X.

Exerciții propuse: 1. mx = 53, my = 130,33; Sxy = 660, Sxx = 868, Syy = 553,33; $r = \frac{660}{\sqrt{868 \times 553{,}33}} = 0{,}952$ (corelație directă puternică); t = 6,24, gl 4, t critic 2,776, p = 0,003, H0 se respinge; $b = \frac{0{,}760 \text{ mmHg}}{\text{an}}$, a = 90,03; TAS = 90,03 + 0,760 x vârsta; la 60 de ani, 135,7 mmHg; $R^2 = 0{,}907$. 2. n = 40: $t = \frac{0{,}30 \times \sqrt{38}}{\sqrt{0{,}91}} = 1{,}94$, gl 38, t critic 2,024, p = 0,060, H0 nu se respinge. n = 200: t = 4,43, p < 0,001, H0 se respinge. Același r devine semnificativ la un eșantion mai mare; după Colton, corelația este slabă în ambele cazuri. 3. $z = 0{,}5 \times \ln\left(\frac{1{,}62}{0{,}38}\right) = 0{,}725$; ES $= \frac{1}{\sqrt{47}} = 0{,}146$; IC pe scara z 0,439-1,011; IC 95% pentru ρ: 0,41-0,77. Limita superioară depășește 0,75, deci corelația din populație ar putea fi puternică, dar limita inferioară (0,41) este compatibilă și cu o corelație slabă. 4. Ranguri VAS: 1, 4, 2, 6, 5, 7, 3; ranguri anxietate: 2, 4, 1, 7, 6, 5, 3; di: -1, 0, 1, -1, -1, 2, 0; $\sum {d_i}^2 = 8$; $r_s = 1 - \frac{6 \times 8}{7 \times 48} = 0{,}857$, p = 0,014; corelație directă puternică. 5. $R^2 = 0{,}738^2 = 0{,}545$: aproximativ 54,5% din variația HbA1c se asociază liniar cu variația glicemiei a jeun; $1 - R^2 = 0{,}455$: 45,5% din variația HbA1c ține de alți factori. După Colton, corelație moderată, foarte aproape de pragul de 0,75. 6. 0,81 directă puternică; -0,573 inversă moderată; 0,256 directă slabă; -0,088 fără corelație; 0,729 directă moderată; -0,832 inversă puternică; 0,107 fără corelație; 0,50 directă moderată (la limita dintre slabă și moderată). 7. Media diferențelor (deplasarea) = 3,0 mmHg; AS = 3,46 mmHg; limitele de concordanță $3{,}0 \pm 1{,}96 \times 3{,}46 = -3{,}8$ până la 9,8 mmHg. Coeficientul de corelație ar arăta doar dacă pacienții sunt ordonați asemănător; o diferență sistematică de câțiva mmHg nu ar scădea r. 8. 5 ani: 1,11^5 = 1,69; 10 ani: 1,11^10 = 2,84 (cu coeficientul nerotunjit, aproximativ 2,9). Nu este corect: OR se referă la șanse (p / (1 - p)), nu la probabilități; cele două sunt apropiate doar pentru evenimente rare. 9. Femeie nefumătoare, 45 de ani, IMC 27: 64,51 + 1,389 x 27 + 0,449 x 45 - 1,119 = 121,1 mmHg. Bărbat fumător, 65 de ani, IMC 35: 64,51 + 1,389 x 35 + 0,449 x 65 - 1,481 + 6,738 = 147,6 mmHg. Diferența de 26,5 mmHg: IMC contribuie cu 8 x 1,389 = 11,1 mmHg (cel mai mult), vârsta cu 20 x 0,449 = 9,0 mmHg, fumatul cu 6,738 + 1,119 = 7,9 mmHg, sexul cu -1,5 mmHg. 10. (a) Fiecare centimetru în plus se asociază cu o greutate mai mare în medie cu 0,95 kg. (b) 1 - 0,38 = 0,62, adică 62%. (c) La 170 cm: -95 + 0,95 x 170 = 66,5 kg. La 120 cm formula dă 19 kg, o extrapolare absurdă în afara domeniului 155-195 cm; modelul nu se aplică.

Capitolul 10

Întrebări: 1 c; 2 c; 3 b; 4 b; 5 c; 6 a, b, d, e; 7 a, b, c, e; 8 a, b, d, e; 9 a, b, c, e; 10 a, b, d, e

Explicații: La întrebarea 1, studiul pornește de la boală (cazuri și martori), deci este caz-control și se calculează doar OR. La întrebarea 2, diferența de risc este 0,20 - 0,10 = 0,10, iar NNH $= \frac{1}{0{,}10} = 10$; varianta "5" apare când se împarte la riscul la expuși. La întrebarea 3, RR = 0,70 înseamnă o reducere relativă a riscului de 30% (1 - 0,70), nu de 70%. La întrebarea 4, boala este frecventă (prevalențe de 45% și 18,75%), deci OR (3,55) este mai mare decât raportul prevalențelor (2,40). La întrebarea 8, factorul de confuzie nu se află pe calea cauzală; o variabilă intermediară (mediator) nu se ajustează ca factor de confuzie. La întrebarea 10, valoarea p nu face parte din criteriile Bradford Hill.

Exerciții propuse: 1. Studiu caz-control (selecție după infertilitate). a = 89 (infertile cu DIU), b = 640 (martore cu DIU), c = 194 (infertile fără DIU), d = 3193 (martore fără DIU). OR =

$\frac{89\times3193}{640\times194} = 2{,}29(;\ \text{IC}\,95\%\ 1{,}76 - 2{,}98)$; χ^2 = 39,35, p < 0,001. Șansa de a fi utilizat un DIU este de 2,3 ori mai mare la femeile infertile; asociere semnificativă. 2. a = 23, b = 84, c = 117, d = 800; $\text{OR} = \frac{23\times800}{84\times117} = 1{,}87(;\ \text{IC}\,95\%\ 1{,}14 - 3{,}09)$; χ^2 = 6,20, p = 0,013. Sindromul Down este rar în populație, deci, dacă eșantionul a fost construit ca studiu caz-control, OR aproximează RR; RR nu se poate calcula direct din acest design. 3. Brațul B: 16 / 53 = 30,2%; brațul A: 6 / 53 = 11,3%; RR = 2,67 (IC 95% 1,13-6,29); diferența absolută 18,9 puncte procentuale (IC 95% 3,9-33,9); NNT = 5,3, rotunjit la 6. Rezultat foarte apropiat de cel al brațului C față de A (RR 2,62; NNT 6). 4. a = 10, b = 43, c = 19, d = 88. $\text{RP} = \frac{18{,}9\%}{17{,}8\%} = 1{,}06(;\ \text{IC}\,95\%\ 0{,}53 - 2{,}12)$; OR = 1,08 (IC 95% 0,46-2,52); χ^2 = 0,03, p = 0,86. H0 nu se respinge; în mg2026 fumatul nu se asociază cu diabetul. Concluzia clinică: datele nu arată o asociere, dar intervalul larg nu exclude nici un efect moderat. 5. a = 21, b = 33, c = 20, d = 86. $\text{RP} = \frac{38{,}9\%}{18{,}9\%} = 2{,}06(;\ \text{IC}\,95\%\ 1{,}23 - 3{,}46)$; OR = 2,74 (IC 95% 1,32-5,69); χ^2 = 7,52, p = 0,006. OR depășește RP pentru că rezultatul este frecvent (25,6% au TAS ≥ 140 mmHg). 6. Expuși: 10 / 400 = 2,5 cazuri la 100 de persoane-an; neexpuși: 8 / 800 = 1,0 la 100 de persoane-an; IRR = 2,5. 7. P ≈ I x D = 0,005 x 4 = 0,02, adică 2%. Dacă durata se dublează la 8 ani, la aceeași incidență prevalența crește la aproximativ 4%, deși tratamentul este un progres. 8. Diferența de risc = 6% - 2% = 4 puncte procentuale; RR = 3; $\text{NNH} = \frac{1}{0{,}04} = 25$: la fiecare 25 de pacienți tratați apare un efect advers în plus față de placebo. 9. FAe = (3 - 1) / 3 = 0,667 (66,7%); $\text{FAP} = \frac{0{,}40\times2}{1+0{,}40\times2} = \frac{0{,}8}{1{,}8} = 0{,}444(44{,}4\%)$. 10. OR strat $1 = \frac{20\times60}{80\times10} = 1{,}5$; OR strat $2 = \frac{30\times60}{20\times60} = 1{,}5$. Tabelul combinat: a = 50, b = 100, c = 70, d = 120, OR brut = 0,86. $\text{OR(MH)} = \frac{\frac{20\times60}{170}+\frac{30\times60}{170}}{\frac{80\times10}{170}+\frac{20\times60}{170}} = \frac{17{,}65}{11{,}76} = 1{,}5$. Măsurile din straturi sunt egale între ele și diferă mult de cea brută: factorul de stratificare este un factor de confuzie, care în analiza brută inversează chiar sensul asocierii. 11. a = 10, b = 55, c = 19, d = 76; OR = 0,73 (IC 95% 0,31-1,69); χ^2 = 0,55, p = 0,46. Intervalul conține 1: nu se poate afirma că mediul rural protejează; rezultatul este nesemnificativ.

Capitolul 11

Întrebări: 1 c; 2 b; 3 b; 4 d; 5 c; 6 a, c, d, e; 7 a, b, d; 8 a, b, e; 9 a, b, c; 10 a, b, c, d

Explicații: La întrebarea 2, din 1000 de pacienți, 100 sunt bolnavi (80 de teste pozitive) și 900 sănătoși (90 de pozitive false), deci $\text{VPP} = \frac{80}{170} = 47\%$; varianta a) confundă VPP cu sensibilitatea. La întrebarea 5, BF10 = 0,05 înseamnă BF01 = 20, adică dovezi puternice pentru H0; factorul Bayes nu este o valoare p. La întrebarea 7, varianta c) este falsă: rapoartele de verosimilitate provin din sensibilitate și specificitate, care nu depind în principiu de prevalență (dar se pot schimba cu spectrul pacienților). La întrebarea 9, biasul de verificare este un bias de design al studiilor diagnostice (nu toți pacienții primesc standardul de referință), nu o eroare cognitivă la patul bolnavului.

Exerciții propuse:

1. a = 45, b = 15, c = 88, d = 352, N = 500. $p(B^+) = \frac{60}{500} = 0{,}12$; $p(S^+/B^+) = \frac{45}{60} = 0{,}75$; $p(S^+/B^-) = \frac{88}{440} = 0{,}20$; $p(B^+/S^+) = \frac{45}{133} = 0{,}338$. Bayes: $\frac{0{,}75\times0{,}12}{0{,}75\times0{,}12+0{,}20\times0{,}88} = \frac{0{,}090}{0{,}266} = 0{,}338$.
2. Din 10.000 de femei: 80 cu cancer (72 pozitive, 8 negative false), 9920 fără cancer (aproximativ 694 pozitive false, 9226 negative). $\text{VPP} = \frac{72}{766} = 9{,}4\%$; $\text{VPN} = \frac{9226}{9234} = 99{,}9\%$. Pentru fiecare cancer găsit se investighează suplimentar aproximativ 766/72 = 10,6, adică circa 11 femei cu mamografie pozitivă.
3. 0,40 -> 0,40/0,60 = 0,667; 0,05 -> 0,0526; șansa 3 -> 3/4 = 0,75; șansa 0,25 -> 0,25/1,25 = 0,20.

4.

   (a) $RV^+ = \frac{0,703}{0,340} = 2,07$; $RV^- = \frac{0,297}{0,660} = 0,45$. (b) Șansa pre-test 0,429; după dNLR peste prag 0,429 x 2,07 = 0,886, probabilitate 47,0%; sub prag 0,429 x 0,45 = 0,193, probabilitate 16,2%. Testul mută probabilitatea puțin, în acord cu AUC 0,624 (satisfăcător pe scala din curs). (c) Trei zecimale (70,312%) sugerează o precizie pe care un eșantion de 336 de pacienți nu o oferă; o zecimală (70,3%) și un interval de încredere ar fi fost potrivite.

5. $RV^- = \frac{0,07}{0,86} = 0,081$. La 33,8%: 0,512 x 0,081 = 0,042, probabilitate post-test 4,0%. La 5%: 0,0526 x 0,081 = 0,0043, probabilitate post-test 0,43%.

6. În serie: Se = 0,90 x 0,85 = 0,765; Sp = 1 - 0,20 x 0,05 = 0,99. În paralel: Se = 1 - 0,10 x 0,15 = 0,985; Sp = 0,80 x 0,95 = 0,76. Triajul la gardă: paralel (nu ratăm bolnavi). Confirmarea înaintea unui tratament toxic: serie (puține pozitive false).

7. Produse: migrenă 0,70 x 0,05 x 0,05 x 0,70 = 0,001225; tensională 0,25 x 0,02 x 0,10 x 0,10 = 0,00005; HSA 0,05 x 0,80 x 0,60 x 0,70 = 0,0168. Suma 0,018075. Probabilități posterioare: migrenă 6,8%; tensională 0,3%; HSA 92,9%. Decizie: investigație imagistică de urgență (CT cerebral), indiferent de rezervele privind independența semnelor.

8. Șansa la prag 0,10/0,90 = 0,111. Pragul de testare: 0,111/4 = 0,0278, adică 2,7%. Pragul de testare-tratament: 0,111/0,25 = 0,444, adică 30,8%.

9.

   (a) SOFA: 0,771 bine, 0,873 foarte bine; NEWS2: 0,892 foarte bine, 0,729 bine; APACHE II: 0,698 satisfăcător, 0,806 la limita dintre bine și foarte bine (0,81 după rotunjire). (b) NEWS2 la internare, SOFA în ziua a cincea; performanța unui scor depinde de momentul aplicării. (c) Regresia logistică produce rapoarte ale șanselor (OR); HR provin dintr-un model Cox, care cere date de timp până la eveniment. Raportarea HR dintr-o regresie logistică este un punct de discutat la lectura critică.

10.

   (a) $BF_{01} = \frac{1}{0,16} = 6,25$: dovezi moderate pentru absența corelației. (b) Șansa a posteriori 1 x 0,16 = 0,16; probabilitatea 0,16/1,16 = 13,8%.

11. P: adulți neinvestigați cu risc cardiometabolic din medicina de familie; I: HbA1c (de exemplu prag ≥ 6,5%); C: TTGO ca standard de referință; O: sensibilitate, specificitate, rapoarte de verosimilitate pentru diabetul de tip 2. Design potrivit: studiu transversal de acurateț diagnostică pe pacienți consecutivi, la care toți primesc ambele teste, raportat după STARD 2015.

Capitolul 12

Întrebări: 1 c; 2 b; 3 d; 4 c; 5 c; 6 a, b, d; 7 a, b, c, e; 8 a, b, d; 9 a, b, d, e; 10 a, c, e

Explicații: La întrebarea 1, varianta b) este formula greșită care apare uneori în notițe; d conține adevărații negativi, adică persoane clasificate corect. La întrebarea 4, AUC nu este acuratețea și nici sensibilitatea; este probabilitatea ca, într-o pereche bolnav-sănătos aleasă la întâmplare, bolnavul să aibă valoarea mai mare. La întrebarea 6, varianta c) descrie biasul de verificare, iar varianta e) este falsă pentru că un eșantion mai mare cu același spectru nepotrivit dă doar o estimare mai precisă a unei valori deplasate. La întrebarea 8, rata erorii este (7 + 33)/160 = 0,25, iar specificitatea este 98/131 (98/105 este VPN).

Exerciții propuse:

1. a = 45, b = 5, c = 30, d = 120. Prevalența 50/200 = 0,25; Se 0,90; $\frac{\text{Sp}\,120}{150} = 0,80$; $\frac{\text{VPP}\,45}{75} = 0,60$; $\frac{\text{VPN}\,120}{125} = 0,96$; $\frac{\text{AC}\,165}{200} = 0,825$; $\frac{\text{RE}\,35}{200} = 0,175$; J 0,70; RV+ 4,5; RV- 0,125; $\frac{\text{DOR}(45\times120)}{5\times30} = 36$.

2. a = 80, b = 20; din $\frac{80}{80+c} = \frac{4}{5}$, c = 20; din $\frac{d}{20+d} = \frac{9}{10}$, d = 180. Se 0,80; $\frac{\text{Sp}\,180}{200} = 0,90$; N = 300.

3. d = 150 și Sp $= \frac{3}{4}$, deci c + d = 200 și c = 50; din $\frac{150}{b+150} = \frac{15}{16}$, b = 10; a = 40. $\frac{\text{Se } 40}{50} = 0{,}80$; $\frac{\text{VPP } 40}{90} = 0{,}444$; N = 250.
4. 100 de bolnavi: 75 AP, 25 FN; 900 de nebolnavi: 855 AN, 45 FP. $\frac{\text{VPP } 75}{120} = 62{,}5\%$; $\frac{\text{VPN } 855}{880} = 97{,}2\%$.
5. HPV: a = 17, b = 1, c = 19, d = 27; $\frac{\text{AC } 44}{64} = 68{,}8\%$; J 0,531. p16/Ki-67: a = 12, b = 6, c = 7, d = 39; $\frac{\text{AC } 51}{64} = 79{,}7\%$; J 0,514. HPV este mai potrivit pentru excludere (Se 94,4%, RV- 0,095), p16/Ki-67 pentru confirmare (Sp 84,8%, RV+ 4,38). Indicii Youden aproape egali ascund aceste roluri diferite.
6. Trapeze: 0,010 + 0,025 + 0,070 + 0,170 + 0,282 + 0,297 = 0,854; foarte bun (0,81-0,90).
7. Po $= \frac{175}{200} = 0{,}875$; A "pneumonie" 45/200 = 0,225, $\frac{B\ 40}{200} = 0{,}20$; Pe = 0,225 x 0,20 + 0,775 x 0,80 = 0,665; κ = 0,210/0,335 = 0,63, acord substanțial.
8.
   (a) Decese 159 x 0,2515 = 40, supraviețuitori 119; AP 0,878 x 40 = 35, FN 5; AN 0,7372 x 119 = 88, FP 31. (b) VPP $= \frac{35}{66} = 53{,}0\%$, apropiat de 53,73% raportat. (c) Wilson pentru 35/40: aproximativ 73,9-94,5%, de aproape 80 de ori mai larg decât intervalul raportat; intervalul din rezumat nu poate proveni dintr-o metodă obișnuită pentru o proporție estimată pe 40 de decese.
9. $\frac{\text{Se } 40}{50} = 80\%$; $\frac{\text{Sp } 360}{450} = 80\%$; precizia 40/130 = 30,8%; $F_1 = \frac{2\times0{,}308\times0{,}80}{0{,}308+0{,}80} = 0{,}44$; $\frac{\text{AC } 400}{500} = 80\%$. Un clasificator care declară totul benign ar avea acuratețea 450/500 = 90%, mai mare decât algoritmul, ceea ce arată de ce acuratețea nu se folosește singură la clase dezechilibrate; precizia mică reflectă prevalența de 10%.
10. 0,624 satisfăcător; 0,752 bun. Testul H0: AUC = 0,5 arată doar că markerul discriminează mai bine decât întâmplarea; pe 336 de pacienți, chiar o discriminare modestă devine semnificativă. Utilitatea clinică se judecă după mărimea AUC, rapoartele de verosimilitate la prag și consecințele deciziilor.
11. Șansa pre-test 0,111; $\text{RV}^+ = \frac{0{,}801}{0{,}143} = 5{,}60$, deci VPP $= \frac{0{,}622}{1{,}622} = 38{,}4\%$; $\text{RV}^- = \frac{0{,}199}{0{,}857} = 0{,}232$, probabilitatea post-test 2,5%, deci VPN 97,5%.
12. $\text{RV}^+ = \frac{0{,}379}{0{,}053} = 7{,}1$; $\text{RV}^- = \frac{0{,}621}{0{,}947} = 0{,}66$. Un rezultat ≥ 126 mg/dL crește șansa de aproximativ 7 ori (bun pentru confirmare), dar un rezultat sub prag aproape nu schimbă probabilitatea (RV- 0,66), deci pragul ratează majoritatea diabeticilor identificați prin TTGO și nu este potrivit pentru screening.

Capitolul 13

Întrebări: 1 c; 2 b; 3 c; 4 c; 5 b; 6 a, b, d; 7 a, c, d; 8 a, b, d, e; 9 a, b, c; 10 a, c, d

Explicații: La întrebarea 4, NNT $= \frac{1}{0{,}183} = 5{,}5$ și se rotunjește întotdeauna în sus, la 6; valorile 3 și 29 sunt limitele intervalului de încredere, nu estimarea punctuală. La întrebarea 6, varianta c este falsă: ITT diluează efectul cu pacienții neaderenți, deci tinde să subestimeze efectul biologic, nu să-l supraestimeze; varianta e descrie analiza per protocol. La întrebarea 9, TAS 1540 și înălțimea 1,72 sunt erori în variabile numerice și se depistează prin minim, maxim și box-plot, nu printr-un tabel de frecvențe al variabilelor categoriale. La întrebarea 10, testul Bartlett trebuie să fie semnificativ (itemii corelați), iar alfa măsoară consistența internă, nu validitatea.

Exerciții propuse: 1. $n = \frac{2\times(1{,}96+1{,}28)^2\times12^2}{8^2} = \frac{2\times10{,}50\times144}{64} = 47{,}2$, deci 48 de pacienți pe grup; cu 10% pierderi, 48/0,9 = 53,3, deci 54 pe grup. 2. p1 = 0,35, p2 = 0,20, $\bar{p}$ = 0,275; $n =$

$\frac{[1{,}96\times\sqrt{2\times0{,}275\times0{,}725}+0{,}84\times\sqrt{0{,}35\times0{,}65+0{,}20\times0{,}80}]^2}{0{,}15^2} = 137{,}9$, deci 138 de pacienți pe grup. 3. Efectul de design = 1 + (20 - 1) x 0,02 = 1,38; 64 x 1,38 = 88,3, deci 89 de pacienți pe braț (aproximativ 5 cabinete de 20 de pacienți pe braț). 4. $\alpha = \frac{5}{4} \times \left(1 - \frac{5{,}5}{18{,}0}\right) = 1{,}25 \times 0{,}694 = 0{,}868$; consistență internă bună. 5. Risc 30/150 = 20% cu programul, 45/150 = 30% în grupul de control; RR = 0,67 (reducere relativă a riscului 33%); reducerea absolută a riscului 10 puncte procentuale; NNT = $\frac{1}{0{,}10} = 10$. 6. 1 - 0,95^8 = 0,337 (aproximativ 34%); pragul Bonferroni este 0,05/8 = 0,00625 pentru fiecare test. 7. a) limita inferioară -2,4 este deasupra lui -3: non-inferioritate demonstrată; b) limita inferioară -4,1 coboară sub -3: rezultat neconcludent, non-inferioritatea nu este demonstrată (dar nici inferioritatea); c) întregul interval este deasupra lui 0: non-inferioritate și chiar superioritate, dacă testarea superiorității a fost prevăzută în protocol. 8. a) încrucișat (cross-over), analiză pentru date pereche; b) randomizat pe clustere (școli); c) pilot cu un singur braț, măsurători repetate (ANOVA cu măsurători repetate sau Friedman); d) factorial 2 x 2; e) cohortă observațională (expunerea nu a fost atribuită de investigator). 9. 1,75: unitate greșită (metri); 1680: cifră în plus; "1 m 70": text în coloană numerică. Toate se verifică în documentul sursă și se corectează doar de acolo; dacă sursa nu este disponibilă, valoarea se declară lipsă. Prevenție: validare în formular "număr între 120 și 220", cu mesajul "în centimetri". Valorile 168, 172, 159, 165, 181 sunt plauzibile. 10. 3/300 = 0,01, adică limita superioară a IC 95% este aproximativ 1%; efectul advers ar putea apărea până la circa 1 caz la 100 de pacienți fără să fi fost observat. Efectele mai rare pot fi detectate doar în faza IV, pe mii sau zeci de mii de pacienți. 11. $p = \frac{16}{53} = 30{,}2\%$; $ES = \sqrt{\frac{0{,}302\times0{,}698}{53}} = 0{,}063$; IC 95% Wald 17,8-42,5%; IC 95% Wilson 19,5-43,5% (asimetric, recomandat pentru eșantioane mici sau proporții departe de 50%). 12. $RR = \frac{\frac{16}{53}}{\frac{6}{53}} = 2{,}67 (; \text{IC } 95\% \; 1{,}13 - 6{,}29)$; diferența de risc 18,9 puncte procentuale (IC 95% 3,9-33,9); NNT = 5,3, adică 6. Rezultate practic identice cu cele pentru C față de A (RR 2,62, NNT 6): pe acest criteriu teleconsultația nu adaugă beneficiu peste aplicație (RR C față de B 0,98, IC 95% 0,55-1,75).

Capitolul 14

Întrebări: 1 b; 2 c; 3 b; 4 b; 5 c; 6 b; 7 b; 8 c; 9 b; 10 b; 11 b; 12 b; 13 b; 14 b; 15 b; 16 a, b, c; 17 a, c, e; 18 a, b, c, e; 19 a, b, c; 20 a, b, c, e; 21 a, b, d; 22 a, b, c, d; 23 a, b, c, e; 24 a, b, c, e; 25 a, b, c, e; 26 a, b, d, e; 27 a, b, c, e; 28 a, b, c, d; 29 a, b, d, e; 30 a, b, c, e

Explicații: La întrebarea 2, variantele a și b sunt capcanele clasice: intervalul de încredere privește media populației, nu valorile individuale, iar media eșantionului se află întotdeauna în interval (este centrul lui). La întrebarea 9, din 150 de bolnavi 135 sunt pozitivi, iar din 850 de sănătoși 170 sunt fals pozitivi; $VPP = \frac{135}{305} = 44{,}3\%$. La întrebarea 13, schimbarea testului după rezultat (varianta d) este pescuit de semnificație, iar varianta a folosește formularea greșită "se acceptă H0". La întrebarea 14, $RR = \frac{\frac{18}{150}}{\frac{10}{250}} = \frac{0{,}12}{0{,}04} = 3{,}0$; 3,27 este OR-ul. La întrebarea 21, interceptul corespunde unui IMC de 0, valoare imposibilă și situată departe de domeniul observat (23,5-48,3), deci nu are interpretare clinică. La întrebarea 24, pragul Youden maximizează Se + Sp, nu VPP. La întrebarea 29, r = 0 exclude doar o relație liniară; o relație curbilinie (de exemplu, în formă de U) poate da r aproape de 0.

Exerciții propuse: 1. Total 250: nefumători 120 (48,0% din total; 52,2% valid), fumători 70 (28,0%; 30,4%), foști fumători 40 (16,0%; 17,4%), lipsă 20 (8,0%); procentele valide se

calculează la 230. 2. OR $= \frac{56\times96}{64\times24} = 3{,}5$; IC 95% 1,97-6,21; χ^2 = 19,20, p < 0,001. Cota de fumat este de 3,5 ori mai mare la pacienții cu BPOC. 3. Risc 30/200 = 15% la expuși, 15/300 = 5% la neexpuși; RR = 3,0 (IC 95% 1,66-5,43); DR = 10 puncte procentuale; NNH = 10; χ^2 = 14,65, p < 0,001. 4. Bolnavi 200: AP 180, FN 20; sănătoși 9800: AN 9310, FP 490. VPP $= \frac{180}{670} = 26{,}9\%$; VPN $= \frac{9310}{9330} = 99{,}8\%$. La o prevalență mică, aproape trei sferturi dintre rezultatele pozitive sunt fals pozitive, deși testul are Sp 95%; un pozitiv cere confirmare, un negativ exclude boala aproape sigur. 5. ES $= \frac{18}{\sqrt{36}} = 3$ mg/dL; IC $95\% = 220 \pm 2{,}03 \times 3 = 213{,}9 - 226{,}1$ mg/dL. 6. $R^2 = 0{,}36$; $t = \frac{0{,}6\times\sqrt{25}}{\sqrt{1-0{,}36}} = \frac{3}{0{,}8} = 3{,}75 > 2{,}06$, deci r este semnificativ (p ≈ 0,001). 7. TAS = 89,47 + 1,435 x 27 = 128,2 mmHg. IMC 15 se află în afara domeniului observat (23,5-48,3 kg/m²), iar extrapolarea dreptei în afara datelor nu este justificată. 8. Cota pre-test = 0,30/0,70 = 0,429; cota post-test = 0,429 x 6 = 2,57; probabilitatea post-test = 2,57/3,57 = 72%. 9. a) ANOVA unifactorială (cu Tukey); b) testul Wilcoxon pentru date pereche (scorul de durere este ordinal); c) chi-pătrat (Fisher la frecvențe așteptate mici); d) corelația Pearson (sau Spearman) și regresia liniară; e) testul Mann-Whitney; f) testul Friedman. 10. Model: "HDL-colesterolul mediu a fost 53,9 mg/dL (AS 11,7) la femei și 44,4 mg/dL (AS 9,7) la bărbați. Diferența de 9,5 mg/dL (IC 95% 6,2-12,9) a fost semnificativă statistic (testul t Welch, ales pentru variațe inegale, Levene p = 0,034; t = 5,63, gl 157,7, p < 0,001)." 11. Variantă corectă: "Prevalența diabetului a fost 33,3% la pacienții obezi și 10,4% la cei neobezi (χ^2 = 12,70, p < 0,001; raportul prevalențelor 3,21, IC 95% 1,64-6,31; OR 4,32, IC 95% 1,86-10,03). Fiind un studiu transversal, rezultatul arată o asociere, nu o relație cauzală." Greșelile inițiale: cauzalitate, confuzia OR cu riscul, lipsa IC. 12. Abaterile standard sunt aproape egale cu mediile, pentru o variabilă care nu poate fi negativă: distribuția este puternic asimetrică. Recomandări: descriere prin mediană și IQR, comparație cu testul Mann-Whitney (sau test t pe date logaritmate), raportarea diferenței medianelor.

Anexa B. Glosar

Glosarul cuprinde termenii folosiți în curs, în ordinea alfabetică a limbii române (a, ă, â, b,..., i, î,..., s, ș, t, ț,...). Pentru fiecare termen sunt date abrevierea folosită în volum, atunci când există, echivalentul englezesc (util la citirea articolelor și la lucrul cu programele statistice, care au interfața în limba engleză), o definiție scurtă și capitolul în care noțiunea este explicată pe larg. Definițiile sunt formulate pentru recapitulare; ele nu înlocuiesc explicațiile, exemplele și capcanele discutate în capitole.

A

- Abatere medie absolută (engl. mean absolute deviation): media valorilor absolute ale abaterilor față de medie. Este intuitivă, dar se folosește mai rar decât abaterea standard. (Cap. 2)
- Abatere standard (AS, DS; engl. standard deviation, SD): rădăcina pătrată a variației; arată cât de împrăștiate sunt valorile individuale în jurul mediei, în unitatea de măsură a variabilei. Nu scade când crește eșantionul. (Cap. 2)
- Absența dovezii (engl. absence of evidence): un rezultat nesemnificativ nu dovedește că efectul lipsește; poate reflecta doar un studiu prea mic. Intervalul de încredere arată ce efecte rămân compatibile cu datele. (Cap. 5)
- Acord între metode (engl. method agreement): măsura în care două metode de măsurare dau aceleași valori la aceiași pacienți. Pentru variabile numerice se evaluează cu metoda Bland-Altman, nu cu coeficientul de corelație. (Cap. 12)
- Acuratețe (engl. accuracy): proporția persoanelor clasificate corect de test, (a + d)/N. Depinde de prevalență și poate fi înșelătoare când boala este rară. (Cap. 12)
- Adevărat negativ (AN; engl. true negative, TN): persoană fără boală cu test negativ (celula d). (Cap. 12)
- Adevărat pozitiv (AP; engl. true positive, TP): bolnav cu test pozitiv (celula a din tabelul diagnostic). (Cap. 12)
- Algoritmul de alegere a testului (engl. test selection algorithm): succesiunea de întrebări care duce la testul potrivit: ce tip are variabila rezultat, câte grupuri sau măsurători se compară, sunt ele independente sau pereche, sunt îndeplinite condițiile testului parametric. (Cap. 8)
- Amplitudine (engl. range): diferența dintre valoarea maximă și cea minimă. Este ușor de calculat, dar depinde doar de două valori, deci este foarte sensibilă la valorile extreme. (Cap. 2)
- Analiza după intenția de tratament (ITT; engl. intention-to-treat analysis): analiza în care fiecare participant este inclus în brațul în care a fost randomizat, indiferent dacă a primit sau nu intervenția. Păstrează avantajele randomizării și este analiza principală în studiile de superioritate. (Cap. 13)
- Analiza integrată a datelor (engl. integrated data analysis): parcurgerea completă a unui set de date, de la curățare și statistică descriptivă la alegerea testelor, interpretare și raportare, într-un raport coerent care răspunde întrebării de cercetare. (Cap. 13)
- Analiza per protocol (PP; engl. per-protocol analysis): analiza limitată la participanții care au respectat protocolul. Poate supraestima efectul, pentru că cei care abandonează diferă adesea de cei care rămân. (Cap. 13)
- Analiza variației (ANOVA; engl. analysis of variance): test care compară mediile a trei sau mai multe grupuri independente, împărțind variabilitatea totală în variabilitate între grupuri și în interiorul grupurilor. Un rezultat semnificativ arată doar că cel puțin o medie diferă. (Cap. 6)

- Analiză interimară (engl. interim analysis): analiză efectuată înainte de finalul studiului, după reguli stabilite în protocol, care poate duce la oprirea timpurie pentru beneficiu, lipsă de efect sau risc. (Cap. 13)
- Anonimizare (engl. anonymisation): prelucrarea datelor astfel încât persoana să nu mai poată fi identificată prin niciun mijloc rezonabil. Datele cu adevărat anonime nu mai sunt date cu caracter personal. (Cap. 1)
- ANOVA cu măsurători repetate (engl. repeated measures ANOVA): extensia testului t pereche pentru trei sau mai multe măsurători pe aceiași indivizi (de exemplu glicemia la 0, 3 și 6 luni). Presupune, între altele, sfericitatea. (Cap. 6)
- Aria de sub curba ROC (AUC; engl. area under the curve): indicator global al capacității de discriminare a testului, între 0,5 (test inutil) și 1 (test perfect). Este probabilitatea ca un bolnav ales la întâmplare să aibă un rezultat mai "anormal" decât un nebolnav ales la întâmplare. (Cap. 12)
- Ascunderea alocării (engl. allocation concealment): măsurile care împiedică persoana care înrolează pacientul să cunoască dinainte brațul în care va fi alocat (plicuri sigilate opace, randomizare centralizată). Este diferită de orbire. (Cap. 13)
- Asimetrie (engl. skewness): măsură a lipsei de simetrie a unei distribuții. Asimetria pozitivă (coadă spre dreapta) este frecventă la variabilele biologice care nu pot scădea sub zero, ca trigliceridele; atunci media depășește mediana. (Cap. 3)

B

- Bias de informație (engl. information bias): eroare sistematică în măsurarea expunerii sau a rezultatului, de exemplu prin instrumente necalibrate sau întrebări formulate diferit pentru cazuri și martori. (Cap. 10)
- Bias de memorie (engl. recall bias): formă de bias de informație specifică studiilor caz-control, în care bolnavii își amintesc expunerile trecute mai bine (sau diferit) decât martorii sănătoși. (Cap. 10)
- Bias de selecție (engl. selection bias): eroare sistematică apărută atunci când indivizii incluși diferă sistematic de populația țintă sau când grupurile comparate sunt formate diferit (de exemplu, voluntarii sunt mai sănătoși decât media). (Cap. 10)
- Bias de spectru (engl. spectrum bias): estimarea eronată a sensibilității și specificității când pacienții studiați nu reflectă spectrul celor din practică (de exemplu doar cazuri severe și voluntari perfect sănătoși). (Cap. 12)
- Bias de verificare (engl. verification bias): eroare sistematică în studiile de acuratețe diagnostică apărută când standardul de referință se aplică doar unora dintre pacienți, de obicei celor cu test pozitiv. (Cap. 12)
- Biostatistică (engl. biostatistics): ramura statisticii care se ocupă cu colectarea, descrierea, analiza și interpretarea datelor din biologie și medicină. Pornește de la variabilitatea biologică și oferă metode prin care incertitudinea se măsoară, în loc să fie ignorată. (Cap. 1)
- Box-plot (diagramă cutie cu mustăți; engl. box-and-whisker plot): grafic care arată mediana, cuartilele, mustățile până la cele mai îndepărtate valori neaberante și, separat, valorile aberante. Este ideal pentru compararea distribuțiilor mai multor grupuri. (Cap. 2)
- Braț (al studiului) (engl. study arm): fiecare dintre grupurile de intervenție dintr-un studiu clinic (de exemplu brațele A, B și C din setul mg2026). (Cap. 13)
- Buna practică în studiul clinic (BPC; engl. Good Clinical Practice, GCP): standard internațional de etică și calitate științifică pentru proiectarea, desfășurarea, înregistrarea și raportarea studiilor clinice, descris în ghidul ICH E6. (Cap. 13)

C

- Cauzalitate (engl. causality): relația în care modificarea expunerii produce modificarea rezultatului. Se demonstrează cel mai convingător prin studii randomizate; în studiile observaționale se argumentează prin criterii precum cele ale lui Bradford Hill. (Cap. 10)
- Clasificator (engl. classifier): regulă sau model care atribuie fiecare individ uneia dintre clase (bolnav sau sănătos) pe baza unor caracteristici; performanța lui se descrie cu sensibilitatea, specificitatea și curba ROC. (Cap. 12)
- Codificare (engl. coding): transformarea răspunsurilor în valori standardizate (de exemplu 0/1 pentru absent/prezent, 1-4 pentru nivelul de educație), descrisă în dicționarul de date. (Cap. 2)
- Coeficient de determinare (R^2; engl. coefficient of determination): proporția din variabilitatea variabilei dependente explicată de model; în regresia simplă este egal cu r^2. Nu spune nimic despre sensul relației. (Cap. 9)
- Coeficient de determinare ajustat (R^2 ajustat; engl. adjusted R-squared): variantă a lui R^2 care penalizează adăugarea de predictori inutili. Se folosește la compararea modelelor cu număr diferit de variabile. (Cap. 9)
- Coeficient de variație (CV; engl. coefficient of variation): abaterea standard exprimată ca procent din medie (100 x s/m). Permite compararea variabilității unor măsurători cu unități sau ordine de mărime diferite. (Cap. 2)
- Coeficientul de corelație Pearson (r; engl. Pearson correlation coefficient): măsură a intensității și sensului relației liniare dintre două variabile numerice, cu valori între -1 și 1. Presupune liniaritate, distribuții aproximativ normale și absența valorilor extreme influente. (Cap. 9)
- Coeficientul de corelație Spearman (rho; engl. Spearman rank correlation): coeficientul Pearson calculat pe ranguri; măsoară o relație monotonă (nu neapărat liniară). Se folosește pentru variabile ordinale sau numerice nenormale. (Cap. 9)
- Coeficientul kappa al lui Cohen (κ; engl. Cohen's kappa): măsură a acordului dintre doi evaluatori pentru o variabilă categorială, corectată pentru acordul datorat întâmplării. Valorile peste 0,60 indică un acord substanțial, cele peste 0,80 un acord aproape perfect. (Cap. 12)
- Coeficientul tau Kendall (engl. Kendall's tau): coeficient neparametric de corelație bazat pe numărul de perechi concordante și discordante. Este preferat de unii autori la eșantioane mici și la multe valori egale. (Cap. 9)
- Cohen d (engl. Cohen's d): diferența dintre două medii exprimată în abateri standard. Valori orientative: 0,2 efect mic, 0,5 mediu, 0,8 mare. (Cap. 6)
- Comisie de etică a cercetării (engl. research ethics committee): organism independent care evaluează și aprobă protocolul unui studiu pe oameni înainte de începerea lui, verificând respectarea drepturilor, siguranței și bunăstării participanților. (Cap. 13)
- Comitet independent de monitorizare a datelor (engl. data monitoring committee): grup de experți independenți de sponsor și investigatori care urmărește datele de siguranță și eficacitate pe parcursul studiului și recomandă continuarea, modificarea sau oprirea lui. (Cap. 13)
- Comparații multiple (engl. multiple comparisons): efectuarea mai multor teste pe aceleași date, care crește probabilitatea de a obține cel puțin un rezultat fals pozitiv. Se controlează prin corecții (Bonferroni) sau prin teste post-hoc. (Cap. 5)
- Consimțământ informat (engl. informed consent): acordul liber, dat în scris, al unei persoane de a participa la un studiu, după ce a primit informații clare despre scop, proceduri, riscuri, beneficii și dreptul de a se retrage oricând. (Cap. 1)
- Corecția Bonferroni (engl. Bonferroni correction): împărțirea pragului α la numărul de comparații efectuate (α/k). Este simplă și conservatoare. (Cap. 5)
- Corecția Yates (corecția de continuitate; engl. Yates continuity correction): ajustare a testului chi-pătrat pentru tabele 2 x 2, care scade puțin valoarea χ^2. Este conservatoare și astăzi se preferă adesea testul exact Fisher la frecvențe mici. (Cap. 7)

- Corelația biserială a rangurilor (engl. rank-biserial correlation): mărime a efectului pentru testul Mann-Whitney, între -1 și 1; arată cât de des valorile dintr-un grup le depășesc pe cele din celălalt. (Cap. 7)
- Corelație (engl. correlation): asocierea dintre două variabile, caracterizată prin sens (pozitiv sau negativ) și intensitate. Corelația nu dovedește o relație cauză-efect. (Cap. 9)
- Corelație falsă (corelație spurioasă; engl. spurious correlation): asociere statistică între două variabile care nu au legătură directă, produsă de o a treia variabilă comună sau de întâmplare. (Cap. 9)
- Cota pre-test și cota post-test (engl. pre-test odds, post-test odds): probabilitățile pre-test și post-test exprimate sub formă de cote; cota post-test = cota pre-test x raportul de verosimilitate. (Cap. 11)
- Cotă (odds; engl. odds): raportul dintre probabilitatea ca un eveniment să se producă și probabilitatea să nu se producă, p/(1 - p). O probabilitate de 0,20 corespunde unei cote de 0,25 (1 la 4). (Cap. 10)
- Criterii de includere și de excludere (engl. eligibility criteria): condițiile pe care trebuie să le îndeplinească (sau să nu le îndeplinească) o persoană pentru a putea participa la studiu. Definesc populația căreia i se pot aplica rezultatele. (Cap. 13)
- Criteriile Bradford Hill (engl. Bradford Hill criteria): repere pentru judecarea cauzalității unei asocieri: tăria asocierii, consistența, specificitatea, succesiunea temporală, gradientul doză-răspuns, plauzibilitatea biologică, coerența, dovezile experimentale și analogia. Dintre ele, doar succesiunea temporală este obligatorie. (Cap. 10)
- Criteriu secundar de evaluare (engl. secondary outcome): rezultat suplimentar, analizat pentru a completa imaginea efectelor intervenției; concluziile bazate doar pe criterii secundare au caracter explorator. (Cap. 13)
- Criteriu surogat (engl. surrogate endpoint): rezultat intermediar (tensiunea arterială, HbA1c, colesterolul) folosit în locul unui rezultat clinic important (infarct, deces), presupunând că modificarea lui prezice efectul clinic. (Cap. 13)
- Criteriul principal de evaluare (engl. primary outcome, primary endpoint): rezultatul principal pe baza căruia se judecă eficacitatea intervenției și se calculează volumul eșantionului. Se stabilește în protocol, înainte de colectarea datelor. (Cap. 13)
- Cuartile (engl. quartiles): cele trei valori (Q1, Q2 = mediana, Q3) care împart seria ordonată în patru părți egale. Între Q1 și Q3 se află jumătatea centrală a observațiilor. (Cap. 2)
- Curățarea datelor (engl. data cleaning): verificarea sistematică a fișierului înainte de analiză: categorii codificate neuniform, valori imposibile, text în coloane numerice, valori lipsă, duplicate, incoerențe între variabile. Fiecare corectură se face după documentul-sursă și se notează într-un jurnal. (Cap. 2)
- Curba ROC (engl. receiver operating characteristic curve): graficul sensibilității în funcție de 1 - specificitate pentru toate pragurile posibile ale unui test numeric. Cu cât curba se apropie mai mult de colțul din stânga sus, cu atât testul este mai bun. (Cap. 12)

D

- Date cu caracter personal (engl. personal data): orice informație referitoare la o persoană identificată sau identificabilă. Datele privind sănătatea fac parte din categoriile speciale, cu protecție sporită prin Regulamentul general privind protecția datelor. (Cap. 1)
- Date pereche (date legate; engl. paired data): observații care formează perechi naturale: același individ măsurat de două ori, doi ochi ai aceluiași pacient, gemeni sau pacienți potriviți în perechi. Se analizează cu teste pentru date dependente. (Cap. 6)

- Declarația de la Helsinki (engl. Declaration of Helsinki): set de principii etice pentru cercetarea medicală pe subiecți umani, adoptat de Asociația Medicală Mondială și revizuit periodic (ultima revizie în 2024). (Cap. 13)
- Densitate de incidență (rata de incidență; engl. incidence rate): numărul cazurilor noi raportat la suma timpului de urmărire al persoanelor la risc (persoane-ani). Ține cont de faptul că participanții sunt urmăriți perioade diferite. (Cap. 10)
- Diagramă circulară (engl. pie chart): grafic în care un cerc este împărțit în sectoare proporționale cu frecvențele categoriilor. Este lizibil doar pentru puține categorii și nu se folosește pentru variabile numerice. (Cap. 2)
- Diagramă cu bare (engl. bar chart): grafic pentru variabile categoriale, în care lungimea fiecărei bare este proporțională cu frecvența categoriei. Barele sunt separate, spre deosebire de histogramă. (Cap. 2)
- Diagramă de dispersie (engl. scatter plot): grafic în care fiecare individ este un punct cu coordonatele date de două variabile numerice. Este obligatoriu înainte de calculul unui coeficient de corelație. (Cap. 9)
- Diagramă de flux (engl. flow diagram): schema care arată numărul de participanți la fiecare etapă a studiului și motivele excluderilor și pierderilor. (Cap. 13)
- Dicționar de date (engl. data dictionary): document care descrie fiecare variabilă dintr-un fișier: numele, semnificația, tipul, scala, unitatea de măsură sau codificarea, domeniul valid și modul de notare a valorilor lipsă. (Cap. 13)
- Diferența de risc (DR; engl. risk difference): diferența absolută dintre riscul la expuși (sau tratați) și riscul la neexpuși (sau martori). Exprimă efectul în puncte procentuale și stă la baza NNT. (Cap. 10)
- Diferența minimă importantă clinic (engl. minimal clinically important difference, MCID): cea mai mică diferență a unui rezultat pe care pacienții sau medicii o consideră relevantă. Se fixează la planificarea studiului și servește la calculul volumului eșantionului. (Cap. 5)
- Dihotomizare (engl. dichotomisation): transformarea unei variabile numerice într-una cu două categorii, printr-un prag (de exemplu obez dacă IMC $\geq$ 30). Simplifică interpretarea, dar pierde informație și scade puterea statistică. (Cap. 2)
- Dimensiunea eșantionului (engl. sample size): numărul de unități statistice necesar pentru a estima un parametru cu precizia dorită sau pentru a detecta un efect dat cu o putere fixată. Se calculează înainte de studiu și se rotunjește în sus. (Cap. 4)
- Distribuția Fisher (distribuția F; engl. F distribution): distribuția raportului a două variațe, definită de două numere de grade de libertate (numărător și numitor). Se folosește în ANOVA, în regresie și în testele de comparare a variațelor. (Cap. 6)
- Distribuția Student (distribuția t; engl. Student t distribution): distribuție simetrică, asemănătoare cu cea normală, dar cu cozi mai groase, folosită când abaterea standard a populației este necunoscută. Forma ei depinde de gradele de libertate și se apropie de normală când acestea cresc. (Cap. 4)
- Distribuție bimodală (engl. bimodal distribution): distribuție cu două vârfuri, care sugerează adesea amestecul a două subpopulații (de exemplu înălțimea la bărbați și femei luate împreună). (Cap. 2)
- Distribuție binomială (engl. binomial distribution): distribuția numărului de "succese" în n încercări independente, fiecare cu aceeași probabilitate p (de exemplu numărul de pacienți cu diabet într-un eșantion de 20). Pentru n mare se aproximează prin distribuția normală. (Cap. 3)
- Distribuție de eșantionare (engl. sampling distribution): distribuția pe care ar avea-o o statistică (de exemplu media) dacă am extrage un număr foarte mare de eșantioane de același volum din aceeași populație. Abaterea ei standard este eroarea standard. (Cap. 4)

- Distribuție de probabilitate (engl. probability distribution): descrierea tuturor valorilor posibile ale unei variabile aleatoare și a probabilităților cu care apar. Pentru variabilele continue se descrie printr-o curbă de densitate, sub care aria totală este 1. (Cap. 3)
- Distribuție lognormală (engl. lognormal distribution): distribuția unei variabile al cărei logaritm are distribuție normală. Este asimetrică la dreapta și descrie bine multe concentrații biologice. (Cap. 3)
- Distribuție normală (distribuția Gauss; engl. normal distribution): distribuție continuă, simetrică, în formă de clopot, complet definită de medie și abaterea standard. Media, mediana și moda coincid, iar cozile se întind teoretic la infinit. (Cap. 3)
- Distribuție normală standard (engl. standard normal distribution): distribuția normală cu media 0 și abaterea standard 1, la care se ajunge prin transformarea z. Ariile ei sunt tabelate (anexa E). (Cap. 3)
- Distribuție Poisson (engl. Poisson distribution): distribuția numărului de evenimente rare petrecute într-un interval de timp sau spațiu (numărul de cazuri noi pe lună într-un spital). Media și varianța ei sunt egale. (Cap. 3)

E

- Epi Info (engl. Epi Info): program gratuit dezvoltat de Centrele pentru Controlul și Prevenirea Bolilor din SUA, folosit în epidemiologie pentru formulare, tabele 2 x 2, RR, OR și testul chi-pătrat. (Cap. 10)
- Epidemiologie (engl. epidemiology): știința care studiază frecvența, distribuția și determinanții bolilor în populații și aplică aceste cunoștințe în prevenție și controlul bolilor. (Cap. 10)
- Eroare aleatoare (engl. random error): variația întâmplătoare a rezultatelor de la un eșantion la altul, fără direcție preferată. Scade odată cu creșterea volumului eșantionului și este cuantificată de eroarea standard și de intervalul de încredere. (Cap. 4)
- Eroare de tip I (engl. type I error): respingerea ipotezei nule atunci când ea este adevărată (rezultat fals pozitiv). Probabilitatea ei este α. (Cap. 5)
- Eroare de tip II (engl. type II error): nerespingerea ipotezei nule atunci când ea este falsă (rezultat fals negativ, efect real ratat). Probabilitatea ei este β. (Cap. 5)
- Eroare sistematică (bias; engl. bias): abatere constantă, în aceeași direcție, a rezultatelor față de valoarea reală, produsă de defecte de selecție, de măsurare sau de analiză. Spre deosebire de eroarea aleatoare, nu scade când crește eșantionul. (Cap. 1)
- Eroarea standard a estimării (SEE; engl. standard error of the estimate): abaterea standard a reziduurilor, exprimată în unitatea variabilei dependente; arată cât de departe de dreaptă sunt, în medie, punctele. (Cap. 9)
- Eroarea standard a mediei (ES; engl. standard error of the mean, SEM): abaterea standard a distribuției mediilor eșantioanelor, estimată prin $\frac{s}{\sqrt{n}}$. Măsoară precizia mediei și scade când crește volumul eșantionului. (Cap. 4)
- Eroarea standard a proporției (engl. standard error of a proportion): $\sqrt{\frac{p_i(1-p_i)}{n}}$; măsoară precizia cu care o proporție observată estimează proporția din populație. (Cap. 4)
- Estimare prin interval (engl. interval estimation): indicarea unui domeniu de valori plauzibile pentru parametru, însoțit de un nivel de încredere. Forma uzuală este intervalul de încredere. (Cap. 4)
- Estimare punctuală (engl. point estimate): o singură valoare calculată din eșantion, folosită ca cea mai bună aproximare a parametrului (de exemplu media TAS de 131,13 mmHg). (Cap. 4)
- Estimator nedeplasat (engl. unbiased estimator): estimator a cărui valoare medie, pe toate eșantioanele posibile, este egală cu parametrul estimat. Varianța calculată cu numitorul n - 1 este un estimator nedeplasat al variației populației. (Cap. 4)

- Eșantioane independente (engl. independent samples): grupuri formate din indivizi diferiți, la care valorile dintr-un grup nu sunt legate de valorile din celălalt. (Cap. 6)
- Eșantion (engl. sample): submulțimea de indivizi din populație pe care se fac efectiv măsurătorile. Calitatea concluziilor depinde mai mult de modul în care a fost ales eșantionul decât de mărimea lui. (Cap. 1)
- Eșantion reprezentativ (engl. representative sample): eșantion care reproduce structura populației (pe vârstă, sex, mediu etc.) în proporții apropiate de cele reale. Se obține prin selecție aleatoare, în care fiecare individ are aceeași șansă de a fi ales, iar alegerile sunt independente. (Cap. 1)
- Eșantionare aleatoare simplă (engl. simple random sampling): metodă de selecție în care fiecare individ din populație are aceeași probabilitate de a intra în eșantion, de exemplu prin tragere la sorți cu un tabel de cifre aleatoare sau cu un generator de numere aleatoare. (Cap. 1)
- Eșantionare de conveniență (engl. convenience sampling): includerea pacienților disponibili ușor (cei care vin la cabinet într-o anumită perioadă, respondenții unui chestionar online). Este frecventă, dar nu garantează reprezentativitatea. (Cap. 1)
- Eșantionare pe grupuri (engl. cluster sampling): se extrag aleator grupuri întregi (cabinete, școli, sate), iar în interiorul lor se studiază toți indivizii sau un subeșantion. Este practică pe teren, dar indivizii din același grup se aseamănă, ceea ce scade precizia. (Cap. 1)
- Eșantionare sistematică (engl. systematic sampling): se alege aleator un punct de pornire, apoi fiecare al k-lea individ dintr-o listă ordonată (de exemplu fiecare a zecea fișă). Este simplă, dar poate introduce erori dacă lista are o periodicitate ascunsă. (Cap. 1)
- Eșantionare stratificată (engl. stratified sampling): populația se împarte în straturi omogene (de exemplu urban și rural), iar din fiecare strat se extrage aleator un eșantion, de obicei proporțional cu mărimea stratului. Crește precizia și garantează prezența tuturor straturilor. (Cap. 1)
- Eta pătrat (η^2; engl. eta squared): proporția din variabilitatea totală explicată de factorul de grupare în ANOVA (SSB/SST). Valori orientative: 0,01 mic, 0,06 mediu, 0,14 mare. (Cap. 6)
- Etapele studiului statistic (engl. stages of a statistical study): succesiunea formulare a întrebării, planificare (design, eșantion, variabile), colectarea datelor, verificare și curățare, analiză descriptivă, analiză inferențială, interpretare și raportare. (Cap. 1)
- Eveniment (engl. event): rezultat posibil al unei observații sau experiment (pacientul are diabet, testul este pozitiv). Evenimentele se combină prin "și" (intersecție) și "sau" (reuniune). (Cap. 1)
- Eveniment advers (engl. adverse event): orice eveniment medical nefavorabil apărut la un participant în timpul studiului, indiferent dacă are sau nu legătură cu intervenția. Evenimentele grave se raportează imediat. (Cap. 13)
- Eveniment contrar (engl. complementary event): evenimentul "non-A", care se produce exact atunci când A nu se produce; $P(\text{non} - A) = 1 - P(A)$. (Cap. 1)
- Evenimente incompatibile (engl. mutually exclusive events): evenimente care nu se pot produce simultan (un pacient nu poate fi în același timp fumător și nefumător). Pentru ele, $P(A \text{ sau} B) = P(A) + P(B)$. (Cap. 1)
- Evenimente independente (engl. independent events): două evenimente pentru care producerea unuia nu modifică probabilitatea celuilalt; atunci $P(A \text{ și } B) = P(A) \times P(B)$. (Cap. 1)
- Exces (boltire; engl. kurtosis): măsură a "greutății" cozilor unei distribuții comparativ cu distribuția normală. Excesul pozitiv indică mai multe valori extreme decât s-ar aştepta la o distribuție normală. (Cap. 3)
- Expunere (engl. exposure): factorul al cărui efect asupra bolii se studiază (un comportament, un tratament, un mediu, o caracteristică genetică). Poate fi punctuală sau cronică, caz în care contează doza și durata. (Cap. 10)

- Extrapolare (engl. extrapolation): folosirea unui model pentru a prezice valori în afara domeniului de valori ale predictorului observat în date. Este riscantă, pentru că relația poate să nu mai fie liniară acolo. (Cap. 9)

F

- Factor de confuzie (engl. confounder): variabilă asociată atât cu expunerea, cât și cu rezultatul, fără a fi o verigă intermediară între ele, care poate crea sau ascunde o asociere. Se controlează prin randomizare, restricție, potrivire, stratificare sau ajustare în modele multivariate. (Cap. 10)
- Factor de risc (engl. risk factor): caracteristică, comportament sau expunere care crește probabilitatea apariției unei boli (fumatul pentru BPOC, obezitatea pentru diabet). Asocierea nu înseamnă automat cauzalitate. (Cap. 10)
- Factor protector (engl. protective factor): expunere asociată cu o probabilitate mai mică de boală, cu RR sau OR sub 1 (activitatea fizică pentru diabet). (Cap. 10)
- Fals negativ (FN; engl. false negative): bolnav cu test negativ (celula b). Consecința este întârzierea diagnosticului și a tratamentului. (Cap. 12)
- Fals pozitiv (FP; engl. false positive): persoană fără boală cu test pozitiv (celula c). Consecințele sunt anxietatea și investigațiile inutile. (Cap. 12)
- Fazele studiilor clinice (engl. clinical trial phases): etapele dezvoltării unui medicament: faza I (siguranță, doză, voluntari puțini), faza II (eficacitate preliminară), faza III (comparație randomizată pe loturi mari), faza IV (supraveghere după autorizare). (Cap. 13)
- Foaie de calcul (engl. spreadsheet): program de tip tabel (Google Sheets, Excel) în care datele se organizează cu un rând pentru fiecare unitate statistică și o coloană pentru fiecare variabilă. (Cap. 2)
- Formular de raportare a cazului (FRC; engl. case report form, CRF): formularul, pe hârtie sau electronic, în care se înregistrează datele fiecărui participant cerute de protocol. Un formular bine proiectat previne o mare parte din erorile de introducere. (Cap. 13)
- Fracție atribuibilă în populație (FAP; engl. population attributable fraction): proporția din toate cazurile din populație care ar dispărea dacă expunerea ar fi eliminată, presupunând o relație cauzală. Depinde atât de RR, cât și de cât de frecventă este expunerea. (Cap. 10)
- Fracție atribuibilă la expuși (engl. attributable fraction among the exposed): proporția din cazurile apărute la expuși care s-ar datora expunerii, (RR - 1)/RR. (Cap. 10)
- Frecvență absolută (engl. absolute frequency): numărul de indivizi care au o anumită valoare sau aparțin unei anumite categorii. Testul chi-pătrat se calculează întotdeauna pe frecvențe absolute, niciodată pe procente. (Cap. 2)
- Frecvență așteptată (engl. expected frequency): numărul de observații care s-ar afla într-o celulă dacă variabilele ar fi independente, calculat ca total rând x total coloană / N. Condiția uzuală a testului chi-pătrat este ca frecvențele așteptate să fie de cel puțin 5. (Cap. 7)
- Frecvență cumulată (engl. cumulative frequency): suma frecvențelor tuturor categoriilor până la categoria curentă inclusiv. Are sens doar pentru variabile ordonate (ordinale sau numerice grupate în clase). (Cap. 2)
- Frecvență relativă (engl. relative frequency): raportul dintre numărul de apariții ale unei valori și numărul total de observații, exprimat ca fracție sau procent. Suma frecvențelor relative ale tuturor categoriilor este 1 (100%). (Cap. 2)
- Frecvență validă (engl. valid percent): procentul calculat la numărul de răspunsuri valide, excluzând valorile lipsă. Când lipsesc date, procentul valid diferă de procentul calculat la total, iar raportul trebuie să precizeze care dintre ele s-a folosit. (Cap. 2)

G

- Ghidul CONSORT (engl. Consolidated Standards of Reporting Trials): ghid de raportare a studiilor clinice randomizate, cu o listă de verificare și o diagramă de flux a participanților (evaluați, randomizați, alocați, urmăriți, analizați). (Cap. 13)
- Grade de libertate (gl; engl. degrees of freedom, df): numărul de valori care pot varia liber după ce s-au fixat anumite restricții (de exemplu media). Determină forma distribuțiilor t, chi-pătrat și F; pentru testul t pe un eșantion, gl = n - 1. (Cap. 4)
- Grafic forest (engl. forest plot): reprezentarea grafică a estimărilor (RR, OR, diferențe de medii) și a intervalelor lor de încredere pentru mai multe studii sau subgrupuri, cu o linie verticală la valoarea "fără efect". (Cap. 10)
- Grafic Q-Q (engl. quantile-quantile plot): grafic care pune cuantilele observate față de cuantilele teoretice ale distribuției normale. Dacă datele sunt normale, punctele se așază aproximativ pe o dreaptă. (Cap. 3)
- Grup de control (engl. control group): grupul care primește intervenția de comparație (placebo, tratamentul standard sau nicio intervenție), față de care se măsoară efectul intervenției studiate. (Cap. 13)

H

- Heterogenitate (engl. heterogeneity): variația rezultatelor între studiile incluse într-o meta-analiză, mai mare decât ar explica întâmplarea; se cuantifică de obicei prin indicele I^2. (Cap. 11)
- Histogramă (engl. histogram): grafic din dreptunghiuri alăturate care arată distribuția unei variabile numerice grupate în clase. Forma ei sugerează simetria, asimetria sau existența mai multor vârfuri. (Cap. 2)

I

- Ierarhia dovezilor (engl. hierarchy of evidence): ordonarea tipurilor de studii după cât de bine controlează erorile sistematice, de la opinia expertului și seriile de cazuri, prin studiile observaționale, până la studiile randomizate și reviziile sistematice cu meta-analiză. (Cap. 11)
- Incidență cumulativă (risc; engl. cumulative incidence): proporția persoanelor inițial sănătoase care dezvoltă boala într-o perioadă dată. Se exprimă întotdeauna împreună cu perioada (de exemplu 3% în 5 ani). (Cap. 10)
- Indicator robust (engl. robust statistic): indicator puțin influențat de valorile extreme sau de abaterile de la normalitate, cum sunt mediana și intervalul intercuartilic. (Cap. 2)
- Indicele Youden (J; engl. Youden index): Se + Sp - 1, între 0 și 1. Pragul care maximizează J este un criteriu frecvent de alegere a pragului optim, când costurile celor două tipuri de erori sunt considerate egale. (Cap. 12)
- Integritatea cercetării (engl. research integrity): respectarea onestității în toate etapele cercetării. Abaterile grave sunt fabricarea datelor, falsificarea lor și plagiatul; la acestea se adaugă practici discutabile, precum raportarea selectivă a rezultatelor semnificative. (Cap. 1)
- Interpretarea critică a unui articol (engl. critical appraisal): evaluarea sistematică a validității, rezultatelor și aplicabilității unui studiu publicat: designul potrivit întrebării, controlul biasului, adecvarea testelor, raportarea IC și a mărimii efectului. (Cap. 14)
- Interval de încredere (IC; engl. confidence interval, CI): interval calculat din eșantion care, dacă studiul s-ar repeta de multe ori, ar conține valoarea reală a parametrului într-un procent dat de cazuri (de obicei 95%). Lățimea lui arată precizia estimării. (Cap. 4)
- Interval de încredere pentru o diferență (engl. confidence interval for a difference): IC calculat pentru diferența a două medii sau proporții. Dacă nu conține 0, diferența este semnificativă la pragul corespunzător; pentru rapoarte (RR, OR), valoarea de referință este 1. (Cap. 4)
- Interval de referință (valori normale; engl. reference interval): intervalul în care se află 95% centrale din valorile unei populații sănătoase, calculat ca media ± 1,96 abateri standard (date normale) sau între percentilele 2,5 și 97,5. Descrie indivizii, nu precizia mediei. (Cap. 3)

- Interval intercuartilic (IQR; engl. interquartile range): diferența Q3 - Q1, care cuprinde 50% centrale din valori. Însoțește mediana în descrierea distribuțiilor asimetrice și stă la baza regulii de detectare a valorilor aberante pe box-plot. (Cap. 2)
- Intervalul Wald pentru proporție (engl. Wald interval): IC calculat ca $p_i \pm z \times \sqrt{\frac{p_i(1-p_i)}{n}}$. Este simplu, dar inexact pentru eșantioane mici sau proporții apropiate de 0 ori 1, unde poate da chiar limite negative. (Cap. 4)
- Intervalul Wilson pentru proporție (engl. Wilson score interval): IC pentru proporție cu proprietăți mai bune decât intervalul Wald, recomandat mai ales la eșantioane mici și proporții extreme. Nu este simetric în jurul lui p. (Cap. 4)
- Ipoteza alternativă (H1, Ha; engl. alternative hypothesis): ipoteza care contrazice H0, susținând existența unei diferențe sau asocieri. Poate fi bilaterală ($\mu1 \neq \mu2$) sau unilaterală ($\mu1 > \mu2$). (Cap. 5)
- Ipoteza nulă (H0; engl. null hypothesis): ipoteza statistică a absenței diferenței sau a asocierii ($\mu1 = \mu2$, RR = 1, $\rho = 0$). Testul calculează cât de compatibile sunt datele cu ea; H0 se respinge sau nu se respinge, dar nu se "acceptă". (Cap. 5)
- Ipoteză de cercetare (engl. research hypothesis): afirmația pe care cercetătorul vrea să o verifice, formulată în termeni clinici (de exemplu, "teleconsultația scade glicemia mai mult decât consilierea standard"). Din ea derivă ipotezele statistice. (Cap. 5)

Î

- Întrebarea PICO (engl. PICO question): formularea structurată a unei întrebări clinice: pacientul sau problema (P), intervenția sau expunerea (I), comparația (C) și rezultatul (O). Ușurează căutarea literaturii. (Cap. 11)

J

- JASP (engl. JASP): program statistic gratuit, cu interfață grafică, folosit la lucrările practice pentru statistica descriptivă, testele t, ANOVA, testele neparametrice, corelație și regresie. (Cap. 2)

L

- Legea numerelor mari (engl. law of large numbers): pe măsură ce numărul observațiilor crește, frecvența relativă a unui eveniment se apropie de probabilitatea lui, iar media eșantionului de media populației. (Cap. 1)
- Letalitate (engl. case fatality rate): proporția bolnavilor de o anumită boală care decedează din cauza ei într-o perioadă dată. Măsoară gravitatea bolii, nu frecvența ei în populație. (Cap. 10)
- Liniaritate (engl. linearity): condiția ca relația dintre două variabile să poată fi descrisă de o dreaptă. Se verifică pe diagrama de dispersie și pe graficul reziduurilor. (Cap. 9)

M

- Marja de eroare (engl. margin of error): jumătatea lățimii intervalului de încredere, adică distanța de la estimarea punctuală la fiecare limită. Este mărimea fixată la calculul volumului eșantionului. (Cap. 4)
- Mărime a efectului clinic (efect absolut; engl. absolute effect): exprimarea rezultatului în unități ușor de înțeles pentru decizia clinică (mmHg, puncte procentuale, NNT), complementară măsurilor relative precum RR. (Cap. 10)
- Mărimea efectului (engl. effect size): măsură a amplitudinii unei diferențe sau asocieri, independentă de volumul eșantionului (Cohen d, eta pătrat, r, OR). Trebuie raportată alături de valoarea p. (Cap. 5)
- MedCalc (engl. MedCalc): program statistic orientat spre cercetarea medicală, folosit în curs în special pentru curbe ROC, tabele 2 x 2, riscuri relative și calculul volumului eșantionului. (Cap. 12)

- Media aritmetică (engl. arithmetic mean): suma valorilor împărțită la numărul lor. Este indicatorul de tendință centrală preferat pentru distribuții simetrice, dar este sensibil la valorile extreme. (Cap. 2)
- Media geometrică (engl. geometric mean): rădăcina de ordin n a produsului celor n valori, egală cu exponențiala mediei logaritmilor. Este potrivită pentru variabile pozitive cu asimetrie la dreapta, cum sunt trigliceridele sau titrurile de anticorpi. (Cap. 2)
- Media ponderată (engl. weighted mean): media în care fiecare valoare contribuie proporțional cu o pondere, de exemplu volumul subgrupului. Se folosește la combinarea mediilor unor grupuri de mărimi diferite. (Cap. 2)
- Mediana (engl. median): valoarea care împarte seria ordonată în două jumătăți egale ca număr de observații. Nu este influențată de valorile extreme și se preferă la distribuțiile asimetrice și la variabilele ordinale. (Cap. 2)
- Medicina bazată pe dovezi (MBD; engl. evidence-based medicine): folosirea conștientă și explicită a celor mai bune dovezi disponibile din cercetare, integrate cu experiența clinică și cu valorile pacientului, în deciziile despre îngrijirea pacienților individuali. (Cap. 11)
- Meta-analiză (engl. meta-analysis): combinarea statistică a rezultatelor mai multor studii independente care răspund aceleiași întrebări, pentru a obține o estimare globală mai precisă a efectului. (Cap. 11)
- Metoda celor mai mici pătrate (engl. least squares method): metoda prin care se aleg coeficienții dreptei de regresie astfel încât suma pătratelor reziduurilor să fie minimă. (Cap. 9)
- Moda (engl. mode): valoarea sau categoria cea mai frecventă. Este singurul indicator de tendință centrală aplicabil variabilelor nominale. (Cap. 2)
- Modificare de efect (interacțiune; engl. effect modification): situația în care mărimea asocierii dintre expunere și boală diferă în funcție de nivelul unei a treia variabile (de exemplu, efectul unui tratament diferă între bărbați și femei). (Cap. 10)
- Morbiditate (engl. morbidity): frecvența bolilor într-o populație, exprimată prin prevalență sau incidență. (Cap. 10)
- Mortalitate (engl. mortality): frecvența deceselor într-o populație într-o perioadă, raportată la întreaga populație (de exemplu, decese la 100.000 de locuitori pe an). (Cap. 10)
- Multicoliniaritate (engl. multicollinearity): corelație puternică între predictorii unui model de regresie, care face coeficienții instabili și greu de interpretat (de exemplu greutatea și IMC în același model). (Cap. 9)

N

- Nivel de încredere (engl. confidence level): procentul (1 - α) x 100 asociat unui interval de încredere, de obicei 95% sau 99%. Un nivel mai ridicat produce un interval mai larg. (Cap. 4)
- Nomograma Fagan (engl. Fagan nomogram): grafic cu trei axe (probabilitate pre-test, raport de verosimilitate, probabilitate post-test) pe care o linie dreaptă face, fără calcule, conversia bayesiană. (Cap. 11)
- Normalitate (engl. normality): proprietatea datelor de a urma, cu aproximație, o distribuție normală. Este o condiție a testelor parametrice și se verifică grafic (histogramă, grafic Q-Q) și prin teste (Shapiro-Wilk). (Cap. 3)
- Numărul necesar de tratat (NNT; engl. number needed to treat): numărul de pacienți care trebuie tratați pentru ca un pacient în plus să aibă beneficiu, egal cu 1/RAR și rotunjit în sus. Un NNT mic indică un tratament eficient. (Cap. 10)
- Numărul necesar pentru a produce un efect advers (NNH; engl. number needed to harm): numărul de pacienți care trebuie tratați pentru ca un pacient în plus să sufere un efect advers, egal cu inversul creșterii absolute a riscului de efect advers. (Cap. 10)

O

- Omogenitatea varianțelor (homoscedasticitate; engl. homogeneity of variance): condiția ca variabilitatea să fie aproximativ aceeași în grupurile comparate (sau, în regresie, de-a lungul dreptei). Este o condiție a testului t clasic, a ANOVA și a regresiei liniare. (Cap. 6)
- Orbire (mascare; engl. blinding): ascunderea intervenției primite față de participant (simplu orb), față de participant și investigator (dublu orb) sau și față de cel care evaluează rezultatele și analizează datele. Reduce biasul de performanță și de evaluare. (Cap. 13)
- Ordonata la origine (a; engl. intercept): valoarea prezisă a lui y când x = 0. Adesea nu are interpretare clinică, pentru că x = 0 se află în afara datelor (IMC 0 nu există). (Cap. 9)

P

- Panta dreptei de regresie (coeficient de regresie, b; engl. slope, regression coefficient): modificarea medie a lui y când x crește cu o unitate. Semnul ei coincide cu semnul coeficientului de corelație. (Cap. 9)
- Parametru (engl. parameter): valoare numerică ce descrie populația (media μ, abaterea standard σ, proporția π). De regulă este necunoscut și se estimează din eșantion. (Cap. 4)
- Percentilă (engl. percentile): valoarea sub care se află un anumit procent din observații; percentila 90 lasă sub ea 90% din valori. Curbele de creștere din pediatrie sunt construite pe percentile. (Cap. 2)
- Persoane-timp (engl. person-time): suma timpilor de urmărire ai tuturor participanților până la boală, pierdere din urmărire sau sfârșitul studiului. Este numitorul densității de incidență. (Cap. 10)
- Pierdere din urmărire (engl. loss to follow-up): situația în care rezultatul unui participant nu mai poate fi măsurat (s-a mutat, nu mai răspunde). Pierderile mari sau inegale între brațe pot introduce bias. (Cap. 13)
- Placebo (engl. placebo): intervenție inactivă, identică în aparență cu tratamentul activ, folosită în grupul de control pentru a permite orbirea și pentru a separa efectul specific de efectul placebo. (Cap. 13)
- Plan de analiză statistică (PAS; engl. statistical analysis plan, SAP): documentul care stabilește dinainte metodele statistice pentru fiecare rezultat, tratarea valorilor lipsă și analizele de subgrup, protejând studiul de alegerile făcute după vederea datelor. (Cap. 13)
- Populație (engl. population): ansamblul tuturor indivizilor care au cel puțin o caracteristică comună și despre care vrem să tragem concluzii, delimitat în spațiu și timp. Exemplu: toți adulții cu risc cardiometabolic înscriși la medicii de familie dintr-o regiune. (Cap. 1)
- Populație țintă (engl. target population): populația căreia îi sunt destinate concluziile studiului, adesea mai largă decât populația din care s-a extras efectiv eșantionul (populația sursă). Diferența dintre cele două limitează validitatea externă. (Cap. 1)
- Prag de semnificație (α; engl. significance level): probabilitatea maximă de eroare de tip I acceptată, fixată înaintea analizei, de obicei 0,05. Dacă $p < \alpha$, rezultatul se declară statistic semnificativ. (Cap. 5)
- Precizie (engl. precision): gradul de apropiere între rezultatele unor măsurători sau estimări repetate; în estimare, este reflectată de lățimea intervalului de încredere. O estimare poate fi precisă, dar deplasată, dacă există bias. (Cap. 4)
- Predicție (engl. prediction): estimarea valorii variabilei dependente pentru un individ nou, pe baza modelului. Intervalul de predicție pentru un individ este mult mai larg decât intervalul de încredere pentru media lui y. (Cap. 9)
- Prevalență (engl. prevalence): proporția persoanelor dintr-o populație care au boala la un moment dat (prevalență de moment) sau într-o perioadă (prevalență de perioadă). Include atât cazurile noi, cât și pe cele vechi. (Cap. 10)

- Probabilitate (engl. probability): număr între 0 și 1 care exprimă cât de posibil este un eveniment; 0 înseamnă imposibil, 1 sigur. Se poate estima prin frecvența relativă a evenimentului într-un număr mare de observații. (Cap. 1)
- Probabilitate condiționată (engl. conditional probability): probabilitatea unui eveniment știind că un alt eveniment s-a produs, notată $P(A \mid B) = \frac{P(A \text{ și } B)}{P(B)}$. Sensibilitatea, specificitatea și valorile predictive sunt probabilități condiționate. (Cap. 1)
- Probabilitate post-test (probabilitate a posteriori; engl. post-test probability): probabilitatea bolii după cunoașterea rezultatului testului. Pentru un test pozitiv este egală cu VPP corespunzătoare probabilității pre-test. (Cap. 11)
- Probabilitate pre-test (probabilitate a priori; engl. pre-test probability): probabilitatea ca pacientul să aibă boala înainte de rezultatul testului, estimată din prevalență și din datele clinice ale pacientului. (Cap. 11)
- Protocolul studiului (engl. study protocol): documentul care descrie în detaliu întrebarea, designul, populația, intervențiile, rezultatele, volumul eșantionului, randomizarea, colectarea datelor, planul de analiză statistică și aspectele etice ale unui studiu. (Cap. 13)
- Pseudonimizare (engl. pseudonymisation): înlocuirea identificatorilor direcți (nume, CNP) cu un cod, păstrând separat cheia de legătură. Datele pseudonimizate rămân date cu caracter personal și trebuie protejate ca atare. (Cap. 1)
- Puterea testului (engl. statistical power): probabilitatea de a respinge H0 atunci când ea este falsă, egală cu 1 - β. Crește cu mărimea efectului, cu volumul eșantionului și cu pragul α și scade când variabilitatea este mare; se planifică de obicei la 80% sau 90%. (Cap. 5)

R

- Randomizare (engl. randomisation): alocarea participanților în brațele studiului printr-un procedeu aleator, astfel încât fiecare să aibă o șansă cunoscută (de obicei egală) de a primi fiecare intervenție. Este principala protecție împotriva factorilor de confuzie. (Cap. 13)
- Randomizare în blocuri (engl. block randomisation): randomizare în care alocările se fac în blocuri mici (de exemplu de 4 sau 6), fiecare bloc conținând un număr egal de alocări pentru fiecare braț. Menține grupurile echilibrate numeric pe tot parcursul recrutării. (Cap. 13)
- Randomizare stratificată (engl. stratified randomisation): randomizare făcută separat în straturi definite de factori prognostici importanți (centru, sex, severitate), pentru a echilibra acești factori între brațe. (Cap. 13)
- Rang (engl. rank): poziția unei valori în seria ordonată crescător. Valorile egale primesc rangul mediu al pozițiilor pe care le ocupă. (Cap. 7)
- Raport de verosimilitate (RV; engl. likelihood ratio, LR): de câte ori este mai probabil un anumit rezultat al testului la un bolnav decât la un nebolnav. $\text{RV}^+ = \frac{\text{Se}}{1-\text{Sp}}$, $\text{RV}^- = \frac{1-\text{Se}}{\text{Sp}}$; un RV+ peste 10 sau un RV- sub 0,1 modifică puternic probabilitatea bolii. (Cap. 11)
- Raportare selectivă (engl. selective reporting): prezentarea doar a rezultatelor semnificative sau favorabile dintre toate analizele efectuate. Distorsionează literatura și este prevenită prin înregistrarea prealabilă a studiilor. (Cap. 5)
- Raportul cotelor (OR; engl. odds ratio): raportul dintre cota bolii la expuși și cota bolii la neexpuși, ad/bc într-un tabel 2 x 2. Este singura măsură de asociere validă în studiile caz-control și aproximează RR când boala este rară. (Cap. 10)
- Raportul cotelor de diagnostic (DOR; engl. diagnostic odds ratio): raportul dintre cota unui test pozitiv la bolnavi și cota unui test pozitiv la nebolnavi, egal cu ad/bc sau RV+/RV-. Rezumă performanța testului într-un singur număr. (Cap. 12)
- Raportul prevalențelor (RP; engl. prevalence ratio): analogul riscului relativ în studiile transversale, calculat ca raport între prevalența bolii la expuși și la neexpuși. (Cap. 10)

- Rata (engl. rate): raport care include timpul în numitor (cazuri pe persoane-ani). În limbajul curent, termenul este folosit adesea impropriu pentru proporții. (Cap. 10)
- Rata erorii de clasificare (engl. misclassification rate): proporția persoanelor clasificate greșit, $\frac{b+c}{N} = \frac{\text{FN+FP}}{N}$, adică 1 - acuratețea. (Cap. 12)
- Recensământ (engl. census): studiul tuturor indivizilor dintr-o populație. Este exact, dar costisitor și rar posibil în cercetarea medicală, unde se lucrează aproape întotdeauna pe eșantioane. (Cap. 1)
- Reducerea absolută a riscului (RAR; engl. absolute risk reduction): diferența dintre riscul din grupul de control și riscul din grupul tratat, pentru un tratament benefic. Inversul ei este NNT. (Cap. 10)
- Reducerea relativă a riscului (RRR; engl. relative risk reduction): scăderea riscului exprimată ca proporție din riscul grupului de control, 1 - RR. Pare impresionantă chiar și când beneficiul absolut este mic. (Cap. 10)
- Registru de studii clinice (engl. trial registry): bază de date publică în care studiile se înregistrează înainte de înrolarea primului participant, cu criteriile de evaluare prespecificate. Previne raportarea selectivă și publicarea doar a rezultatelor favorabile. (Cap. 13)
- Regiune de respingere (engl. rejection region): mulțimea valorilor statisticii testului pentru care H0 se respinge. La un test bilateral este împărțită în două cozi, fiecare cu aria α/2. (Cap. 5)
- Regresie liniară multiplă (engl. multiple linear regression): model cu o variabilă dependentă numerică și mai mulți predictori; fiecare coeficient arată efectul predictorului respectiv la valori constante ale celorlalți (ajustare). (Cap. 9)
- Regresie liniară simplă (engl. simple linear regression): model care descrie variabila dependentă y ca funcție liniară de un predictor x, y = a + b x x, și permite predicția valorilor lui y. (Cap. 9)
- Regresie logistică (engl. logistic regression): model pentru o variabilă dependentă dihotomică (bolnav sau sănătos), ai cărui coeficienți, exponențiați, dau rapoarte ale cotelor ajustate pentru ceilalți predictori. (Cap. 9)
- Regula 68-95-99,7 (engl. empirical rule): într-o distribuție normală, aproximativ 68,3% din valori se află la cel mult o abatere standard de medie, 95,4% la cel mult două și 99,7% la cel mult trei. Exact 95% se află în media ± 1,96 abateri standard. (Cap. 3)
- Regula adunării (engl. addition rule): $P(A \text{ sau} B) = P(A) + P(B) - P(A \text{ și } B)$. Termenul scăzut evită numărarea de două ori a indivizilor care au ambele caracteristici. (Cap. 1)
- Regula înmulțirii (engl. multiplication rule): $P(A \text{ și } B) = P(B) \times P(A \mid B)$; pentru evenimente independente devine P(A) x P(B). (Cap. 1)
- Regulamentul (UE) 536/2014 (engl. EU Clinical Trials Regulation): regulamentul european care reglementează autorizarea și desfășurarea studiilor clinice cu medicamente de uz uman în Uniunea Europeană. (Cap. 13)
- Regulamentul general privind protecția datelor (GDPR; engl. General Data Protection Regulation): Regulamentul (UE) 2016/679, care stabilește condițiile în care pot fi prelucrate datele cu caracter personal, inclusiv cele medicale folosite în cercetare. (Cap. 1)
- Reproductibilitate (engl. reproducibility): posibilitatea ca alți cercetători să obțină aceleași rezultate pornind de la aceleași date și aceleași metode. Presupune date documentate, scripturi sau pași de analiză descriși complet și versiuni de program precizate. (Cap. 14)
- Revizie sistematică (engl. systematic review): sinteză a tuturor studiilor relevante pentru o întrebare precisă, cu căutare exhaustivă, criterii explicite de selecție și evaluarea calității fiecărui studiu. (Cap. 11)
- Reziduu (engl. residual): diferența dintre valoarea observată și valoarea prezisă de model pentru un individ. Analiza reziduurilor verifică liniaritatea, omogenitatea varianței și normalitatea. (Cap. 9)

- Risc atribuibil (engl. attributable risk): excesul de risc la expuși datorat expunerii, egal cu diferența de risc, presupunând o relație cauzală. (Cap. 10)
- Risc relativ (RR; engl. relative risk, risk ratio): raportul dintre riscul (incidența) bolii la expuși și riscul la neexpuși. RR = 1 indică absența asocierii, RR > 1 un factor de risc, RR < 1 un factor protector. Se calculează în studiile de cohortă și în studiile clinice. (Cap. 10)
- Robustețe (engl. robustness): proprietatea unui test de a da rezultate corecte chiar când condițiile lui sunt încălcate moderat. Testul t este destul de robust la abateri ușoare de la normalitate pe eșantioane mari. (Cap. 8)

S

- Scală de interval (engl. interval scale): scală numerică cu distanțe egale între valori, dar fără zero absolut (temperatura în grade Celsius). Diferențele au sens, rapoartele nu. (Cap. 2)
- Scală de raport (engl. ratio scale): scală numerică cu distanțe egale și zero absolut (greutatea, concentrația unei substanțe), pentru care și rapoartele au sens: 100 kg înseamnă de două ori mai mult decât 50 kg. (Cap. 2)
- Scală Likert (engl. Likert scale): scală de răspuns ordonată, de obicei cu 5 sau 7 trepte (de la "deloc de acord" la "total de acord"), folosită în chestionare. Datele obținute sunt ordinale și se analizează de preferință cu teste neparametrice. (Cap. 7)
- Scor z (scor standardizat; engl. z-score): numărul de abateri standard la care se află o valoare față de medie, $z = \frac{x-m}{s}$. Permite compararea unor valori măsurate pe scale diferite și citirea ariilor din tabelul normal. (Cap. 3)
- Semnificație clinică (engl. clinical significance): importanța practică a unui efect pentru pacient sau pentru decizia medicală. Se judecă după mărimea efectului și după intervalul de încredere, comparate cu diferența minimă importantă clinic. (Cap. 5)
- Semnificație statistică (engl. statistical significance): situația în care $p < \alpha$, adică datele sunt puțin compatibile cu H0. Nu garantează că efectul este mare, important clinic sau cauzal. (Cap. 5)
- Sensibilitate (Se; engl. sensitivity): proporția bolnavilor la care testul este pozitiv, a/(a + b). Un test foarte sensibil ratează puțini bolnavi; dacă este negativ, exclude boala (SnNOut). (Cap. 12)
- Serie de cazuri (engl. case series): descrierea unui grup de pacienți cu aceeași boală sau același tratament, fără grup de comparație. Poate genera ipoteze, dar nu le poate testa. (Cap. 10)
- Sfericitate (engl. sphericity): condiția ANOVA cu măsurători repetate ca variațele diferențelor dintre toate perechile de momente să fie egale; se verifică prin testul Mauchly și, dacă nu este îndeplinită, se aplică o corecție (Greenhouse-Geisser). (Cap. 6)
- Sistemul GRADE (engl. GRADE approach): metodă de evaluare a certitudinii dovezilor (ridicată, moderată, scăzută, foarte scăzută) și a tăriei recomandărilor din ghiduri. (Cap. 11)
- SnNOut (engl. SnNOut): regulă mnemotehnică: un test foarte sensibil (Sn), negativ (N), exclude boala (Out, rule out). (Cap. 12)
- Specificitate (Sp; engl. specificity): proporția nebolnavilor la care testul este negativ, d/(c + d). Un test foarte specific dă puține rezultate fals pozitive; dacă este pozitiv, confirmă boala (SpPIn). (Cap. 12)
- SpPIn (engl. SpPIn): regulă mnemotehnică: un test foarte specific (Sp), pozitiv (P), confirmă boala (In, rule in). (Cap. 12)
- Standard de referință (gold standard; engl. reference standard): cea mai bună metodă disponibilă pentru stabilirea prezenței bolii, față de care se evaluează testul studiat (de exemplu biopsia pentru fibroza hepatică, TTGO pentru diabet). (Cap. 12)
- Standardizarea ratelor (engl. rate standardisation): metodă de comparare a ratelor între populații cu structuri diferite pe vârstă, prin raportarea la o populație standard comună. (Cap. 10)

- STARD (engl. Standards for Reporting of Diagnostic Accuracy Studies): ghid cu elementele care trebuie raportate într-un studiu de acurateţe diagnostică, inclusiv diagrama de flux a pacienţilor şi tabelul 2 x 2. (Cap. 12)
- Statistica F (engl. F statistic): raportul dintre media pătratelor între grupuri şi media pătratelor în interiorul grupurilor. Valori mult peste 1 arată că mediile diferă mai mult decât ar explica variabilitatea întâmplătoare. (Cap. 6)
- Statistica testului (engl. test statistic): valoarea calculată din date (t, z, F, χ^2, U) care măsoară cât de departe sunt datele de ceea ce ar prezice H0. Din ea şi din distribuţia ei sub H0 se obţine valoarea p. (Cap. 5)
- Statistică (indicator de eşantion) (engl. statistic): valoare calculată din datele eşantionului (media m, abaterea standard s, proporţia p), folosită ca estimare a parametrului corespunzător. Variază de la un eşantion la altul. (Cap. 4)
- Statistică bayesiană (engl. Bayesian statistics): abordarea în care cunoştinţele anterioare (distribuţia a priori) se actualizează cu datele, producând o distribuţie a posteriori a parametrului. Permite afirmaţii directe de tipul "probabilitatea ca efectul să fie pozitiv este 95%". (Cap. 11)
- Statistică descriptivă (engl. descriptive statistics): partea statisticii care rezumă datele unui eşantion prin tabele de frecvenţe, indicatori (medie, mediană, abatere standard) şi grafice, fără a face afirmaţii despre populaţie. Este primul pas al oricărei analize. (Cap. 2)
- Statistică frecventistă (engl. frequentist statistics): abordarea în care probabilitatea este frecvenţa pe termen lung a unui eveniment în repetări ipotetice ale studiului; pe ea se bazează valorile p şi intervalele de încredere din acest curs. (Cap. 11)
- Statistică inferenţială (engl. inferential statistics): partea statisticii care generalizează de la eşantion la populaţie, prin estimarea parametrilor cu intervale de încredere şi prin testarea ipotezelor. Concluziile ei sunt probabilistice, nu certe. (Cap. 1)
- STROBE (engl. Strengthening the Reporting of Observational Studies in Epidemiology): ghid de raportare pentru studiile observaţionale (cohortă, caz-control, transversale). (Cap. 10)
- Studiu caz-control (engl. case-control study): studiu care porneşte de la bolnavi (cazuri) şi persoane fără boală (martori) şi compară frecvenţa expunerii anterioare. Este eficient pentru boli rare, dar permite doar calculul OR şi este expus biasului de memorie. (Cap. 10)
- Studiu clinic (engl. clinical trial): studiu experimental pe oameni care evaluează efectele unei intervenţii (medicament, dispozitiv, procedură, program educaţional) asupra sănătăţii. Se desfăşoară după un protocol aprobat de comisia de etică. (Cap. 13)
- Studiu clinic randomizat (SCR; engl. randomised controlled trial, RCT): studiu în care participanţii sunt alocaţi la întâmplare în grupuri diferite de intervenţie, ceea ce face grupurile comparabile atât pentru factorii cunoscuţi, cât şi pentru cei necunoscuţi. (Cap. 13)
- Studiu de cohortă (engl. cohort study): studiu în care un grup de persoane fără boală, clasificate după expunere, este urmărit în timp pentru apariţia bolii. Permite calculul incidenţei şi al RR; poate fi prospectiv sau retrospectiv (istoric). (Cap. 10)
- Studiu de echivalenţă (engl. equivalence trial): studiu care urmăreşte să arate că două intervenţii nu diferă în niciun sens cu mai mult decât o marjă prestabilită. (Cap. 13)
- Studiu de non-inferioritate (engl. non-inferiority trial): studiu care urmăreşte să arate că o intervenţie nouă nu este mai rea decât tratamentul standard cu mai mult decât o marjă prestabilită, de obicei pentru că are alte avantaje (cost, siguranţă, administrare). (Cap. 13)
- Studiu de superioritate (engl. superiority trial): studiu care urmăreşte să arate că o intervenţie este mai bună decât comparatorul. (Cap. 13)
- Studiu ecologic (engl. ecological study): studiu în care unitatea de analiză este un grup (ţară, judeţ), nu individul. Asocierile observate la nivel de grup nu se pot transfera automat la indivizi (eroarea ecologică). (Cap. 10)

- Studiu experimental (engl. experimental study): studiu în care cercetătorul alocă intervenția participanților, ideal prin randomizare. Studiul clinic randomizat este forma de referință. (Cap. 13)
- Studiu explicativ (engl. explanatory trial): studiu clinic în condiții ideale, controlate strict, care răspunde la întrebarea "poate funcționa intervenția?" (eficacitate, nu eficiență). (Cap. 13)
- Studiu încrucișat (engl. crossover trial): studiu în care fiecare participant primește succesiv ambele intervenții, în ordine randomizată, separate de o perioadă de spălare. Fiecare pacient este propriul control, ceea ce cere teste pentru date pereche. (Cap. 13)
- Studiu observațional (engl. observational study): studiu în care cercetătorul doar observă expunerea și rezultatul, fără să le influențeze (transversal, cohortă, caz-control). Este expus factorilor de confuzie. (Cap. 10)
- Studiu pilot (engl. pilot study): studiu de mici dimensiuni care testează fezabilitatea procedurilor (recrutare, formulare, aderență) înaintea studiului principal; nu este conceput pentru a testa eficacitatea. (Cap. 13)
- Studiu pragmatic (engl. pragmatic trial): studiu clinic desfășurat în condiții apropiate de practica obișnuită, cu criterii de includere largi, care răspunde la întrebarea "funcționează intervenția în viața reală?". (Cap. 13)
- Studiu transversal (engl. cross-sectional study): studiu în care expunerea și boala se măsoară în același moment, ca o fotografie a populației. Estimează prevalența, dar nu poate stabili succesiunea în timp a expunerii și bolii. (Cap. 10)

T

- Tabel 2 x 2 (engl. two-by-two table): tabel de contingență cu două rânduri și două coloane (celulele a, b, c, d), folosit atât în analiza riscului (expunere x boală), cât și în evaluarea testelor diagnostice (boală x rezultatul testului). (Cap. 10)
- Tabel de contingență (engl. contingency table): tabel cu dublă intrare care încrucișează categoriile a două variabile și dă frecvența fiecărei combinații. Tabelul 2 x 2 este cazul cel mai simplu. (Cap. 7)
- Tabel de frecvențe (engl. frequency table): tabel care arată, pentru fiecare valoare sau categorie, frecvența absolută, procentul, procentul valid și, pentru variabilele ordonate, procentul cumulat. Este cel mai simplu instrument de detectare a erorilor de codificare. (Cap. 2)
- Teorema limită centrală (engl. central limit theorem): media unor eșantioane suficient de mari are o distribuție aproximativ normală, indiferent de forma distribuției variabilei în populație. Ea justifică folosirea testelor bazate pe medie la eșantioane mari. (Cap. 4)
- Teorema lui Bayes (engl. Bayes' theorem): relația care permite calculul probabilității unei cauze (boala) cunoscând efectul observat (testul pozitiv): $P(B \mid T) = \frac{P(T|B) \times P(B)}{P(T)}$. În medicină, transformă probabilitatea pre-test în probabilitate post-test. (Cap. 11)
- Test bilateral (engl. two-sided test): test în care ipoteza alternativă admite diferențe în ambele direcții ($\mu 1 \neq \mu 2$). Este alegerea implicită în cercetarea medicală. (Cap. 5)
- Test de screening (engl. screening test): test aplicat unor persoane fără simptome pentru a identifica boala într-un stadiu precoce. Trebuie să fie foarte sensibil, ieftin și acceptabil; rezultatele pozitive se confirmă cu un test specific. (Cap. 12)
- Test diagnostic (engl. diagnostic test): orice procedură (analiză de laborator, imagistică, scor clinic, întrebare) folosită pentru a stabili dacă un pacient are sau nu o anumită boală. (Cap. 12)
- Test neparametric (engl. nonparametric test): test care nu presupune o anumită distribuție a datelor (de obicei cea normală) și lucrează adesea cu rangurile valorilor. Se folosește la variabile ordinale, la distribuții asimetrice cu eșantioane mici și în prezența valorilor aberante. (Cap. 7)

- Test parametric (engl. parametric test): test care presupune o anumită distribuție a datelor, de regulă normală, și compară parametri precum media (testele t, ANOVA, corelația Pearson). Când condițiile sunt îndeplinite, este mai puternic decât echivalentul neparametric. (Cap. 8)
- Test statistic (engl. statistical test): procedeu prin care se decide, cu un risc de eroare cunoscut, dacă datele sunt compatibile cu ipoteza nulă. Alegerea testului depinde de întrebare, de tipul variabilelor, de numărul grupurilor și de independența lor. (Cap. 5)
- Test unilateral (engl. one-sided test): test în care ipoteza alternativă admite o singură direcție ($\mu 1 > \mu 2$). Se justifică numai dacă o diferență în sens opus este imposibilă sau lipsită de interes și dacă direcția a fost stabilită înaintea analizei. (Cap. 5)
- Teste în paralel (engl. parallel testing): aplicarea simultană a două teste; pacientul este considerat pozitiv dacă oricare test este pozitiv. Crește sensibilitatea și scade specificitatea. (Cap. 12)
- Teste în serie (engl. serial testing): aplicarea celui de-al doilea test doar la cei pozitivi la primul; pacientul este considerat bolnav doar dacă ambele sunt pozitive. Crește specificitatea și scade sensibilitatea. (Cap. 12)
- Teste post-hoc (engl. post hoc tests): comparații două câte două efectuate după o ANOVA semnificativă, cu controlul erorii de tip I (Tukey, Bonferroni, Games-Howell). Ele arată care grupuri diferă între ele. (Cap. 6)
- Testul chi-pătrat de asociere (testul χ^2 de independență; engl. chi-square test of independence): test care verifică dacă două variabile categoriale sunt asociate, comparând frecvențele observate dintr-un tabel de contingență cu cele așteptate în absența asocierii. (Cap. 7)
- Testul chi-pătrat de concordanță (engl. chi-square goodness-of-fit test): test care compară distribuția observată a unei singure variabile categoriale cu o distribuție teoretică (de exemplu, proporții egale pe trei brațe). (Cap. 7)
- Testul Dunn (engl. Dunn test): test post-hoc pentru comparații două câte două după un test Kruskal-Wallis semnificativ, cu ajustarea valorilor p pentru comparații multiple. (Cap. 7)
- Testul exact Fisher (engl. Fisher exact test): test pentru tabele 2 x 2 (și mai mari) care calculează exact probabilitatea configurației observate și a celor mai extreme. Se folosește când frecvențele așteptate sunt mici. (Cap. 7)
- Testul Friedman (engl. Friedman test): alternativa neparametrică a ANOVA cu măsurători repetate, pentru trei sau mai multe măsurători pe aceiași indivizi. Rangurile se atribuie în interiorul fiecărui individ. (Cap. 7)
- Testul Kolmogorov-Smirnov (KS; engl. Kolmogorov-Smirnov test): test care compară distribuția observată cu o distribuție teoretică, folosit și pentru normalitate. Pentru verificarea normalității este mai puțin puternic decât testul Shapiro-Wilk. (Cap. 3)
- Testul Kruskal-Wallis (engl. Kruskal-Wallis test): alternativa neparametrică a ANOVA cu un factor, pentru trei sau mai multe grupuri independente. Un rezultat semnificativ se continuă cu comparații post-hoc (de exemplu testul Dunn). (Cap. 7)
- Testul Levene (engl. Levene test): test al egalității varianțelor în două sau mai multe grupuri, cu H0 "varianțele sunt egale". Un $p < 0,05$ orientează spre testul Welch sau spre ANOVA Welch. (Cap. 6)
- Testul Mann-Whitney U (testul Wilcoxon al sumei rangurilor; engl. Mann-Whitney U test): alternativa neparametrică a testului t pentru eșantioane independente; compară distribuțiile a două grupuri pe baza rangurilor. Rezultatul se raportează cu mediane și intervale intercuartilice. (Cap. 7)
- Testul McNemar (engl. McNemar test): test pentru două variabile dihotomice măsurate pe aceiași indivizi (de exemplu, rezultatul a două teste diagnostice la aceiași pacienți); folosește doar perechile discordante. (Cap. 7)

- Testul Shapiro-Wilk (engl. Shapiro-Wilk test): test de normalitate în care H0 afirmă că datele provin dintr-o distribuție normală. Un $p < 0,05$ indică abaterea de la normalitate; un $p \geq 0,05$ nu dovedește normalitatea, ci doar că abaterea nu a fost detectată. (Cap. 3)
- Testul t pentru eșantioane independente (engl. independent samples t-test): compară mediile a două grupuri formate din indivizi diferiți. Varianta clasică (Student) presupune variațe egale; varianta Welch nu. (Cap. 6)
- Testul t pentru un eșantion (engl. one-sample t-test): compară media unui eșantion cu o valoare de referință fixată (de exemplu o normă publicată), când σ este necunoscut. Are n - 1 grade de libertate. (Cap. 6)
- Testul t pereche (engl. paired t-test): compară mediile a două serii de măsurători legate (același pacient înainte și după tratament), analizând media diferențelor individuale. Condiția principală este normalitatea diferențelor. (Cap. 6)
- Testul t Welch (engl. Welch t-test): variantă a testului t pentru eșantioane independente care nu presupune variațe egale și ajustează gradele de libertate. Este o alegere sigură în majoritatea situațiilor. (Cap. 6)
- Testul Tukey (engl. Tukey HSD test): test post-hoc care compară toate perechile de medii, păstrând eroarea globală de tip I la α; dă diferențele cu IC ajustate. (Cap. 6)
- Testul Wilcoxon al rangurilor cu semn (engl. Wilcoxon signed-rank test): alternativa neparametrică a testului t pereche; analizează rangurile valorilor absolute ale diferențelor, ținând cont de semnul lor. (Cap. 7)
- Testul z (engl. z-test): test bazat pe distribuția normală, folosit pentru o medie când abaterea standard a populației este cunoscută sau eșantionul este mare, și pentru proporții. (Cap. 5)
- Transformare logaritmică (engl. log transformation): înlocuirea fiecărei valori cu logaritmul ei, pentru a reduce asimetria pozitivă și a apropia distribuția de cea normală. Rezultatele se readuc pe scala inițială prin exponențiere, obținându-se media geometrică. (Cap. 3)

U

- Unitate statistică (engl. statistical unit): elementul individual asupra căruia se fac observațiile (un pacient, un ochi, o probă de sânge). Stabilirea corectă a unității contează: doi ochi ai aceluiași pacient nu sunt observații independente. (Cap. 1)

V

- Validarea datelor la introducere (engl. data validation rules): reguli impuse în formularul electronic (domenii permise, câmpuri obligatorii, liste de opțiuni) care împiedică introducerea valorilor imposibile sau codificate neuniform. (Cap. 13)
- Valoare aberantă (engl. outlier): observație care se abate puternic de restul datelor. Poate fi o eroare de introducere (TAS 1540 mmHg) sau o valoare reală extremă; în primul caz se corectează după sursă, în al doilea se păstrează și se analizează influența ei. (Cap. 2)
- Valoare critică (engl. critical value): valoarea statisticii testului care separă regiunea de nerespingere de regiunea de respingere a lui H0, pentru un α dat. Se citește din tabelele statistice (anexa E). (Cap. 5)
- Valoare de referință a testului (μ_0; engl. reference value, null value): valoarea parametrului specificată de ipoteza nulă, de exemplu media de 120 mmHg într-un test t pe un eșantion, 1 pentru RR și OR sau 0 pentru o diferență. (Cap. 5)
- Valoare influentă (engl. influential point): observație a cărei eliminare schimbă mult coeficientul de corelație sau dreapta de regresie, de obicei un punct extrem pe axa x. (Cap. 9)
- Valoare lipsă (engl. missing value): observație care ar fi trebuit înregistrată, dar lipsește. Valorile lipsă se raportează pentru fiecare variabilă și nu se completează după bunul plac, ci doar printr-o metodă de imputare prevăzută în planul de analiză. (Cap. 2)

- Valoare prag (cut-off; engl. cut-off value): valoarea unui test numeric (scor, concentrație) peste care rezultatul se consideră pozitiv. Mutarea pragului crește sensibilitatea pe seama specificității sau invers. (Cap. 12)
- Valoare predictivă și prevalență (engl. predictive values and prevalence): relația prin care, la sensibilitate și specificitate constante, VPP crește și VPN scade când prevalența bolii crește. Explică de ce un test bun la spital dă multe rezultate fals pozitive în screeningul populației generale. (Cap. 12)
- Valoarea p (engl. p-value): probabilitatea de a obține un rezultat cel puțin la fel de extrem ca cel observat, dacă H0 ar fi adevărată. Nu este probabilitatea ca H0 să fie adevărată și nu măsoară mărimea sau importanța clinică a efectului. (Cap. 5)
- Valoarea predictivă negativă (VPN; engl. negative predictive value, NPV): proporția persoanelor fără boală dintre cele cu test negativ, d/(b + d). Crește când prevalența scade. (Cap. 12)
- Valoarea predictivă pozitivă (VPP; engl. positive predictive value, PPV): proporția bolnavilor dintre persoanele cu test pozitiv, a/(a + c). Depinde puternic de prevalență: același test are VPP mai mică într-o populație cu boală rară. (Cap. 12)
- Variabilă (engl. variable): caracteristică ce poate lua valori diferite de la un individ la altul (vârstă, sex, glicemie). Tipul variabilei determină indicatorii descriptivi, graficele și testele care se pot folosi. (Cap. 2)
- Variabilă aleatoare (engl. random variable): variabilă a cărei valoare depinde de rezultatul unui proces aleator, de exemplu valoarea glicemiei unui pacient extras la întâmplare. (Cap. 3)
- Variabilă calitativă (categorială) (engl. categorical variable): variabilă ale cărei valori sunt categorii, nu măsurători (sexul, grupa sanguină, statusul de fumător). Se descrie prin frecvențe și procente. (Cap. 2)
- Variabilă continuă (engl. continuous variable): variabilă numerică ce poate lua, teoretic, orice valoare dintr-un interval (glicemia, înălțimea); precizia ei este limitată doar de instrumentul de măsură. (Cap. 2)
- Variabilă dependentă (variabilă rezultat; engl. dependent variable, outcome): variabila al cărei comportament vrem să-l explicăm sau să-l prezicem (de exemplu TAS sau prezența diabetului). (Cap. 8)
- Variabilă derivată (engl. derived variable): variabilă calculată din alte variabile (IMC din greutate și înălțime, scăderea în greutate din două măsurători). Se recalculează, nu se introduce de mână, iar compararea cu valoarea înregistrată ajută la descoperirea erorilor. (Cap. 2)
- Variabilă dihotomică (binară) (engl. binary variable): variabilă nominală cu exact două categorii (bolnav sau sănătos, 0 sau 1). Este tipul de variabilă din tabelele 2 x 2 de risc și de diagnostic. (Cap. 2)
- Variabilă discretă (engl. discrete variable): variabilă numerică ce ia doar valori separate, de obicei numere întregi rezultate din numărare (numărul de consultații, numărul de sarcini). (Cap. 2)
- Variabilă independentă (predictor; engl. independent variable, predictor): variabila folosită pentru a explica sau prezice variabila rezultat (de exemplu IMC, vârsta, brațul de tratament). (Cap. 8)
- Variabilă indicator (variabilă dummy; engl. dummy variable): variabilă 0/1 creată pentru a introduce o categorie într-un model de regresie; o variabilă cu k categorii necesită k - 1 indicatori, față de o categorie de referință. (Cap. 9)
- Variabilă nominală (engl. nominal variable): variabilă categorială fără ordine naturală între categorii (mediul urban sau rural, brațul de tratament). Codurile numerice atribuite categoriilor sunt simple etichete. (Cap. 2)
- Variabilă numerică (cantitativă) (engl. numerical variable): variabilă exprimată printr-un număr cu unitate de măsură, pentru care operațiile aritmetice au sens (TAS în mmHg, greutatea în kg). (Cap. 2)

- Variabilă ordinală (engl. ordinal variable): variabilă categorială ale cărei categorii au o ordine naturală, dar distanțele dintre ele nu sunt neapărat egale (stadiul unei boli, nivelul de educație, scorul de satisfacție 1-5). Se descrie prin mediană și cuartile. (Cap. 2)
- Variabilitate biologică (engl. biological variability): faptul că aceeași caracteristică diferă între indivizi și, la același individ, de la un moment la altul. Este motivul pentru care medicina are nevoie de statistică. (Cap. 1)
- Variabilitate în interiorul grupurilor (engl. within-group variability): partea din variabilitatea totală datorată diferențelor dintre indivizii aceluiași grup (suma pătratelor reziduală). (Cap. 6)
- Variabilitate între grupuri (engl. between-group variability): partea din variabilitatea totală explicată de diferențele dintre mediile grupurilor (suma pătratelor intergrupală). (Cap. 6)
- Varianță (engl. variance): media pătratelor abaterilor față de medie, calculată pe eșantion cu numitorul n - 1. Are ca unitate pătratul unității variabilei, de aceea în raportare se folosește abaterea standard. (Cap. 2)

Anexa C. Setul de date didactic mg2026: scenariu, dicționar de date și erorile din fișierul brut

Aproape toate exemplele numerice din acest volum, de la media tensiunii sistolice din capitolul 2 până la curba ROC a scorului FINDRISC din capitolul 12, sunt calculate pe același set de date, numit mg2026. Am ales această soluție dintr-un motiv simplu: un student care recunoaște variabilele de la un capitol la altul se poate concentra pe metodă, fără să reînvețe de fiecare dată contextul clinic. Anexa de față adună într-un singur loc tot ce trebuie știut despre fișiere: scenariul, dicționarul variabilelor, erorile lăsate intenționat în varianta brută și modul în care se citează setul într-un referat sau într-o temă.

C.1 Scenariul

Setul descrie un studiu randomizat pragmatic, simulat, desfășurat în 12 cabinete de medicină de familie din vestul țării. Au fost evaluați pentru eligibilitate 212 adulți, dintre care 52 au fost excluși (31 nu îndeplineau criteriile, 21 au refuzat participarea), iar 160 au fost randomizați (diagrama de flux CONSORT este prezentată în capitolul 13). Criteriul principal de includere a fost riscul cardiometabolic crescut: suprapondere sau obezitate, prediabet sau suspiciune de diabet nediagnosticat.

Participanții au fost repartizați aleator în trei brațe:

- brațul A (n = 53): consiliere standard privind stilul de viață, făcută de medicul de familie;
- brațul B (n = 53): aceeași consiliere, plus o aplicație mobilă de automonitorizare (greutate, pași, alimentație);
- brațul C (n = 54): consiliere, aplicație și o teleconsultație lunară cu medicul.

Măsurătorile s-au făcut la includere (momentul 0), la 3 luni și la 6 luni. La includere, toți pacienții au efectuat un test oral de toleranță la glucoză (TTGO), considerat aici standardul de referință pentru diagnosticul de diabet, ceea ce permite evaluarea scorului FINDRISC ca test de screening (capitolul 12). Rezultatul principal al studiului este scăderea glicemiei a jeun la 6 luni; rezultatele secundare sunt scăderea în greutate (inclusiv proporția celor cu scădere de cel puțin 5%), tensiunea arterială sistolică la 6 luni și satisfacția față de intervenție.

Același set permite și analize observaționale, transversale, pe datele de la includere: asocierea dintre obezitate și diabet, corelația dintre IMC și tensiunea arterială, comparația dintre sexe. Aceasta este o alegere didactică deliberată. Într-un studiu real, analizele secundare ar trebui declarate în protocol înainte de colectarea datelor, iar rezultatele lor ar fi raportate ca explorative.

C.2 Fișierele

Există două fișiere, disponibile pe platforma Moodle a disciplinei [52]:

- **mg2026_curat.csv**: 160 de rânduri (un rând pentru fiecare pacient) și 28 de coloane, fără valori lipsă și fără erori de introducere. Este fișierul pe care se fac toate analizele din curs și din lucrările practice.
- **mg2026_brut.csv**: 161 de rânduri și aceleași 28 de coloane, cu 11 erori de introducere lăsate intenționat și un rând duplicat. Este materialul pentru lucrarea practică de curățare a datelor (capitolul 2 și primele lucrări practice).

Ambele fișiere folosesc virgula ca separator de câmpuri și punctul ca separator zecimal (convenția internațională CSV). La importul în Google Sheets cu setările regionale pentru România sau în

Excel în limba română, programul poate interpreta greșit zecimalele; în JASP și în MedCalc importul funcționează direct. Codificarea caracterelor este UTF-8, iar numele variabilelor sunt scrise fără diacritice, pentru compatibilitate cu toate programele.

C.3 Dicționarul de date

Tabelul C.1 descrie fiecare variabilă. Coloana "Domeniu valid" dă regula de validare pe care o recomandăm la verificarea fișierului brut (valori plauzibile clinic), urmată, între paranteze, de intervalul efectiv observat în fișierul curat. În fișierul curat nu există valori lipsă; coloana "Lipsă" se referă la fișierul brut.

Tabelul C.1. Dicționarul variabilelor din setul mg2026.

Variabila	Descriere	Tip	Scala	Unitate / codificare	Domeniu valid (observat)	Lipsă în brut
id	identificatorul pacientului	text	nominală (identificator)	P001-P160	unic pe fiecare rând (P001-P160)	0; un id apare de două ori
brat	brațul de randomizare	categorial	nominală	A, B, C (vezi C.1)	A, B, C (A 53, B 53, C 54)	0
sex	sexul	categorial	nominală dihotomică	F, M	F, M (F 86, M 74)	0
varsta	vârsta la includere	numeric discret	raport	ani împliniți	18-100 (35-79)	1
mediu	mediul de rezidență	categorial	nominală dihotomică	urban, rural	urban, rural (95 / 65)	0
educatie	nivelul de educație	categorial ordonat	ordinală	1 gimnaziu; 2 liceu sau școală profesională; 3 universitar; 4 postuniversitar	1-4 (1-4)	0
fumat	statusul de fumător	categorial	nominală (3 categorii)	nefumator, fost fumator, fumator	cele trei etichete (74 / 33 / 53)	0
activitate	activitatea fizică autoevaluată	categorial ordonat	ordinală	1 foarte redusă... 5 foarte intensă	1-5 (1-5)	0
inaltime	înălțimea	numeric continuu	raport	cm, întreg	130-210 (145-192)	0
greutate0	greutatea la includere	numeric continuu	raport	kg, o zecimală	35-250 (58,1-159,9)	0
imc0	indicele de masă corporală la includere	numeric continuu (derivat)	raport	kg/m², o zecimală; greutate0 / (inaltime/100)²	15-70 (23,5-48,3)	0
talie	circumferința taliei	numeric continuu	raport	cm, întreg	50-180 (65-140)	0
tas0	tensiunea arterială sistolică la includere	numeric	raport	mmHg	70-260 (102-169)	0

Variabila	Descriere	Tip	Scala	Unitate / codificare	Domeniu valid (observat)	Lipsă în brut
tad0	tensiunea arterială diastolică la includere	numeric	raport	mmHg	40-150 (58-103)	0
fc0	frecvența cardiacă la includere	numeric	raport	bătăi/min	35-180 (51-102)	0
colesterol	colesterolul total	numeric	raport	mg/dL	80-450 (119-306)	0
hdl	HDL-colesterolul	numeric	raport	mg/dL	10-120 (20-78)	0
trigliceride	trigliceridele serice	numeric	raport	mg/dL	20-1500 (44-442)	0
glicemie0	glicemia a jeun la includere	numeric	raport	mg/dL, întreg	40-400 (75-154)	0
glicemie3	glicemia a jeun la 3 luni	numeric	raport	mg/dL	40-400 (49-144)	0
glicemie6	glicemia a jeun la 6 luni	numeric	raport	mg/dL	40-400 (45-150)	1
hba1c0	hemoglobina glicată la includere	numeric continuu	raport	%, o zecimală	3,5-15 (4,5-7,0)	0
findrisc	scorul FINDRISC de risc de diabet	numeric discret (scor)	ordinală tratată ca numerică	puncte	0-26 (0-25)	0
diabet_ogtt	diabet confirmat prin TTGO la includere	categorial	nominală dihotomică	0 nu, 1 da	0, 1 (29 cu diabet)	0
greutate6	greutatea la 6 luni	numeric continuu	raport	kg, o zecimală	35-250 (50,7-155,6)	0
tas6	tensiunea arterială sistolică la 6 luni	numeric	raport	mmHg	70-260 (85-166)	0
scadere5	scădere în greutate de cel puțin 5% la 6 luni	categorial (derivat)	nominală dihotomică	0 nu, 1 da; 1 dacă (greutate0 - greutate6)/greutate0 ≥ 0,05	0, 1 (38 cu 1)	0
satisfactie	satisfacția față de intervenție la 6 luni	categorial ordonat	ordinală (Likert)	1 deloc mulțumit... 5 foarte mulțumit	1-5 (1-5)	0

Câteva observații despre scale, pentru că exact aici apar greșelile la examen. Variabilele educatie, activitate și satisfactie sunt codificate cu cifre, dar sunt ordinale: media lor are un sens limitat, iar pentru comparații se folosesc mediana, intervalul intercuartilic și testele neparametrice (capitolul 7). Scorul FINDRISC este, strict vorbind, tot ordinal, dar cu 27 de valori posibile este tratat în practică drept numeric; îl descriem cu media și cu mediana. Tensiunea, glicemia și greutatea sunt variabile de raport (au un zero absolut), deși pentru statistica din acest curs distincția dintre scala de interval și cea de raport contează puțin. În sfârșit, imc0 și scadere5 sunt variabile derivate, calculate din alte coloane; ele nu se introduc de mână, ci se recalculează, iar

diferența dintre valoarea înregistrată și cea recalculată este un instrument excelent de control al erorilor (vezi C.4).

C.4 Erorile din fișierul brut și corectarea lor

Fișierul mg2026_brut.csv imită ce se întâmplă în orice cabinet sau secție unde datele sunt copiate de mână din fișe într-o foaie de calcul. Erorile sunt de patru feluri: codificare neuniformă a categoriilor, valori imposibile (de obicei o cifră în plus, o virgulă mutată sau o unitate de măsură greșită), text introdus într-o coloană numerică și valori lipsă. La acestea se adaugă un rând copiat de două ori. Fiecare eroare este o deformare a valorii adevărate, așa că, odată detectată, poate fi corectată; totuși, corectura se confirmă întotdeauna în documentul-sursă (fișa pacientului, formularul de raportare a cazului, buletinul de laborator), al cărui rol îl joacă aici fișierul curat.

Tabelul C.2. Erorile din mg2026_brut.csv, modul de detectare și valoarea corectă.

Nr.	Pacient (rândul de date)	Variabila	Valoarea din fișierul brut	Cum se detectează	Natura erorii	Valoarea corectă (fișierul curat)
1	P005 (5)	sex	f	tabel de frecvențe pentru sex: apare o a treia categorie	literă mică în loc de majusculă	F
2	P021 (21)	sex	Feminin	același tabel de frecvențe: a patra categorie	cuvânt întreg în loc de cod	F
3	P034 (34)	tas0	1540	minim/maxim, box-plot; valoarea depășește regula 70-260	un zero tastat în plus	154
4	P042 (42)	varsta	celulă goală	numărul de valori valide (160 în loc de 161); COUNTBLANK	valoare lipsă	70 (recuperată din fișă)
5	P057 (57)	inaltime	1.72	minim < 130; valoare în metri, nu în centimetri	unitate de măsură greșită	172
6	P064 (64)	glicemie0	$\frac{87\text{mg}}{\text{dl}}$	coloana este importată ca text; ISNUMBER dă FALSE	unitate scrisă în celulă	87
7	P084 (84)	fumat	"Fumator " (majusculă și spațiu final)	tabel de frecvențe: o a patra categorie; LEN arată un caracter în plus	majusculă și spațiu invizibil	fumator
8	P089 (89)	hba1c0	5,2	text cu virgulă într-o coloană cu punct	separator zecimal greșit	5,2 (scris 5.2 în CSV)

Nr.	Pacient (rândul de date)	Variabila	Valoarea din fișierul brut	Cum se detectează	Natura erorii	Valoarea corectă (fișierul curat)
				zecimal; coloana devine text		
9	P102 (102)	greutate0	7.26	minim < 35 kg; IMC recalculat ar fi 2,7 kg/m²	virgulă zecimală mutată	72,6
10	P121 (121)	mediu	Urban	tabel de frecvențe: "Urban" separat de "urban"	majusculă	urban
11	P134 (134)	glicemie6	celulă goală	numărul de valori valide pentru glicemie6	valoare lipsă	135 (recuperată din fișă)
12	P091 (91 și 161)	toate	rând identic repetat la sfârșit	COUNTIF pe id > 1; în pandas, duplicated()	rând duplicat	un singur rând P091

Câteva comentarii la tabel, în ordinea în care le-ar descoperi un student atent.

Primul pas este întotdeauna tabelul de frecvențe pentru fiecare variabilă categorială. În fișierul brut, sexul are patru categorii (F 84, M 75, f 1, Feminin 1), fumatul are patru (nefumator 75, fumator 52, fost fumator 33, "Fumator " 1), iar mediul de rezidență trei (urban 95, rural 65, Urban 1). Totalurile depășesc cu 1 valorile din fișierul curat din cauza rândului duplicat. Programele tratează "f" și "F" ca pe două valori diferite, la fel "Urban" și "urban", iar un spațiu invizibil la sfârșitul cuvântului este suficient ca "Fumator " să fie numărat separat. Uniformizarea etichetelor se face automat (litere mici sau majuscule după dicționar, eliminarea spațiilor cu funcția TRIM), apoi rezultatul se compară cu fișa.

Al doilea pas este statistica descriptivă cu minimul și maximul fiecărei variabile numerice. Trei erori sar imediat în ochi: TAS 1540 mmHg, înălțimea 1,72 și greutatea 7,26 kg. Valoarea de 1540 mmHg nu este o curiozitate inofensivă. Dacă o lăsăm în fișier, media TAS calculată pe cele 161 de rânduri devine 139,70 mmHg în loc de 131,13, iar abaterea standard urcă de la 13,88 la 111,89 mmHg; mediana rămâne 131, ceea ce ilustrează perfect robustețea medianei față de valorile aberante (capitolul 2). La înălțime și greutate, cea mai elegantă verificare folosește variabila derivată imc0, care în fișierul brut a rămas corectă. Pentru P057, IMC înregistrat este 26,3 kg/m², iar greutatea 77,7 kg; înălțimea compatibilă cu aceste două valori este $\sqrt{\frac{77,7}{26,3}} = 1,72m$, ceea ce confirmă că s-a scris înălțimea în metri și că valoarea corectă este 172 cm. Pentru P102, IMC 27,0 și înălțimea 164 cm dau o greutate de 27,0 x 1,64² = 72,6 kg: virgula zecimală a fost mutată cu o poziție. Verificarea încrucișată între variabile transformă o bănuială într-o certitudine și este mai sigură decât orice ghicire.

Al treilea pas privește tipul coloanelor. Când într-o coloană numerică apare un singur text ("87 mg/dl" sau "5,2"), programul de import declară întreaga coloană drept text, iar media nu mai poate fi calculată deloc. Simptomul este ușor de recunoscut: JASP afișează variabila cu pictograma de variabilă nominală, iar în Google Sheets funcția ISNUMBER întoarce FALSE. Corectura (eliminarea unității, respectiv înlocuirea virgulei cu punctul) se face în celula respectivă, după care coloana se reimportă ca numerică.

Valorile lipsă (vârsta pentru P042 și glicemia la 6 luni pentru P134) nu se completează cu media sau cu o valoare "probabilă" fără un plan de analiză care să prevadă explicit metoda de imputare. În practica unui studiu, se caută întâi valoarea în fișă; dacă nu există, se raportează câte valori lipsesc pentru fiecare variabilă, iar analiza se face pe cazurile disponibile sau, în studiile clinice, după strategia prevăzută în protocol (capitolul 13). În scenariul nostru, sursa conține ambele valori, așa că ele au fost recuperate.

Rândul duplicat (P091, care apare pe rândul 91 și din nou pe rândul 161) umflă volumul eșantionului la 161 și dă o pondere dublă unui singur pacient. Se detectează verificând unicitatea identificatorului: în Google Sheets, =COUNTIF(A:A; A2) > 1; în JASP, sortând după id; în Python, df.duplicated(). Una dintre copii se șterge, după ce se confirmă că cele două rânduri sunt identice în toate coloanele.

> **Regula care previne greșeala**
>
> Corectați datele numai după documentul-sursă, nu după intuiție, chiar și atunci când corectura pare evidentă. Păstrați fișierul brut neatins, lucrați pe o copie și notați fiecare modificare într-un jurnal al curățării (rândul, variabila, valoarea veche, valoarea nouă, motivul). Un referat care spune "am corectat 11 erori și am eliminat un duplicat, conform jurnalului din anexă" este verificabil; unul care spune doar "datele au fost curățate" nu este.

C.5 Cum a fost generat setul

Datele sunt în întregime simulate. Pacienții sunt fictivi, nu există nicio legătură cu persoane reale, iar setul nu provine din niciun cabinet sau spital. Prin urmare, nu se pune problema protecției datelor personale în sensul Regulamentului general privind protecția datelor [32], iar fișierele pot fi folosite liber în scop didactic.

Simularea a fost făcută în Python, cu generatorul de numere aleatoare din biblioteca numpy (default_rng) inițializat cu o sămânță fixă (2026). Sămânța fixă face ca setul să fie reproductibil: rulând din nou scriptul se obțin exact aceleași 160 de rânduri. Variabilele nu au fost generate independent, ci prin relații plauzibile clinic: tensiunea sistolică crește cu vârsta, cu IMC și la fumători; HbA1c depinde de glicemie; diabetul confirmat prin TTGO și scorul FINDRISC depind de o aceeași variabilă latentă de risc (vârstă, IMC, talie, activitate fizică); efectul intervenției asupra glicemiei și greutății crește de la brațul A la brațul C. Distribuția trigliceridelor a fost generată lognormal, pentru a avea un exemplu de variabilă asimetrică care se normalizează prin logaritmare (capitolul 3). Mărimile efectelor au fost alese astfel încât setul să conțină, pentru scopuri didactice, atât rezultate clar semnificative (ANOVA pe brațe, corelația IMC-TAS), cât și rezultate nesemnificative (TAS pe sexe) sau aflate la limită (testul Wilcoxon pentru greutate în brațul A, $p = 0{,}052$).

Consecința practică este importantă și trebuie spusă cinstit: cifrele din mg2026 nu reprezintă o estimare a niciunei realități epidemiologice. Prevalența de 18,1% a diabetului, efectul aplicației mobile sau aria de sub curba ROC a scorului FINDRISC (0,852) sunt proprietăți ale simulării, nu rezultate care pot fi citate ca dovezi clinice. Pentru dovezi reale, capitolele volumului folosesc studiile publicate ale colegilor din disciplină și literatura de specialitate.

C.6 Cum se citează setul în referate și teme

Când folosiți mg2026 într-un referat, într-o temă sau la examenul practic, precizați trei lucruri: că datele sunt simulate, ce fișier ați folosit (curat sau brut) și ce prelucrări ați făcut înainte de analiză. Un model de frază pentru secțiunea de metode:

"Analiza a folosit setul de date didactic mg2026 (date simulate, 160 de pacienți fictivi, fișierul mg2026_curat.csv), pus la dispoziție pe platforma Moodle a disciplinei de Informatică

medicală și biostatistică, UMF "Victor Babeș" Timișoara. Prelucrarea statistică s-a făcut în JASP 0.98.1, cu pragul de semnificație $\alpha = 0,05$."

Model de intrare în lista bibliografică:

Muntean C. Setul de date didactic mg2026 (date simulate pentru disciplina Biostatistică, Medicină, anul I). În: Muntean C. Biostatistică aplicată în medicina generală. Orizont 2026. Timișoara: Editura "Victor Babeș"; 2026. Fișierele mg2026_curat.csv și mg2026_brut.csv, disponibile pe moodle.umft.ro.

Programele se citează cu versiunea folosită: JASP 0.98.1 [47], MedCalc 24.0 [50], Epi Info 7.2.7 [51]. Dacă ați lucrat pe fișierul brut, adăugați în anexa referatului jurnalul curățării (tabelul C.2 este un model) și raportați numărul final de rânduri analizate. Nu prezentați niciodată rezultatele pe mg2026 ca și cum ar proveni de la pacienți reali, nici măcar într-un exercițiu de redactare a unui rezumat de congres: obiceiul de a declara corect proveniența datelor se formează din anul I.

De reținut

- mg2026 este un set simulat, reproductibil (sămânța 2026), cu 160 de pacienți fictivi în trei brațe randomizate (A 53, B 53, C 54) și măsurători la 0, 3 și 6 luni.
- Fișierul curat (160 de rânduri) servește analizelor; fișierul brut (161 de rânduri) conține 11 erori și un duplicat, pentru exercițiul de curățare.
- Ordinea verificărilor: frecvențe pentru categorii, minim și maxim pentru numerice, tipul coloanelor, valori lipsă, unicitatea identificatorului, coerența variabilelor derivate (IMC).
- Fiecare eroare este o deformare a valorii adevărate, dar corectura se confirmă în documentul-sursă; coerența cu variabilele derivate (IMC) este cea mai bună verificare.
- În orice lucrare, datele se declară ca simulate, iar prelucrările se documentează.

Anexa D. Fișa de formule

Fișa adună formulele folosite în curs, ordonate după capitolul în care apar. Pentru fiecare formulă găsiți semnificația simbolurilor, condițiile în care se poate folosi și un exemplu numeric de un rând, de cele mai multe ori pe setul didactic mg2026 (anexa C). Formulele sunt scrise în text simplu: x înseamnă înmulțire, √ rădăcina pătrată, Σ suma tuturor termenilor, ln logaritmul natural, iar e^x funcția exponențială. Rezultatele intermediare sunt rotunjite doar la final, ca să nu acumulați erori de rotunjire; la examen păstrați cel puțin trei zecimale în calculele intermediare.

Notațiile generale, valabile pentru toată fișa: n este volumul eșantionului, N volumul total (sau al populației, după context), xi o valoare individuală, m media eșantionului, μ media populației, s abaterea standard a eșantionului, σ abaterea standard a populației, p o proporție observată, α pragul de semnificație, β probabilitatea erorii de tip II.

Pentru tabelele 2 x 2 se folosesc două convenții, una pentru analiza riscului și una pentru testele diagnostice. Le repetăm aici, pentru că jumătate din greșelile de la examen vin din confuzia lor.

Tabelul D.1. Tabelul 2 x 2 în analiza riscului (capitolul 10).

	Boală prezentă	**Boală absentă**	**Total**
Expuși	a	b	a + b
Neexpuși	c	d	c + d
Total	a + c	b + d	N

Tabelul D.2. Tabelul 2 x 2 pentru un test diagnostic (capitolul 12), cu boala pe rânduri, ca în lucrările practice ale disciplinei.

	Test pozitiv	**Test negativ**	**Total**
Bolnavi (B+)	a = adevărat pozitivi (AP)	b = fals negativi (FN)	a + b
Nebolnavi (B-)	c = fals pozitivi (FP)	d = adevărat negativi (AN)	c + d
Total	a + c	b + d	N

D.1 Capitolul 1. Probabilități

Probabilitatea clasică. P(A) = numărul cazurilor favorabile / numărul cazurilor posibile. Valoarea este între 0 și 1. Condiție: toate cazurile sunt egal probabile (de exemplu, extragerea aleatoare a unei fișe). Exemplu: probabilitatea ca un pacient extras la întâmplare din mg2026 să fie fumător este 53/160 = 0,331.

Probabilitatea evenimentului contrar. $P(\text{non} - A) = 1 - P(A)$. Exemplu: $P(\text{ nefumător saufost fumător }) = 1 - 0{,}331 = 0{,}669$.

Regula adunării. $P(A\,\text{sau}B) = P(A) + P(B) - P(A\ \text{și}\ B)$. Dacă A și B se exclud reciproc, P(A și B) = 0 și termenul dispare. Exemplu: $P(\text{ femeie sau fumător }) = \frac{86}{160} + \frac{53}{160} - \frac{30}{160} = \frac{109}{160} = 0{,}681$.

Probabilitatea condiționată. $P(A \mid B) = \frac{P(A\ \text{și}\ B)}{P(B)}$, citită "probabilitatea lui A știind că B s-a produs". Exemplu: $P(\text{ diabet } \mid \text{obez}) = \frac{\frac{18}{160}}{\frac{54}{160}} = \frac{18}{54} = 0{,}333$.

Regula înmulțirii. $P(A\ \text{și}\ B) = P(B) \times P(A \mid B)$. Dacă A și B sunt independente, P(A | B) = P(A) și $P(A\ \text{și}\ B) = P(A) \times P(B)$. Exemplu: sub ipoteza de independență, P(femeie și

fumător) ar fi 0,5375 x 0,331 = 0,178; observat este 30/160 = 0,1875, foarte aproape, deci sexul și fumatul par independente în acest set.

Frecvența relativă ca estimare a probabilității. $f = \frac{k}{n}$, unde k este numărul de apariții ale evenimentului în n observații. Exemplu: 29 de pacienți cu diabet confirmat din 160, f = 0,181.

D.2 Capitolul 2. Descrierea datelor

Frecvența absolută, relativă și cumulată. $f_i = \frac{n_i}{n}$; frecvența procentuală = 100 x ni / n; frecvența cumulată la categoria k = suma frecvențelor categoriilor 1...k (doar pentru variabile ordinale sau numerice grupate). Când există valori lipsă, frecvența "validă" se calculează la numărul de răspunsuri valide, nu la total. Exemplu: satisfacție 4 sau 5: (55 + 28)/160 = 51,9%; frecvența cumulată până la scorul 3 inclusiv: (4 + 19 + 54)/160 = 48,1%.

Media aritmetică. $m = \frac{\sum x_i}{n}$. Condiții: variabilă numerică; sensibilă la valori extreme, deci se preferă la distribuții aproximativ simetrice. Exemplu: TAS la cinci pacienți 120, 135, 128, 142, 130 mmHg, $m = \frac{655}{5} = 131$ mmHg.

Media ponderată. mp $= \frac{\sum(w_i \times x_i)}{\sum w_i}$, unde wi este ponderea (de obicei volumul subgrupului). Se folosește când combinăm medii ale unor grupuri de mărimi diferite. Exemplu: scăderea medie a glicemiei pe cele trei brațe, $\frac{53 \times 3{,}0 + 53 \times 11{,}1 + 54 \times 16{,}3}{160} = 10{,}17$ mg/dL (media calculată direct pe toți pacienții este 10,19; diferența vine din rotunjirea mediilor de grup).

Media geometrică. mg $= (x_1 \times x_2 \times ... \times \text{xn})^{\frac{1}{n}} = e$^(media lui ln x). Condiții: toate valorile strict pozitive; potrivită pentru variabile asimetrice la dreapta, lognormale (trigliceride, titruri de anticorpi, încărcături virale). Exemplu: pentru 100, 150 și 225, mg $= (100 \times 150 \times 225)^{\frac{1}{3}} = 150$; pentru trigliceridele din mg2026, media logaritmilor este 4,970 și mg $= e^{4{,}970} = 144{,}0$ mg/dL (media aritmetică este 161,18).

Poziția medianei. Se ordonează valorile crescător; mediana se află pe poziția (n + 1)/2. Dacă n este par, mediana este media celor două valori centrale (pozițiile n/2 și n/2 + 1). Exemplu: pentru n = 160, poziția este 80,5, deci mediana TAS este media valorilor de pe locurile 80 și 81: (131 + 131)/2 = 131 mmHg.

Pozițiile cuartilelor (metoda simplă). Q1 pe poziția (n + 1)/4, Q3 pe poziția 3(n + 1)/4, cu interpolare între valorile vecine. Programele folosesc metode de interpolare ușor diferite, deci pot da rezultate care diferă cu o zecimală. Exemplu: în mg2026, Q1 al TAS = 122 și Q3 = 140 mmHg.

Amplitudinea (domeniul de variație). A = xmax - xmin. Exemplu: TAS 169 - 102 = 67 mmHg.

Intervalul intercuartilic. IQR = Q3 - Q1; conține 50% centrale din valori. Exemplu: TAS 140 - 122 = 18 mmHg; IMC $31{,}4 - 26{,}2 = \frac{5{,}2\ \text{kg}}{m^2}$.

Limitele pentru valori aberante (regula lui Tukey, folosită la box-plot). Limita inferioară = Q1 - 1,5 x IQR; limita superioară = Q3 + 1,5 x IQR. Exemplu: IMC, $31{,}4 + 1{,}5 \times 5{,}2 = \frac{39{,}2\ \text{kg}}{m^2}$; valoarea maximă 48,3 apare ca punct separat pe box-plot.

Abaterea medie absolută. AMA $= \frac{\sum|x_i - m|}{n}$. Exemplu: pentru 120, 135, 128, 142, 130 (m = 131), abaterile absolute sunt 11, 4, 3, 11, 1, iar AMA $= \frac{30}{5} = 6$ mmHg.

Varianța eșantionului. $s^2 = \frac{\sum(x_i - m)^2}{n-1}$. Numitorul n - 1 (gradele de libertate) corectează tendința de subestimare a varianței populației. Exemplu: pentru aceleași cinci valori, $\sum(x_i - m)^2 = 121 + 16 + 9 + 121 + 1 = 268$, $s^2 = \frac{268}{4} = 67$ mmHg^2.

Abaterea standard. $s = \sqrt{s^2}$. Are aceeași unitate de măsură ca datele. Exemplu: $s = \sqrt{67} =$ 8,19 mmHg; în mg2026, s(TAS) = 13,88 mmHg.

Coeficientul de variație. CV = 100 x s / m (%). Condiții: variabilă de raport, cu medie pozitivă; permite compararea variabilității unor variabile cu unități diferite. Orientativ, CV sub 10% indică o variabilitate mică, 10-30% moderată, peste 30% mare. Exemplu: TAS, $\frac{100 \times 13,88}{131,13} =$ 10,6%; trigliceride, 47,3%.

Eroarea standard a mediei. ES $= \frac{s}{\sqrt{n}}$. Măsoară precizia mediei ca estimare a lui μ, nu împrăștierea pacienților. Exemplu: TAS, 13,88/√160 = 13,88/12,649 = 1,097 mmHg.

D.3 Capitolul 3. Distribuția normală, scorul z, transformări

Scorul z (standardizarea). $z = \frac{x-m}{s}$ (pe eșantion) sau $z = \frac{x-\mu}{\sigma}$ (pe populație). Arată la câte abateri standard se află o valoare față de medie. Condiție pentru citirea ariilor din tabelul normal: distribuție aproximativ normală. Exemplu: TAS 160 mmHg, $z = \frac{160-131,13}{13,88} = 2,08$; din tabelul E.1, Φ(2,08) = 0,9812, deci aproximativ 1,9% din valori ar depăși 160 mmHg.

Transformarea inversă. x = m + z x s. Exemplu: percentila 97,5 a TAS (z = 1,96) este 131,13 + 1,96 x 13,88 = 158,3 mmHg.

Intervale pentru valori individuale (regula 68-95-99,7). Pentru o distribuție normală, aproximativ 68,3% din valori se află în $m \pm s$, 95,4% în $m \pm 2s$, 99,7% în $m \pm 3s$, iar exact 95% în $m \pm 1,96s$. Exemplu (TAS): $m \pm s = 117,3 - 145,0$; $m \pm 1,96s = 103,9 - 158,3$ (conține 93,8% din observațiile reale); $m \pm 2s = 103,4 - 158,9$; $m \pm 3s = 89,5 - 172,8$ mmHg.

Intervale pentru media populației (pe baza erorii standard). μ se află, cu probabilitățile corespunzătoare, în $m \pm \text{ES}(68,3\%)$, $m \pm 2\text{ES}(95,4\%)$, $m \pm 3\text{ES}(99,7\%)$ sau $m \pm 1,96\text{ES}(95\%)$. Condiție: n suficient de mare (teorema limită centrală) sau date normale. Exemplu (TAS, ES = 1,097): $m \pm \text{ES} = 130,03 - 132,23$; $m \pm 2\text{ES} = 128,94 - 133,32$; $m \pm 3\text{ES} = 127,84 - 134,42$ mmHg.

Transformarea logaritmică. y = ln(x); după analiza pe scala logaritmică, rezultatele se readuc pe scala inițială cu $x = e^y$. Condiție: x > 0; utilă la asimetrie pozitivă. Exemplu: asimetria trigliceridelor scade de la 0,95 la -0,28 după logaritmare, iar IC 95% al mediei geometrice este $e^{4,970 \pm \frac{1,975 \times 0,488}{\sqrt{160}}} = 133,44 - 155,42$ mg/dL.

Asimetria și excesul (interpretare). Asimetria (skewness) apropiată de 0 indică simetrie; pozitivă, coadă spre dreapta. Excesul (kurtosis minus 3) apropiat de 0 indică o formă apropiată de cea normală. Regulă orientativă: valori absolute peste 1 sugerează o abatere importantă. Exemplu: IMC, asimetrie 1,42 și exces 3,68, deci distribuție nenormală, confirmată de Shapiro-Wilk (p < 0,001).

D.4 Capitolul 4. Estimare, intervale de încredere, dimensiunea eșantionului

IC pentru medie cu distribuția normală (σ cunoscut sau n mare). $\text{IC}(1-\alpha) = m \pm \frac{z\left(1-\frac{\alpha}{2}\right)\times\sigma}{\sqrt{n}}$; z = 1,645 (90%), 1,960 (95%), 2,576 (99%). Exemplu: TAS, $131{,}13 \pm 1{,}96 \times 1{,}097 = 128{,}98 - 133{,}28$ mmHg.

IC pentru medie cu distribuția Student (σ necunoscut, caz obișnuit). $\text{IC}(1-\alpha) = m \pm \frac{t(\alpha;\, n-1)\times s}{\sqrt{n}}$. Condiție: date aproximativ normale sau n mare. Exemplu: TAS, t(0,05; 159) = 1,975, $\text{IC}\,95\% = 131{,}13 \pm 1{,}975 \times 1{,}097 = 128{,}96 - 133{,}30$ mmHg; IC 99% = 128,27-133,99 mmHg.

Lățimea IC și efectul lui n. Lățimea = $\frac{2\times t\times s}{\sqrt{n}}$; dacă n crește de 4 ori, lățimea scade aproximativ la jumătate. Exemplu: IC al mediei TAS are lățimea 27,2 mmHg la n = 10, 10,6 la n = 40 și 4,3 la n = 160.

Eroarea standard a unei proporții. $\text{ES}(p_i) = \sqrt{\frac{p_i\times(1-p_i)}{n}}$. Exemplu: diabet 29/160, p_i = 0,181, $\text{ES} = \sqrt{\frac{0{,}181\times0{,}819}{160}} = 0{,}0305$.

IC pentru proporție, metoda Wald. $\text{IC}\,95\% = p_i \pm 1{,}96 \times \sqrt{\frac{p_i(1-p_i)}{n}}$. Condiții: n x $p_i \geq 5$ și n x (1 - p_i) ≥ 5 (unii autori cer 10); la proporții mici sau eșantioane mici dă intervale prea înguste sau limite negative. Exemplu: $0{,}181 \pm 1{,}96 \times 0{,}0305 = 0{,}122 - 0{,}241$, adică 12,2-24,1%.

IC pentru proporție, metoda Wilson (recomandată). Centrul = (p_i + z^2/(2n)) / (1 + z^2/n); semilățimea = $\left[\frac{z}{1+\frac{z^2}{n}}\right] \times \sqrt{\frac{p_i(1-p_i)}{n} + \frac{z^2}{4\,n^2}}$; IC = centrul ± semilățimea. Funcționează bine și la proporții apropiate de 0 sau 1. Exemplu: pentru 29/160 și z = 1,96, centrul = 0,1887, semilățimea = 0,0595, IC 95% = 12,9-24,8%.

IC al mediei geometrice. $\text{IC} = e^{\text{ml}\pm\frac{t\times\text{sl}}{\sqrt{n}}}$, unde ml și sl sunt media și abaterea standard ale lui ln x. Exemplu: trigliceride, 133,44-155,42 mg/dL (vezi D.3).

Volumul eșantionului pentru estimarea unei medii. $n = \left(\frac{z\left(1-\frac{\alpha}{2}\right)\times\sigma}{E}\right)^2$, unde E este marja de eroare dorită (semilățimea IC). Se rotunjește întotdeauna în sus. Exemplu: σ = 14 mmHg, $E = \pm 3$ mmHg, $n = \left(\frac{1{,}96\times14}{3}\right)^2 = 83{,}7$, deci 84 de pacienți.

Volumul eșantionului pentru estimarea unei proporții. $n = \frac{z^2\times p_i(1-p_i)}{E^2}$. Dacă p_i este necunoscut, se ia p_i = 0,5 (varianta cea mai acoperitoare). Exemplu: p_i = 0,20, E = 0,05, $n = \frac{1{,}96^2\times0{,}2\times0{,}8}{0{,}0025} = 245{,}9$, deci 246.

Volumul pe grup pentru compararea a două medii (testul t independent). n pe grup = 2 x (z(1-α/2) + z(1-β))² / d², unde $d = \frac{\mu_1-\mu_2}{\sigma}$ este mărimea efectului; z(1-β) = 0,842 pentru putere 80% și 1,282 pentru putere 90%. Exemplu: d = 0,5, α = 0,05, putere 80%: $\frac{2\times(1{,}96+0{,}842)^2}{0{,}25} = 62{,}8$, adică 63 pe grup; programele care folosesc distribuția t (G*Power) dau 64 pe grup, iar pentru d = 0,3 dau 176.

Volumul pe grup pentru compararea a două proporții. n pe grup = (z(1-α/2) + z(1-β))² x [p1(1 - p1) + p2(1 - p2)] / (p1 - p2)². Exemplu: p1 = 0,30, p2 = 0,11, α = 0,05, putere 80%: $\frac{7,849\times(0,21+0,0979)}{0,0361} = 66,9$, deci 67 pe grup.

Corecția pentru abandon. n ajustat $= \frac{n}{1-r}$, unde r este proporția estimată de pierderi din urmărire. Exemplu: 64 pe grup și 15% abandon, 64/0,85 = 75,3, deci 76 pe grup.

D.5 Capitolul 5. Bazele testării ipotezelor

Erorile și puterea. α = P(respingem H0 | H0 adevărată), eroarea de tip I; β = P(nu respingem H0 | H0 falsă), eroarea de tip II; puterea = 1 - β; nivelul de încredere = 1 - α. Exemplu: un studiu planificat cu α = 0,05 și putere 80% acceptă un risc de 5% de rezultat fals pozitiv și de 20% de a rata un efect real de mărimea presupusă.

Regula de decizie cu p. Dacă p < α, H0 se respinge (diferență statistic semnificativă); dacă p ≥ α, H0 nu se respinge (nu s-a demonstrat o diferență). Exemplu: TAS femei vs bărbați, p = 0,81 > 0,05, deci H0 nu se respinge.

Regula de decizie cu valoarea critică. H0 se respinge dacă |statistica calculată| ≥ valoarea critică din tabel (anexa E) pentru α și gradele de libertate respective. Exemplu: t = 11,14 cu gl = 159 > t critic 1,975, deci p < 0,05.

Testul z pentru o medie (σ cunoscut). $z = \frac{m-\mu_0}{\frac{\sigma}{\sqrt{n}}}$. Condiții: σ cunoscut din populație sau n mare, date aproximativ normale. Exemplu: media TAS 131,13 față de o valoare de referință μ0 = 128 mmHg, cu σ = 14: $z = \frac{3,13}{\frac{14}{\sqrt{160}}} = 2,83$, p bilateral = 0,005.

Testul t pentru un eșantion. $t = \frac{m-\mu_0}{\frac{s}{\sqrt{n}}}$, cu gl = n - 1. Condiții: date aproximativ normale (sau n mare), observații independente. Exemplu: aceeași comparație cu s = 13,88: $t = \frac{3,13}{1,097} = 2,85$, gl = 159, p = 0,005.

Testul z pentru o proporție. $z = \frac{p-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$. Condiții: n x p0 ≥ 5 și n x (1 - p0) ≥ 5. Exemplu: diabet 0,181 față de o prevalență presupusă p0 = 0,12: $z = \frac{0,061}{\sqrt{\frac{0,12\times0,88}{160}}} = \frac{0,061}{0,0257} = 2,38$, p bilateral = 0,017.

Corecția Bonferroni pentru comparații multiple. α' = α / k, unde k este numărul de comparații. Exemplu: trei comparații două câte două între brațele A, B, C, α' = 0,05/3 = 0,0167.

D.6 Capitolul 6. Teste parametrice: t și ANOVA

Testul t pereche. d = diferența pentru fiecare pacient; md = media diferențelor, sd = abaterea lor standard; $t = \frac{\text{md}}{\frac{\text{sd}}{\sqrt{n}}}$, gl = n - 1. Condiții: măsurători pereche (același pacient, două momente sau două metode), diferențe aproximativ normale. Exemplu: glicemia la 0 față de 6 luni, md = 10,19, sd = 11,57, n = 160: $t = \frac{10,19}{\frac{11,57}{12,649}} = 11,14$, gl = 159, p < 0,001.

IC pentru media diferențelor pereche. $\text{IC}\,95\% = \text{md} \pm \frac{t(0,05;\, n-1)\times\text{sd}}{\sqrt{n}}$. Exemplu: $10,19 \pm 1,975 \times 0,915 = 8,38 - 11,99$ mg/dL.

Varianța comună (pooled) pentru două grupuri independente. $\text{sp}^2 = \frac{(n_1-1)\times {s_1}^2+(n_2-1)\times {s_2}^2}{n_1+n_2-2}$. Exemplu: TAS, bărbați (130,85; 13,25; 74) și femei (131,37; 14,47; 86): $\text{sp}^2 = \frac{73\times 175{,}56+85\times 209{,}38}{158} = 193{,}76$, sp = 13,92 mmHg.

Testul t independent clasic (Student, varianțe egale). $t = \frac{m_1-m_2}{\text{sp}\times\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$, gl = n1 + n2 - 2. Condiții: grupuri independente, date aproximativ normale în fiecare grup, varianțe egale (test Levene nesemnificativ). Exemplu: $t = -\frac{0{,}52}{13{,}92\times\sqrt{\frac{1}{74}+\frac{1}{86}}} = -0{,}24$, gl = 158, p = 0,81.

Testul t Welch (varianțe inegale). $t = \frac{m_1-m_2}{\sqrt{\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2}}}$; $\text{gl} = \frac{\left(\frac{{s_1}^2}{n_1}+\frac{{s_2}^2}{n_2}\right)^2}{\frac{\left(\frac{{s_1}^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{{s_2}^2}{n_2}\right)^2}{n_2-1}}$. Se poate folosi oricând, nu doar la varianțe inegale; JASP îl oferă ca opțiune. Exemplu: TAS pe sexe, $t = -\frac{0{,}52}{2{,}193} = -0{,}24$, gl = 157,4, p = 0,81.

IC pentru diferența a două medii independente. $\text{IC}\,95\% = (m_1 - m_2) \pm t \times \text{ES}(m_1 - m_2)$, cu $\text{ES} = \text{sp} \times \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ sau ES Welch. Exemplu: diferența M - F a TAS, -0,5 mmHg, IC 95% de la -4,9 la 3,8 mmHg (intervalul conține 0).

Mărimea efectului Cohen d (grupuri independente). $d = \frac{m_1-m_2}{\text{sp}}$. Repere orientative [76]: 0,2 mic, 0,5 mediu, 0,8 mare. Exemplu: TAS M vs F, $d = -\frac{0{,}52}{13{,}92} = -0{,}04$ (efect neglijabil).

Mărimea efectului pentru date pereche. $\text{dz} = \frac{\text{md}}{\text{sd}}$. Exemplu: glicemia 0 vs 6 luni, $\text{dz} = \frac{10{,}19}{11{,}57} = 0{,}88$.

ANOVA cu un factor: sumele de pătrate. Media generală mG; mj și nj media și volumul grupului j; k numărul de grupuri; N totalul. SST (totală) = $\sum(x_i - m_G)^2$SSB (între grupuri, intergrupală) = $\sum n_j \times \left(m_j - m_G\right)^2$SSW (în interiorul grupurilor, reziduală) = $\sum\left(x_i - m_j\right)^2 = \sum\left(n_j - 1\right) \times {s_j}^2$SST = SSB + SSW. Exemplu: scăderea glicemiei la 6 luni pe brațe, SSB = 4806,9, SSW = 16459,5, SST = 21266,4.

ANOVA cu un factor: tabelul și statistica F. gl intergrupal = k - 1; gl rezidual = N - k; $\text{MSB} = \frac{\text{SSB}}{k-1}$; $\text{MSW} = \frac{\text{SSW}}{N-k}$; $F = \frac{\text{MSB}}{\text{MSW}}$. Condiții: grupuri independente, normalitate în fiecare grup (sau a reziduurilor), varianțe omogene (Levene). Exemplu: $\text{MSB} = \frac{4806{,}9}{2} = 2403{,}4$; $\text{MSW} = \frac{16459{,}5}{157} = 104{,}84$; F(2; 157) = 22,93, p < 0,001.

Tabelul D.3. Structura tabelului ANOVA, cu valorile exemplului.

Sursa variației	**SS**	**gl**	$\text{MS} = \frac{\text{SS}}{\text{gl}}$	**F**
Între grupuri	4806,9	2	2403,4	22,93
În grupuri (rezidual)	16459,5	157	104,84	
Total	21266,4	159		

Eta pătrat. $\eta^2 = \frac{\text{SSB}}{\text{SST}}$; proporția din variabilitatea totală explicată de apartenența la grup. Repere orientative: 0,01 mic, 0,06 mediu, 0,14 mare. Exemplu: 4806,9/21266,4 = 0,226.

D.7 Capitolul 7. Teste neparametrice și testul chi-pătrat

Frecvențele așteptate. E = (totalul rândului x totalul coloanei) / N. Exemplu: în tabelul obezitate x diabet, $E(\text{ obezi cu diabet }) = \frac{54\times29}{160} = 9{,}79$; celelalte: 44,21; 19,21; 86,79.

Statistica chi-pătrat (Pearson). $\chi^2 = \frac{\sum(O-E)^2}{E}$, însumat pe toate celulele, cu gl = (r - 1) x (c - 1). Condiții: observații independente, frecvențe absolute (nu procente); cel mult 20% din celule cu E < 5 și nicio celulă cu E < 1; altfel, testul exact Fisher (la 2 x 2) sau gruparea categoriilor. Exemplu: obezitate x diabet, χ^2 = 12,70, gl = 1, p < 0,001.

Formula rapidă pentru tabelul 2 x 2. $\chi^2 = \frac{N\times(\text{ad}-\text{bc})^2}{(a+b)\,(c+d)\,(a+c)\,(b+d)}$. Exemplu: $\frac{160\times(18\times95-36\times11)^2}{54\times106\times29\times131} = 12{,}70$.

Corecția de continuitate Yates (2 x 2). $\chi^2Y = \frac{N\times\left(|\text{ad}-\text{bc}|-\frac{N}{2}\right)^2}{(a+b)\,(c+d)\,(a+c)\,(b+d)}$. Exemplu: 11,20, p = 0,0008.

Gradele de libertate. gl = (numărul de rânduri - 1) x (numărul de coloane - 1). Exemplu: braț (3) x scădere ≥5% (2), gl = 2 x 1 = 2; χ^2 = 6,77, p = 0,034.

Testul Mann-Whitney U. Se ordonează toate valorile împreună și se atribuie ranguri (la egalități, rangul mediu). R1 = suma rangurilor din grupul 1. $U_1 = R_1 - \frac{n_1(n_1+1)}{2}$; U2 = n1 x n2 - U1; U = min(U1, U2) pentru tabel. Condiții: două grupuri independente, variabilă cel puțin ordinală. Exemplu: grupul A {12, 15, 18}, grupul B {14, 20, 22, 25}; rangurile lui A sunt 1, 3, 4, R1 = 8, U1 = 8 - 6 = 2, U2 = 12 - 2 = 10, U = 2.

Aproximarea normală pentru U (eșantioane mari). $z = \frac{U-\frac{\text{n1n2}}{2}}{\sqrt{\frac{n_1\times n_2\times(n_1+n_2+1)}{12}}}$. Exemplu: satisfacția brațul C vs A, U = 2265,5, n1 = 54, n2 = 53: $z = \frac{2265{,}5-1431}{\sqrt{\frac{2862\times108}{12}}} = \frac{834{,}5}{160{,}5} = 5{,}2$, p < 0,001 (fără corecția pentru ranguri egale).

Corelația biserială a rangurilor (mărimea efectului la Mann-Whitney). r = 2 x U1/(n1 x n2) - 1, unde U1 corespunde grupului cu ranguri mai mari. Exemplu: 2 x 2265,5/2862 - 1 = 0,58.

Testul Wilcoxon pentru date pereche (testul rangurilor cu semn). Se calculează diferențele, se elimină cele nule, se ordonează valorile absolute și se atribuie ranguri; W+ = suma rangurilor diferențelor pozitive, W- = suma celor negative, iar $W^+ + W^- = \frac{n(n+1)}{2}$. Statistica testului este min(W+, W-). Exemplu: pentru 6 diferențe nenule, $W^+ + W^- = \frac{6\times7}{2} = 21$.

Testul Kruskal-Wallis. $H = \left[\frac{12}{N(N+1)}\right]\times\sum\left(\frac{{R_j}^2}{n_j}\right) - 3(N+1)$, cu gl = k - 1 (aproximare chi-pătrat). Condiții: k ≥ 3 grupuri independente, variabilă cel puțin ordinală. Exemplu: scăderea glicemiei pe brațe, H = 36,63, gl = 2, p < 0,001.

Testul Friedman. $\chi^2F = \left[\frac{12}{n\times k\times(k+1)}\right]\times\sum{R_j}^2 - 3n(k+1)$, unde Rj este suma rangurilor pentru momentul j (rangurile se atribuie în interiorul fiecărui pacient), gl = k - 1. Exemplu: glicemia la 0, 3 și 6 luni, χ^2F = 74,67, gl = 2, p < 0,001.

Coeficientul de corelație Spearman. $\text{rho} = 1 - \frac{6\times\sum d^2}{n(n^2-1)}$, unde d este diferența dintre rangul lui x și rangul lui y la același individ (formula exactă când nu există ranguri egale). Exemplu: rangurile x 1, 2, 3, 4, 5 și y 2, 1, 4, 3, 5 dau d = -1, 1, -1, 1, 0, $\sum d^2 = 4$, $\text{rho} = 1 - \frac{24}{120} = 0{,}80$.

D.8 Capitolul 8. Alegerea testului (repere numerice)

Verificarea normalității. Testul Shapiro-Wilk: p ≥ 0,05 înseamnă că nu s-a demonstrat abaterea de la normalitate (se pot folosi testele parametrice); p < 0,05 orientează spre transformare sau spre teste neparametrice. Exemplu: TAS, p = 0,303 (parametric); IMC, p < 0,001 (neparametric sau transformare).

Verificarea omogenității variațelor. Testul Levene: p ≥ 0,05 permite testul t clasic sau ANOVA clasică; p < 0,05 cere Welch. Exemplu: TAS pe sexe, Levene p = 0,529.

Regula frecvențelor așteptate. Chi-pătrat dacă toate E ≥ 5 (sau cel mult 20% din celule sub 5); Fisher altfel. Exemplu: braț x scădere ≥5%, cea mai mică E = 12,6, deci chi-pătrat.

D.9 Capitolul 9. Corelație și regresie

Coeficientul de corelație Pearson. $r = \frac{\sum(x_i - m_x)(y_i - m_y)}{\sqrt{\left[\sum(x_i - m_x)^2 \times \sum(y_i - m_y)^2\right]}}$, cu valori între -1 și 1. Condiții: ambele variabile numerice, relație liniară, distribuție aproximativ normală, fără valori extreme influente. Exemplu: IMC și TAS în mg2026, r = 0,405.

Testul t pentru coeficientul de corelație. $t = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}}$, gl = n - 2; H0: ρ = 0. Exemplu: $\frac{0{,}405 \times \sqrt{158}}{\sqrt{1-0{,}164}} = 5{,}57$, p < 0,001.

Coeficientul de determinare. $R^2 = r^2$ (în regresia simplă); proporția din variabilitatea lui y explicată de x. Exemplu: 0,405² = 0,164, adică IMC explică 16,4% din variabilitatea TAS.

Panta dreptei de regresie (metoda celor mai mici pătrate). $b = \frac{\sum(x_i - m_x)(y_i - m_y)}{\sum(x_i - m_x)^2} = \frac{r \times s_y}{s_x}$. Semnificație: modificarea medie a lui y la creșterea lui x cu o unitate. Exemplu: $b = \frac{0{,}405 \times 13{,}88}{3{,}91} = 1{,}44$ (programul, cu date nerotunjite, dă 1,435 mmHg pe kg/m²).

Ordonata la origine. a = my - b x mx. Exemplu: a = 131,13 - 1,435 x 29,04 = 89,46 (programul: 89,47 mmHg).

Ecuația de regresie și predicția. ŷ = a + b x x. Condiție: predicția are sens doar în intervalul valorilor lui x observate. Exemplu: TAS = 89,47 + 1,435 x IMC; la IMC 30, ŷ = 132,5 mmHg; la IMC 35, ŷ = 139,7 mmHg.

Reziduul. ei = yi - ŷi. Exemplu: un pacient cu IMC 30 și TAS 140 are reziduul 140 - 132,5 = 7,5 mmHg.

Eroarea standard a estimării (SEE). $\text{SEE} = \sqrt{\left[\frac{\sum e_i^2}{n-2}\right]} \approx s_y \times \sqrt{1-r^2} \times \sqrt{\frac{n-1}{n-2}}$. Exprimată în unitățile lui y, arată cât de departe de dreaptă sunt, în medie, punctele. Exemplu: $13{,}88 \times \sqrt{0{,}836} \times \sqrt{\frac{159}{158}} = 12{,}73$ mmHg.

IC pentru pantă. $b \pm t(0{,}05;\ n-2) \times \text{ES}(b)$. Exemplu: 1,435, IC 95% 0,925-1,944; intervalul nu conține 0, deci panta este semnificativă.

Coeficientul de determinare ajustat (regresie multiplă). $R^2_{\text{aj}} = 1 - \frac{(1-R^2) \times (n-1)}{n-p-1}$, unde p este numărul de predictori (variabile din model, inclusiv variabilele indicator). Exemplu: TAS ~ IMC + vârstă + sex + fumat (5 termeni), $R^2 = 0{,}308$, $R^2_{\text{aj}} = 1 - \frac{0{,}692 \times 159}{154} = 0{,}286$.

Regresia logistică: din coeficient în OR. $\mathrm{OR} = e^{b}$; $\mathrm{IC}\,95\% = e^{b \pm 1{,}96 \times \mathrm{ES}(b)}$. Exemplu: pentru vârstă, OR = 1,11 pe an, adică șansa de diabet crește cu aproximativ 11% pentru fiecare an în plus, la obezitate și sex constante.

D.10 Capitolul 10. Epidemiologie și analiza riscului

Prevalența. P = numărul cazurilor existente la un moment dat / populația la acel moment. Exemplu: diabet confirmat la includere, 29/160 = 18,1%.

Incidența cumulativă (riscul). IC = numărul cazurilor noi într-o perioadă / populația la risc la începutul perioadei. Exemplu: dacă 12 din 400 de persoane fără diabet dezvoltă boala în 5 ani, IC = 3,0% în 5 ani.

Densitatea de incidență (rata). DI = numărul cazurilor noi / suma persoane-timp de urmărire. Exemplu: 12 cazuri noi în 1850 de persoane-ani, DI = 6,5 la 1000 de persoane-ani.

Relația aproximativă prevalență-incidență (boală stabilă, rară). P ≈ DI x D, unde D este durata medie a bolii. Exemplu: DI = 2 la 1000 pe an și durata 10 ani, P ≈ 2%.

Riscul relativ (RR) sau raportul prevalențelor (RP). $\mathrm{RR} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$. Se calculează din studii de cohortă și din studii clinice (RR), respectiv din studii transversale (RP). RR = 1 înseamnă absența asocierii, RR > 1 factor de risc, RR < 1 factor protector. Exemplu: obezitate și diabet, (18/54)/(11/106) = 0,333/0,104 = 3,21.

IC pentru RR pe scală logaritmică. $\mathrm{ES}[\ln(\mathrm{RR})] = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}$; $\mathrm{IC}\,95\% = e^{\ln \mathrm{RR} \pm 1{,}96 \times \mathrm{ES}}$. Exemplu: ln 3,21 = 1,167, ES = 0,344, $\mathrm{IC}\,95\% = e^{1{,}167 \pm 0{,}675} = 1{,}64 - 6{,}31$.

Raportul cotelor (odds ratio, OR). $\mathrm{OR} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a \times d}{b \times c}$. Singura măsură de asociere validă în studiile caz-control; aproximează RR când boala este rară (sub aproximativ 10%). Exemplu: $\frac{18 \times 95}{36 \times 11} = \frac{1710}{396} = 4{,}32$.

IC pentru OR pe scală logaritmică (metoda Woolf). $\mathrm{ES}[\ln(\mathrm{OR})] = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$; $\mathrm{IC}\,95\% = e^{\ln \mathrm{OR} \pm 1{,}96 \times \mathrm{ES}}$. Exemplu: ln 4,32 = 1,463, ES = 0,430, IC 95% = 1,86-10,03. Dacă IC nu conține 1, asocierea este semnificativă la 5%.

Diferența de risc (riscul atribuibil la expuși, reducerea absolută a riscului). $\mathrm{DR} = p_1 - p_0 = \frac{a}{a+b} - \frac{c}{c+d}$. $\mathrm{IC}\,95\% = \mathrm{DR} \pm 1{,}96 \times \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_0(1-p_0)}{n_0}}$. Exemplu: scădere ≥5% în brațul C față de A, 16/54 - 6/53 = 0,296 - 0,113 = 0,183, IC 95% 0,034-0,332, adică 18,3 puncte procentuale (3,4-33,2).

Numărul necesar de tratat (NNT). $\mathrm{NNT} = \frac{1}{|\mathrm{DR}|}$, rotunjit în sus; limitele IC se obțin inversând limitele IC al DR. Pentru un efect nociv, aceeași formulă dă numărul necesar pentru a produce un efect advers (NNH). Exemplu: 1/0,183 = 5,5, deci 6 pacienți tratați în brațul C în loc de A pentru ca încă unul să slăbească cel puțin 5%; IC 95% 3,0-29,1.

Reducerea relativă a riscului. $\mathrm{RRR} = \frac{p_0 - p_1}{p_0} = 1 - \mathrm{RR}$ (pentru un tratament care scade riscul). Exemplu: dacă mortalitatea scade de la 10% la 7%, RR = 0,7, RRR = 30%, DR = 3 puncte procentuale, NNT = 34.

Fracția atribuibilă la expuși. FAe = (RR - 1) / RR. Exemplu: (3,21 - 1)/3,21 = 0,689; 68,9% din cazurile de diabet de la obezi ar fi atribuibile obezității, dacă relația ar fi cauzală și neconfundată.

Fracția atribuibilă în populație (formula Levin). FAP = pe x (RR - 1) / [1 + pe x (RR - 1)], unde pe este proporția expușilor în populație; echivalent, FAP $= \frac{\mathrm{Pt}-P_0}{\mathrm{Pt}}$. Exemplu: pe = 54/160 = 0,3375, FAP $= \frac{0{,}3375\times 2{,}21}{1+0{,}3375\times 2{,}21} = 0{,}427$; verificare: (0,1813 - 0,1038)/0,1813 = 0,427.

Testul ipotezei pentru RR și OR. H0: RR = 1 (sau OR = 1); se folosește testul chi-pătrat pe tabelul 2 x 2 (D.7) sau testul exact Fisher. Exemplu: χ^2 = 12,70, p < 0,001, concordant cu IC al OR care exclude 1.

D.11 Capitolul 11. Bayes și medicina bazată pe dovezi

Teorema lui Bayes. $P(B \mid T) = \frac{P(T|B)\times P(B)}{P(T)}$, cu $P(T) = P(T \mid B) \times P(B) + P(T \mid \mathrm{non}-B) \times P(\mathrm{non}-B)$. Exemplu: vezi formula VPP de mai jos.

Probabilitatea post-test a unui test pozitiv (VPP din Se, Sp și prevalență). VPP $= \frac{\mathrm{Se}\times P}{\mathrm{Se}\times P+(1-\mathrm{Sp})\times(1-P)}$. Exemplu: prevalență 10%, Se 80%, Sp 90%: 0,08/(0,08 + 0,09) = 0,471 (47%).

VPN din Se, Sp și prevalență. VPN $= \frac{\mathrm{Sp}\times(1-P)}{\mathrm{Sp}\times(1-P)+(1-\mathrm{Se})\times P}$. Exemplu: aceleași valori, 0,81/(0,81 + 0,02) = 0,976.

Conversia probabilitate-cotă (odds). cota = p / (1 - p); p = cota / (1 + cota). Exemplu: p = 0,181, cota = 0,181/0,819 = 0,221; cota 0,666 corespunde lui $p = \frac{0{,}666}{1{,}666} = 0{,}40$.

Raportul de verosimilitate pozitiv. $\mathrm{RV}^+ = \frac{\mathrm{Se}}{1-\mathrm{Sp}}$. Interpretare orientativă: peste 10, creștere mare a probabilității bolii; 5-10, moderată; 2-5, mică; 1, testul nu aduce informație. Exemplu: FINDRISC ≥ 12, 0,759/0,252 = 3,01.

Raportul de verosimilitate negativ. $\mathrm{RV}^- = \frac{1-\mathrm{Se}}{\mathrm{Sp}}$. Sub 0,1, scădere mare a probabilității bolii. Exemplu: 0,241/0,748 = 0,32.

Cota post-test. cota post-test = cota pre-test x RV; apoi probabilitatea post-test = cota/(1 + cota). Este forma practică a teoremei lui Bayes, pe care se bazează nomograma Fagan [122]. Exemplu: probabilitatea pre-test 18,1% (cota 0,221), test pozitiv cu RV+ 3,01: cota post-test 0,666, probabilitatea post-test 40,0%, identică cu VPP din tabel.

D.12 Capitolul 12. Teste diagnostice și curba ROC

Toate formulele de mai jos folosesc notația din tabelul D.2. Exemplul comun este scorul FINDRISC cu pragul ≥ 12 față de diabetul confirmat prin TTGO: a = 22, b = 7, c = 33, d = 98, N = 160.

Prevalența (în eșantionul studiului). $P = \frac{a+b}{N}$. Exemplu: 29/160 = 18,1%.

Sensibilitatea. Se $= \frac{a}{a+b} = P(T^+ \mid B^+)$. Exemplu: 22/29 = 75,9%.

Specificitatea. Sp $= \frac{d}{c+d} = P(T^- \mid B^-)$. Exemplu: 98/131 = 74,8%.

Valoarea predictivă pozitivă. VPP $= \frac{a}{a+c} = P(B^+ \mid T^+)$. Depinde de prevalență. Exemplu: 22/55 = 40,0%.

Valoarea predictivă negativă. VPN $= \frac{d}{b+d} = P(B^- \mid T^-)$. Exemplu: 98/105 = 93,3%.

Acuratețea (proporția clasificărilor corecte). Ac $= \frac{a+d}{N}$. Exemplu: (22 + 98)/160 = 75,0%.

Rata erorii de clasificare. RE $= \frac{b+c}{N} = \frac{\text{FN+FP}}{N} = 1 - \text{Ac}$. Atenție: în unele materiale circulă formula greșită (b + d)/N; corect este (b + c)/N. Exemplu: (7 + 33)/160 = 25,0%.

Proporția fals pozitivilor și a fals negativilor. Rata fals pozitivă = $\frac{c}{c+d} = 1 - \text{Sp}$; rata fals negativă = $\frac{b}{a+b} = 1 - \text{Se}$. Exemplu: 33/131 = 25,2%; 7/29 = 24,1%.

Indicele Youden. J = Se + Sp - 1, între 0 (test inutil) și 1 (test perfect). Pragul optim după acest criteriu este cel care maximizează J [134]. Exemplu: la pragul FINDRISC ≥ 11, J = 0,931 + 0,664 - 1 = 0,595; la pragul ≥ 12, J = 0,759 + 0,748 - 1 = 0,507.

Raportul cotelor de diagnostic (DOR). DOR $= \frac{\text{RV}^+}{\text{RV}^-} = \frac{a \times d}{b \times c}$. Exemplu: $\frac{22 \times 98}{7 \times 33} = 9{,}33$.

Aria de sub curba ROC (AUC), interpretare. AUC = probabilitatea ca un bolnav ales la întâmplare să aibă un scor mai mare decât un nebolnav ales la întâmplare; 0,5 = test inutil, 1 = test perfect. Repere folosite în curs: 0,91-1 excelent, 0,81-0,90 foarte bun, 0,71-0,80 bun, 0,61-0,70 satisfăcător, sub 0,60 slab. Exemplu: FINDRISC, AUC = 0,852 (IC 95% 0,762-0,943).

Reconstituirea tabelului 2 x 2 din indicatori. Dacă se cunosc N, prevalența P, Se și Sp: a + b = P x N; a = Se x (a + b); b = (a + b) - a; c + d = N - (a + b); d = Sp x (c + d); c = (c + d) - d. Dacă se cunosc VPP și a: $a + c = \frac{a}{\text{VPP}}$. Dacă se cunosc VPN și b: $b + d = \frac{b}{1-\text{VPN}}$. Exemplu: N = 1000, P = 10%, Se = 80%, Sp = 90% dau a = 80, b = 20, c = 90, d = 810, iar VPP $= \frac{80}{170} = 47{,}1\%$.

Coeficientul kappa al lui Cohen (acordul dintre doi evaluatori). κ = (po - pe) / (1 - pe), unde po este proporția acordului observat, iar pe = Σ(proporția marginală a evaluatorului 1 x proporția marginală a evaluatorului 2), suma pe categorii. Interpretare după Landis și Koch [103]: 0,21-0,40 slab, 0,41-0,60 moderat, 0,61-0,80 substanțial, peste 0,80 aproape perfect. Exemplu: doi medici citesc 100 de ECG; ambii "patologic" la 20, ambii "normal" la 60, dezacord la 20 (10 + 10). po = 0,80; fiecare medic a declarat patologice 30% din trasee, deci pe = 0,3 x 0,3 + 0,7 x 0,7 = 0,58; κ = 0,22/0,42 = 0,52 (acord moderat).

D.13 Capitolul 13. Studii clinice (formule de planificare)

Formulele de planificare (volumul eșantionului pentru două medii sau două proporții și corecția pentru abandon) sunt în secțiunea D.4. În raportarea unui studiu clinic se folosesc, pentru rezultatul principal binar, RR, DR și NNT (D.10), iar pentru rezultatul numeric diferența mediilor cu IC 95% (D.6).

Rata de abandon și rata de răspuns. Rata de abandon = pierduți din urmărire / randomizați; rata de participare = randomizați / evaluați pentru eligibilitate. Exemplu: în scenariul mg2026, 160 randomizați din 212 evaluați, rata de participare = 75,5%.

Randomizarea în blocuri (numărul de secvențe posibile). Pentru un bloc de 4 cu 2 alocări A și 2 alocări B există $\frac{4!}{2! \times 2!} = 6$ secvențe posibile (AABB, ABAB, ABBA, BAAB, BABA, BBAA).

De reținut

- În tabelul diagnostic cu boala pe rânduri: $\text{Se} = \frac{a}{a+b}$, $\text{Sp} = \frac{d}{c+d}$, $\text{VPP} = \frac{a}{a+c}$, $\text{VPN} = \frac{d}{b+d}$, rata erorii = (b + c)/N.
- În tabelul de risc cu expunerea pe rânduri: $\text{RR} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$, $\text{OR} = \frac{\text{ad}}{\text{bc}}$; IC se calculează pe scală logaritmică, deci este asimetric în jurul estimării.
- $\text{ES} = \frac{s}{\sqrt{n}}$ descrie precizia mediei; s descrie împrăștierea pacienților. Intervalul $m \pm 1{,}96s$ este pentru valori individuale, $m \pm 1{,}96\text{ES}$ este pentru medie.
- $\text{NNT} = \frac{1}{\text{DR}}$, rotunjit în sus; RR și OR nu spun nimic despre mărimea absolută a beneficiului.
- Volumele de eșantion se rotunjesc întotdeauna în sus și se corectează pentru abandon.

Anexa E. Tabele statistice

Tabelele din această anexă au fost calculate cu biblioteca scipy (Python), prin funcțiile de repartiție inversă ale distribuțiilor normală, Student, chi-pătrat și Fisher, și nu au fost copiate din alte manuale. Programele statistice (JASP, MedCalc, Epi Info) dau direct valoarea p, așa că în practica de zi cu zi tabelele sunt rareori necesare. La examen și la seminar însă ele rămân utile: vă obligă să înțelegeți ce înseamnă o valoare critică, de ce un test bilateral cere un prag mai mare decât unul unilateral și cum scade valoarea critică a lui t pe măsură ce crește numărul gradelor de libertate. Toate zecimalele sunt scrise cu virgulă.

Regula generală de citire este aceeași pentru tabelele 3-7: calculați statistica testului din date, găsiți în tabel valoarea critică pentru gradele de libertate și pragul α ales, apoi comparați. Dacă valoarea calculată (în modul) este mai mare sau egală cu valoarea critică, $p \leq \alpha$ și H0 se respinge. Dacă este mai mică, H0 nu se respinge.

E.1 Distribuția normală standard: funcția de repartiție Φ(z)

Tabelul dă Φ(z), adică aria de sub curba normală standard la stânga valorii z (probabilitatea $P(Z \leq z)$). Rândul dă partea întreagă și prima zecimală a lui z, iar coloana dă a doua zecimală. Din Φ(z) se obțin celelalte arii de care aveți nevoie:

- aria dintre 0 și z: Φ(z) - 0,5;
- aria din coada dreaptă ($P(Z > z)$): 1 - Φ(z);
- aria dintre -z și +z: 2 x Φ(z) - 1;
- valoarea p bilaterală pentru un scor z observat: 2 x (1 - Φ(|z|));
- pentru z negativ: Φ(-z) = 1 - Φ(z), deoarece curba este simetrică.

Exemplu. În setul mg2026, scorul z al unei tensiuni sistolice de 160 mmHg este 2,08. Pe rândul 2,0, coloana 0,08, citim Φ(2,08) = 0,9812. Prin urmare, aproximativ 1 - 0,9812 = 0,0188, adică 1,9% dintre pacienți ar fi de așteptat să aibă TAS peste 160 mmHg dacă distribuția ar fi exact normală, iar aria dintre medie și 160 mmHg este 0,9812 - 0,5 = 0,4812.

Tabelul E.1. Funcția de repartiție Φ(z) a distribuției normale standard, z = 0,00-3,49.

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

E.2 Valori critice ale distribuției normale standard

Valoarea critică z este punctul dincolo de care se află o arie egală cu α (test unilateral) sau două arii de câte α/2 (test bilateral). Aceleași valori servesc la construirea intervalelor de încredere: pentru un IC de 95% se folosește z bilateral pentru α = 0,05, adică 1,960; pentru 99%, 2,576; pentru 90%, 1,645.

Exemplu. Un test z bilateral dă z = 2,30. Valoarea critică pentru α = 0,05 bilateral este 1,960, iar pentru α = 0,01 bilateral este 2,576. Cum 1,960 < 2,30 < 2,576, rezultă 0,01 < p < 0,05 (valoarea exactă, din tabelul E.1, este 2 x (1 - 0,9893) = 0,0214).

Tabelul E.2. Valori critice z pentru praguri uzuale.

α	z critic unilateral (z1-α)	z critic bilateral (z1-α/2)	Nivelul IC bilateral (1 - α)
0,200	0,842	1,282	80,0%
0,100	1,282	1,645	90,0%
0,050	1,645	1,960	95,0%
0,025	1,960	2,241	97,5%
0,020	2,054	2,326	98,0%
0,010	2,326	2,576	99,0%
0,005	2,576	2,807	99,5%
0,001	3,090	3,291	99,9%

E.3 Valori critice ale distribuției Student (t)

Tabelul dă valoarea t pentru care aria din cele două cozi luate împreună este egală cu α (test bilateral). Pentru un test unilateral cu prag α citiți coloana 2α (de exemplu, testul unilateral la α = 0,05 folosește coloana 0,10). Gradele de libertate (gl) sunt n - 1 pentru testul t pe un eșantion și pentru testul t pereche, n1 + n2 - 2 pentru testul t independent clasic și n - 2 pentru testul coeficientului de corelație. Rândul ∞ coincide cu valorile critice z din tabelul E.2. Dacă numărul gradelor de libertate lipsește din tabel, folosiți rândul imediat inferior (rezultatul este conservator).

Exemplu. Pentru IC 95% al mediei TAS pe un subeșantion de n = 10 pacienți, gl = 9 și t(0,05; 9) = 2,262. Cu media 130,0 și abaterea standard s, IC $= 130{,}0 \pm \frac{2{,}262 \times s}{\sqrt{10}}$. Pentru n = 160 (gl = 159) se folosește rândul 120, t = 1,980, foarte aproape de 1,96.

Tabelul E.3. Valori critice t bilaterale.

gl	$\alpha = 0{,}20$	$\alpha = 0{,}10$	$\alpha = 0{,}05$	$\alpha = 0{,}02$	$\alpha = 0{,}01$	$\alpha = 0{,}001$
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,920	4,303	6,965	9,925	31,599
3	1,638	2,353	3,182	4,541	5,841	12,924
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,869
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,408
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
11	1,363	1,796	2,201	2,718	3,106	4,437
12	1,356	1,782	2,179	2,681	3,055	4,318
13	1,350	1,771	2,160	2,650	3,012	4,221
14	1,345	1,761	2,145	2,624	2,977	4,140
15	1,341	1,753	2,131	2,602	2,947	4,073
16	1,337	1,746	2,120	2,583	2,921	4,015
17	1,333	1,740	2,110	2,567	2,898	3,965
18	1,330	1,734	2,101	2,552	2,878	3,922
19	1,328	1,729	2,093	2,539	2,861	3,883
20	1,325	1,725	2,086	2,528	2,845	3,850
21	1,323	1,721	2,080	2,518	2,831	3,819
22	1,321	1,717	2,074	2,508	2,819	3,792
23	1,319	1,714	2,069	2,500	2,807	3,768
24	1,318	1,711	2,064	2,492	2,797	3,745
25	1,316	1,708	2,060	2,485	2,787	3,725
26	1,315	1,706	2,056	2,479	2,779	3,707
27	1,314	1,703	2,052	2,473	2,771	3,690
28	1,313	1,701	2,048	2,467	2,763	3,674
29	1,311	1,699	2,045	2,462	2,756	3,659
30	1,310	1,697	2,042	2,457	2,750	3,646
40	1,303	1,684	2,021	2,423	2,704	3,551
50	1,299	1,676	2,009	2,403	2,678	3,496
60	1,296	1,671	2,000	2,390	2,660	3,460
80	1,292	1,664	1,990	2,374	2,639	3,416
100	1,290	1,660	1,984	2,364	2,626	3,390
120	1,289	1,658	1,980	2,358	2,617	3,373
∞	1,282	1,645	1,960	2,326	2,576	3,291

E.4 Valori critice ale distribuției chi-pătrat (χ^2)

Testul chi-pătrat de asociere este întotdeauna unilateral pe coada dreaptă: valorile mari ale lui χ^2 arată o abatere mare a frecvențelor observate de la cele așteptate. Tabelul dă valoarea χ^2 care lasă în dreapta o arie egală cu α. Pentru un tabel de contingență cu r rânduri și c coloane, gl = (r - 1) x (c - 1); un tabel 2 x 2 are gl = 1.

Exemplu. În mg2026, tabelul braț x scădere ponderală de cel puțin 5% are 3 rânduri și 2 coloane, deci gl = 2, iar χ^2 calculat = 6,77. Valoarea critică pentru α = 0,05 și gl = 2 este 5,991, iar

pentru α = 0,01 este 9,210. Prin urmare 0,01 < p < 0,05 (programul dă p = 0,034) și H0 se respinge la pragul de 5%.

Tabelul E.4. Valori critice χ^2 (aria din dreapta = α).

gl	$\alpha = 0{,}10$	$\alpha = 0{,}05$	$\alpha = 0{,}01$	$\alpha = 0{,}001$
1	2,706	3,841	6,635	10,828
2	4,605	5,991	9,210	13,816
3	6,251	7,815	11,345	16,266
4	7,779	9,488	13,277	18,467
5	9,236	11,070	15,086	20,515
6	10,645	12,592	16,812	22,458
7	12,017	14,067	18,475	24,322
8	13,362	15,507	20,090	26,124
9	14,684	16,919	21,666	27,877
10	15,987	18,307	23,209	29,588
11	17,275	19,675	24,725	31,264
12	18,549	21,026	26,217	32,909
13	19,812	22,362	27,688	34,528
14	21,064	23,685	29,141	36,123
15	22,307	24,996	30,578	37,697
16	23,542	26,296	32,000	39,252
17	24,769	27,587	33,409	40,790
18	25,989	28,869	34,805	42,312
19	27,204	30,144	36,191	43,820
20	28,412	31,410	37,566	45,315
21	29,615	32,671	38,932	46,797
22	30,813	33,924	40,289	48,268
23	32,007	35,172	41,638	49,728
24	33,196	36,415	42,980	51,179
25	34,382	37,652	44,314	52,620
26	35,563	38,885	45,642	54,052
27	36,741	40,113	46,963	55,476
28	37,916	41,337	48,278	56,892
29	39,087	42,557	49,588	58,301
30	40,256	43,773	50,892	59,703

E.5 Valori critice ale distribuției Fisher (F) pentru α = 0,05

Distribuția F are două numere de grade de libertate. În ANOVA cu un factor, gl1 (numărătorul) = k - 1, unde k este numărul de grupuri, iar gl2 (numitorul) = N - k, unde N este numărul total de observații. Coloanele tabelului corespund lui gl1, rândurile lui gl2. Dacă F calculat este mai mare sau egal cu valoarea din tabel, p ≤ 0,05. Același tabel servește, cu gl1 = numărul de predictori și gl2 = n - p - 1, pentru testul F global al unei regresii multiple.

Exemplu. ANOVA pentru scăderea glicemiei pe cele trei brațe din mg2026: k = 3, N = 160, deci gl1 = 2 și gl2 = 157. Rândul 157 nu există în tabel; folosim rândul 120 (conservator): F critic = 3,072. Valoarea calculată, F = 22,93, depășește cu mult pragul, deci p < 0,05 (programul dă p < 0,001).

Tabelul E.5. Valori critice F pentru α = 0,05 (gl1 pe coloane, gl2 pe rânduri).

gl2 \ gl1	1	2	3	4	5	6	7	8	9	10
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79

gl2 gl1	1	2	3	4	5	6	7	8	9	10
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83

E.6 Valori critice ale coeficientului de corelație Pearson (r)

Pentru a decide dacă un coeficient de corelație observat diferă semnificativ de zero, comparați |r| cu valoarea critică pentru volumul eșantionului n. Valorile provin din relația dintre r și testul t, cu gl = n - 2:

$$r_{critic} = t_{critic} / \sqrt{(gl + t_{critic}^2)}$$

Tabelul este bilateral (H1: $\rho \neq 0$). Observați cât de mult scade pragul pe măsură ce crește n: cu 10 pacienți aveți nevoie de $|r| \geq 0{,}632$ pentru $p < 0{,}05$, pe când cu 100 de pacienți ajunge $|r| \geq 0{,}197$. De aici și avertismentul din capitolul 9: pe eșantioane mari, corelații slabe, fără importanță clinică, devin "semnificative".

Exemplu. Corelația dintre IMC și TAS în mg2026 este r = 0,405, cu n = 160. Chiar și pragul pentru n = 100 la $\alpha = 0{,}01$ (0,256) este depășit, deci $p < 0{,}01$. La n = 160 pragul este și mai mic.

Tabelul E.6. Valori critice bilaterale ale lui r Pearson.

n	$\text{gl} = n - 2$	$\alpha = 0{,}05$	$\alpha = 0{,}01$
5	3	0,878	0,959
6	4	0,811	0,917
7	5	0,754	0,875
8	6	0,707	0,834
9	7	0,666	0,798
10	8	0,632	0,765
11	9	0,602	0,735
12	10	0,576	0,708
13	11	0,553	0,684
14	12	0,532	0,661
15	13	0,514	0,641
16	14	0,497	0,623
17	15	0,482	0,606
18	16	0,468	0,590
19	17	0,456	0,575
20	18	0,444	0,561
21	19	0,433	0,549
22	20	0,423	0,537
23	21	0,413	0,526
24	22	0,404	0,515
25	23	0,396	0,505
26	24	0,388	0,496
27	25	0,381	0,487
28	26	0,374	0,479
29	27	0,367	0,471
30	28	0,361	0,463
35	33	0,334	0,430
40	38	0,312	0,403
45	43	0,294	0,380
50	48	0,279	0,361
55	53	0,266	0,345
60	58	0,254	0,330
65	63	0,244	0,317
70	68	0,235	0,306
75	73	0,227	0,296
80	78	0,220	0,286
85	83	0,213	0,278
90	88	0,207	0,270
95	93	0,202	0,263
100	98	0,197	0,256

E.7 Valori critice aproximative ale coeficientului de corelație Spearman (rho)

Pentru eșantioane foarte mici, distribuția lui rho sub H0 nu este continuă și valorile critice se obțin prin enumerarea tuturor permutărilor rangurilor. În tabelul de mai jos, valorile pentru n = 5-10 sunt exacte (calculate prin enumerarea completă a celor n! permutări), iar pentru n = 11-30 sunt aproximate prin formula bazată pe distribuția t cu n - 2 grade de libertate, aceeași ca la r Pearson. Liniuța înseamnă că, la acel n, nicio valoare a lui rho nu poate atinge pragul cerut. Aproximarea t subestimează puțin pragul exact pentru n între 11 și 15 (cu aproximativ 0,01-0,02), deci un rho

aflat chiar la limită trebuie judecat după valoarea p exactă. Tabelul este bilateral; regula de decizie: |rho| ≥ valoarea critică înseamnă p ≤ α. Pentru raportare folosiți valoarea p exactă dată de program.

Exemplu. Un medic de familie corelează, la 12 pacienți, scorul de activitate fizică (1-5) cu numărul de consultații pe an și obține rho = -0,61. Valoarea critică pentru n = 12 și α = 0,05 este 0,576; cum |-0,61| ≥ 0,576, corelația negativă este semnificativă la pragul de 5%, dar nu la 1% (pragul este 0,708).

Tabelul E.7. Valori critice bilaterale ale lui rho Spearman (exacte pentru n ≤ 10, aproximative pentru n ≥ 11).

n	$\alpha = 0{,}05$	$\alpha = 0{,}01$	Metoda
5	1,000	-	exactă
6	0,886	1,000	exactă
7	0,786	0,929	exactă
8	0,738	0,881	exactă
9	0,700	0,833	exactă
10	0,648	0,794	exactă
11	0,602	0,735	aproximare t
12	0,576	0,708	aproximare t
13	0,553	0,684	aproximare t
14	0,532	0,661	aproximare t
15	0,514	0,641	aproximare t
16	0,497	0,623	aproximare t
17	0,482	0,606	aproximare t
18	0,468	0,590	aproximare t
19	0,456	0,575	aproximare t
20	0,444	0,561	aproximare t
21	0,433	0,549	aproximare t
22	0,423	0,537	aproximare t
23	0,413	0,526	aproximare t
24	0,404	0,515	aproximare t
25	0,396	0,505	aproximare t
26	0,388	0,496	aproximare t
27	0,381	0,487	aproximare t
28	0,374	0,479	aproximare t
29	0,367	0,471	aproximare t
30	0,361	0,463	aproximare t

E.8 Tabel de 1000 de cifre aleatoare

Cifrele au fost generate cu generatorul numpy (default_rng) inițializat cu sămânța 2026, astfel încât tabelul poate fi reprodus oricând. Sunt așezate pe 25 de rânduri a câte 40 de cifre, grupate câte cinci pentru ușurința citirii. Tabelul se folosește la exercițiile de eșantionare aleatoare simplă și de randomizare din capitolele 1, 4 și 13.

Mod de folosire. (1) Numerotați unitățile din populație, de exemplu fișele pacienților de la 001 la N. (2) Alegeți un punct de pornire fără să vă uitați la cifre (de exemplu, închideți ochii și puneți creionul pe tabel) și o direcție de citire (de la stânga la dreapta, apoi rândul următor). (3) Citiți grupuri de atâtea cifre câte are N (trei cifre pentru N = 160). (4) Păstrați numerele cuprinse între 001 și N, ignorați-le pe cele din afara intervalului și pe cele repetate, până obțineți volumul dorit al eșantionului.

Exemplu. Vrem să extragem 5 pacienți din cei 160 ai setului mg2026, pornind de la începutul rândului 01 și citind câte trei cifre. Reținem doar numerele 001-160, în ordinea în care apar, fără repetiții. Rezultatul pentru acest punct de pornire este dat sub tabel, ca să vă puteți verifica.

Tabelul E.8. 1000 de cifre aleatoare (sămânța 2026).

Rând	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40
01	81063	40363	87797	18602	19792	66715	68644	31212
02	85042	68044	93114	53462	92989	75623	49755	44903
03	46430	21742	54456	11795	46033	47484	06759	47007
04	85737	43019	49302	37773	35641	75998	96211	92002
05	92160	40030	86945	64558	39112	92756	06057	99275
06	78417	61417	30265	83693	06635	68142	30409	54295
07	96987	37383	01462	82435	19965	74802	46779	89366
08	66541	10339	22016	36172	09330	52266	11272	41167
09	85241	24723	00595	05164	32134	56909	50882	13892
10	47172	81972	86476	02643	51950	47556	04495	18293
11	87760	28747	67504	28828	32252	26460	02323	94571
12	14641	69584	43551	91538	60028	85564	53427	67334
13	67098	92451	30143	08122	59013	10926	87030	94945
14	09852	44726	82752	08805	00095	97714	55106	19394
15	47814	17618	98872	93577	34433	92417	84012	12557
16	97438	07541	23055	39400	31889	54868	11037	55010
17	38782	39765	17337	41838	43883	35657	84889	48999
18	39498	11627	75755	25167	71771	06440	26623	43235
19	56213	78739	53584	03942	83309	83429	91282	38286
20	10159	41014	43622	95551	57032	59306	36123	85676
21	82625	39050	36236	76037	82381	93643	30979	59035
22	40148	62882	92537	00918	66700	00941	31970	36424
23	47546	80472	81077	29606	39208	23633	41589	75280
24	78458	90542	81342	66013	84467	16883	78189	88248
25	02231	19480	09902	46801	05552	79698	07652	33131

Verificare pentru exemplul de mai sus: citind de la începutul tabelului câte trei cifre, primele cinci numere distincte cuprinse între 001 și 160 sunt 036, 021, 121, 042, 034. Dacă ați pornit din alt punct al tabelului, eșantionul vostru va fi altul, la fel de valid.

Anexa F. Teste-grilă de antrenament pentru examen

Examenul teoretic de Biostatistică este un test-grilă cu 50 de întrebări, corectat în sistem rezidențiat. Această anexă vă oferă trei teste complete, construite după același model, ca să vă puteți antrena în condiții cât mai apropiate de cele reale. Fiecare test acoperă toate cele 14 capitole ale cursului, aproximativ proporțional cu volumul lor, și amestecă întrebări de teorie, de interpretare și de calcul.

Cum sunt construite întrebările și cum se punctează

Întrebările sunt de două tipuri, marcate la începutul enunțului:

- **CS (complement simplu)**: dintre cele cinci variante (a-e), exact una este corectă.
- **CM (complement multiplu)**: dintre cele cinci variante, sunt corecte cel puțin două și cel mult patru.

Regulile exacte de punctare (câte puncte valorează fiecare tip de întrebare și cum se tratează un răspuns CM parțial) sunt stabilite de metodologia universității pentru examenele corectate în sistem rezidențiat și sunt anunțate de disciplină înaintea sesiunii; vă rog să le citiți acolo, nu în această carte. Principiul pe care trebuie să-l rețineți este simplu: un răspuns este considerat corect atunci când corespunde întocmai grilei. La CS, aceasta înseamnă varianta corectă și numai ea. La CM, strategia sigură este să marcați toate variantele corecte și nicio variantă greșită; o variantă greșită bifată "de siguranță" sau o variantă corectă omisă vă poate costa punctajul întrebării.

Câteva sfaturi practice, din experiența corectării acestor examene:

- Citiți enunțul până la capăt. Diferența dintre "interval pentru valori individuale" și "interval de încredere pentru medie", sau dintre "procent valid" și "procent din total", se află adesea într-un singur cuvânt.
- La calcule, scrieți pe ciornă tabelul 2 x 2 complet (a, b, c, d și totalurile) înainte de a calcula orice indicator. Majoritatea greșelilor vin din citirea celulei greșite.
- Verificați ordinul de mărime al rezultatului: o probabilitate nu poate depăși 1, o sensibilitate de 90% nu poate da un VPP de 150%, un volum de eșantion se rotunjește în sus.
- La CM, judecați fiecare variantă separat, ca și cum ar fi o întrebare de tip adevărat/fals.
- Pentru calcule aveți nevoie doar de operații simple și, uneori, de tabelele din anexa E.

Lucrați fiecare test într-o singură ședință, fără întrerupere și fără să vă uitați la răspunsuri. Notați-vă timpul. Abia apoi verificați grila din secțiunea "Răspunsuri și explicații", de la finalul anexei, și citiți explicațiile pentru întrebările greșite. Explicațiile sunt date pentru toate întrebările de calcul și pentru cele care conțin o capcană; pentru întrebările de definiție simplă, recitiți secțiunea corespunzătoare din capitol sau glosarul (anexa B). Formulele sunt adunate în anexa D.

Numerele marcate ca provenind din setul mg2026 sunt cele calculate în capitolele cursului pe setul didactic simulat (anexa C); restul datelor din enunțuri sunt construite pentru exercițiu.

Testul 1

1. (CS) Care dintre următoarele operații este un exemplu de inferență statistică?

a) Calculul mediei TAS a celor 160 de pacienți din setul mg2026

b) Construirea histogramei glicemiei a jeun a pacienților din eșantion

c) Numărarea fumătorilor din eșantion

d) Estimarea, pe baza eșantionului, a intervalului în care se află media TAS a tuturor adulților cu risc cardiometabolic din regiune

e) Întocmirea tabelului de frecvențe pentru nivelul de educație

2. (CM) Despre metodele de eșantionare, sunt corecte afirmațiile:

a) În eșantionarea aleatoare simplă, fiecare individ din populație are aceeași probabilitate de a fi selectat

b) Un eșantion de conveniență devine reprezentativ dacă are peste 1000 de persoane

c) Eșantionarea stratificată garantează prezența tuturor straturilor populației în eșantion

d) Eroarea sistematică de selecție scade pe măsură ce crește volumul eșantionului

e) În eșantionarea pe grupuri (clustere) se extrag aleator grupuri întregi, de exemplu cabinete

3. (CS) Într-un cabinet de medicină de familie, 30% dintre pacienți sunt fumători, 20% au hipertensiune arterială, iar 8% sunt fumători hipertensivi. Probabilitatea ca un pacient ales la întâmplare să fie fumător sau hipertensiv este:

a) 0,50

b) 0,42

c) 0,06

d) 0,58

e) 0,38

4. (CS) Clasa funcțională NYHA a insuficienței cardiace (I, II, III, IV) este o variabilă:

a) nominală

b) de raport

c) continuă

d) de interval

e) ordinală

5. (CM) Duratele de spitalizare (zile) ale cinci pacienți sunt 4, 7, 7, 9, 13. Sunt corecte afirmațiile:

a) Media este 8 zile

b) Mediana este 7 zile

c) Moda este 7 zile

d) Amplitudinea este 13 zile

e) Dacă valoarea 13 ar fi înlocuită cu 43, mediana ar crește

6. (CS) O variabilă are media 80 și abaterea standard 12. Coeficientul de variație este:

a) 6,7%

b) 12%

c) 15%

d) 66,7%

e) 0,15%

7. (CM) Despre diagrama box-plot, sunt corecte afirmațiile:

a) Linia din interiorul cutiei reprezintă media aritmetică

b) Marginile cutiei corespund primei și celei de-a treia cuartile

c) Înălțimea cutiei este intervalul intercuartilic

d) Punctele desenate separat, dincolo de mustăți, sunt valori aberante după regula 1,5 x IQR

e) Este graficul recomandat pentru variabile nominale

8. (CS) Colesterolul total are, într-o populație, o distribuție normală cu media 200 mg/dL și abaterea standard 40 mg/dL. Procentul aproximativ al persoanelor cu colesterol peste 260 mg/dL este:

a) 6,7%

b) 2,3%

c) 13,4%

d) 15,9%

e) 43,3%

9. (CM) Despre distribuția normală, sunt corecte afirmațiile:

a) Este simetrică față de medie

b) Media, mediana și moda coincid

c) Aproximativ 95,4% din valori se află în intervalul $m \pm 2s$

d) Aria totală de sub curbă depinde de abaterea standard

e) Un test Shapiro-Wilk cu p = 0,30 dovedește că datele sunt normale

10. (CS) Trigliceridele din setul mg2026 au asimetria 0,95 și testul Shapiro-Wilk p < 0,001. Cel mai potrivit mod de a descrie tendința centrală și împrăștierea, fără transformare, este:

a) media ± abaterea standard

b) mediana și intervalul intercuartilic

c) moda și amplitudinea

d) media și eroarea standard

e) doar coeficientul de variație

11. (CS) În setul mg2026, TAS are media 131,13 mmHg și abaterea standard 13,88 mmHg. Scorul z al unei valori de 160 mmHg este:

a) 0,21

b) 1,96

c) 2,08

d) 2,88

e) 28,87

12. (CS) Pe un eșantion de 100 de pacienți, media unei variabile este 150, iar abaterea standard 20. Intervalul de încredere de 95% pentru media populației (folosind z = 1,96) este:

a) 110,8-189,2

b) 148,0-152,0

c) 149,6-150,4

d) 146,1-153,9

e) 130,0-170,0

13. (CM) Despre abaterea standard (AS) și eroarea standard a mediei (ES), sunt corecte afirmațiile:

a) $\text{ES} = \frac{\text{AS}}{\sqrt{n}}$

b) AS descrie împrăștierea pacienților, iar ES descrie precizia mediei

c) Dacă volumul eșantionului crește de 4 ori, ES se reduce aproximativ la jumătate

d) Dacă volumul eșantionului crește, AS scade proporțional cu $\sqrt{n}$

e) Intervalul $m \pm 1{,}96\text{ES}$ conține aproximativ 95% dintre pacienți

14. (CS) Într-un eșantion de 200 de gravide, 40 au anemie. Intervalul de încredere de 95% pentru prevalența anemiei, calculat prin metoda Wald, este aproximativ:

a) 18,0-22,0%

b) 14,5-25,5%

c) 10,0-30,0%

d) 16,0-24,0%

e) 19,4-20,6%

15. (CS) Ce volum de eșantion este necesar pentru a estima o prevalență complet necunoscută cu o marjă de eroare de ±5% și un nivel de încredere de 95%?

a) 100

b) 196

c) 246

d) 385

e) 1537

16. (CS) Valoarea p a unui test statistic reprezintă:

a) probabilitatea de a obține un rezultat cel puțin la fel de extrem ca cel observat, dacă ipoteza nulă ar fi adevărată

b) probabilitatea ca ipoteza nulă să fie adevărată

c) probabilitatea ca ipoteza alternativă să fie adevărată

d) probabilitatea erorii de tip II

e) mărimea efectului tratamentului

17. (CM) Despre erorile statistice și puterea testului, sunt corecte afirmațiile:

a) Eroarea de tip I constă în respingerea ipotezei nule atunci când aceasta este adevărată

b) β este probabilitatea erorii de tip II

c) Puterea testului este 1 - α

d) Puterea crește odată cu volumul eșantionului

e) Scăderea pragului α de la 0,05 la 0,01, la același volum al eșantionului, crește puterea

18. (CS) În mg2026, diferența TAS dintre bărbați și femei este -0,5 mmHg (IC 95% de la -4,9 la 3,8 mmHg), p = 0,81. Concluzia corectă este:

a) s-a demonstrat că TAS medie este egală la cele două sexe

b) ipoteza nulă se acceptă

c) ipoteza nulă nu se respinge; datele sunt compatibile atât cu absența diferenței, cât și cu diferențe de câțiva mmHg în oricare sens

d) bărbații au o TAS semnificativ mai mică

e) puterea studiului este de 81%

19. (CM) Pentru un test z bilateral cu pragul α = 0,05, sunt corecte afirmațiile:

a) Valoarea critică este 1,96

b) Valoarea critică este 1,645

c) Un z calculat de 2,30 corespunde unui p < 0,05

d) Un z calculat de 2,30 corespunde unui p < 0,01

e) Dacă IC 95% al diferenței nu conține 0, atunci p < 0,05

20. (CS) La 16 pacienți se măsoară TAS înainte și după un tratament. Media diferențelor este 4 mmHg, iar abaterea standard a diferențelor 8 mmHg. Statistica t a testului t pereche este:

a) 0,5

b) 1,0

c) 8,0

d) 16,0

e) 2,0

21. (CM) Pentru situația de la întrebarea 20 (t pereche, n = 16), sunt corecte afirmațiile:

a) Numărul gradelor de libertate este 15

b) Valoarea critică bilaterală a lui t pentru $\alpha = 0,05$ este 2,131

c) Ipoteza nulă se respinge la pragul bilateral de 0,05

d) Valoarea p bilaterală este cuprinsă între 0,05 și 0,10

e) La un test unilateral, cu direcția stabilită înaintea studiului și $\alpha = 0,05$, rezultatul ar fi semnificativ

22. (CS) Pentru a compara media scăderii glicemiei între trei brațe independente ale unui studiu, cu date normale și variante omogene, se folosește:

a) testul t pereche

b) trei teste t independente, fără nicio corecție

c) ANOVA cu un factor

d) testul chi-pătrat

e) testul Friedman

23. (CM) În mg2026, ANOVA pentru scăderea glicemiei la 6 luni pe brațe dă $F(2; 157) = 22,93$, $p < 0,001$, $\eta^2 = 0,226$, iar testul Tukey dă pentru diferența C-B $p = 0,024$. Sunt corecte afirmațiile:

a) ANOVA singură arată care brațe diferă între ele

b) Gradele de libertate corespund la 3 grupuri și 160 de pacienți

c) Aproximativ 22,6% din variabilitatea scăderii glicemiei este explicată de braț

d) Diferența dintre brațele C și B este semnificativă la pragul de 0,05

e) Rezultatul obligă la repetarea analizei cu testul Kruskal-Wallis

24. (CS) Alternativa neparametrică a testului t pereche este:

a) testul Mann-Whitney

b) testul Wilcoxon al rangurilor cu semn

c) testul Kruskal-Wallis

d) testul Friedman

e) testul chi-pătrat

25. (CS) Un tabel de contingență are 3 rânduri și 4 coloane. Numărul gradelor de libertate pentru testul chi-pătrat este:

a) 6

b) 7

c) 12

d) 5

e) 2

26. (CS) Într-un tabel de contingență cu N = 200, totalul unui rând este 60, iar totalul unei coloane este 50. Frecvența așteptată în celula de la intersecția lor este:

a) 10

b) 12

c) 15

d) 25

e) 30

27. (CM) Despre testul chi-pătrat de asociere, sunt corecte afirmațiile:

a) Se calculează pe frecvențe absolute

b) Poate fi aplicat direct pe procentele din tabel

c) La un tabel 2 x 2 cu frecvențe așteptate sub 5 se preferă testul exact Fisher

d) O valoare mare a lui χ^2 indică o abatere mare a frecvențelor observate de la cele așteptate

e) Valoarea lui χ^2 arată direct sensul și mărimea asocierii

28. (CS) Satisfacția față de intervenție (scală Likert 1-5) se compară între brațul C și brațul A. Testul potrivit este:

a) testul t pentru eșantioane independente

b) testul Wilcoxon al rangurilor cu semn

c) testul t pereche

d) testul Mann-Whitney

e) coeficientul de corelație Pearson

29. (CM) Care dintre următoarele situații cer un test pentru date pereche (dependente)?

a) Glicemia acelorași pacienți la includere și la 6 luni

b) TAS la bărbați comparativ cu femeile

c) Acuitatea vizuală a ochiului drept comparativ cu ochiul stâng, la aceiași pacienți

d) Rezultatul (pozitiv sau negativ) a două teste diagnostice aplicate acelorași pacienți

e) IMC al pacienților din trei cabinete diferite

30. (CS) Un student compară o variabilă numerică între două grupuri independente de câte 25 de pacienți; testul Shapiro-Wilk dă p = 0,003 într-unul dintre grupuri, iar histograma arată o asimetrie puternică. Cea mai potrivită alegere este:

a) testul t clasic, fără rezerve

b) testul Mann-Whitney

c) testul chi-pătrat

d) coeficientul Pearson

e) ANOVA cu măsurători repetate

31. (CS) Între două variabile, coeficientul de corelație Pearson este r = 0,6. Coeficientul de determinare este:

a) 0,06

b) 0,36

c) 0,60

d) 0,77

e) 1,20

32. (CM) În mg2026, dreapta de regresie a TAS în funcție de IMC este TAS = 89,47 + 1,435 x IMC (r = 0,405). Sunt corecte afirmațiile:

a) La un IMC de 30 kg/m², TAS prezisă este aproximativ 132,5 mmHg

b) TAS crește în medie cu 1,435 mmHg pentru fiecare kg/m² în plus

c) 89,47 mmHg este TAS reală a unui pacient cu IMC egal cu 0

d) IMC explică aproximativ 16,4% din variabilitatea TAS

e) Valoarea lui r demonstrează că obezitatea cauzează creșterea TAS

33. (CS) Coeficientul de corelație Spearman este preferat coeficientului Pearson atunci când:

a) ambele variabile sunt normale și relația este liniară

b) volumul eșantionului depășește 1000

c) ambele variabile sunt nominale

d) se dorește o ecuație de predicție

e) cel puțin una dintre variabile este ordinală sau relația este monotonă, dar nu liniară

34. (CM) Într-un studiu pe 210 pacienți se obține r = 0,15, p = 0,03. Sunt corecte afirmațiile:

a) Corelația este statistic semnificativă la pragul de 0,05

b) Intensitatea asocierii este foarte mică, deși rezultatul este semnificativ

c) Variabila x explică aproximativ 2,25% din variabilitatea lui y

d) Rezultatul demonstrează o relație cauzală

e) Un coeficient mic poate deveni semnificativ pe un eșantion mare

35. (CS) Într-un studiu de cohortă, 30 din 100 de expuși și 10 din 100 de neexpuși dezvoltă boala. Riscul relativ este:

a) 2,0

b) 3,0

c) 3,86

d) 0,33

e) 20

36. (CM) Pentru datele de la întrebarea 35, sunt corecte afirmațiile:

a) OR este aproximativ 3,86

b) Diferența de risc este de 20 de puncte procentuale

c) Numărul necesar pentru a produce un caz în plus (NNH) este 5

d) OR este mai mic decât RR

e) Fracția atribuibilă la expuși este aproximativ 66,7%

37. (CS) Măsura de asociere care poate fi calculată corect într-un studiu caz-control este:

a) raportul cotelor (OR)

b) riscul relativ (RR)

c) incidența cumulativă

d) prevalența bolii în populație

e) numărul necesar de tratat

38. (CS) Într-un studiu randomizat, evenimentul nefavorabil apare la 20% dintre pacienții din grupul de control și la 15% dintre cei tratați. Numărul necesar de tratat (NNT) este:

a) 5

b) 15

c) 20

d) 25

e) 33

39. (CM) Despre factorul de confuzie, sunt corecte afirmațiile:

a) Este asociat cu expunerea

b) Este asociat cu rezultatul (boala)

c) Este o verigă intermediară pe calea cauzală dintre expunere și boală

d) Influența lui poate fi controlată prin randomizare

e) Este eliminat automat prin creșterea volumului eșantionului

40. (CS) O boală are prevalența 2%. Un test are sensibilitatea 95% și specificitatea 90%. Probabilitatea ca o persoană cu test pozitiv să aibă boala este aproximativ:

a) 95%

b) 90%

c) 2%

d) 16%

e) 50%

41. (CM) Probabilitatea pre-test a unei boli este 10%, iar testul are raportul de verosimilitate pozitiv RV+ = 9. Sunt corecte afirmațiile:

a) Cota pre-test este aproximativ 0,11

b) După un test pozitiv, cota post-test este 1,0

c) Probabilitatea post-test după un test pozitiv este 50%

d) Probabilitatea post-test după un test pozitiv este 90%

e) Un RV+ de 9 corespunde, de exemplu, unui test cu Se 90% și Sp 90%

42. (CS) În formularea unei întrebări clinice după modelul PICO, litera O desemnează:

a) observația

b) rezultatul urmărit (outcome)

c) obiectivul studiului

d) cota (odds)

e) orbirea

43. (CM) Un test este aplicat la 50 de bolnavi și 150 de nebolnavi: 45 de bolnavi au test pozitiv, iar 130 de nebolnavi au test negativ. Sunt corecte afirmațiile:

a) Sensibilitatea este 90%

b) Specificitatea este aproximativ 86,7%

c) Valoarea predictivă pozitivă este aproximativ 69,2%

d) Rata erorii de clasificare este (5 + 130)/200

e) Acuratețea testului este 75%

44. (CS) Pentru testul de la întrebarea 43, rata erorii de clasificare este:

a) 2,5%

b) 10%

c) 13,3%

d) 67,5%

e) 12,5%

45. (CS) Într-o populație de 200 de persoane, prevalența bolii este 30%. Testul are sensibilitatea 80% și specificitatea 90%. Valoarea predictivă pozitivă este aproximativ:

a) 30,0%

b) 77,4%

c) 80,0%

d) 90,0%

e) 91,3%

46. (CM) Despre curba ROC, sunt corecte afirmațiile:

a) Pe axa verticală se reprezintă sensibilitatea

b) Pe axa orizontală se reprezintă 1 - specificitatea

c) O arie de sub curbă de 0,5 corespunde unui test fără valoare discriminativă

d) O arie de 0,852 înseamnă că 85,2% dintre pacienți sunt clasificați corect la orice prag

e) Pragul care maximizează indicele Youden este întotdeauna cel cu sensibilitatea maximă

47. (CS) Un test are sensibilitatea 85% și specificitatea 75%. Indicele Youden este:

a) 0,10

b) 0,40

c) 0,60

d) 0,64

e) 1,60

48. (CS) Într-un studiu clinic randomizat de superioritate, analiza principală recomandată este:

a) analiza după intenția de tratament

b) analiza per protocol

c) analiza doar a pacienților care au terminat studiul

d) analiza unui subgrup ales după vederea datelor

e) analiza interimară

49. (CM) Despre studiile clinice randomizate, sunt corecte afirmațiile:

a) Randomizarea tinde să echilibreze între brațe și factorii de confuzie necunoscuți

b) Ascunderea alocării și orbirea sunt același lucru

c) Într-un studiu dublu orb, nici pacientul, nici investigatorul nu cunosc tratamentul primit

d) Criteriul principal de evaluare se poate alege după analiza datelor

e) Ghidul CONSORT cere o diagramă de flux a participanților

50. (CS) Un articol raportează, pentru un studiu transversal pe 80 de pacienți, "OR = 2,1; p = 0,04", fără alte detalii. Informația care lipsește și este cea mai importantă pentru interpretare este:

a) numele programului statistic

b) valoarea exactă a pragului α

c) media vârstei pacienților

d) intervalul de încredere al OR

e) numărul autorilor

Testul 2

1. (CS) Media tensiunii arteriale sistolice a tuturor adulților din România, valoare necunoscută pe care încercăm să o estimăm, este:

a) un parametru

b) o statistică de eșantion

c) un eșantion

d) o variabilă calitativă

e) o eroare standard

2. (CM) Despre protecția datelor și etica cercetării, sunt corecte afirmațiile:

a) Datele pseudonimizate rămân date cu caracter personal

b) Consimțământul informat poate fi retras oricând de participant

c) Anonimizarea înseamnă înlocuirea numelui cu un cod, păstrând o listă de corespondență

d) Datele privind sănătatea fac parte din categoriile speciale de date protejate de GDPR

e) Un set de date simulat, cu pacienți fictivi, intră sub incidența GDPR ca orice set clinic

3. (CS) Evenimentele A și B sunt independente, cu P(A) = 0,3 și P(B) = 0,4. Probabilitatea ca ambele să se producă este:

a) 0,70

b) 0,12

c) 0,10

d) 0,58

e) 0,75

4. (CS) Numărul de consultații pe an la șase pacienți este 3, 5, 8, 10, 12, 20. Mediana este:

a) 8

b) 9,67

c) 9

d) 10

e) 11,5

5. (CM) Un chestionar a fost completat de 200 de persoane. La întrebarea "Fumați?", 20 de persoane nu au răspuns, iar 54 au răspuns "da". Sunt corecte afirmațiile:

a) Procentul valid al fumătorilor este 30%

b) Procentul fumătorilor raportat la totalul respondenților este 27%

c) Procentul valid se calculează raportând la 200

d) Raportul trebuie să precizeze numărul de valori lipsă

e) Răspunsurile lipsă se înlocuiesc cu "nu", pentru a nu pierde date

6. (CS) Pentru o variabilă, Q1 = 20 și Q3 = 30. După regula lui Tukey folosită la box-plot, limita superioară peste care o valoare este considerată aberantă este:

a) 35

b) 40

c) 50

d) 45

e) 60

7. (CS) Graficul potrivit pentru a arăta forma distribuției glicemiei a jeun la 160 de pacienți este:

a) diagrama circulară

b) diagrama cu bare

c) diagrama de dispersie

d) graficul forest

e) histograma

8. (CS) Hemoglobina are, la femeile sănătoase dintr-o populație, distribuție normală cu media 14 g/dL și abaterea standard 1,2 g/dL. Procentul aproximativ al femeilor cu hemoglobina sub 11,6 g/dL este:

a) 0,13%

b) 2,3%

c) 5%

d) 15,9%

e) 47,7%

9. (CM) O variabilă are distribuție normală cu media 120 și abaterea standard 10. Sunt corecte afirmațiile:

a) Aproximativ 95% dintre indivizi au valori între 100,4 și 139,6

b) Aproximativ 68,3% dintre indivizi au valori între 110 și 130

c) Aproximativ 99,7% dintre indivizi au valori între 90 și 150

d) Intervalul 100,4-139,6 este intervalul de încredere de 95% pentru media populației

e) Aproximativ 2,5% dintre indivizi au valori sub 100,4

10. (CS) După analiza unei variabile pe scală logaritmică, media logaritmilor readusă pe scala inițială prin exponențiere este:

a) media aritmetică

b) mediana

c) media geometrică

d) moda

e) abaterea standard

11. (CS) Teorema limită centrală afirmă că:

a) mediile unor eșantioane suficient de mari au o distribuție aproximativ normală, oricare ar fi distribuția variabilei în populație

b) toate variabilele biologice au distribuție normală

c) mediana și media sunt întotdeauna egale în eșantioanele mari

d) abaterea standard scade când crește volumul eșantionului

e) orice test devine semnificativ dacă eșantionul depășește 30 de persoane

12. (CS) Pe 16 pacienți, media unei variabile este 50, iar abaterea standard 8. Știind că t(0,05; 15) = 2,131, intervalul de încredere de 95% pentru media populației este:

a) 46,1-53,9

b) 45,7-54,3

c) 34,0-66,0

d) 48,0-52,0

e) 42,0-58,0

13. (CS) Ce volum de eșantion este necesar pentru a estima media unei variabile cu $\sigma = 10$, cu o marjă de eroare de ±2 unități și un nivel de încredere de 95%?

a) 10

b) 49

c) 96

d) 97

e) 385

14. (CM) În mg2026, intervalul de încredere de 95% pentru media TAS este 128,96-133,30 mmHg. Sunt corecte afirmațiile:

a) Dacă studiul s-ar repeta de multe ori, aproximativ 95% dintre intervalele construite la fel ar conține media reală

b) 95% dintre pacienți au TAS între 128,96 și 133,30 mmHg

c) Intervalul de încredere de 99%, calculat pe aceleași date, este mai larg

d) La un eșantion de 4 ori mai mare, intervalul ar fi aproximativ de două ori mai îngust

e) Intervalul este centrat pe media eșantionului

15. (CS) Într-un studiu pilot, 2 din 25 de pacienți au o reacție adversă. Intervalul de încredere Wald pentru proporție are limita inferioară negativă. Atitudinea corectă este:

a) se raportează limita negativă așa cum a ieșit

b) se înlocuiește limita negativă cu 0 și se păstrează limita superioară Wald

c) se renunță la interval și se raportează doar proporția

d) se mărește artificial numărul de cazuri

e) se folosește intervalul Wilson, potrivit pentru eșantioane mici și proporții extreme

16. (CM) Despre ipotezele statistice, sunt corecte afirmațiile:

a) Pentru compararea a două medii, H0 este $\mu1 = \mu2$

b) Ipoteza alternativă bilaterală este $\mu1 \neq \mu2$

c) Când $p > 0,05$, ipoteza nulă se acceptă ca fiind demonstrată

d) Pentru un risc relativ, H0 este $RR = 1$

e) Pentru un coeficient de corelație, H0 este $\rho = 1$

17. (CS) Se fac 5 comparații pe aceleași date și se aplică corecția Bonferroni pentru un α global de 0,05. Pragul folosit pentru fiecare comparație este:

a) 0,05

b) 0,025

c) 0,01

d) 0,005

e) 0,25

18. (CS) Într-un studiu pe 20.000 de pacienți, un medicament scade TAS cu 1,2 mmHg față de placebo (IC 95% 0,8-1,6 mmHg), $p < 0,001$. Interpretarea cea mai potrivită este:

a) efectul este statistic semnificativ, dar probabil fără importanță clinică

b) efectul este important clinic, pentru că p este foarte mic

c) rezultatul nu este semnificativ statistic

d) intervalul de încredere conține 0

e) studiul a avut putere insuficientă

19. (CM) Puterea unui test statistic crește atunci când:

a) crește volumul eșantionului

b) crește mărimea efectului căutat

c) crește variabilitatea (abaterea standard) a variabilei

d) pragul α crește de la 0,01 la 0,05

e) un design pereche este înlocuit cu unul cu grupuri independente, la același număr de măsurători

20. (CS) Se compară, cu testul t clasic pentru eșantioane independente, două grupuri de 25 și 30 de pacienți. Numărul gradelor de libertate este:

a) 55

b) 54

c) 53

d) 29

e) 24

21. (CM) În mg2026, glicemia la 0 și 6 luni se compară cu testul t pereche: media diferențelor 10,19 mg/dL, abaterea standard a diferențelor 11,57, $n = 160$, $t = 11,14$, IC 95% 8,38-11,99. Sunt corecte afirmațiile:

a) Eroarea standard a mediei diferențelor este aproximativ 0,915

b) Numărul gradelor de libertate este 159

c) Intervalul de încredere nu conține 0, în acord cu $p < 0,05$

d) Testul presupune varianțe egale ale glicemiei la cele două momente

e) Mărimea efectului dz = 0,88 indică un efect mare

22. (CS) Două grupuri independente cu date normale au varianțe diferite (testul Levene $p = 0,002$). Testul recomandat pentru compararea mediilor este:

a) testul t pereche

b) testul chi-pătrat

c) testul Friedman

d) testul t Welch

e) testul McNemar

23. (CS) Într-o ANOVA cu un factor, suma pătratelor între grupuri este 300, iar suma pătratelor totală este 1200. Eta pătrat este:

a) 0,25

b) 0,33

c) 0,75

d) 4,00

e) 0,0025

24. (CS) Două grupuri au mediile 136 și 130 mmHg, iar abaterea standard comună (pooled) este 12 mmHg. Mărimea efectului Cohen d și interpretarea ei sunt:

a) 2,0, efect mare

b) 0,5, efect mic

c) 72, efect foarte mare

d) 0,05, efect neglijabil

e) 0,5, efect mediu

25. (CM) Într-un tabel 2 x 2, a = 20, b = 30, c = 10, d = 40 (N = 100). Știind că valorile critice χ^2 pentru gl = 1 sunt 3,841 ($\alpha = 0,05$) și 6,635 ($\alpha = 0,01$), sunt corecte afirmațiile:

a) χ^2 este aproximativ 4,76

b) Numărul gradelor de libertate este 1

c) $p < 0,05$

d) $p < 0,01$

e) Frecvența așteptată pentru celula a este 15

26. (CS) Două teste de screening (A și B) se aplică acelorași 100 de pacienți, iar rezultatul fiecăruia este pozitiv sau negativ. Pentru a compara proporțiile de rezultate pozitive se folosește:

a) testul chi-pătrat de independență

b) testul McNemar

c) testul Mann-Whitney

d) testul t pentru eșantioane independente

e) testul Kruskal-Wallis

27. (CS) Grupul A are valorile 12, 15, 18, iar grupul B valorile 14, 20, 22, 25. Statistica U a testului Mann-Whitney (cea mai mică dintre U1 și U2) este:

a) 2

b) 8

c) 10

d) 12

e) 20

28. (CM) Despre testele Kruskal-Wallis și Friedman, sunt corecte afirmațiile:

a) Testul Kruskal-Wallis este alternativa neparametrică a ANOVA cu un factor

b) Testul Friedman se folosește pentru măsurători repetate pe aceiași indivizi

c) Un test Kruskal-Wallis semnificativ arată exact care grupuri diferă între ele

d) Testul Kruskal-Wallis se bazează pe ranguri

e) Testul Friedman se aplică grupurilor independente

29. (CS) Pentru a studia asocierea dintre fumat (da/nu) și BPOC (da/nu) pe 300 de pacienți, cu toate frecvențele așteptate peste 5, se folosește:

a) coeficientul Pearson

b) testul t pentru eșantioane independente

c) ANOVA

d) testul chi-pătrat

e) testul Wilcoxon

30. (CM) Sunt corecte asocierile dintre situație și test:

a) trei momente de măsurare, aceiași pacienți, date normale: ANOVA cu măsurători repetate

b) două grupuri independente, variabilă ordinală: testul Mann-Whitney

c) două variabile nominale: testul chi-pătrat

d) două variabile numerice normale, relație liniară: coeficientul Spearman este obligatoriu

e) trei grupuri independente, date normale: testul Friedman

31. (CS) La 6 pacienți, suma pătratelor diferențelor de rang dintre două variabile este $\sum d^2 = 14$. Coeficientul Spearman este:

a) 0,4

b) 0,6

c) 0,8

d) 0,93

e) -0,6

32. (CM) Pe 27 de pacienți se obține r = 0,5 între două variabile. Valoarea critică t pentru gl = 25 și α = 0,05 bilateral este 2,060. Sunt corecte afirmațiile:

a) Numărul gradelor de libertate este 25

b) Statistica t este aproximativ 2,89

c) Corelația este semnificativă la pragul de 0,05

d) Coeficientul de determinare este 0,25

e) Corelația este perfectă

33. (CS) În mg2026, TAS = 108,3 + 0,406 x vârsta. TAS prezisă pentru un pacient de 60 de ani este aproximativ:

a) 132,7 mmHg

b) 24,4 mmHg

c) 108,7 mmHg

d) 168,3 mmHg

e) 141,8 mmHg

34. (CS) În regresia multiplă TAS ~ IMC + vârstă + sex + fumat din mg2026, coeficientul IMC este 1,389 ($p < 0,001$). Interpretarea corectă este:

a) IMC explică 1,389% din variabilitatea TAS

b) corelația dintre IMC și TAS este 1,389

c) TAS crește în medie cu 1,389 mmHg pentru fiecare kg/m² în plus, la vârstă, sex și status de fumător constante

d) un pacient cu IMC 0 are TAS 1,389 mmHg

e) IMC crește cu 1,389 kg/m² pentru fiecare mmHg de TAS

35. (CS) Într-un studiu caz-control, 40 din 100 de cazuri și 20 din 100 de martori au fost expuși. Raportul cotelor este aproximativ:

a) 2,0

b) 1,5

c) 0,375

d) 20

e) 2,67

36. (CM) Un studiu raportează RR = 1,8 (IC 95% 0,9-3,6). Sunt corecte afirmațiile:

a) Asocierea nu este statistic semnificativă la pragul de 0,05

b) Datele sunt compatibile cu absența efectului

c) Datele sunt compatibile și cu o dublare a riscului

d) S-a demonstrat că expunerea nu are niciun efect

e) Valoarea p este sub 0,05

37. (CS) Într-o cohortă de 2000 de persoane fără diabet urmărite 2 ani apar 50 de cazuri noi de diabet. Incidența cumulativă este:

a) 2,5% în 2 ani

b) 1,25% în 2 ani

c) 25% în 2 ani

d) 50 la 1000 de persoane-ani

e) 2,5 la 1000 de persoane-ani

38. (CS) Se aleg 120 de pacienți cu cancer pulmonar și 240 de persoane fără cancer și se compară frecvența fumatului în trecut. Tipul de studiu este:

a) studiu de cohortă prospectiv

b) studiu clinic randomizat

c) studiu ecologic

d) studiu caz-control

e) serie de cazuri

39. (CM) În mg2026 (analiză transversală la includere), 18 din 54 de obezi și 11 din 106 neobezi au diabet. Sunt corecte afirmațiile:

a) Prevalența diabetului la obezi este 33,3%

b) Raportul prevalențelor este aproximativ 3,21

c) OR este aproximativ 4,32

d) Studiul demonstrează că obezitatea a precedat apariția diabetului

e) OR este mai mare decât raportul prevalențelor pentru că boala nu este rară în acest eșantion

40. (CS) Un test are raportul de verosimilitate pozitiv RV+ = 8. Probabilitatea pre-test a bolii este 20%. Probabilitatea post-test după un rezultat pozitiv este aproximativ:

a) 20%

b) 66,7%

c) 80%

d) 88,9%

e) 160%

41. (CM) Despre medicina bazată pe dovezi, sunt corecte afirmațiile:

a) Reviziile sistematice cu meta-analiză ale studiilor randomizate de calitate se află în vârful ierarhiei dovezilor

b) O serie de cazuri poate genera ipoteze, dar nu le poate testa

c) Opinia expertului este cea mai solidă formă de dovadă

d) Medicina bazată pe dovezi integrează dovezile din cercetare cu experiența clinică și valorile pacientului

e) Un studiu observațional mare este întotdeauna superior unui studiu randomizat

42. (CS) Același test diagnostic (sensibilitate și specificitate neschimbate) este folosit într-o populație în care prevalența bolii este mai mică. Ce se întâmplă cu valorile predictive?

a) VPP crește, VPN scade

b) ambele rămân neschimbate

c) VPP scade, VPN crește

d) ambele scad

e) ambele cresc

43. (CM) Despre un test diagnostic se știe: 40 de rezultate adevărat pozitive, VPP = 0,8, 10 rezultate fals negative și VPN = 0,9. Sunt corecte afirmațiile:

a) Numărul rezultatelor fals pozitive este 10

b) Numărul rezultatelor adevărat negative este 90

c) Sensibilitatea este 80%

d) Specificitatea este 90%

e) Numărul total de persoane testate este 200

44. (CS) Doi radiologi clasifică aceleași radiografii; acordul observat este 0,85, iar acordul așteptat prin întâmplare este 0,50. Coeficientul kappa este:

a) 0,35

b) 0,50

c) 0,70

d) 0,85

e) 1,70

45. (CS) Un test cu sensibilitate foarte mare are rezultat negativ la un pacient. Interpretarea corectă (regula SnNOut) este:

a) boala este practic exclusă

b) boala este confirmată

c) rezultatul nu aduce nicio informație

d) trebuie calculată specificitatea pentru a decide

e) pacientul are boala într-un stadiu incipient

46. (CS) Pentru scorul FINDRISC din mg2026, pragul ≥ 12 are Se 75,9% și Sp 74,8%, iar pragul ≥ 16 are Se 41,4% și Sp 94,7%. Ridicarea pragului de la 12 la 16 are ca efect:

a) creșterea sensibilității și a specificității

b) scăderea ambelor

c) creșterea sensibilității și scăderea specificității

d) scăderea sensibilității și creșterea specificității

e) modificarea ariei de sub curba ROC

47. (CM) Despre testele de screening și combinarea testelor, sunt corecte afirmațiile:

a) La un test de screening se preferă o sensibilitate mare

b) Rezultatele pozitive la screening trebuie confirmate cu un test specific

c) În populația generală, VPP a unui test este de regulă mai mică decât într-o clinică de specialitate

d) Testarea în serie crește sensibilitatea

e) Testarea în paralel crește sensibilitatea

48. (CS) Calculul arată că sunt necesari 64 de pacienți pe grup. Se estimează un abandon de 20%. Câți pacienți trebuie recrutați pe grup?

a) 51

b) 77

c) 80

d) 84

e) 128

49. (CM) Despre protocolul studiului și formularele de colectare a datelor, sunt corecte afirmațiile:

a) Protocolul se scrie și se aprobă înainte de colectarea datelor

b) Regulile de validare a răspunsurilor într-un formular electronic reduc erorile de introducere

c) Un formular bine proiectat cere TAS ca text liber, cu unitatea de măsură scrisă de utilizator

d) Planul de analiză statistică se stabilește înaintea analizei datelor

e) Studiile clinice se înregistrează în registrele publice după publicarea rezultatelor

50. (CS) JASP afișează pentru un test "p <.001", iar un student scrie în referat "p = 0,000". Formularea corectă în raport este:

a) $p = 0$

b) $p = 0{,}000$

c) p este nul, deci H0 este falsă cu certitudine

d) $p = 0{,}001$

e) p < 0,001

Testul 3

1. (CS) Prima etapă a unui studiu statistic este:

a) analiza statistică a datelor

b) colectarea datelor

c) formularea întrebării de cercetare și planificarea studiului

d) curățarea fișierului de date

e) alegerea programului statistic

2. (CM) Despre erorile din cercetare, sunt corecte afirmațiile:

a) Eroarea aleatoare scade când crește volumul eșantionului

b) Eroarea sistematică (biasul) scade când crește volumul eșantionului

c) Biasul de memorie este caracteristic studiilor caz-control

d) Un eșantion format din voluntari poate introduce un bias de selecție

e) Intervalul de încredere cuantifică mărimea biasului

3. (CM) Pentru două evenimente, P(A) = 0,3, P(B) = 0,4 și P(A și B) = 0,12. Sunt corecte afirmațiile:

a) A și B sunt independente

b) $P(A \text{ sau} B) = 0{,}58$

c) $P(A \mid B) = 0{,}3$

d) A și B sunt incompatibile

e) $P(\text{non} - A) = 0{,}7$

4. (CS) Titrurile de anticorpi la trei pacienți sunt 10, 100 și 1000. Media geometrică este:

a) 370

b) 100

c) 1000

d) 30

e) 10

5. (CS) Variabila A are media 50 și abaterea standard 5; variabila B are media 200 și abaterea standard 30. Care afirmație este corectă?

a) A are variabilitatea relativă mai mare, pentru că media ei este mai mică

b) Cele două variabile au aceeași variabilitate relativă

c) Variabilitatea relativă nu se poate compara

d) B are variabilitatea absolută mai mică

e) B are variabilitatea relativă mai mare (CV 15% față de 10%)

6. (CM) Sunt corecte asocierile dintre variabilă și tipul ei:

a) Glicemia (mg/dL): numerică continuă

b) Numărul de sarcini: numerică discretă

c) Grupa sanguină: ordinală

d) Scorul de satisfacție 1-5: ordinală

e) Temperatura în grade Celsius: scală de raport

7. (CS) Indicatorul de tendință centrală cel mai puțin influențat de valorile extreme este:

a) mediana

b) media aritmetică

c) media ponderată

d) amplitudinea

e) abaterea standard

8. (CS) Pe graficul Q-Q al unei variabile, punctele se așază aproximativ pe dreaptă. Aceasta sugerează că:

a) variabila este ordinală

b) există multe valori aberante

c) distribuția este puternic asimetrică

d) distribuția este apropiată de cea normală

e) variabilele sunt corelate

9. (CM) În mg2026, IMC are media 29,04, mediana 28,25, asimetria 1,42 și testul Shapiro-Wilk $p < 0,001$. Sunt corecte afirmațiile:

a) Distribuția este asimetrică la dreapta

b) Media depășește mediana, în acord cu asimetria pozitivă

c) Testul Shapiro-Wilk indică abaterea de la normalitate

d) Intervalul $m \pm 1,96s$ conține exact 95% dintre pacienți

e) Pentru compararea IMC între grupuri pot fi preferate testele neparametrice

10. (CS) Valoarea scorului z care lasă în dreapta ei 2,5% din aria de sub curba normală standard este:

a) 1,645

b) 1,96

c) 2,33

d) 2,576

e) 3,00

11. (CS) Greutatea unui copil se află la percentila 90 pentru vârstă și sex. Aceasta înseamnă că:

a) copilul are cu 90% mai mult decât greutatea medie

b) copilul este obez

c) 10% dintre copii au greutatea mai mică decât el

d) greutatea lui este de 90 de ori mai mare decât media

e) aproximativ 90% dintre copiii de aceeași vârstă și sex au greutatea mai mică

12. (CS) Într-un eșantion de 60 de pacienți, 18 au hipertensiune necontrolată. Intervalul de încredere de 95% pentru proporție (metoda Wald) este aproximativ:

a) 18,4-41,6%

b) 24,1-35,9%

c) 28,0-32,0%

d) 10,0-50,0%

e) 0,18-0,42%

13. (CM) Volumul necesar al eșantionului pentru estimarea unui parametru crește atunci când:

a) marja de eroare dorită scade

b) nivelul de încredere crește de la 95% la 99%

c) variabilitatea variabilei (σ) este mai mare

d) proporția estimată se apropie de 0,5

e) populația din care se extrage eșantionul crește de la 1 milion la 2 milioane de persoane

14. (CS) Volumul eșantionului crește de la 40 la 160 de pacienți, iar abaterea standard rămâne aproximativ aceeași. Lățimea intervalului de încredere pentru medie devine aproximativ:

a) de patru ori mai mică

b) neschimbată

c) de două ori mai mică

d) de două ori mai mare

e) de patru ori mai mare

15. (CS) Câți pacienți sunt necesari pentru a estima o prevalență presupusă de 10%, cu o marjă de eroare de ±3% și un nivel de încredere de 95%?

a) 100

b) 196

c) 400

d) 385

e) 1000

16. (CS) Pe 64 de pacienți, media unei variabile este 128, abaterea standard 16, iar valoarea de referință este 120. Statistica t a testului t pentru un eșantion este:

a) 0,5

b) 2

c) 4

d) 8

e) 32

17. (CM) Despre valoarea p, sunt corecte afirmațiile:

a) p = 0,03 înseamnă că probabilitatea ca H0 să fie adevărată este 3%

b) Un p foarte mic poate apărea și pentru un efect clinic neglijabil, dacă eșantionul este mare

c) Un p > 0,05 nu dovedește absența efectului

d) La același efect, valoarea p depinde de volumul eșantionului

e) Valorile p = 0,049 și p = 0,051 reprezintă dovezi radical diferite

18. (CS) Un test unilateral este justificat atunci când:

a) direcția efectului a fost stabilită înaintea studiului, iar un efect în sens opus este imposibil sau lipsit de interes

b) testul bilateral a dat p = 0,08

c) cercetătorul dorește un rezultat semnificativ

d) eșantionul este mic

e) datele nu sunt normale

19. (CS) Se efectuează 20 de teste independente, iar pentru toate H0 este adevărată. La $\alpha = 0,05$, probabilitatea de a obține cel puțin un rezultat fals semnificativ este aproximativ:

a) 0,05

b) 0,20

c) 0,36

d) 0,64

e) 1,00

20. (CS) Într-o ANOVA cu trei grupuri și 60 de pacienți, media pătratelor între grupuri este 150, iar media pătratelor reziduală este 30. Statistica F este:

a) 0,2

b) 5

c) 4,5

d) 120

e) 180

21. (CM) Pentru ANOVA de la întrebarea 20, știind că F critic pentru gl (2; 57) și $\alpha = 0,05$ este aproximativ 3,16, sunt corecte afirmațiile:

a) Numărul gradelor de libertate între grupuri este 2

b) Numărul gradelor de libertate rezidual este 57

c) $p < 0,05$

d) S-a demonstrat că toate cele trei medii diferă între ele

e) Pasul următor este un test post-hoc, de exemplu Tukey

22. (CS) Testul Tukey, aplicat după o ANOVA semnificativă, are rolul de a:

a) verifica normalitatea datelor

b) verifica egalitatea varianțelor

c) înlocui ANOVA când datele sunt ordinale

d) calcula mărimea efectului globală

e) identifica perechile de grupuri care diferă, controlând eroarea de tip I

23. (CS) Greutatea a 30 de pacienți se măsoară la începutul și la sfârșitul unui program de slăbire; diferențele au distribuție normală. Testul potrivit este:

a) testul t pereche

b) testul t pentru eșantioane independente

c) testul Mann-Whitney

d) testul chi-pătrat

e) ANOVA cu un factor

24. (CM) Condițiile testului t clasic (Student) pentru eșantioane independente sunt:

a) observații independente

b) distribuție aproximativ normală în fiecare grup (sau eșantioane mari)

c) varianțe egale în cele două grupuri

d) volume egale ale celor două grupuri

e) variabilă rezultat ordinală

25. (CS) Pentru un tabel de contingență cu 2 rânduri și 3 coloane, numărul gradelor de libertate al testului chi-pătrat este:

a) 1

b) 2

c) 3

d) 5

e) 6

26. (CS) La un test Mann-Whitney, grupul 1 are 4 pacienți, grupul 2 are 5 pacienți, iar suma rangurilor din grupul 1 este 12. Valoarea $U_1 = R_1 - \frac{n_1(n_1+1)}{2}$ este:

a) 2

b) 12

c) 18

d) 20

e) 10

27. (CM) În mg2026, scăderea în greutate de cel puțin 5% a apărut la 6/53 pacienți în brațul A, 16/53 în brațul B și 16/54 în brațul C; $\chi^2 = 6,77$, gl = 2, p = 0,034, iar cea mai mică frecvență așteptată este 12,6. Sunt corecte afirmațiile:

a) Condiția privind frecvențele așteptate este îndeplinită

b) Proporția pacienților cu scădere de cel puțin 5% diferă semnificativ între brațe

c) Testul arată că brațul C diferă semnificativ de brațul B

d) Gradele de libertate se calculează ca (3 - 1) x (2 - 1)

e) Testul exact Fisher era obligatoriu

28. (CS) Glicemia a fost măsurată la aceiași pacienți la 0, 3 și 6 luni, iar distribuția este puternic asimetrică pe un eșantion mic. Testul potrivit este:

a) ANOVA cu un factor

b) testul Kruskal-Wallis

c) testul chi-pătrat

d) testul Friedman

e) testul Mann-Whitney

29. (CS) HbA1c (distribuție normală în ambele grupuri) se compară între pacienții cu și fără diabet confirmat; testul Levene dă p = 0,40. Testul potrivit este:

a) testul t pereche

b) testul Wilcoxon

c) testul t pentru eșantioane independente (varianta clasică sau Welch)

d) testul chi-pătrat

e) testul Friedman

30. (CM) Stadiul fibrozei hepatice (F0-F4) este o variabilă ordinală. Sunt metode potrivite:

a) testul Mann-Whitney pentru compararea a două grupuri

b) testul Kruskal-Wallis pentru compararea a trei grupuri

c) coeficientul Spearman pentru asocierea cu o variabilă numerică

d) testul t independent aplicat codurilor 0-4, ca și cum ar fi o variabilă continuă normală

e) testul chi-pătrat pentru tabelul stadiu x grup, acceptând că se pierde informația despre ordine

31. (CS) Într-o regresie liniară simplă, $R^2 = 0{,}25$, iar panta dreptei este negativă. Coeficientul de corelație Pearson este:

a) 0,25

b) -0,25

c) 0,5

d) -0,5

e) 0,0625

32. (CM) Despre regresia liniară, sunt corecte afirmațiile:

a) Reziduul este diferența dintre valoarea observată și valoarea prezisă

b) Metoda celor mai mici pătrate minimizează suma pătratelor reziduurilor

c) Eroarea standard a estimării se exprimă în unitățile variabilei dependente

d) Predicția pentru valori ale lui x mult în afara domeniului observat este la fel de sigură

e) Un R^2 mare dovedește o relație cauzală

33. (CS) Într-o regresie logistică, coeficientul pentru fumat este b = 0,693. Raportul cotelor corespunzător este aproximativ:

a) 0,693

b) 2,0

c) 1,693

d) 0,5

e) 6,93

34. (CS) Pe o diagramă de dispersie, un singur pacient are valori foarte mari pe ambele axe, departe de ceilalți. Acest punct:

a) nu poate influența coeficientul de corelație

b) scade întotdeauna coeficientul de corelație

c) transformă automat corelația în una neliniară

d) trebuie șters fără verificare

e) poate modifica mult coeficientul de corelație și trebuie verificat

35. (CS) Într-un studiu randomizat, evenimentul apare la 8 din 200 de pacienți tratați și la 16 din 200 de pacienți cu placebo. Riscul relativ este:

a) 0,5

b) 2,0

c) 0,04

d) 0,08

e) 0,96

36. (CM) Pentru datele de la întrebarea 35, sunt corecte afirmațiile:

a) Reducerea absolută a riscului este de 4 puncte procentuale

b) Reducerea relativă a riscului este 50%

c) NNT este 25

d) OR este exact egal cu RR

e) Raportarea doar a reducerii relative a riscului poate exagera impresia de beneficiu

37. (CS) O expunere are RR = 3, iar 25% din populație este expusă. Fracția atribuibilă în populație (formula Levin) este aproximativ:

a) 0,25

b) 0,333

c) 0,50

d) 0,667

e) 0,75

38. (CS) O boală cronică are incidență mică, dar bolnavii supraviețuiesc foarte mulți ani. În această situație:

a) prevalența este întotdeauna egală cu incidența

b) prevalența este mai mică decât incidența anuală

c) prevalența este relativ mare comparativ cu incidența anuală

d) incidența nu se poate calcula

e) letalitatea este mare

39. (CM) Despre tipurile de studii epidemiologice, sunt corecte afirmațiile:

a) Studiul de cohortă permite calculul incidenței

b) Studiul caz-control este eficient pentru bolile rare

c) Studiul caz-control permite calculul direct al riscului relativ

d) Un studiu de cohortă prospectiv poate fi costisitor și de lungă durată

e) Studiul transversal stabilește succesiunea în timp a expunerii și a bolii

40. (CS) Un test are sensibilitatea 90% și specificitatea 95%. Într-o populație cu prevalența 1%, VPP este aproximativ:

a) 15,4%

b) 90%

c) 95%

d) 1%

e) 50%

41. (CM) Despre rapoartele de verosimilitate, sunt corecte afirmațiile:

a) $RV^{-} = \frac{1-Se}{Sp}$

b) Pentru Se = 80% și Sp = 90%, RV- este aproximativ 0,22

c) Un RV egal cu 1 nu modifică probabilitatea bolii

d) Nomograma Fagan leagă probabilitatea pre-test, raportul de verosimilitate și probabilitatea post-test

e) RV+ depinde de prevalența bolii în populația testată

42. (CS) Cota (odds) unui eveniment este 3. Probabilitatea evenimentului este:

a) 0,25

b) 0,33

c) 0,75

d) 3,00

e) 0,67

43. (CS) Un test diagnostic dă: a = 80 adevărat pozitivi, b = 20 fals negativi, c = 10 fals pozitivi, d = 90 adevărat negativi. Raportul cotelor de diagnostic (DOR) este:

a) 4

b) 8

c) 9

d) 36

e) 72

44. (CM) Pentru testul de la întrebarea 43, sunt corecte afirmațiile:

a) Sensibilitatea este 80%

b) Specificitatea este 90%

c) Rata erorii de clasificare este (20 + 10)/200 = 15%

d) Rata erorii de clasificare este (20 + 90)/200

e) Acuratețea este 85%

45. (CS) În scenariul setului mg2026, standardul de referință pentru diagnosticul de diabet, față de care se evaluează scorul FINDRISC, este:

a) glicemia a jeun

b) HbA1c

c) scorul FINDRISC însuși

d) declarația pacientului

e) testul oral de toleranță la glucoză

46. (CM) În mg2026, aria de sub curba ROC a scorului FINDRISC este 0,852 (IC 95% 0,762-0,943), iar a glicemiei a jeun 0,844. Sunt corecte afirmațiile:

a) FINDRISC are o capacitate de discriminare foarte bună

b) IC al ariei nu conține 0,5, deci scorul discriminează mai bine decât întâmplarea

c) Diferența dintre 0,852 și 0,844 demonstrează superioritatea FINDRISC

d) Aria de sub curbă nu depinde de pragul ales

e) Aria este probabilitatea ca un bolnav ales la întâmplare să aibă un scor mai mare decât un nebolnav ales la întâmplare

47. (CS) Într-un studiu de acuratețe, biopsia (standardul de referință) se efectuează doar la pacienții cu test pozitiv. Eroarea introdusă se numește:

a) bias de memorie

b) bias de verificare

c) eroare de tip II

d) factor de confuzie

e) bias de publicare

48. (CS) Faza unui studiu clinic în care medicamentul este comparat randomizat, pe loturi mari, cu tratamentul standard sau cu placebo, înaintea autorizării, este:

a) faza I

b) faza II

c) faza IV

d) faza III

e) faza preclinică

49. (CM) Despre randomizare și orbire, sunt corecte afirmațiile:

a) Randomizarea în blocuri menține brațele echilibrate numeric pe parcursul recrutării

b) Ascunderea alocării împiedică cunoașterea brațului înainte de înrolarea pacientului

c) Un studiu deschis (fără orbire) nu poate fi randomizat

d) Folosirea unui placebo permite orbirea participanților

e) Analiza per protocol păstrează întotdeauna avantajele randomizării

50. (CS) Un cercetător vrea să afle dacă proporția pacienților cu scădere în greutate de cel puțin 5% diferă între cele trei brațe și apoi să cuantifice efectul brațului C față de brațul A. Succesiunea corectă de analize este:

a) ANOVA, apoi testul Tukey

b) testul t pereche, apoi corelația Pearson

c) testul chi-pătrat pe tabelul 3 x 2, apoi RR, diferența de risc și NNT cu IC 95% pentru C față de A

d) regresia liniară simplă, apoi R^2

e) testul Kruskal-Wallis, apoi testul Friedman

Răspunsuri și explicații

Verificați-vă întâi răspunsurile cu grila, fără să citiți explicațiile, și notați-vă scorul. Apoi citiți explicațiile pentru toate întrebările greșite și pentru cele la care ați ghicit. Toate calculele de mai jos au fost verificate cu un program (Python, biblioteca scipy).

Testul 1: grila de răspunsuri

1. d
2. a, c, e
3. b

4. e
5. a, b, c
6. c
7. b, c, d
8. a
9. a, b, c
10. b
11. c
12. d
13. a, b, c
14. b
15. d
16. a
17. a, b, d
18. c
19. a, c, e
20. e
21. a, b, d, e
22. c
23. b, c, d
24. b
25. a
26. c
27. a, c, d
28. d
29. a, c, d
30. b
31. b
32. a, b, d
33. e
34. a, b, c, e
35. b
36. a, b, c, e
37. a
38. c
39. a, b, d
40. d
41. a, b, c, e
42. b
43. a, b, c
44. e
45. b

46. a, b, c
47. c
48. a
49. a, c, e
50. d

Testul 1: explicații

Întrebarea 1. Toate celelalte variante descriu eșantionul pe care îl avem (statistică descriptivă). Inferența înseamnă generalizarea la populație, iar varianta d face exact asta, printr-un interval de încredere.

Întrebarea 2. Capcana este varianta d: biasul de selecție este o eroare sistematică și nu dispare la eșantioane mari; un eșantion de convenienț de 10.000 de persoane rămâne la fel de nereprezentativ. Varianta b greșește din același motiv.

Întrebarea 3. Regula adunării: $P(F \text{ sau} \text{HTA}) = 0{,}30 + 0{,}20 - 0{,}08 = 0{,}42$. Cine adună doar 0,30 + 0,20 = 0,50 îi numără de două ori pe fumătorii hipertensivi.

Întrebarea 4. Clasele NYHA au o ordine naturală (I mai ușor decât IV), dar distanțele dintre clase nu sunt egale, deci variabila este ordinală.

Întrebarea 5. Media = (4 + 7 + 7 + 9 + 13)/5 = 40/5 = 8. Valorile ordonate au în mijloc (poziția 3) valoarea 7, deci mediana este 7; moda este tot 7. Amplitudinea este 13 - 4 = 9, nu 13. Dacă 13 devine 43, media urcă la 14, dar mediana rămâne 7: mediana este robustă la valori extreme.

Întrebarea 6. $\text{CV} = \frac{100 \times 12}{80} = 15\%$. Varianta e (0,15%) este greșeala tipică de a uita înmulțirea cu 100, iar 6,7% este raportul inversat (80/12 = 6,67).

Întrebarea 7. Linia din cutie este mediana, nu media (capcana a). Box-plotul descrie variabile numerice sau ordinale; pentru variabile nominale se folosesc diagrama cu bare sau cea circulară.

Întrebarea 8. $z = \frac{260-200}{40} = 1{,}5$. Din tabelul E.1, Φ(1,5) = 0,9332, deci aria din dreapta este 1 - 0,9332 = 0,0668, adică 6,7%. Varianta b (2,3%) corespunde lui z = 2.

Întrebarea 9. Aria totală de sub orice curbă de densitate este 1, indiferent de abaterea standard (d este falsă). Un p = 0,30 la Shapiro-Wilk nu "dovedește" normalitatea, ci arată doar că abaterea de la normalitate nu a fost detectată (e este falsă).

Întrebarea 10. La o distribuție asimetrică, media și abaterea standard sunt trase de coada lungă. Mediana și intervalul intercuartilic (151,5 mg/dL; 107-203,25) descriu corect datele. O alternativă corectă, care nu apare printre variante, ar fi media geometrică după logaritmare.

Întrebarea 11. $z = \frac{160-131{,}13}{13{,}88} = \frac{28{,}87}{13{,}88} = 2{,}08$. Varianta e (28,87) este doar numărătorul.

Întrebarea 12. $\text{ES} = \frac{20}{\sqrt{100}} = 2$; $\text{IC } 95\% = 150 \pm 1{,}96 \times 2 = 150 \pm 3{,}92$, adică 146,1-153,9. Varianta a (110,8-189,2) este intervalul pentru valorile individuale ($m \pm 1{,}96s$), nu pentru medie, cea mai frecventă confuzie la examen.

Întrebarea 13. $\text{ES} = \frac{\text{AS}}{\sqrt{n}}$; la n de 4 ori mai mare, $\sqrt{n}$ se dublează și ES se înjumătățește. AS nu scade sistematic cu n: ea descrie variabilitatea biologică, care nu depinde de câți pacienți măsurăm. Intervalul $m \pm 1{,}96\text{ES}$ conține media populației cu 95% încredere, nu 95% dintre pacienți.

Întrebarea 14. $p = \frac{40}{200} = 0{,}20$; $\text{ES} = \sqrt{\frac{0{,}2 \times 0{,}8}{200}} = 0{,}0283$; $\text{IC} = 0{,}20 \pm 1{,}96 \times 0{,}0283 = 0{,}20 \pm 0{,}055$, adică 14,5-25,5%.

Întrebarea 15. Când prevalența este necunoscută, se folosește p = 0,5 (varianta cea mai acoperitoare): $n = \frac{1{,}96^2 \times 0{,}5 \times 0{,}5}{0{,}05^2} = 384{,}16$, rotunjit în sus la 385. Valoarea 246 corespunde lui p = 0,20.

Întrebarea 16. Definiția din varianta a este singura corectă. Variantele b și c sunt interpretările greșite cele mai răspândite: valoarea p se calculează presupunând H0 adevărată, deci nu poate fi probabilitatea ca H0 să fie adevărată.

Întrebarea 17. Puterea este 1 - β, nu 1 - α (c falsă). Un α mai strict (0,01) face mai greu de respins H0 și, la același n, scade puterea (e falsă).

Întrebarea 18. Este exemplul clasic de rezultat nesemnificativ din mg2026. H0 nu se "acceptă" și egalitatea nu este "demonstrată": intervalul de la -4,9 la 3,8 mmHg arată că datele sunt compatibile și cu diferențe de câțiva mmHg. Valoarea 0,81 este p, nu puterea.

Întrebarea 19. Valoarea 1,645 este critică pentru un test unilateral la α = 0,05. Pentru z = 2,30, p bilateral = 2 x (1 - 0,9893) = 0,021: mai mic decât 0,05, dar mai mare decât 0,01, deci d este falsă.

Întrebarea 20. ES al diferențelor = 8/√16 = 8/4 = 2; $t = \frac{4}{2} = 2{,}0$. Varianta a (0,5) apare dacă se împarte media la abaterea standard (4/8), adică se calculează dz, nu t.

Întrebarea 21. gl = n - 1 = 15; t critic bilateral (α = 0,05) = 2,131 (tabelul E.3). Cum 2,0 < 2,131, H0 nu se respinge (c falsă). Valoarea critică pentru α = 0,10 bilateral este 1,753 < 2,0, deci 0,05 < p < 0,10 (exact, p = 0,064). La un test unilateral prespecificat, pragul este 1,753, iar p unilateral = 0,032, deci rezultatul ar fi semnificativ (e corectă). Această variantă nu înseamnă că avem voie să trecem la test unilateral după ce am văzut datele.

Întrebarea 22. Trei teste t fără corecție cresc riscul global de eroare de tip I. Pentru trei grupuri independente cu condițiile îndeplinite se folosește ANOVA cu un factor.

Întrebarea 23. gl1 = k - 1 = 2, deci k = 3; gl2 = N - k = 157, deci N = 160. $\eta^2 = 0{,}226$ înseamnă 22,6% din variabilitate explicată de braț. ANOVA singură spune doar că cel puțin o medie diferă; perechile se identifică prin testul post-hoc (Tukey). Testul Levene p = 0,163 nu cere alternativa neparametrică.

Întrebarea 25. gl = (3 - 1) x (4 - 1) = 2 x 3 = 6. Varianta c (12) este numărul celulelor.

Întrebarea 26. $E = \frac{60 \times 50}{200} = \frac{3000}{200} = 15$.

Întrebarea 27. Testul se aplică pe frecvențe absolute; pe procente, valoarea χ^2 ar depinde de baza arbitrară 100. χ^2 arată cât de departe sunt datele de independență, dar nu sensul și nici mărimea asocierii; pentru acestea se folosesc RR, OR sau diferența proporțiilor.

Întrebarea 28. Satisfacția pe scala Likert este ordinală, iar brațele sunt grupuri independente, deci testul Mann-Whitney (în mg2026: U = 2265,5, p < 0,001).

Întrebarea 29. Ochii aceluiași pacient și rezultatele a două teste aplicate acelorași pacienți sunt observații pereche (pentru variabile dihotomice pereche se folosește testul McNemar). Varianta e descrie trei grupuri independente.

Întrebarea 30. Cu 25 de pacienți pe grup, o asimetrie puternică și Shapiro-Wilk p = 0,003, alternativa sigură pentru două grupuri independente este testul Mann-Whitney (sau, alternativ, o transformare logaritmică urmată de testul t).

Întrebarea 31. $R^2 = 0{,}6^2 = 0{,}36$.

Întrebarea 32. TAS prezisă = 89,47 + 1,435 x 30 = 89,47 + 43,05 = 132,5 mmHg. $R^2 = 0{,}405^2 = 0{,}164$. Ordonata la origine corespunde unui IMC de 0, valoare imposibilă și aflată departe de date, deci nu este TAS "reală" a vreunui pacient. Corelația nu demonstrează cauzalitatea.

Întrebarea 33. Spearman lucrează pe ranguri și măsoară o relație monotonă; este potrivit pentru variabile ordinale sau pentru relații monotone neliniare. Pentru variabile nominale se folosesc tabele de contingență, iar pentru o ecuație de predicție, regresia.

Întrebarea 34. $R^2 = 0{,}15^2 = 0{,}0225$, adică 2,25%. Cu n = 210, chiar un r de 0,15 devine semnificativ (tabelul E.6: pentru n = 100 pragul la α = 0,05 este 0,197, iar la n = 210 pragul este mai mic). După regulile empirice prezentate în curs, un |r| sub 0,25 înseamnă o asociere practic neglijabilă. Semnificația statistică nu înseamnă relevanță și nici cauzalitate.

Întrebarea 35. Riscul la expuși = 30/100 = 0,30; la neexpuși = 10/100 = 0,10; RR $= \frac{0{,}30}{0{,}10} =$ 3,0.

Întrebarea 36. OR $= \frac{30\times90}{70\times10} = \frac{2700}{700} = 3{,}86$, mai mare decât RR (d falsă: când boala nu este rară, OR depășește RR pentru un factor de risc). Diferența de risc = 0,30 - 0,10 = 0,20; NNH $= \frac{1}{0{,}20} = 5$. Fracția atribuibilă la expuși = (3 - 1)/3 = 0,667.

Întrebarea 37. În studiul caz-control proporția bolnavilor este fixată de cercetător, deci riscurile și incidențele nu se pot calcula; OR este singura măsură validă.

Întrebarea 38. RAR = 20% - 15% = 5% = 0,05; NNT $= \frac{1}{0{,}05} = 20$. Varianta 33 ar fi 1/0,03 (o eroare de scădere), iar 5 este 1/0,20.

Întrebarea 39. Un factor de confuzie este legat atât de expunere, cât și de rezultat, dar nu este o verigă intermediară (aceea se numește mediator). Randomizarea îl controlează; creșterea eșantionului nu, pentru că este o eroare sistematică.

Întrebarea 40. La 1000 de persoane: 20 bolnavi, din care 19 pozitivi (95%); 980 nebolnavi, din care 98 fals pozitivi (10%). VPP $= \frac{19}{19+98} = \frac{19}{117} = 0{,}162$, adică aproximativ 16%. Este exemplul tipic: cu o boală rară, chiar un test bun dă mai mulți fals pozitivi decât adevărat pozitivi.

Întrebarea 41. Cota pre-test = 0,10/0,90 = 0,111; cota post-test = 0,111 x 9 = 1,0; probabilitatea post-test = 1/(1 + 1) = 0,50. Un test cu Se 90% și Sp 90% are $RV^+ = \frac{0{,}90}{0{,}10} = 9$.

Întrebarea 43. Tabelul complet: a = 45, b = 5, c = 20, d = 130, N = 200. Se $= \frac{45}{50} = 90\%$; Sp $= \frac{130}{150} = 86{,}7\%$; VPP $= \frac{45}{65} = 69{,}2\%$; VPN $= \frac{130}{135} = 96{,}3\%$; acuratețea = (45 + 130)/200 = 87,5%. Rata erorii este $\frac{b+c}{N} = \frac{5+20}{200}$, nu (5 + 130)/200.

Întrebarea 44. Rata erorii de clasificare $= \frac{\mathrm{FN+FP}}{N} = \frac{5+20}{200} = \frac{25}{200} = 12{,}5\% = 1 - 0{,}875$.

Întrebarea 45. Bolnavi = 0,30 x 200 = 60; a = 0,80 x 60 = 48, b = 12. Nebolnavi = 140; d = 0,90 x 140 = 126, c = 14. VPP $= \frac{48}{48+14} = \frac{48}{62} = 77{,}4\%$. Varianta e (91,3%) este VPN $= \frac{126}{138}$.

Întrebarea 46. AUC nu este procentul pacienților clasificați corect: este probabilitatea ca un bolnav ales la întâmplare să aibă un scor mai mare decât un nebolnav ales la întâmplare. Pragul optim după Youden maximizează Se + Sp - 1, nu doar sensibilitatea.

Întrebarea 47. J = 0,85 + 0,75 - 1 = 0,60.

Întrebarea 48. Analiza după intenția de tratament păstrează grupurile așa cum le-a format randomizarea și este analiza principală în studiile de superioritate.

Întrebarea 49. Ascunderea alocării (nu se știe dinainte brațul următor) și orbirea (nu se știe tratamentul primit, pe durata studiului) sunt lucruri diferite. Criteriul principal se fixează în protocol, înainte de colectarea datelor.

Întrebarea 50. Fără IC nu putem judeca precizia estimării: un OR de 2,1 pe 80 de pacienți poate avea un interval foarte larg, compatibil atât cu un efect mic, cât și cu unul mare.

Testul 2: grila de răspunsuri

1. a
2. a, b, d
3. b
4. c
5. a, b, d
6. d
7. e
8. b
9. a, b, c, e
10. c
11. a
12. b
13. d
14. a, c, d, e
15. e
16. a, b, d
17. c
18. a
19. a, b, d
20. c
21. a, b, c, e
22. d
23. a
24. e
25. a, b, c, e
26. b
27. a
28. a, b, d
29. d
30. a, b, c
31. b
32. a, b, c, d
33. a
34. c

35. e
36. a, b, c
37. a
38. d
39. a, b, c, e
40. b
41. a, b, d
42. c
43. a, b, c, d
44. c
45. a
46. d
47. a, b, c, e
48. c
49. a, b, d
50. e

Testul 2: explicații

Întrebarea 1. O valoare care descrie populația și pe care o estimăm este un parametru (μ); media calculată pe eșantion ar fi statistica (m).

Întrebarea 2. Anonimizarea presupune că persoana nu mai poate fi reidentificată; înlocuirea numelui cu un cod, cu păstrarea cheii de legătură, este pseudonimizare (c falsă). Un set simulat cu pacienți fictivi nu conține date ale unor persoane reale, deci nu sunt date cu caracter personal (e falsă).

Întrebarea 3. Pentru evenimente independente, $P(A \text{ și } B) = 0{,}3 \times 0{,}4 = 0{,}12$. Valoarea 0,70 ar fi suma, iar 0,58 ar fi P(A sau B).

Întrebarea 4. n = 6 (par), deci mediana este media valorilor de pe pozițiile 3 și 4: (8 + 10)/2 = 9. Media este 58/6 = 9,67, capcana din varianta b.

Întrebarea 5. Răspunsuri valide = 200 - 20 = 180; procentul valid = 54/180 = 30%; procentul din total $= \frac{54}{200} = 27\%$. Diferența arată de ce raportul trebuie să precizeze baza de calcul și numărul valorilor lipsă. Valorile lipsă nu se înlocuiesc arbitrar.

Întrebarea 6. IQR = 30 - 20 = 10; limita superioară = Q3 + 1,5 x IQR = 30 + 15 = 45.

Întrebarea 7. Histograma arată forma distribuției unei variabile numerice. Diagrama cu bare și cea circulară sunt pentru variabile categoriale, diagrama de dispersie pentru două variabile numerice, iar graficul forest pentru estimări cu IC din mai multe studii.

Întrebarea 8. $z = \frac{11{,}6-14}{1{,}2} = -2{,}0$; Φ(-2) = 1 - Φ(2) = 1 - 0,9772 = 0,0228, adică 2,3%.

Întrebarea 9. m ± 1,96s = 120 ± 19,6 = 100,4-139,6 (95% dintre indivizi); $m \pm s = 110 - 130(68{,}3\%)$; $m \pm 3s = 90 - 150(99{,}7\%)$. Sub 100,4 se află 2,5% (jumătate din cele 5% rămase în afara intervalului). Intervalul 100,4-139,6 descrie indivizii, nu media populației: un IC pentru medie ar folosi ES $= \frac{s}{\sqrt{n}}$ și ar fi mult mai îngust.

Întrebarea 10. Exponențiala mediei logaritmilor este media geometrică; în mg2026, pentru trigliceride, ea este 144,0 mg/dL, față de media aritmetică de 161,18 mg/dL.

Întrebarea 11. Teorema se referă la distribuția mediilor eșantioanelor, nu la distribuția variabilei. Varianta d este falsă: abaterea standard nu scade cu n; eroarea standard scade.

Întrebarea 12. $\text{ES} = \frac{8}{\sqrt{16}} = 2$; $\text{IC}\,95\% = 50 \pm 2{,}131 \times 2 = 50 \pm 4{,}26$, adică 45,7-54,3. Varianta a (46,1-53,9) folosește z = 1,96, inadecvat la n = 16 cu σ necunoscut.

Întrebarea 13. $n = \left(\frac{1{,}96 \times 10}{2}\right)^2 = 9{,}8^2 = 96{,}04$. Volumul eșantionului se rotunjește întotdeauna în sus, deci 97; cu 96 de pacienți marja ar fi puțin peste 2 unități.

Întrebarea 14. Intervalul de încredere se referă la media populației, nu la valorile pacienților (b falsă). Un nivel de încredere mai mare cere un interval mai larg; la un n de 4 ori mai mare, ES se înjumătățește.

Întrebarea 15. $p = \frac{2}{25} = 0{,}08$; $\text{ES} = \sqrt{\frac{0{,}08 \times 0{,}92}{25}} = 0{,}054$; limita inferioară Wald = 0,08 - 1,96 x 0,054 = -0,026, imposibilă pentru o proporție. Intervalul Wilson rezolvă problema și rămâne între 0 și 1.

Întrebarea 16. H0 nu se "acceptă" și nu se "demonstrează" (c falsă). Pentru corelație, H0 este ρ = 0 (absența corelației), nu ρ = 1.

Întrebarea 17. α' = 0,05/5 = 0,01.

Întrebarea 18. Cu 20.000 de pacienți, chiar o scădere de 1,2 mmHg devine extrem de semnificativă statistic. Limita superioară a IC (1,6 mmHg) arată însă că efectul este, cel mult, foarte mic din punct de vedere clinic.

Întrebarea 19. Puterea crește cu n, cu mărimea efectului și cu α, și scade cu variabilitatea. Designul pereche elimină variabilitatea dintre indivizi; renunțarea la el, la același număr de măsurători, scade de regulă puterea.

Întrebarea 20. gl = n1 + n2 - 2 = 25 + 30 - 2 = 53.

Întrebarea 21. $\text{ES} = \frac{11{,}57}{\sqrt{160}} = \frac{11{,}57}{12{,}649} = 0{,}915$; $t = \frac{10{,}19}{0{,}915} = 11{,}14$; gl = 159; $\text{dz} = \frac{10{,}19}{11{,}57} = 0{,}88$ (efect mare după reperele lui Cohen). Testul t pereche nu cere varianțe egale la cele două momente: el lucrează pe o singură serie, cea a diferențelor, a cărei normalitate trebuie verificată (în mg2026, Shapiro-Wilk p = 0,725).

Întrebarea 22. La varianțe inegale se folosește testul t Welch, care ajustează gradele de libertate.

Întrebarea 23. $\eta^2 = \frac{\text{SSB}}{\text{SST}} = \frac{300}{1200} = 0{,}25$.

Întrebarea 24. $d = \frac{136-130}{12} = 0{,}5$, efect mediu după reperele lui Cohen. Varianta b are cifra corectă, dar interpretarea greșită.

Întrebarea 25. $\chi^2 = \frac{N(\text{ad}-\text{bc})^2}{(a+b)(c+d)(a+c)(b+d)} = \frac{100 \times (800-300)^2}{50 \times 50 \times 30 \times 70} = 25.000.000/5.250.000 = 4{,}76$, cu gl = 1. Cum 3,841 < 4,76 < 6,635, rezultă 0,01 < p < 0,05 (exact, p = 0,029). Frecvența așteptată pentru $a = \frac{(a+b)(a+c)}{N} = \frac{50 \times 30}{100} = 15$.

Întrebarea 26. Rezultatele celor două teste provin de la aceiași pacienți (date pereche dihotomice), deci testul McNemar. Testul chi-pătrat de independență ar presupune grupuri independente.

Întrebarea 27. Ordonăm toate cele 7 valori: 12 (A, rangul 1), 14 (B, 2), 15 (A, 3), 18 (A, 4), 20 (B, 5), 22 (B, 6), 25 (B, 7). $R(A) = 1 + 3 + 4 = 8$; $U(A) = 8 - \frac{3 \times 4}{2} = 2$; $U(B) = 3 \times 4 -$

$2 = 10$. Cea mai mică valoare este U = 2 (p exact = 0,23, nesemnificativ la eșantioane atât de mici).

Întrebarea 28. Un test Kruskal-Wallis semnificativ spune doar că cel puțin un grup diferă; perechile se identifică prin teste post-hoc (Dunn). Testul Friedman este pentru măsurători repetate, nu pentru grupuri independente.

Întrebarea 29. Ambele variabile sunt dihotomice, iar frecvențele așteptate sunt peste 5: testul chi-pătrat pe tabelul 2 x 2 (urmat de calculul OR sau RR cu IC).

Întrebarea 30. Pentru două variabile numerice normale cu relație liniară, coeficientul Pearson este alegerea potrivită; Spearman nu este "obligatoriu". Pentru trei grupuri independente cu date normale se folosește ANOVA cu un factor, nu Friedman.

Întrebarea 31. rho $= 1 - \frac{6\times14}{6\times(36-1)} = 1 - \frac{84}{210} = 1 - 0{,}4 = 0{,}6$.

Întrebarea 32. gl = 27 - 2 = 25; $t = \frac{0{,}5\times\sqrt{25}}{\sqrt{1-0{,}25}} = \frac{2{,}5}{0{,}866} = 2{,}89 > 2{,}060$, deci p < 0,05 (exact, p = 0,008). $R^2 = 0{,}25$. Tabelul E.6 confirmă: pentru n = 27, r critic la 0,05 este sub 0,5.

Întrebarea 33. 108,3 + 0,406 x 60 = 108,3 + 24,36 = 132,7 mmHg. Varianta b (24,4) este doar produsul.

Întrebarea 34. Într-un model multiplu, coeficientul unui predictor este efectul lui ajustat pentru ceilalți predictori: +1,389 mmHg pentru fiecare kg/m², la vârstă, sex și status de fumător egale.

Întrebarea 35. Cota expunerii la cazuri = 40/60; la martori = 20/80; OR $= \frac{40\times80}{60\times20} = \frac{3200}{1200} = 2{,}67$. Varianta a (2,0) este raportul proporțiilor de expuși (40%/20%), care nu are sens de risc într-un studiu caz-control.

Întrebarea 36. IC conține 1, deci p > 0,05 și asocierea nu este semnificativă. Dar IC se întinde până la 3,6: datele nu exclud un efect important. Rezultatul nesemnificativ nu demonstrează absența efectului (d falsă).

Întrebarea 37. Incidența cumulativă = 50/2000 = 0,025 = 2,5% în 2 ani. Densitatea de incidență ar cere persoane-timp (aproximativ 50/4000 = 12,5 la 1000 de persoane-ani, dacă toți au fost urmăriți 2 ani); variantele d și e nu corespund niciunui calcul corect.

Întrebarea 38. Selecția pornește de la boală (cazuri și martori), iar expunerea se caută retrospectiv: studiu caz-control.

Întrebarea 39. Prevalențe: 18/54 = 33,3% și 11/106 = 10,4%; RP = 3,21; OR $= \frac{18\times95}{36\times11} = \frac{1710}{396} = 4{,}32$. Un studiu transversal măsoară expunerea și boala simultan, deci nu poate stabili succesiunea lor (d falsă). Diabetul nu este rar în eșantion (18,1%), de aceea OR se îndepărtează de RP.

Întrebarea 40. Cota pre-test = 0,2/0,8 = 0,25; cota post-test = 0,25 x 8 = 2,0; probabilitatea post-test = 2/3 = 66,7%. Varianta e (160%) apare când se înmulțește direct probabilitatea cu RV, o eroare de neiertat: probabilitățile nu pot depăși 100%.

Întrebarea 41. Opinia expertului se află la baza ierarhiei dovezilor, iar un studiu observațional, oricât de mare, nu elimină factorii de confuzie așa cum o face randomizarea.

Întrebarea 42. La Se și Sp constante, VPP scade și VPN crește când prevalența scade. În mg2026, pentru FINDRISC ≥ 12, VPP ar fi 5,8% la prevalență 2% și 75,1% la prevalență 50%.

Întrebarea 43. $a + c = \frac{a}{\text{VPP}} = \frac{40}{0{,}8} = 50$, deci c = 10. $b + d = \frac{b}{1-\text{VPN}} = \frac{10}{0{,}1} = 100$, deci d = 90. Se $= \frac{40}{40+10} = 80\%$; Sp $= \frac{90}{10+90} = 90\%$. N = 40 + 10 + 10 + 90 = 150, nu 200.

Întrebarea 44. κ = (0,85 - 0,50)/(1 - 0,50) = 0,35/0,50 = 0,70 (acord substanțial). Varianta a (0,35) este doar numărătorul.

Întrebarea 45. SnNOut: un test foarte sensibil ratează puțini bolnavi, deci un rezultat negativ exclude practic boala.

Întrebarea 46. Un prag mai ridicat face testul mai "sever": mai puțini pozitivi, deci sensibilitate mai mică și specificitate mai mare. Aria de sub curba ROC caracterizează testul pe toate pragurile și nu se schimbă la mutarea pragului.

Întrebarea 47. Testarea în serie (pozitiv doar dacă ambele teste sunt pozitive) crește specificitatea și scade sensibilitatea (d falsă); testarea în paralel face invers.

Întrebarea 48. n ajustat $= \frac{64}{1-0{,}20} = \frac{64}{0{,}8} = 80$. Varianta b (77) provine din calculul greșit 64 x 1,2 = 76,8: adăugarea a 20% la 64 nu compensează o pierdere de 20% din numărul recrutat.

Întrebarea 49. Variabilele numerice se colectează ca numere, cu unitatea fixată în formular și cu domeniu de validare; textul liber cu unitate ("87 mg/dl") este exact eroarea din fișierul brut mg2026. Studiile clinice se înregistrează înainte de înrolarea primului pacient.

Întrebarea 50. O valoare p nu este niciodată exact 0; programul afișează "<.001" pentru că valoarea este foarte mică. Se raportează p < 0,001.

Testul 3: grila de răspunsuri

1. c
2. a, c, d
3. a, b, c, e
4. b
5. e
6. a, b, d
7. a
8. d
9. a, b, c, e
10. b
11. e
12. a
13. a, b, c, d
14. c
15. d
16. c
17. b, c, d
18. a
19. d
20. b
21. a, b, c, e
22. e

23. a
24. a, b, c
25. b
26. a
27. a, b, d
28. d
29. c
30. a, b, c, e
31. d
32. a, b, c
33. b
34. e
35. a
36. a, b, c, e
37. b
38. c
39. a, b, d
40. a
41. a, b, c, d
42. c
43. d
44. a, b, c, e
45. e
46. a, b, d, e
47. b
48. d
49. a, b, d
50. c

Testul 3: explicații

Întrebarea 2. Biasul nu scade cu volumul eșantionului (b falsă), iar intervalul de încredere cuantifică doar eroarea aleatoare, nu biasul (e falsă). Un IC îngust dintr-un studiu cu bias este o estimare precisă a unei valori greșite.

Întrebarea 3. $P(A) \times P(B) = 0{,}3 \times 0{,}4 = 0{,}12 = P(A \text{ și } B)$, deci A și B sunt independente; $P(A \mid B) = \frac{0{,}12}{0{,}4} = 0{,}3 = P(A)$, aceeași concluzie. $P(A \text{ sau} B) = 0{,}3 + 0{,}4 - 0{,}12 = 0{,}58$. Evenimentele nu sunt incompatibile, pentru că P(A și B) ≠ 0. $P(\text{non} - A) = 1 - 0{,}3 = 0{,}7$.

Întrebarea 4. Media geometrică = (10 x 100 x 1000)^(1/3) = (1.000.000)^(1/3) = 100. Media aritmetică (370) este dominată de valoarea cea mai mare.

Întrebarea 5. $\text{CV}(A) = \frac{5}{50} = 10\%$; $\text{CV}(B) = \frac{30}{200} = 15\%$. B are atât variabilitatea absolută (s), cât și cea relativă mai mari.

Întrebarea 6. Grupa sanguină este nominală (categoriile nu au ordine). Temperatura în grade Celsius este pe scală de interval: 0 °C nu înseamnă absența temperaturii, iar 20 °C nu este "de două ori mai cald" decât 10 °C.

Întrebarea 9. Asimetria pozitivă (1,42) trage media (29,04) peste mediană (28,25). La o distribuție nenormală, intervalul $m \pm 1{,}96s$ nu conține exact 95% dintre pacienți.

Întrebarea 10. Aria din dreapta lui 1,96 este 2,5% (Φ(1,96) = 0,9750); 1,645 lasă 5% în dreapta, iar 2,576 lasă 0,5%.

Întrebarea 11. Percentila 90 înseamnă că 90% dintre copiii de referință au valori mai mici; nu este un diagnostic de obezitate și nu exprimă un procent din medie.

Întrebarea 12. $p = \frac{18}{60} = 0{,}30$; $\text{ES} = \sqrt{\frac{0{,}3 \times 0{,}7}{60}} = 0{,}0592$; $\text{IC} = 0{,}30 \pm 1{,}96 \times 0{,}0592 = 0{,}30 \pm 0{,}116$, adică 18,4-41,6%. Varianta e confundă proporția cu procentul.

Întrebarea 13. $n = \frac{z^2 \times p_i(1-p_i)}{E^2}$ crește când E scade, când z crește (99% față de 95%), iar pentru medii $n = \left(\frac{z\sigma}{E}\right)^2$ crește cu σ. Produsul $p_i(1 - p_i)$ este maxim la $p_i = 0{,}5$. Mărimea populației contează doar când eșantionul reprezintă o fracțiune mare din ea; între 1 și 2 milioane nu schimbă practic nimic.

Întrebarea 14. Lățimea IC este proporțională cu $\frac{1}{\sqrt{n}}$: √(40/160) = √0,25 = 0,5, deci intervalul devine de aproximativ două ori mai îngust.

Întrebarea 15. $n = \frac{1{,}96^2 \times 0{,}10 \times 0{,}90}{0{,}03^2} = \frac{3{,}8416 \times 0{,}09}{0{,}0009} = 384{,}16$, deci 385.

Întrebarea 16. $\text{ES} = \frac{16}{\sqrt{64}} = 2$; $t = \frac{128-120}{2} = 4$, cu gl = 63 ($p < 0{,}001$).

Întrebarea 17. Valoarea p nu este probabilitatea ca H0 să fie adevărată (a falsă). Valorile 0,049 și 0,051 aduc practic aceeași dovadă; pragul 0,05 este o convenție, nu o graniță naturală (e falsă).

Întrebarea 18. Alegerea unui test unilateral după ce s-au văzut datele (de exemplu pentru a "salva" un p bilateral de 0,08) este o practică incorectă.

Întrebarea 19. Probabilitatea ca niciun test să nu fie fals semnificativ = 0,95^20 = 0,358; probabilitatea de cel puțin un rezultat fals pozitiv = 1 - 0,358 = 0,64. De aici nevoia corecțiilor pentru comparații multiple.

Întrebarea 20. $F = \frac{\text{MSB}}{\text{MSW}} = \frac{150}{30} = 5$.

Întrebarea 21. gl între grupuri = k - 1 = 2; gl rezidual = N - k = 60 - 3 = 57. Cum 5 > 3,16, $p < 0{,}05$ (exact, p = 0,010). ANOVA semnificativă arată doar că cel puțin o medie diferă; nu că toate diferă între ele.

Întrebarea 24. Testul t nu cere volume egale ale grupurilor; cere observații independente, normalitate (sau eșantioane mari) și, în varianta clasică, varianțe egale.

Întrebarea 25. gl = (2 - 1) x (3 - 1) = 2.

Întrebarea 26. $U_1 = 12 - \frac{4 \times 5}{2} = 12 - 10 = 2$; U2 = n1 x n2 - U1 = 20 - 2 = 18.

Întrebarea 27. Frecvența așteptată minimă (12,6) depășește 5, deci testul chi-pătrat este potrivit, iar Fisher nu este necesar. Testul global pe tabelul 3 x 2 arată doar că proporțiile nu sunt toate egale; în plus, brațele B și C au proporții aproape identice (16/53 și 16/54), deci varianta c este falsă.

Întrebarea 28. Măsurători repetate (aceiași pacienți, trei momente) și date nenormale pe un eșantion mic: testul Friedman. Kruskal-Wallis ar fi potrivit pentru trei grupuri independente.

Întrebarea 29. Două grupuri independente, date normale, variațe omogene (Levene p = 0,40): testul t pentru eșantioane independente; varianta Welch este și ea acceptabilă.

Întrebarea 30. Pentru o variabilă ordinală se folosesc teste pe ranguri sau, pierzând informația despre ordine, testul chi-pătrat. Tratarea codurilor 0-4 ca valori continue normale presupune distanțe egale între stadii, ceea ce nu este justificat.

Întrebarea 31. |r| = √0,25 = 0,5; semnul lui r este semnul pantei, deci r = -0,5.

Întrebarea 32. Extrapolarea în afara domeniului observat este riscantă, iar R^2 mare arată o asociere puternică, nu cauzalitate.

Întrebarea 33. $\text{OR} = e^{0,693} = 2,0$. Coeficientul b este pe scala logaritmului cotei și trebuie exponențiat.

Întrebarea 34. O valoare extremă pe ambele axe poate crea sau umfla o corelație. Punctul se verifică în documentul-sursă și se analizează cu și fără el, dar nu se șterge doar pentru că "strică" rezultatul.

Întrebarea 35. Risc tratați = 8/200 = 0,04; risc placebo = 16/200 = 0,08; $\text{RR} = \frac{0,04}{0,08} = 0,5$.

Întrebarea 36. RAR = 0,08 - 0,04 = 0,04 (4 puncte procentuale); $\text{RRR} = \frac{0,04}{0,08} = 50\%$; $\text{NNT} = \frac{1}{0,04} = 25$. $\text{OR} = \frac{8\times184}{192\times16} = \frac{1472}{3072} = 0,48$, apropiat de RR (evenimentul este rar), dar nu egal. Un "risc redus la jumătate" sună spectaculos, deși înseamnă doar 4 evenimente evitate la 100 de pacienți tratați.

Întrebarea 37. $\text{FAP} = \frac{0,25\times(3-1)}{1+0,25\times(3-1)} = \frac{0,5}{1,5} = 0,333$.

Întrebarea 38. Prevalența este aproximativ incidența înmulțită cu durata bolii; o durată lungă face ca prevalența să fie mare chiar la o incidență mică.

Întrebarea 39. În studiul caz-control nu se pot calcula riscuri, deci nici RR direct (c falsă). Studiul transversal nu stabilește succesiunea în timp (e falsă).

Întrebarea 40. La 10.000 de persoane: 100 de bolnavi, 90 pozitivi; 9900 de nebolnavi, 495 fals pozitivi (5%). $\text{VPP} = \frac{90}{90+495} = \frac{90}{585} = 15,4\%$. La prevalența 20%, același test ar avea VPP de aproximativ 82%.

Întrebarea 41. $\text{RV}^- = \frac{1-0,8}{0,9} = \frac{0,2}{0,9} = 0,22$. Rapoartele de verosimilitate se calculează doar din Se și Sp, deci nu depind de prevalență (e falsă); de aceea sunt mai ușor de transferat de la un studiu la altul decât valorile predictive.

Întrebarea 42. p = cota/(1 + cota) = 3/4 = 0,75. Varianta b (0,33) este inversul cotei.

Întrebarea 43. $\text{DOR} = \frac{a\times d}{b\times c} = \frac{80\times90}{20\times10} = \frac{7200}{200} = 36$. Verificare: $\text{RV}^+ = \frac{0,8}{0,1} = 8$, $\text{RV}^- = \frac{0,2}{0,9} = 0,222$, iar 8/0,222 = 36.

Întrebarea 44. $\text{Se} = \frac{80}{100} = 80\%$; $\text{Sp} = \frac{90}{100} = 90\%$; acuratețea = (80 + 90)/200 = 85%; rata erorii = $\frac{b+c}{N} = \frac{20+10}{200} = 15\% = 1 - 0,85$. Formula (b + d)/N, prezentă în unele materiale, este greșită.

Întrebarea 45. În scenariul mg2026, diabetul a fost confirmat prin testul oral de toleranță la glucoză (TTGO), standardul de referință față de care se calculează Se și Sp ale scorului FINDRISC.

Întrebarea 46. O arie între 0,81 și 0,90 indică o discriminare foarte bună. IC 0,762-0,943 exclude 0,5. Diferența de 0,008 dintre cele două arii este mult mai mică decât incertitudinea fiecărei estimări; superioritatea s-ar putea afirma doar după un test formal de comparare a ariilor (c falsă).

Întrebarea 47. Când standardul de referință se aplică selectiv, în funcție de rezultatul testului studiat, apare biasul de verificare, care de obicei supraestimează sensibilitatea.

Întrebarea 48. Faza III compară randomizat, pe loturi mari, noul tratament cu standardul sau cu placebo, înaintea autorizării. Faza IV are loc după autorizare.

Întrebarea 49. Multe studii randomizate sunt deschise (de exemplu, cele cu intervenții educaționale, ca brațele din mg2026); randomizarea nu depinde de orbire. Analiza per protocol exclude pacienții care nu au respectat protocolul și poate distruge echilibrul creat de randomizare.

Întrebarea 50. Rezultatul este dihotomic, iar grupurile sunt trei, deci testul chi-pătrat pe tabelul 3 x 2 ($\chi^2 = 6,77$, $p = 0,034$ în mg2026). Pentru comparația clinică C față de A se raportează RR = 2,62 (IC 95% 1,11-6,17), diferența de risc de 18,3 puncte procentuale (IC 95% 3,4-33,2) și NNT = 5,5, adică 6.

Bibliografie

Referințele sunt numerotate în ordinea primei apariții în text (stil Vancouver). Fiecare articol a fost verificat în PubMed sau CrossRef înainte de a fi citat; paginile oficiale au fost accesate la data indicată.

1. Muntean C. Informatică medicală: metode, sisteme și practică. Orizont 2026. Curs și lucrări practice pentru studenții Facultății de Medicină. Timișoara: Editura Victor Babeș; 2026. ISBN 978-606-786-641-4.
2. Lupușoru R, Sporea I, Rațiu I, Lungeanu D, Popescu A, Dănilă M, et al. Contrast-Enhanced Ultrasonography with Arrival Time Parametric Imaging as a Non-Invasive Diagnostic Tool for Liver Cirrhosis. Diagnostics (Basel). 2022;12(12):3013. doi:10.3390/diagnostics12123013 PMID: 36553020
3. Moga TV, Sporea I, Lupușoru R, Popescu A, Popa A, Bota S, et al. Performance of a Noninvasive Time-Harmonic Elastography Technique for Liver Fibrosis Evaluation Using Vibration Controlled Transient Elastography as Reference Method. Diagnostics (Basel). 2020;10(9):653. doi:10.3390/diagnostics10090653 PMID: 32878078
4. Moleriu RD, Piroș T, Moleriu LC, Muntean C, Lupușoru R, Dicu AM, et al. Convenience-Oriented Dietary Behavioral Patterns Across BMI Classes in University Students: Associations with Overweight and Obesity Risk During the Transition to University Life. Nutrients. 2026;18(14):2368. doi:10.3390/nu18142368 PMID: 42514438
5. Muntean C, Piroș T, Marin RC, Moleriu LC, Lupușoru R, Dicu AM, et al. Chrononutrition Behaviors and BMI Change During the Transition to University Life in Young Adults: A Prospective Study. Nutrients. 2026;18(12):1975. doi:10.3390/nu18121975 PMID: 42356360
6. Piroș T, Lupusoru R, Moleriu LC, Muntean C, Moleriu RD, Cîmpian DM, et al. Dietary Patterns, Cooking Methods, and Their Association with Prediabetes Risk Markers in Romanian University Students: A Cross-Sectional Analysis. Nutrients. 2026;18(6):977. doi:10.3390/nu18060977 PMID: 41901152
7. Frandes M, Gherbon A, Timar B, Muntean C. Digital Engagement in Diabetes Care: A Multi-Domain Analysis of Psychosocial and Clinical Determinants. Healthcare (Basel). 2026;14(6):800. doi:10.3390/healthcare14060800 PMID: 41897253
8. Cioca F, Timar R, Ignuta F, Vlad A, Bratosin F, Rosca O, et al. Comparative Analysis of COVID-19 Outcomes in Type 1 and Type 2 Diabetes: A Three-Year Retrospective Study. Medicina (Kaunas). 2024;60(2):210. doi:10.3390/medicina60020210 PMID: 38399498
9. Popa DI, Bălulescu LA, Mederle OA, Levai CM, Buleu T, Tudor A, et al. Sex-Related Differences in Clinical Characteristics and Outcomes Among Patients Clinically Managed for Suspected Bacterial Infection in the Emergency Department. Medicina (Kaunas). 2026;62(9):1776. doi:10.3390/medicina62091776 PMID: 42796382
10. Marin RC, Păunescu H, Muntean C, Balaban DV, Coman LI, Radu N, et al. Assessing Acceptance and Intention to Use of Home Telemonitoring Among Patients with Liver Cirrhosis: Development and Initial Psychometric Validation of a TAM-Based Questionnaire. Diagnostics (Basel). 2026;16(16):2635. doi:10.3390/diagnostics16162635 PMID: 42651036
11. Cirstea N, Tit DM, Toma MM, Purza AL, Radu A, Bungau GS, et al. Patient-Reported Experiences in Chronic Dermatological Conditions: Validation of the Romanian PSPSQ 2.0 Within Contemporary Dermatologic Care Pathways. Diagnostics (Basel). 2026;16(13):2112. doi:10.3390/diagnostics16132112 PMID: 42449893
12. Focșa M, Tudor A, Cătuș C, Muntean C, Mada L. Informatică medicală: îndrumător de lucrări practice. Timișoara: Editura Victor Babeș; 2010.

13. Vernic C, Mada L, Lungeanu D, Muntean C, Apostol AS, Cătuș CO. Aplicații practice de informatică medicală și biostatistică. Timișoara: Editura Victor Babeș; 2010.

14. Vernic C, Frandeș M, Tudor A, Apostol AS, Voicu A, Șerban C, et al. Prelucrarea datelor biomedicale: aplicații practice. Timișoara: Eurostampa; 2018.

15. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996;312(7023):71-2. doi:10.1136/bmj.312.7023.71 PMID: 8555924

16. Lungeanu D, Petrica A, Lupusoru R, Marza AM, Mederle OA, Timar B. Beyond the Digital Competencies of Medical Students: Concerns over Integrating Data Science Basics into the Medical Curriculum. Int J Environ Res Public Health. 2022;19(23):15958. doi:10.3390/ijerph192315958 PMID: 36498065

17. Lungeanu D, Dragan I, Muntean C, Timar B. Harnessing Information Technology while Teaching in Distress: the Pandemic Toll vs. the Opportunity of Long-Awaited Paradigm Change. In: 2020 International Conference on e-Health and Bioengineering (EHB); Iași, Romania. IEEE; 2020. p. 1-4. doi:10.1109/ehb50910.2020.9280122

18. Muntean C. Metode computaționale în cercetarea medicală, stomatologică și farmaceutică: ghid metodologic pentru studii doctorale și cercetare aplicată. Timișoara: Editura Victor Babeș; 2026. ISBN 978-606-786-644-5.

19. Mihalaș GI, Lungeanu D. Biostatistică. Timișoara: Editura Victor Babeș; 2011. ISBN 978-606-8054-35-3.

20. Mischie A, Jinga V, Nodiți-Cuc AR, Popovici RA, Marian D, Strete GE, et al. Self-Reported Disordered Eating and Weight-Control Behaviours Across BMI Categories in Romanian Young Adults: A Cross-Sectional Online Convenience Survey. Nutrients. 2026;18(16):2677. doi:10.3390/nu18162677 PMID: 42654257

21. Vesa CM, Tit DM, Babes EE, Bungau G, Radu AF, Moleriu RD. Use of Botanical Supplements Among Romanian Individuals with Diabetes: Results from an Online Study on Prevalence, Practices, and Glycemic Control. Nutrients. 2025;17(15):2440. doi:10.3390/nu17152440 PMID: 40806023

22. Popa MD, Sharma A, Kundnani NR, Gag OL, Rosca CI, Mocanu V, et al. Identification of Heavy Tobacco Smoking Predictors-Influence of Marijuana Consuming Peers and Truancy among College Students. Healthcare (Basel). 2021;9(12):1666. doi:10.3390/healthcare9121666 PMID: 34946393

23. Christodorescu RM, Muntean C, Andor M, Lighezan D, Pop Moldovan A, Blajevschi A, et al. Hemoglobin-to-Red Cell Distribution Width Ratio as a Prognostic Marker in Decompensated Heart Failure Patients: A Prospective Observational Study. Life (Basel). 2026;16(4):551. doi:10.3390/life16040551 PMID: 42073361

24. Schiavenato M, Chu F. PICO: What it is and what it is not. Nurse Educ Pract. 2021;56:103194. doi:10.1016/j.nepr.2021.103194

25. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453-7. doi:10.1016/S0140-6736(07)61602-X PMID: 18064739

26. Cordoș AA, Ștefănigă SA, Muntean C, Vernic CV, Bolboacă SD. ALSATION Study Protocol: Romanian Translation of Three Health Literacy Surveys. Stud Health Technol Inform. 2024;321:12-16. doi:10.3233/SHTI241052 PMID: 39575771

27. Jianu N, Buda VO, Căpățână D, Muntean C, Onea TN, Jivulescu MA, et al. Osteoporosis: a problem still faulty addressed by the Romanian healthcare system. Results of a questionnaire survey of people aged 40 years and over. Front Med (Lausanne). 2024;11:1485382. doi:10.3389/fmed.2024.1485382 PMID: 39507714

28. Shortliffe EH, Chiang MF. Biomedical Data: Their Acquisition, Storage, and Use. In: Shortliffe EH, Cimino JJ, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. 5th ed. Cham: Springer; 2021. p. 45-75. doi:10.1007/978-3-030-58721-5_2

29. Melton GB, McDonald CJ, Tang PC, Hripcsak G. Electronic Health Records. In: Shortliffe EH, Cimino JJ, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. 5th ed. Cham: Springer; 2021. p. 467-509. doi:10.1007/978-3-030-58721-5_14

30. World Health Organization. International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10), version 2019 [Internet]. Geneva: WHO; 2019 [cited 2026 Oct 8]. Available from: https://icd.who.int/browse10/2019/en

31. Regenstrief Institute. LOINC Users Guide: 2 Major Parts of a LOINC term [Internet]. Indianapolis (IN): Regenstrief Institute; 2026 [cited 2026 Oct 8]. Available from: https://loinc.org/kb/users-guide/major-parts-of-a-loinc-term/

32. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union. 2016 May 4;L 119:1-88. Available from: https://eur-lex.europa.eu/eli/reg/2016/679/oj

33. Legea nr. 190/2018 privind măsuri de punere în aplicare a Regulamentului (UE) 2016/679. Monitorul Oficial al României, Partea I. 2018 Jul 18;nr. 651. Available from: https://legislatie.just.ro/Public/DetaliiDocument/203151

34. Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847. Official Journal of the European Union. 2025 Mar 5;L 2025/327. Available from: https://eur-lex.europa.eu/eli/reg/2025/327/oj

35. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human participants. JAMA. 2025;333(1):71-4. doi:10.1001/jama.2024.21972 PMID: 39425955

36. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals [Internet]. ICMJE; 2026 [cited 2026 Oct 8]. Available from: https://www.icmje.org/recommendations/

37. Goodman KW, Miller RA. Ethics in Biomedical and Health Informatics: Users, Standards, and Outcomes. In: Shortliffe EH, Cimino JJ, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. 5th ed. Cham: Springer; 2021. p. 391-423. doi:10.1007/978-3-030-58721-5_12

38. Baneu P, Prelipcean A, Buda VO, Jianu N, Tudor A, Andor M, et al. Under-Prescription of Drugs in the Elderly Population of Western Romania: An Analysis Based on STOPP/START Version 2 Criteria. J Clin Med. 2024;13(19):5970. doi:10.3390/jcm13195970 PMID: 39408034

39. Marc L, Mihaescu A, Lupusoru R, Schiller O, Bob F, Chisavu L, et al. Hepatitis C and hepatitis B virus infection in hemodialysis patients after nationwide direct antiviral agents therapy-experience of 10 Romanian HD centers. Int Urol Nephrol. 2023;55(11):2951-8. doi:10.1007/s11255-023-03587-0 PMID: 37024632

40. Gal-Nădășan EG, Popescu IM, Bădițoiu LM, Gal-Nădășan N, Cioca F, Popovici ED, et al. Healthcare Workers' Vulnerability to SARS-CoV-2 in Western Romania: A Study on Incidence and Risk Factors for Non-Vaccination and Reinfection. Int J Gen Med. 2023;16:5621-32. doi:10.2147/IJGM.S442098 PMID: 38045906

41. Mihalaș GI, Lungeanu D. Informatică medicală și biostatistică. Timișoara: Editura Victor Babeș; 2009.

42. Kirkwood BR, Sterne JAC. Essential Medical Statistics. 2nd ed. Oxford: Blackwell Science; 2003.

43. Google. Google Sheets function list [Internet]. Mountain View (CA): Google; 2026 [cited 2026 Oct 8]. Available from: https://support.google.com/docs/table/25273?hl=en

44. Google. Create and use pivot tables. Google Docs Editors Help [Internet]. Mountain View (CA): Google; 2026 [cited 2026 Oct 8]. Available from: https://support.google.com/docs/answer/1272900

45. Google. How to set rules for your form. Google Docs Editors Help [Internet]. Mountain View (CA): Google; 2026 [cited 2026 Oct 8]. Available from: https://support.google.com/docs/answer/3378864

46. Google. Choose where to save form responses. Google Docs Editors Help [Internet]. Mountain View (CA): Google; 2026 [cited 2026 Oct 8]. Available from: https://support.google.com/docs/answer/2917686

47. JASP Team. JASP, version 0.98.1 [software]. Amsterdam: University of Amsterdam; 2026. Available from: https://jasp-stats.org/

48. Borelli M. Medical Statistics with JASP: lecture notes for the ICTP Master of Advanced Studies in Medical Physics. 1st ed. Trieste: ICTP; 2022.

49. Goss-Sampson MA. Statistical Analysis in JASP: A Guide for Students [Internet]. JASP; 2018 [cited 2026 Oct 8]. Available from: https://jasp-stats.org/jasp-materials/

50. MedCalc Software Ltd. MedCalc Statistical Software, version 24.0 [software]. Ostend: MedCalc Software Ltd; 2026. Available from: https://www.medcalc.org/

51. Centers for Disease Control and Prevention. Epi Info 7 for Windows, version 7.2.7 [software]. Atlanta (GA): CDC; 2025. Available from: https://www.cdc.gov/epiinfo/index.html

52. Universitatea de Medicină și Farmacie "Victor Babeș" Timișoara. UMFT eLearning platform [Internet]. Universitatea de Medicină și Farmacie "Victor Babeș" Timișoara; [cited 2026 Oct 9]. Available from: https://moodle.umft.ro/

53. Altman DG. Practical Statistics for Medical Research. London: Chapman and Hall; 1991.

54. Frandes M, Deiac AV, Timar B, Lungeanu D. Instrument for assessing mobile technology acceptability in diabetes self-management: a validation and reliability study. Patient Prefer Adherence. 2017;11:259-269. doi:10.2147/PPA.S127922 PMID: 28243069

55. Feier CVI, Santoro RR, Faur AM, Muntean C, Olariu S. Assessing Changes in Colon Cancer Care during the COVID-19 Pandemic: A Four-Year Analysis at a Romanian University Hospital. J Clin Med. 2023;12(20):6558. doi:10.3390/jcm12206558 PMID: 37892695

56. Hopewell S, Chan AW, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ. 2025;389:e081123. doi:10.1136/bmj-2024-081123 PMID: 40228833

57. Rațiu I, Lupușoru R, Lungeanu D, Popescu A, Sporea I, Goldiș A, et al. Diagnosis of malignant biliary obstruction: pondering over the ERCP, MRCP and histology. J Int Med Res. 2022;50(2):3000605221076924. doi:10.1177/03000605221076924 PMID: 35170356

58. Bende F, Sporea I, Șirli R, Bâldea V, Lazăr A, Lupușoru R, et al. Ultrasound-Guided Attenuation Parameter (UGAP) for the quantification of liver steatosis using the Controlled Attenuation Parameter (CAP) as the reference method. Med Ultrason. 2021;23(1):7-14. doi:10.11152/mu-2688 PMID: 33220028

59. Gherbon A, Frandes M, Dîrpeș D, Timar R, Timar B. Impact of SGLT-2 inhibitors on modifiable cardiovascular risk factors in Romanian patients with type 2 diabetes mellitus. Diabetol Metab Syndr. 2024;16(1):85. doi:10.1186/s13098-024-01326-8 PMID: 38627784

60. Dragomir AC, Buzatu AR, Marin RC, Moleriu RD, Muntean C, Piroș T, et al. Development and Initial Psychometric Validation of the Multidimensional Menopausal Burden Questionnaire (MMBQ-44) in a Sample of Romanian Women. Diagnostics (Basel). 2026;16(16):2574. doi:10.3390/diagnostics16162574 PMID: 42650976

61. Popa DI, Buleu F, Iancu A, Tudor A, Williams CG, Sutoi D, et al. Performance of GFAP and UCH-L1 for Early Acute Stroke Diagnosis in the Emergency Department. J Clin Med. 2025;14(13):4746. doi:10.3390/jcm14134746 PMID: 40649119

62. Popa DI, Buleu F, Williams C, Tudor A, Sutoi D, Trebuian CI, et al. Evaluating Thrombolysis Rates and Emergency Department Time Targets in Acute Ischemic Stroke: Need for Personalized Medicine. J Pers Med. 2024;14(9):955. doi:10.3390/jpm14090955 PMID: 39338208

63. Rosner B. Fundamentals of biostatistics. 8th ed. Boston (MA): Cengage Learning; 2016. ISBN 978-1-305-26892-0.

64. Băicuș C, Bălănescu P. Epidemiologie clinică și biostatistică. București: UMF Carol Davila; 2019.

65. Nica C, Sporea I, Popa A, Lupușoru R, Voicu Stan Erimescu M, Șirli R, et al. The value of Ultrasound Fat Fraction as compared to Controlled Attenuation Parameter for the severity of liver steatosis assessment - preliminary results. Med Ultrason. 2026;28(2):137-44. doi:10.11152/mu-4575 PMID: 41562217

66. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52(3-4):591-611. doi:10.1093/biomet/52.3-4.591

67. Frandes M, Gherbon A, Tudor A, Albai O, Vlad MM, Muntean C. TyG-Centered Endocrine-Metabolic Architecture in Patients with Thyroid Dysfunction: A Systems Phenotyping Study. Metabolites. 2026;16(7):490. doi:10.3390/metabo16070490 PMID: 42506441

68. Marin RC, Moleriu RD, Bungau GS, Tit DM, Muntean C. Discordant Immune-Virologic Responses During Antiretroviral Therapy: Immune Dysregulation Patterns, CD4/CD8 Ratio Inversion, and Clinical Predictors in a Romanian HIV Cohort. Viruses. 2026;18(5):512. doi:10.3390/v18050512 PMID: 42198715

69. Gherbon A, Roman D, Nicula-Neagu M, Frandes M. Demographic Vulnerability and Mental Health in Older Adults with Type 2 Diabetes: A Lifespan-Informed Analysis of Diabetes Distress. Diseases. 2026;14(9):332. doi:10.3390/diseases14090332 PMID: 42783225

70. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31(4):337-50. doi:10.1007/s10654-016-0149-3 PMID: 27209009

71. Moleriu LC, Lupusoru R, Muntean C, Piroș T, Popescu A, Sirli R, et al. Dietary Behaviors, Digestive Symptoms, and Neurovegetative Features in Disorders of Gut-Brain Interaction: A Cross-Sectional Clinical Study. Nutrients. 2026;18(7):1023. doi:10.3390/nu18071023 PMID: 41978074

72. Cirstea N, Tit DM, Purza AL, Toma MM, Radu A, Bungau GS, et al. Associations Between Patient Satisfaction with Pharmacist Services, Social Support, and Medication Adherence in Chronic Dermatological Conditions: A Cross-Sectional Study. Healthcare (Basel). 2026;14(15):2346. doi:10.3390/healthcare14152346 PMID: 42588315

73. Dicu AM, Marin RC, Muntean C, Moleriu RD, Tudor A, Piroș T, et al. Development and Validation of a Multidimensional Consumer Assessment Framework for Alternative-Ingredient Bakery and Confectionery Products: Implications for Nutrition Methodology and Sustainable Diet Research. Foods. 2026;15(17):2950. doi:10.3390/foods15172950 PMID: 42737195

74. Wasserstein RL, Lazar NA. The ASA Statement on p-Values: Context, Process, and Purpose. Am Stat. 2016;70(2):129-33. doi:10.1080/00031305.2016.1154108

75. Albai O, Frandes M, Sima A, Timar B, Vlad A, Timar R. Practical Applicability of the ISARIC-4C Score on Severity and Mortality due to SARS-CoV-2 Infection in Patients with Type 2 Diabetes. Medicina (Kaunas). 2022;58(7):848. doi:10.3390/medicina58070848 PMID: 35888567

76. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale (NJ): Lawrence Erlbaum; 1988.

77. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39(2):175-91. doi:10.3758/bf03193146 PMID: 17695343

78. Popovici SE, Ritiu SA, Miutescu B, Moga TV, Sporea I, Sandesc D, et al. Entropy-Guided Sedation Is Associated with Improved Hemodynamic Stability and Recovery During ERCP: A Prospective Real-World Cohort Study. J Clin Med. 2026;15(10):3665. doi:10.3390/jcm15103665 PMID: 42194633

79. Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ. 1995;311(7003):485. doi:10.1136/bmj.311.7003.485 PMID: 7647644

80. American Diabetes Association Professional Practice Committee for Diabetes. 2. Diagnosis and classification of diabetes: Standards of Care in Diabetes-2026. Diabetes Care. 2026;49(Suppl 1):S27-49. doi:10.2337/dc26-S002 PMID: 41358893

81. Al Barri L, Yasar II, Mercea N, Tudor A, Stanca HT, Roșca C, et al. Interdevice Agreement of Keratometry, Astigmatism Vectors, and Ocular Biometry in Cataract Candidates: SS-OCT (Argos) vs. OLCI (Aladdin) vs. Scheimpflug-Placido (Sirius). Bioengineering (Basel). 2026;13(3):296. doi:10.3390/bioengineering13030296 PMID: 41899827

82. Padurean LL, Frandes M, Stefanescu HE, Muntean C, Horhat ID. Impact of Smart Hearing Aid Technology on Self-Esteem in Patients with Sensorineural Hearing Loss: A Cross-Sectional Multivariate Study. Healthcare (Basel). 2026;14(10):1336. doi:10.3390/healthcare14101336 PMID: 42194428

83. Student. The probable error of a mean. Biometrika. 1908;6(1):1-25. doi:10.2307/2331554

84. Ursoiu F, Moleriu L, Lungeanu D, Puschită M. The association between HIV clinical disease severity and psychiatric disorders as seen in Western Romania. AIDS Care. 2018;30(11):1368-1371. doi:10.1080/09540121.2018.1455959 PMID: 29592527

85. Gherbon A, Frandes M, Timar R, Nicula M. Beneficial effects of Aloe ferox on lipid profile, blood pressure, and glycemic control in obese persons: A CONSORT-clinical study. Medicine (Baltimore). 2021;100(50):e28336. doi:10.1097/MD.0000000000028336 PMID: 34918714

86. Buda V, Andor M, Cristescu C, Tomescu MC, Muntean DM, Bâibâță DE, et al. Thrombospondin-1 Serum Levels In Hypertensive Patients With Endothelial Dysfunction After One Year Of Treatment With Perindopril. Drug Des Devel Ther. 2019;13:3515-3526. doi:10.2147/DDDT.S218428 PMID: 31631975

87. Buleu FN, Luca CT, Tudor A, Badalica-Petrescu M, Caraba A, Pah A, et al. Correlations between Vascular Stiffness Indicators, OPG, and 25-OH Vitamin D3 Status in Heart Failure Patients. Medicina (Kaunas). 2019;55(6):309. doi:10.3390/medicina55060309 PMID: 31242663

88. Turi V, Iurciuc S, Crețu OM, Tit DM, Bungau S, Apostol A, et al. Arterial function in hypertensive pregnant women. Is arterial stiffness a marker for the outcomes in pregnancy?. Life Sci. 2021;264:118723. doi:10.1016/j.lfs.2020.118723 PMID: 33160988

89. Uivarosan D, Tit DM, Iovan C, Nistor-Cseppento DC, Endres L, Lazar L, et al. Effects of combining modern recovery techniques with neurotrophic medication and standard treatment in stroke patients. Sci Total Environ. 2019;679:80-87. doi:10.1016/j.scitotenv.2019.05.070 PMID: 31100563

90. Stoichescu-Hogea G, Buleu FN, Christodorescu R, Sosdean R, Tudor A, Ember A, et al. Contribution of Global and Regional Longitudinal Strain for Clinical Assessment of HFpEF in Coronary and Hypertensive Patients. Medicina (Kaunas). 2021;57(12):1372. doi:10.3390/medicina57121372 PMID: 34946317

91. Miulescu RG, Roșca I, Marin RC, Muntean C, Pavel AN, Stoleru S, et al. Clinical Outcomes Following Personalized Gut Microbiota-Targeted Therapy in Children with Atopic Dermatitis and Laboratory-Confirmed Dysbiosis: A Single-Arm Prospective Pilot Study. Nutrients. 2026;18(15):2572. doi:10.3390/nu18152572 PMID: 42588195

92. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat. 1947;18(1):50-60. doi:10.1214/aoms/1177730491

93. Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bull. 1945;1(6):80-3. doi:10.2307/3001968

94. Munteanu LA, Frandes M, Timar B, Tudorache E, Fildan AP, Oancea C, et al. The efficacy of a mobile phone application to improve adherence to treatment and self-management in people with chronic respiratory disease in Romanian population - a pilot study. BMC Health Serv Res. 2020;20(1):475. doi:10.1186/s12913-020-05340-0 PMID: 32460752

95. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc. 1952;47(260):583-621. doi:10.1080/01621459.1952.10483441

96. Hasan EM, Calma CL, Tudor A, Oancea C, Tudorache V, Petrache IA, et al. Coping, Anxiety, and Pain Intensity in Patients Requiring Thoracic Surgery. J Pers Med. 2021;11(11):1221. doi:10.3390/jpm11111221 PMID: 34834573

97. Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Assoc. 1937;32(200):675-701. doi:10.1080/01621459.1937.10503522

98. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Lond Edinb Dublin Philos Mag J Sci. 1900;50(302):157-75. doi:10.1080/14786440009463897

99. Marin RC, Streinu-Cercel A, Moleriu LC, Bungau SG. Analysis of virological response to therapy and resistance profile in treatment-experienced and naive HIV-1 infected Romanian patients receiving regimens containing darunavir boosted with ritonavir or cobicistat. Biomed Pharmacother. 2022;150:113077. doi:10.1016/j.biopha.2022.113077 PMID: 35658217

100. Popa DI, Buleu F, Iancu A, Tudor A, Williams CG, Militaru M, et al. Long COVID and Acute Stroke in the Emergency Department: An Analysis of Presentation, Reperfusion Treatment, and Early Outcomes. J Clin Med. 2025;14(18):6514. doi:10.3390/jcm14186514 PMID: 41010719

101. Iovanescu D, Frandes M, Lungeanu D, Burlea A, Miutescu BP, Miutescu E. Diagnosis reliability of combined flexible sigmoidoscopy and fecal-immunochemical test in colorectal neoplasia screening. Onco Targets Ther. 2016;9:6819-6828. doi:10.2147/OTT.S122425 PMID: 27853376

102. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307-10. PMID: 2868172

103. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-74. PMID: 843571

104. Masood M, Masood Y, Newton JT. The clustering effects of surfaces within the tooth and teeth within individuals. J Dent Res. 2015;94(2):281-8. doi:10.1177/0022034514559408 PMID: 25421840

105. Danila AI, Cioca F, Gadde ST, Daruvuri SP, Timar R, Hogea E. Prognostic Utility of dNLR, ALRI, APRI, and SII in COVID-19 Patients with Diabetes: A Cross-Sectional Study. Diagnostics (Basel). 2024;14(15):1685. doi:10.3390/diagnostics14151685 PMID: 39125561

106. Ion R, Kimidi JSSK, Kalapala C, Fnu O, Chandrababu VR, Desireddygari OR, et al. Evaluating the Utility of Clinical Scores APACHE, CURB, SOFA, and NEWS2 at Admission and 5-Days after Symptom Onset in Predicting Severe COVID-19 in Patients with Diabetes. J Pers Med. 2024;14(8):868. doi:10.3390/jpm14080868 PMID: 39202059

107. Bâldea V, Sporea I, Tudor A, Popescu A, Bende F, Șirli R. Virtual Touch Quantification using Acoustic Radiation Force Impulse Imaging Technology versus Transient Elastography for the Noninvasive Assessment of Liver Fibrosis in Patients with Chronic Hepatitis B or C using Liver Biopsy as the Gold Standard. J Gastrointestin Liver Dis. 2020;29(2):181-190. doi:10.15403/jgld-2256 PMID: 32530985

108. Hill AB. The environment and disease: association or causation?. Proc R Soc Med. 1965;58(5):295-300. doi:10.1177/003591576505800503 PMID: 14283879

109. Mateoc-Sîrb T, Suciu IM, Gaiță D, Minodora A, Popescu R, Vlad T, et al. Guideline-Directed Medical Therapy Intensity, Ventricular Remodeling, and Clinical Outcomes After Acute Myocardial Infarction: A Single-Center Real-World Retrospective Cohort Study. Biomedicines. 2026;14(5):1067. doi:10.3390/biomedicines14051067 PMID: 42193392

110. Fofiu R, Bende F, Popescu A, Șirli R, Lupușoru R, Ghiuchici AM, et al. Spleen and Liver Stiffness for Predicting High-Risk Varices in Patients with Compensated Liver Cirrhosis. Ultrasound Med Biol. 2021;47(1):76-83. doi:10.1016/j.ultrasmedbio.2020.09.004 PMID: 33067019

111. Wassertheil-Smoller S, Smoller J. Biostatistics and Epidemiology: A Primer for Health and Biomedical Professionals. 5th ed. Cham: Springer; 2024. ISBN 978-3-031-53042-5.

112. Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. Lancet. 2002;359(9303):341-5. doi:10.1016/S0140-6736(02)07500-1 PMID: 11830217

113. Schulz KF, Grimes DA. Case-control studies: research in reverse. Lancet. 2002;359(9304):431-4. doi:10.1016/S0140-6736(02)07605-5 PMID: 11844534

114. Pirtea L, Grigoraş D, Matusz P, Pirtea M, Moleriu L, Tudor A, et al. Age and HPV type as risk factors for HPV persistence after loop excision in patients with high grade cervical lesions: an observational study. BMC Surg. 2016;16(1):70. doi:10.1186/s12893-016-0185-7 PMID: 27716233

115. Pantea-Roşan LR, Pantea VA, Bungau S, Tit DM, Behl T, Vesa CM, et al. No-Reflow after PPCI-A Predictor of Short-Term Outcomes in STEMI Patients. J Clin Med. 2020;9(9):2956. doi:10.3390/jcm9092956 PMID: 32932736

116. Feier CVI, Ardelean MV, Gaborean V, Muntean C, Faur AM, Rus VI, et al. Diagnostic and Prognostic Value of AISI, SII, and SIRI in Predicting Gangrenous Evolution of Acute Lithiasic Cholecystitis. Diagnostics (Basel). 2026;16(3):441. doi:10.3390/diagnostics16030441 PMID: 41681759

117. Petrusan AM, Feier CVI, Muntean C, Gaborean V, Petrusan AS, Nicoara D, et al. Association of BRCA Mutation Status with Clinical Outcomes in High-Grade Serous Ovarian Cancer. Healthcare (Basel). 2026;14(9):1193. doi:10.3390/healthcare14091193 PMID: 42121636

118. Gherbon A, Frandes M, Toderescu CD, Dirpes D, Timar R, Nicula MN, et al. Maternal and Fetal-Neonatal Complications of Romanian Women with Gestational Diabetes: A Retrospective Comparative Study. Medicina (Kaunas). 2025;61(7):1190. doi:10.3390/medicina61071190 PMID: 40731821

119. Alexandru GC, Muntean C, Chioran D, Riviş M, Gligor LN, Pricop MO, et al. Epidemiological Assessment of Oral and Oropharyngeal Squamous Cell Carcinomas Before, During, and After the COVID-19 Pandemic, 2018-2024. J Clin Med. 2026;15(18):6963. doi:10.3390/jcm15186963 PMID: 42795735

120. Owens DK, Goldhaber-Fiebert JD, Sox HC. Biomedical Decision Making: Probabilistic Clinical Reasoning. In: Shortliffe EH, Cimino JJ, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. 5th ed. Cham: Springer; 2021. p. 77-120. doi:10.1007/978-3-030-58721-5_3

121. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ. 2004;329(7458):168-9. doi:10.1136/bmj.329.7458.168 PMID: 15258077

122. Fagan TJ. Nomogram for Bayes's theorem. N Engl J Med. 1975;293(5):257. doi:10.1056/NEJM197507312930513 PMID: 1143310

123. Barbulescu A, Ratiu I, Sporea I, Lungeanu D, Lupusoru R, Miutescu B, et al. Extending and validating the Baveno VI criteria for the exclusion of high-risk varices. Med Ultrason. 2021;23(3):265-70. doi:10.11152/mu-2929 PMID: 33657192

124. Musen MA, Middleton B, Greenes RA. Clinical Decision-Support Systems. In: Shortliffe EH, Cimino JJ, editors. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. 5th ed. Cham: Springer; 2021. p. 795-840. doi:10.1007/978-3-030-58721-5_24

125. Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care. 2003;26(3):725-31. doi:10.2337/diacare.26.3.725 PMID: 12610029

126. Sipos SI, Vlad DC, Enatescu VR, Moleriu RD, Petre I, Balasan CG, et al. The Role of IL-6 in the Diagnosis of Neonatal Sepsis and Its Influence on Maternal Mental Health. Cureus. 2025;17(3):e80693. doi:10.7759/cureus.80693 PMID: 40242671

127. Pavel AL, Kundnani NR, Morariu SI, Tudor A, Man DE, Duda-Seiman DM, et al. Importance of H-FABP in Early Diagnosis of Acute Myocardial Infarction. Int J Gen Med. 2024;17:4335-4346. doi:10.2147/IJGM.S476736 PMID: 39346631

128. Guyatt G, Agoritsas T, Brignardello-Petersen R, Mustafa RA, Rylance J, Foroutan F, et al. Core GRADE 1: overview of the Core GRADE approach. BMJ. 2025;389:e081903. doi:10.1136/bmj-2024-081903

129. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71 PMID: 33782057

130. Suciu BS, Turi VR, Crisan S, Luca CT, Lazar DC, Faur-Grigori AA, et al. Heart Failure with Preserved Ejection Fraction (HFpEF), Pulse Wave Velocity, and Heart Rate Recovery Interconnections-A Brief Literature Review. J Clin Med. 2025;14(24):8781. doi:10.3390/jcm14248781 PMID: 41464682

131. Moleriu LC, Sas I, RaȚiu AC, Margan MM, Secoșan CA, Ilina RȘ, et al. HPV genotyping and p16÷Ki-67 dual staining in the detection of high-grade cervical lesion in patients with LSIL on Pap smear. Rom J Morphol Embryol. 2017;58(2):433-437. PMID: 28730227

132. Golu I, Sporea I, Moleriu L, Tudor A, Cornianu M, Vlad A, et al. 2D-Shear Wave Elastography in the Evaluation of Parathyroid Lesions in Patients with Hyperparathyroidism. Int J Endocrinol. 2017;2017:9092120. doi:10.1155/2017/9092120 PMID: 28845158

133. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. BMJ. 1994;309(6947):102. doi:10.1136/bmj.309.6947.102 PMID: 8038641

134. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32-5. doi:10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3 PMID: 15405679

135. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. BMJ. 1994;308(6943):1552. doi:10.1136/bmj.308.6943.1552 PMID: 8019315

136. Monaghan TF, Rahman SN, Agudelo CW, Wein AJ, Lazar JM, Everaert K, et al. Foundational Statistical Principles in Medical Research: Sensitivity, Specificity, Positive Predictive Value, and Negative Predictive Value. Medicina (Kaunas). 2021;57(5):503. doi:10.3390/medicina57050503

137. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36. doi:10.1148/radiology.143.1.7063747 PMID: 7063747

138. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. BMJ. 1994;309(6948):188. doi:10.1136/bmj.309.6948.188 PMID: 8044101

139. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378

140. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ. 2015;351:h5527. doi:10.1136/bmj.h5527 PMID: 26511519

141. Sounderajah V, Guni A, Liu X, Collins GS, Karthikesalingam A, Markar SR, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nat Med. 2025;31(10):3283-9. doi:10.1038/s41591-025-03953-8

142. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. Lancet. 2002;359(9307):696-700. doi:10.1016/S0140-6736(02)07816-9 PMID: 11879884

143. Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC (Text with EEA relevance). OJ L 158, 27.5.2014, p. 1-76. Available from: https://eur-lex.europa.eu/eli/reg/2014/536/oj

144. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH harmonised guideline: guideline for good clinical practice E6(R3). Final version, adopted on 6 January 2025 [Internet]. ICH; 2025 [cited 2026 Oct 9]. Available from: https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf

145. Putnoky S, Banu AM, Moleriu LC, Putnoky S, Șerban DM, Niculescu MD, et al. Reliability and validity of a General Nutrition Knowledge Questionnaire for adults in a Romanian population. Eur J Clin Nutr. 2020;74(11):1576-1584. doi:10.1038/s41430-020-0616-5 PMID: 32235889

146. Serban DM, Serban CL, Ursoniu S, Putnoky S, Moleriu RD, Putnoky S. Mindful Eating Questionnaire: Validation and Reliability in Romanian Adults. Int J Environ Res Public Health. 2022;19(17):10517. doi:10.3390/ijerph191710517 PMID: 36078231

147. Google. View and manage form responses. Google Docs Editors Help [Internet]. Mountain View (CA): Google; 2026 [cited 2026 Oct 8]. Available from: https://support.google.com/docs/answer/139706

148. Google. Show questions based on answers. Google Docs Editors Help [Internet]. Mountain View (CA): Google; 2026 [cited 2026 Oct 8]. Available from: https://support.google.com/docs/answer/141062

149. Bolboacă SD, Drugan T. Aplicații practice de biostatistică. Cluj-Napoca: Editura Medicală Universitară "Iuliu Hațieganu"; 2023. ISBN 978-606-075-206-6 (PDF). Coordinator role ('coord.') could not be confirmed from a primary source; both sources list them as authors

150. Iancu M, Bolboacă SD. Informatică medicală și biostatistică: noțiuni fundamentale pentru profesioniștii în științele sănătății. Cluj-Napoca: Editura Medicală Universitară "Iuliu Hațieganu"; 2025. ISBN 978-606-075-342-1 (PDF).

151. Țarcă V. Noțiuni fundamentale de biostatistică. Iași: Junimea; 2025. ISBN 978-973-37-2931-0.