



UNIVERSITATEA DE STAT DE MEDICINĂ ȘI FARMACIE
„NICOLAE TESTEMIȚANU” DIN REPUBLICA MOLDOVA



*Catedra de medicină socială și management
„Nicolae Testemițanu”*

Elena Raevschi, Olga Penina, Galina Obreja

Bazele biostatisticii și metodologia cercetării

Chișinău

Centrul Editorial-Poligrafic *Medicina*

2025

1

CZU[311.2+001.891]:614.2(075.8)

R 19

Aprobat de Consiliul de Management al Calității al USMF „Nicolae Testemițanu”,
proces-verbal nr. 2 din 23.10.2024.

Autori:

- Elena RAEVSCHI* – doctor habilitat în științe medicale, profesor universitar, Catedra de medicină socială și management „Nicolae Testemițanu”.
- Olga PENINA* – doctor habilitat în științe medicale, conferențiar universitar, Catedra de medicină socială și management „Nicolae Testemițanu”.
- Galina OBREJA* – doctor în științe medicale, conferențiar universitar, Catedra de medicină socială și management „Nicolae Testemițanu”.

Recenzenți:

- Larisa SPINEI* – doctor habilitat în științe medicale, profesor universitar, Catedra de medicină socială și management „Nicolae Testemițanu”.
- Alexandru CORLĂTEANU* – doctor habilitat în științe medicale, profesor universitar, Departamentul Medicină internă, Disciplina de pneumologie și alergologie.

Redactor: În redacția autorilor

Această publicație este prezentată în forma sa originală.

Prezentul compendiu „Bazele biostatisticii și metodologia cercetării” a fost elaborat având la bază experiența internațională și standardele didactice actualizate. Acesta corespunde în întregime cerințelor curriculumului la disciplina „Biostatistică și metodologia cercetării științifice” pentru studenții de la programele de studii superioare integrate și de licență din cadrul USMF „Nicolae Testemițanu”.

DESCRIEREA CIP A CAMEREI NAȚIONALE A CĂRȚII DIN REPUBLICA MOLDOVA
Raevschi, Elena.

Bazele biostatisticii și metodologia cercetării / Elena Raevschi, Olga Penina, Galina Obreja ; Universitatea de Stat de Medicină și Farmacie „Nicolae Testemițanu” din Republica Moldova, Catedra de medicină socială și management. – Chișinău : CEP *Medicina*, 2025. – 178 p. : fig. tab.

Bibliogr.: p. 174-175. – În red. aut. – 500 ex.

ISBN 978-9975-82-402-6.

[311.2+001.891]:614.2(075.8)

R 19

ISBN 978-9975-82-402-6.

© CEP *Medicina*, 2025

© Elena Raevschi, Olga Penina, Galina Obreja 2025

CUPRINS

Prefață	8
CAPITOLUL 1. INTRODUCERE ÎN BAZELE BIOSTATISTICII ȘI METODOLOGIA CERCETĂRII ÎN SĂNĂTATE	9
CAPITOLUL 2. STATISTICA DESCRIPTIVĂ. PREZENTAREA DATELOR ..	13
Concepte cheie	13
2.1. Concepte de bază	15
2.1.1. Variabilă	15
2.1.2. Populație	16
2.1.3. Eșantion.....	17
2.1.4. Parametri și statistici	17
2.2. Scale de măsurare	18
2.2.1. Scala nominală	18
2.2.2. Scala ordinală	19
2.2.3. Scala numerică.....	20
2.3. Tabele	21
2.3.1. Distribuții de frecvențe	21
2.3.2 Frecvența relativă.....	23
2.4. Grafice	24
2.4.1. Graficele cu linii.....	24
2.4.2. Grafice cu bare	25
2.4.3. Grafice circulare	26
2.4.4. Histograme și poligoane de frecvență.....	27
2.4.5. Graficele cu cutie (box plot)	28
2.4.6. Grafice cu bare de eroare	29
2.4.7. Grafice de dispersie	30
Exerciții de recapitulare.....	31
Întrebări de recapitulare	32
CAPITOLUL 3. STATISTICA DESCRIPTIVĂ. REZUMAREA DATELOR NUMERICE	33
Concepte cheie	33
3.1. Măsuri de tendință centrală	35
3.1.1. Media	35
3.1.2. Mediana	36
3.1.3. Modul.....	36

3.1.4. Relația empirică între măsurile tendinței centrale.....	36
3.2. Măsurile ale variabilității	38
3.2.1. Amplitudinea	39
3.2.2. Intervalul intercuartil	39
3.2.3. Varianța și abaterea standard	40
3.2.4. Coeficientul de variație.....	42
3.3. Distribuția normală și proprietățile acesteia	43
3.4. Asimetria și kurtosis.....	44
3.5. Exemplu de calcul.....	46
Exerciții de recapitulare.....	48
Întrebări de recapitulare.....	50
CAPITOLUL 4. STATISTICA DESCRIPTIVĂ: REZUMAREA DATELOR	
NOMINALE ȘI ORDINALE.....	51
Concepte Cheie.....	51
4.1. Metode de descriere a datelor categoriale.....	52
4.1.1. Proportii și procente	52
4.1.2. Raporturi.....	53
4.1.3. Rate.....	54
4.2. Indicatorii stării de sănătate în statistica descriptivă	55
4.2.1. Ratele de mortalitate.....	55
4.2.2. Ratele de morbiditate	56
4.3. Rate standardizate. Metoda directă de standardizare	57
Exerciții de recapitulare.....	61
Întrebări de recapitulare.....	63
CAPITOLUL 5. CORELAȚIE ȘI REGRESIE.....	64
Concepte cheie	64
5.1. Corelație	65
5.1.1. Tipuri de coeficienți de corelație.....	65
5.1.2 Coeficientul de determinare.....	69
5.2. Regresie. Abordări generale.....	70
5.2.1. Regresie liniară simplă	70
5.2.2. Regresie liniară multiplă	72
5.2.3. Regresie logistică.....	73
Exerciții de recapitulare.....	74
Întrebări de recapitulare.....	76

CAPITOLUL 6. STATISTICA INFERENȚIALĂ. TEORIA PROBABILITĂȚII ȘI TESTAREA IPOTEZELOR.....	77
Concepte cheie	77
6.1. Teoria probabilităților.....	79
6.1.1. Concepte generale.....	79
6.1.2. Legea numerelor mari	79
6.1.3. Teorema limitei centrale	80
6.1.4. Utilizarea erorii standard.....	80
6.2. Eșantionare	81
6.2.1. Definiția eșantionării.....	81
6.2.2. Metode de eșantionare.....	81
6.3. Estimarea și testarea ipotezelor.....	82
6.4. Intervalele de încredere.....	83
6.5. Testarea ipotezelor: concepte teoretice de bază.....	86
6.5.1. Definirea ipotezei	86
6.5.2. Tipuri de ipoteze.....	86
6.5.3. Erori de tip I și de tip II	87
6.5.4. Puterea studiului.....	87
6.5.5. Nivelul de încredere	88
6.5.6. Nivelul de semnificație	89
6.5.7. Valoarea p.....	89
6.6. Procesul de testare a ipotezelor. Abordări generale.....	90
Exerciții de recapitulare.....	91
Întrebări de recapitulare	92
CAPITOLUL 7. TESTAREA IPOTEZELOR. METODE PARAMETRICE ȘI NEPARAMETRICE	94
Concepte cheie	94
7.1. Teste parametrice și neparametrice.....	94
7.2. Abordarea generală a testării ipotezelor	96
7.2.1. Etapele testării ipotezelor.....	96
7.2.2. Testarea ipotezelor. Teste bilaterale, unilaterale stânga și unilaterale dreapta	98
7.3. Teste parametrice	100
7.3.1. Testul t pentru un singur eșantion.....	100
7.3.4. Testul t pentru două eșantioane independente	103

7.3.5. Testul t pentru coeficientul de corelație	106
7.4. Testele neparametrice	109
7.4.1. Testul chi-pătrat.....	109
Exerciții de recapitulare.....	112
Întrebări de recapitulare.....	113
CAPITOLUL 8. INTRODUCERE ÎN METODOLOGIA CERCETĂRII	114
Concepte cheie	114
8.1. Definiția, caracteristicile și tipurile de cercetări.....	115
8.1.1. Definiția și caracteristicile cercetării.....	115
8.1.2. Erori aleatorii și erori sistematice în cercetare.....	116
8.1.3. Tipuri de cercetări.....	118
8.2. Etapele procesului de cercetare	121
8.3. Formularea problemei de cercetare	122
8.4. Sinteza literaturii	124
8.5. Formularea scopului și obiectivelor studiului.....	127
8.6. Pregătirea designului de cercetare și colectarea datelor....	127
8.6.1. Definirea designului de cercetare și pașii necesari	127
8.6.2. Determinarea designului eșantionului	127
8.6.3. Instrumente de colectare a datelor	128
8.6.5. Clasificarea designului de cercetare	130
8.7.3. Puterea dovezilor în designul studiilor.....	134
Exerciții de recapitulare.....	136
Întrebări de recapitulare.....	137
CAPITOLUL 9. STUDII OBSERVAȚIONALE DESCRIPTIVE	139
Concepte cheie	139
9.1. Studiul serie de cazuri / raportare de caz.....	139
9.2. Studiul transversal.....	140
Întrebări de recapitulare.....	142
CAPITOLUL 10. STUDII ANALITICE OBSERVAȚIONALE	143
Concepte cheie	143
10.1 Studiul caz-control	144
10.2. Studiul de cohortă.....	147
Exerciții de recapitulare.....	152
Întrebări de recapitulare.....	153
CAPITOLUL 11. STUDII EXPERIMENTALE	154

Concepte cheie	154
11.1. Clasificarea studiilor clinice.....	155
11.2. Studii clinice cu grupuri de control paralele	156
11.3. Studii clinice cu grupuri de control succesive.....	158
11.4. Studii clinice cu grupuri de control externe	159
11.5. Analiza statistică a studiilor clinice	161
Exerciții de recapitulare.....	165
Întrebări de recapitulare	166
CAPITOLUL 12. REDACTAREA REZULTATELOR CERCETĂRII. METODE	
GENERALE	167
12.1. Redactarea unei lucrări de cercetare.....	167
12.2. Prezentarea publică a rezultatelor cercetărilor medicale	168
12.3. Structura și principiile de elaborare a lucrării de absolvire / tezei de licență la USMF „Nicolae Testemitanu”	169
CAPITOLUL 13. INTRODUCERE ÎN ETICA CERCETĂRII	170
13.1. Definirea și obiectivele eticii în cercetare.....	170
13.2. Principiile eticii în cercetare	170
CAPITOLUL 14. PREVENIREA PLAGIATULUI. PRINCIPII DE BAZĂ.....	172
14.1. Definiția și clasificarea plagiatului.....	172
14.2. Metode de prevenire a plagiatului.....	173
BIBLIOGRAFIE	174
ANEXA A. Valori critice pentru distribuția „t”	176
ANEXA B. Valori critice pentru distribuția chi pătrat.....	177

Prefață

Biostatistica și metodologia cercetării, ca disciplină, constituie un domeniu interdisciplinar care își extrage esența dintr-o gamă variată de contribuții științifice și joacă un rol crucial în efectuarea cercetărilor care respectă standardele globale. Compendiul „Bazele biostatisticii și metodologia cercetării” este aliniat cu curriculumul disciplinei și furnizează informații detaliate și bine structurate despre fiecare pas necesar pentru a desfășura un studiu științific riguros. Acest compendiu servește drept introducere în biostatistică și metodologia cercetării pentru studenții din domeniul științelor sănătății, având ca scop facilitarea realizării lucrărilor lor de absolvire / tezelor de licență.

Prezentul Compendiu este editat în patru limbi (română, rusă, engleză și franceză) și aduce o structură revizuită, un conținut extins și discuții îmbunătățite pe diverse teme din întreaga lucrare. Acesta include figuri și tabele suplimentare pentru elucidarea conceptelor, în special în domeniile statisticii descriptive și inferențiale. Cartea oferă explicații detaliate ale testării ipotezelor utilizând atât teste parametrice, cât și neparametrice, precum și analize cuprinzătoare de corelație și regresie. Tabelele de valori critice pentru distribuția t și distribuția chi-pătrat au fost adăugate în anexa A și anexa B.

Mulțumiri

Exprimăm recunoștință profundă profesorilor care au revizuit manuscrisul: Larisa Spinei, doctor habilitat, profesor universitar la Catedra de medicină socială și management „Nicolae Testemițanu”, și Alexandru Corlăteanu, doctor habilitat, profesor universitar la Departamentul Medicină internă, ambii de la Universitatea de Stat de Medicină și Farmacie „Nicolae Testemițanu”. De asemenea, ne exprimăm aprecierea față de toți profesorii care au contribuit la predarea cursului și au oferit sugestii inestimabile.

Autorii

CAPITOLUL 1. INTRODUCERE ÎN BAZELE BIOSTATISTICII ȘI METODOLOGIA CERCETĂRII ÎN SĂNĂTATE

Compendiul „Bazele biostatisticii și metodologia cercetării” introduce studenții de la programele de studii integrate și de licență din cadrul Universității de Stat de Medicină și Farmacie „Nicolae Testemițanu” în studiul statisticii aplicate în medicină, stomatologie și în alte subdomenii ale domeniului sănătății. Scopul principal este de a dezvolta cunoștințe despre metodele contemporane utilizate în cercetarea practică. Dobândirea cunoștințelor necesare pentru utilizarea metodelor moderne de documentare, asimilarea definițiilor teoretice aplicabile în cercetare și respectarea standardelor de reglementare sunt esențiale pentru evidențierea rezultatelor cercetărilor incluse în lucrările de absolvire /tezele de licență.

Cursul de biostatistică și metodologia cercetării științifice acoperă aspectele teoretice și practice legate de realizarea cercetărilor științifice și analiza datelor statistice. Cursul are un conținut similar cu cel al altor universități europene, incluzând informații actualizate, și le oferă studenților pregătirea necesară pentru a desfășura cercetări științifice în domeniul științelor biomedicale. Cursul promovează o abordare în principal aplicativă a metodelor statistice necesare pentru soluționarea problemelor practice din domeniul biomedical.

Obiectivul principal al cursului

Să ajute studenții să înțeleagă conceptul de bază al Biostatisticii și Metodologiei Cercetării, astfel încât să poată folosi aceste cunoștințe pentru a planifica și analiza date în cercetarea biomedicală.

La nivelul cunoașterii și înțelegerii:

- Să cunoască conceptele teoretice ale metodologiei cercetării științifice în sănătate.
- Dezvoltarea unei gândiri clare și continue, capabile să gestioneze și să proceseze datele.
- Să cunoască principiile, tehnologia, metodele și tehnicile utilizate în cercetarea medicală.
- Să înțeleagă corelația între metodele moderne utilizate în biostatistică și metodologia cercetării în sănătate.
- Să identifice posibilitățile de analiză și interpretare, precum și limitele metodelor moderne utilizate în cercetarea științifică.

La nivelul aplicării:

- Să analizeze definițiile și metodele teoretice și practice ale metodologiei cercetării științifice.
- Să folosească metodele și tehnicile statistice în procesul științific.
- Să demonstreze capacitatea de a analiza, interpreta și prezenta rezultatele cercetării științifice.
- Să utilizeze cunoștințele de bază în biostatistică necesare pentru aplicarea optimă a acestora în obținerea rezultatelor corecte ale cercetării științifice.
- Să stăpânească limbajul și terminologia specifică stilului științific.
- Să evalueze informațiile conținute într-un articol sau raport de specialitate și să aprecieze relevanța acestora.
- Să caute informații științifice folosind metode clasice sau metode informatice pentru căutarea și selecția datelor.
- Să folosească metode moderne de redactare și prezentare a propunerilor științifice și a raportului final de rezultate.

La nivelul integrării:

- Să aprecieze valoarea teoretică și aplicabilă a metodologiei cercetării medicale în diferite discipline din domeniul sănătății.
- Să evalueze locul și rolul biostatisticii și metodologiei cercetării în cariera medicală profesională.
- Să integreze cunoștințele în biostatistică și metodologie a cercetării cu disciplinele clinice.
- Să aplice cunoștințele acumulate în activitățile practice și de cercetare.
- Să fie competent în utilizarea critică a informațiilor din publicațiile științifice, folosind noi tehnologii de informație și comunicare.

Rezultatele studiului:

- Să explice conceptele de bază legate de organizarea cercetării științifice și publicarea rezultatelor.
- Să dezvolte un proiect de cercetare în domeniul biomedical.
- Să prezinte descrierea datelor experimentale în funcție de natura lor și să explice corect rezultatele inferenței statistice.
- Să determine metodele statistice pentru analiza datelor, având în vedere caracteristicile designului studiului, scala de măsurare și numărul de variabile implicate.
- Să caracterizeze trăsăturile de bază ale designurilor studiilor epidemiologice (observaționale și experimentale), avantajele și limitările acestora.
- Să realizeze un studiu epidemiologic (observațional sau experimental) și să interpreteze corect rezultatele acestuia.
- Să dezvolte o lucrare științifică, inclusiv lucrarea de absolvire / teza de licență, și să valorifice rezultatele acesteia.

- Să evalueze rolul și importanța biostatisticii și a metodologiei cercetării în contextul modern al medicinei bazate pe dovezi.
- Să fie deschis la învățarea continuă.

Cercetarea în domeniul sănătății reprezintă un domeniu interdisciplinar, care se bazează pe un spectru extins de cunoștințe. Biostatistica și Metodologia Cercetării Științifice este o disciplină ce permite integrarea și analiza cunoștințelor din studii fundamentale și aplicate. Această disciplină este crucială pentru evaluarea activităților de cercetare conform standardelor contemporane. Fiind un domeniu integrativ, acesta corelează cu alte discipline care utilizează statistica. O înțelegere profundă a matematicii de liceu și a biomedicinei de bază este necesară pentru stăpânirea acestei discipline.

CAPITOLUL 2. STATISTICA DESCRIPTIVĂ. PREZENTAREA DATELOR

Concepte cheie

- ❖ *Statistică* - o caracteristică sau valoare derivată din datele unui eșantion.
- ❖ *Parametru* - o caracteristică sau valoare derivată din datele unei populații.
- ❖ *Variabilă* - o caracteristică măsurată a unei unități de observație.
- ❖ *Variabilă cantitativă (numerică)* - o variabilă pentru care valorile atribuite sunt ordonate și semnificative.
- ❖ *Variabilă calitativă (categorială)* - o variabilă pentru care valorile atribuite au semnificație ca referințe nominale (etichete), dar nu ca valori numerice.
- ❖ *Variabilă alternativă (dihotomică)* - o variabilă calitativă ce are doar două categorii.
- ❖ *Variabilă nealternativă* - o variabilă calitativă ce are mai mult de două categorii.
- ❖ *Variabilă discretă* - o variabilă numerică ce ia doar valori întregi.
- ❖ *Variabilă continuă* - o variabilă numerică ce poate lua orice valoare pe un continuum.
- ❖ *Date nominale* - date clasificate în diferite categorii calitative, care pot fi listate în orice ordine.
- ❖ *Date ordinale* - date clasificate în diferite categorii unde ordinea între categorii este semnificativă, dar nu există informații despre distanța cantitativă între categorii.
- ❖ *Date de interval* - date cu intervale egale între elemente, dar fără un zero absolut.
- ❖ *Date de raport* - date cu intervale egale și un zero absolut.
- ❖ *Distribuție de frecvențe* - colecția valorilor dintr-un eșantion pentru o singură variabilă.

- ❖ *Histogramă și poligonul frecvenței* - grafice utilizate pentru a prezenta distribuția de frecvențe a datelor numerice. Acestea oferă informații despre forma distribuției de frecvențe.
- ❖ *Diagramă de dispersie* - grafic utilizat pentru a ilustra relația dintre două variabile numerice.
- ❖ *Diagramă cu bare și diagramă circulară* - grafice utilizate pentru a prezenta date calitative.

Statistica descriptivă facilitează înțelegerea și sintetizarea unor seturi mari de date prin intermediul câtorva indicatori esențiali. Aceste date numerice oferă o modalitate de a interpreta rapid un volum considerabil de informații. De exemplu, în loc să analizezi individual rezultatele testelor a 100 de studenți, poți examina media acestora pentru a obține o imagine generală despre performanțele grupei. Indicatorii esențiali permit identificarea unor modele și tendințe în date, simplificând astfel înțelegerea și comunicarea informațiilor relevante din cercetarea ta. Totuși, este esențial de reținut că statistica descriptivă nu oferă explicații privind motivele existenței acestor modele sau semnificația lor. Ea furnizează, în schimb, indicii care pot conduce la formularea întrebărilor sau ideilor, cunoscute sub denumirea de ipoteze. O ipoteză reprezintă o explicație posibilă ce poate fi testată prin cercetări suplimentare.

2.1. Concepte de bază

2.1.1. Variabilă

Definiția variabilei. O variabilă este o caracteristică de interes care are valori diferite pentru subiecți sau obiecte diferite incluse într-un studiu.

Exemple: vârsta, data nașterii, etnia, numărul de copii, tensiunea arterială.

Clasificarea variabilelor. Pentru a prezenta corect statisticile descriptive, este necesar să înțelegeți tipurile de date întâlnite frecvent în studiile de cercetare (*figura 2.1*).

Variabile cantitative (numerice). Aceste variabile pot fi cuantificate și se clasifică după cum urmează:

1. *Variabile discrete* - acestea reprezintă cantități care pot lua doar valori specifice, distincte, de obicei numere întregi, fără valori intermediare posibile.

Exemple: numărul de pacienți, numărul de noi cazuri de boli cardiace, numărul de nou-născuți într-un anumit an .

2. *Variabile continue* - acestea reprezintă cantități care pot lua orice valoare într-un interval, fără a fi limitate la valori discrete specifice.

Exemple: vârsta, nivelul de glucoză în sânge, tensiunea arterială.

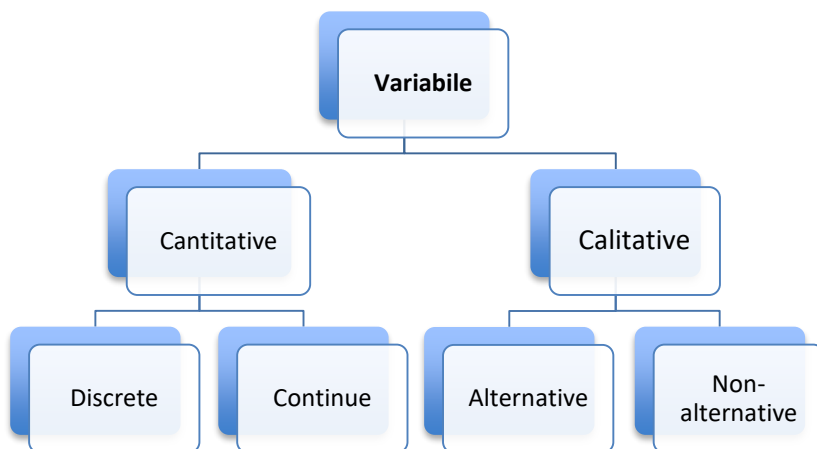


Figura 2.1. Tipuri de variabile

Variabile calitative (catoriale). Aceste variabile nu pot fi cuantificate și se clasifică după cum urmează:

1. *Variabile alternative (dicotomice sau binare)* - reprezintă categorii unde rezultatele pot lua doar una dintre cele două valori posibile.

Exemple: „Da” sau „Nu”; „masculin” sau „feminin” .

2. *Variabile non-alternative* - reprezintă categorii unde rezultatele pot lua mai multe valori.

Exemple: grupa sanguină, nivelul de severitate al unei boli .

În analize, variabilele calitative sunt adesea codificate prin valori numerice.

2.1.2. Populație

O populație se referă la întregul grup de persoane sau elemente care împărtășesc o caracteristică comună și sunt subiecții unui studiu. Aceasta include fiecare membru care îndeplinește criteriile stabilite de

studiu. De exemplu, o populație ar putea fi toți elevii dintr-o anumită școală, toți pacienții dintr-un spital sau toți copacii dintr-o pădure.

- *Populația țintă* - este întregul grup asupra căruia sunt destinate a fi generalizate rezultatele studiului. Este cunoscută și sub denumirea de populație *teoretică*.
- *Populația studiată* - este segmentul specific al populației țintă care este accesibil și din care se colectează efectiv datele pentru studiu. Este cunoscută și sub denumirea de populație *accesibilă*.

2.1.3. Eșantion

Eșantionul reprezintă un subset al populației studiate. *Unitatea de observație*, sau unitatea statistică, este entitatea de la care se colectează informațiile, precum o persoană, o gospodărie, o comunitate, o instituție de învățământ etc. Identificarea clară a unității de observație este crucială pentru un design logic al studiului, pentru colectarea organizată a datelor și pentru analiza obiectivă.

Populația țintă vs. populația de studiu: cercetare în domeniul bolilor cardiovasculare.

- **Populația țintă** - toți adulții cu vârsta de 40 de ani și peste, diagnosticați cu boli cardiovasculare la nivel mondial.
- **Populația de studiu** - adulți cu vârsta de 40 de ani și peste, cu boli cardiovasculare, care sunt pacienți în spitale principale din Republica Moldova.

Populația de studiu servește ca un subset reprezentativ al populației țintă mai largi.

2.1.4. Parametri și statistici

Parametru se referă la o caracteristică sau valoare obținută din datele întregii populații.

Statistica se referă la o caracteristică sau valoare obținută din datele eșantionului.

Tabelul 2.1. Simboluri pentru populație și eșantion

	Parametrii (din datele populației)	Statisticile (din datele eșantionului)
	<i>Simboluri grecești</i>	<i>Simboluri romane</i>
Medie	μ	$\bar{x}$
Deviație standard	σ	S
Varianță	σ^2	S^2
Volum	N	n

2.2. Scale de măsurare

Scalele de măsurare diferă în funcție de natura variabilelor. Ele stabilesc modul în care datele sunt prezentate și rezumate, precum și metodele statistice utilizate pentru analiza acestora. În statistică, există trei scale de măsurare:

- scala nominală (clasificativă);
- scala ordinală (de rang);
- scala numerică (scalele de interval și de raport).

2.2.1. Scala nominală

Datele de pe scala nominală sunt organizate în categorii care nu au o ordine intrinsecă. O variabilă măsurată pe această scală poate avea două sau mai multe categorii, în funcție de variabilitatea caracteristicii calitative.

Exemplul 1. Variabila „gen” are, în general, două categorii: masculin și feminin. Datele nominale care sunt clasificate în una dintre cele două categorii distincte, precum masculin sau feminin, sunt considerate variabile categoriale alternative sau variabile dicotomice.

Exemplul 2. Totuși, nu toate datele nominale sunt dicotomice. Multe variabile nominale au trei sau mai multe categorii. De exemplu, „anemia” poate fi împărțită în mai multe subcategorii: anemia microcitară, anemia macrocitară sau megaloblastică și anemia normocitară. Această variabilă este clasificată ca o variabilă categorială non-alternativă. Ordinea în care sunt prezentate aceste categorii nu este relevantă, deoarece nu există un rang sau o relație între ele.

Datele pe scala nominală sunt utilizate pentru a eticheta variabilele fără a oferi o valoare cantitativă. Fiecare categorie este distinctă, iar secvența categoriilor este neimportantă.

2.2.2. Scala ordinală

Scala ordinală, cunoscută și sub denumirea de scala de rang, clasifică persoanele, obiectele, răspunsurile sau caracteristicile în subgrupuri pe baza unei trăsături comune și le ordonează într-o ordine specifică.

Subgrupurile sunt aranjate fie în ordine crescătoare, fie descrescătoare, în funcție de măsura în care o categorie reflectă amploarea variației variabilei. Când ordinea între categorii devine semnificativă, datele sunt considerate a fi pe o scală ordinală de măsurare.

Exemplul 1. „Venitul” unui pacient poate fi măsurat folosind o variabilă calitativă cu categoriile „peste medie”, „medie” și „sub medie”. „Distanța” dintre aceste subcategorii nu este egală, deoarece nu există o unitate de măsură cantitativă.

Exemplul 2. *Scorul Apgar*, care evaluează maturitatea nou-născuților, variază de la 0 la 10. Scorurile mai mici indică depresia funcționării cardiopulmonare și neurologice, în timp ce scorurile mai mari sugerează o funcționare optimă. Diferența între scorurile de 8 și 9 nu are aceleași implicații clinice ca diferența dintre scorurile de 0 și 1.

Acest exemplu ilustrează faptul că în scalele ordinale, intervalele dintre punctele de măsurare nu sunt egale.

Datele de pe scala ordinală nu doar că clasifică variabilele, ci le și ordonează, deși intervalele între categorii nu sunt neapărat uniforme sau măsurabile.

2.2.3. Scala numerică

Scala numerică se distinge prin intervale egale între punctele succesive de măsurare. Există două tipuri de scale numerice:

⇒ Scala de interval

Scala de interval înglobează toate caracteristicile scalelor nominală și ordinală. În plus, permite aranjarea datelor într-o ordine ierarhică, cu distanțe egale între punctele de măsurare. Datele pe scala de interval nu au un zero absolut (adică un punct care semnifică absența totală a variabilei), ceea ce înseamnă că pot avea valori semnificative atât sub zero, cât și peste zero.

Exemplu:

Scala Celsius: 0°C până la 100°C

Scala Fahrenheit: 32°F până la 212°F

Scările de temperatură Celsius și Fahrenheit măsoară temperatura cu intervale egale între grade. Ambele scări includ valori sub zero (pentru Celsius) sau sub punctul de îngheț al apei (32°F pentru Fahrenheit), indicând temperaturi sub punctul de îngheț al apei.

⇒ Scala de raport

Scala de raport înglobează toate caracteristicile scalelor nominală, ordinală și de interval, având în plus un punct zero fix. Aceasta înseamnă că are o valoare zero absolută, indicând absența totală a variabilei (nu există valori sub zero). Exemple de variabile măsurate pe scala de raport includ nivelul colesterolului seric, venitul, vârsta, înălțimea și greutatea

corporală.

Tabelul de mai jos evidențiază diferențele esențiale dintre cele patru scale de măsurare (*tabelul 2.2*).

Tabelul 2.2. Diferența dintre cele patru scale de măsurare

Scala	Indică diferența	Indică direcția diferenței	Indică mărimea diferenței	Zero absolut
Nominală	+			
Ordinală	+	+		
De interval	+	+	+	
De raport	+	+	+	+

2.3. Tabele

Un tabel reprezintă cea mai simplă modalitate de a sintetiza un set de date și poate fi utilizat pentru toate tipurile de variabile.

2.3.1. Distribuții de frecvențe

Un mod frecvent utilizat pentru a rezuma datele este tabelul de distribuție a frecvențelor. Acest tabel categorizează datele și afișează numerele corespunzătoare pentru diferite tipuri de date. Pentru datele nominale și ordinale, distribuția de frecvențe organizează categoriile împreună cu numerele lor. De exemplu, *tabelul 2.3* prezintă distribuția nou-născuților după sex în Republica Moldova pentru anul 2022, utilizând date nominale.

Tabelul 2.3. Distribuția de frecvențe pentru date nominale. Distribuția nașterilor după sex în Republica Moldova în 2022

Sexul nou-născutului	Numărul de nașteri
Masculin	13.950
Feminin	13.068

Pentru a reprezenta datele numerice, se folosesc atât distribuțiile

de frecvențe negrupate, cât și cele grupate. De exemplu, *tabelul 2.4* ilustrează o distribuție de frecvențe negrupate prin prezentarea numărului anual de nașteri în Republica Moldova din 2000 până în 2020, utilizând date de interval.

Tabelul 2.4. Distribuția de frecvențe pentru date de interval. Numărul anual de nașteri în Republica Moldova, în 2000-2022

Anul	Numărul de nașteri
2000	36.939
2010	40.474
2020	30.834
2022	27.018

Când datele numerice sunt foarte detaliate, distribuțiile de frecvențe grupate sunt deosebit de utile. Această metodă implică împărțirea intervalului de valori în intervale distincte, care nu se suprapun. După stabilirea acestor intervale, numărul de observații din fiecare interval este numărat. *Tabelul 2.5* exemplifică această metodă prin detalierea distribuției de frecvențe grupate a nașterilor în funcție de vârsta mamei în Republica Moldova pentru anul 2022, utilizând date de raport.

Tabelul 2.5. Distribuția de frecvențe grupate pentru date de raport: numărul de nașteri în funcție de vârsta mamei în Republica Moldova, în 2022

Grupă de vârstă a mamei	Numărul de nașteri
Mai puțin de 24 ani	7.117
25-34 ani	15.032
35-44 ani	4.843
45 ani și peste	26

Tabelele sunt cele mai eficiente atunci când sunt clare și bine

organizate. Prin urmare, tabelele și coloanele acestora trebuie întotdeauna etichetate clar, iar unitățile de măsură trebuie specificate acolo unde este necesar.

2.3.2 Frecvența relativă

Uneori, este util să cunoaștem proporția valorilor care se încadrează într-un anumit interval într-o distribuție de frecvențe, în loc de numărul absolut. Frecvența relativă pentru un interval reprezintă proporția numărului total de observații care apar în acel interval. *Frecvența relativă* se calculează prin împărțirea numărului de valori dintr-un interval la numărul total de valori din tabel, exprimată ca procent. Frecvențele relative sunt esențiale pentru compararea seturilor de date cu un număr inegal de observații.

Frecvența relativă cumulată pentru un interval reprezintă procentajul numărului total de observații care au o valoare mai mică sau egală cu limita superioară a intervalului. Frecvența relativă cumulată se calculează prin însumarea frecvențelor relative pentru intervalul specificat și pentru toate intervalele anterioare.

Tabelul 2.6 ilustrează frecvențele absolute, relative și relative cumulate ale valorii indicelui de șoc la 931 de pacienți.

Tabelul 2.6. Frecvențele absolute, relative și relative cumulate ale valorii indicelui de șoc pentru 931 de pacienți

Valoarea indicelui de șoc	Frecvența absolută (numărul de pacienți)	Frecvența relativă (%)	Frecvența relativă cumulată (%)
0,30-0,39	38	4,1	4,1
0,40-0,49	104	11,2	15,3
0,50-0,59	198	21,3	36,6
0,60-0,69	199	21,4	58,0
0,70-0,79	155	16,6	74,6

Valoarea indicelui de șoc	Frecvența absolută (numărul de pacienți)	Frecvența relativă (%)	Frecvența relativă cumulată (%)
0,80-0,89	102	11,0	85,6
0,90-0,99	60	6,4	92,0
1,00-1,09	37	4,0	96,0
1,10-1,19	19	2,0	98,0
1,20-1,29	19	2,0	100,0
Total	931	100,0	

2.4. Grafice

Graficele (diagramele) reprezintă o a doua metodă de a rezuma și prezenta datele. Acestea sunt adesea mai ușor de interpretat decât tabelele, însă pot oferi informații mai puțin detaliate. Cele mai relevante grafice sunt de obicei simple și auto-explicative. La fel ca și în cazul tabelelor, graficele trebuie să fie clar etichetate, iar unitățile de măsură trebuie să fie specificate.

2.4.1. Graficele cu linii

Graficele cu linii (diagrame liniare) sunt frecvent utilizate pentru a ilustra tendințele în timp ale datelor numerice. Fiecare valoare de pe axa x corespunde unei singure valori de pe axa y, iar punctele adiacente sunt conectate prin linii drepte (*figura 2.2*).

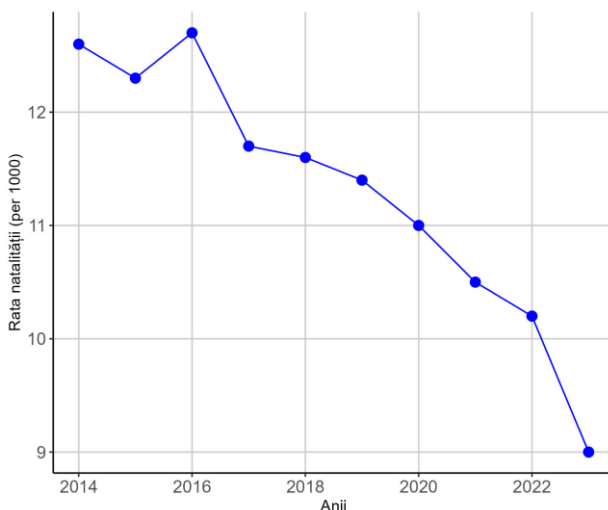


Figura 2.2. Rata natalității în municipiul Chișinău (2014-2023), la 1000 de locuitori

2.4.2. Grafice cu bare

Graficele cu bare (diagramele cu bare/coloane), fie că sunt verticale sau orizontale, sunt un instrument de bază pentru ilustrarea distribuțiilor de frecvențe ale datelor nominale sau ordinale. În aceste grafice, barele trebuie să fie de lățimi uniforme și să fie dispuse la distanță una de cealaltă, pentru a preveni crearea unei impresii false de continuitate. Graficele cu bare sunt eficiente în compararea mai multor valori, categoriile fiind prezentate fie pe axa verticală, fie pe axa orizontală, așa cum este exemplificat în *figura 2.3*.

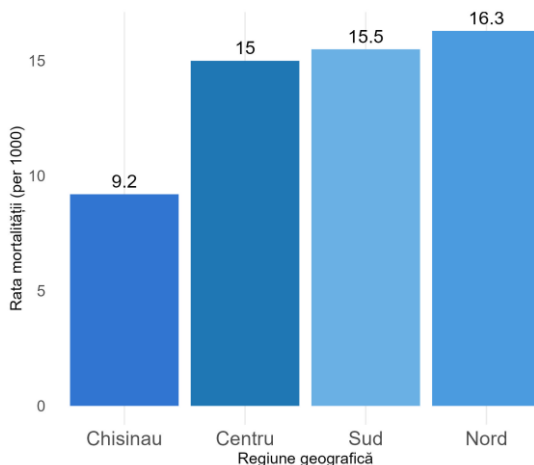


Figura 2.3. Rata brută a mortalității pe regiuni geografice în Republica Moldova în 2023, la 1.000 de locuitori.

2.4.3. Grafice circulare

Graficele circulare ilustrează proporția fiecărei valori în raport cu întregul și sunt utilizate pentru a reprezenta distribuțiile de frecvențe ale datelor nominale.

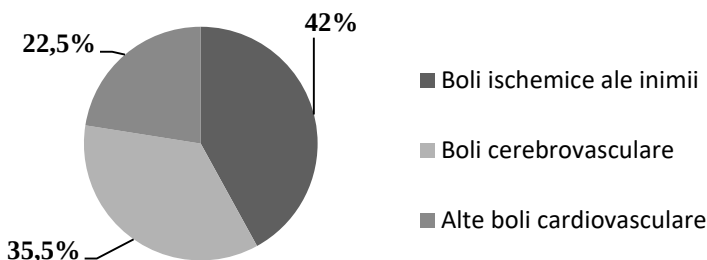


Figura 2.4. Distribuția globală a deceselor cardiovasculare cauzate de infarctul miocardic, accidentele vasculare cerebrale și alte tipuri de boli cardiovasculare în anul 2020

2.4.4. Histograme și poligoane de frecvență

Histogramele sunt metode extrem de eficiente pentru prezentarea frecvențelor absolute și relative ale datelor numerice. Deși prezintă asemănări cu graficele cu bare, ele se disting prin câteva caracteristici esențiale:

- Graficele cu bare ilustrează distribuția de frecvențe pentru date categoricale (nominale sau ordinale), în timp ce histogramele reflectă distribuția de frecvențe pentru date numerice (discrete sau continue).
- Histogramele dezvăluie forma distribuției de frecvențe, spre deosebire de graficele cu bare, care doar numără valorile fără a evidenția forma distribuției.

Într-o histogramă, axa orizontală indică limitele intervalelor, iar axa verticală reprezintă frecvența absolută sau relativă.

Poligonul de frecvență, asemănător cu histograma, dar utilizând puncte conectate prin linii drepte în loc de bare, furnizează o reprezentare grafică a distribuției setului de date. Pentru a construi un poligon de frecvență, trebuie să plasați mijlocul fiecărui interval pe axa orizontală și frecvențele corespunzătoare pe axa verticală, apoi să conectați aceste puncte cu linii drepte. Poligoanele de frecvență și histogramele pot fi suprapuse pentru analize comparative, așa cum este ilustrat în *figura 2.5* cu setul de date din *tabelul 2.6*.

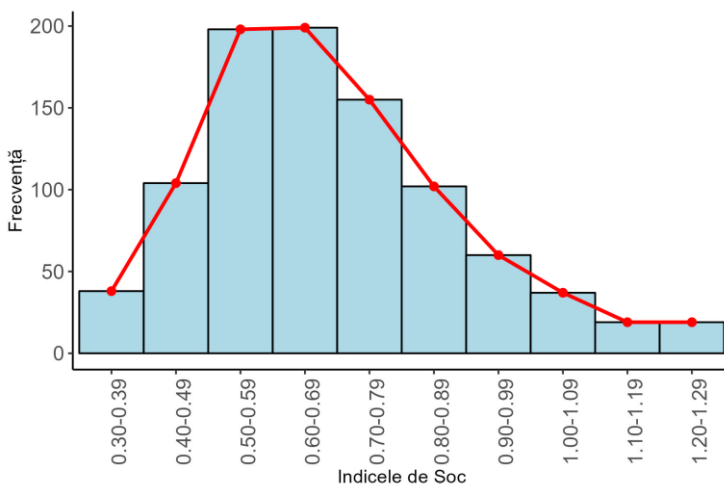


Figura 2.5. Histograma și poligonul de frecvență pentru distribuția indicelui de șoc la 931 de pacienți

2.4.5. Graficele cu cutie (box plot)

Graficele cu cutie, cunoscute și sub denumirea de grafice cu cutie și mustăți (*box-and-whisker plots*), reprezintă un instrument valoros pentru rezumarea distribuției unuia sau mai multor seturi de date numerice. Spre deosebire de alte tipuri de grafice, graficele cu cutie oferă o imagine concisă a distribuției datelor, fără a arăta fiecare punct individual.

Figura 2.6 ilustrează un grafic cu cutie și componentele sale. Cutia centrală, care poate fi orientată vertical sau orizontal, se întinde de la percentila 25 la percentila 75 a setului de date. O linie în interiorul cutiei reprezintă mediana (percentila 50), reflectând tendința centrală a datelor. Atunci când mediana se află aproape de centru între quartile, aceasta sugerează o distribuție simetrică a datelor.

Liniile care se extind de la cutie, cunoscute sub numele de mustăți, indică intervalul valorilor tipice din setul de date. Aceste mustăți se

extind cu până la 1,5 ori în raport cu intervalul interquartil (diferența dintre percentila 75 și percentila 25) deasupra și dedesubtul cutiei. Punctele de date care se află în afara acestor mustăți sunt marcate cu cercuri, indicând valori atipice sau valori care depășesc intervalul tipic.

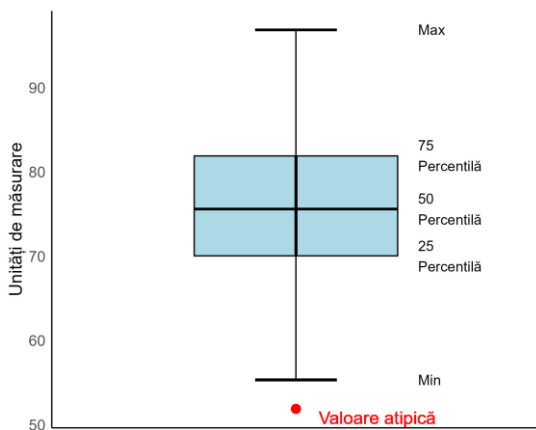


Figura 2.6. Grafic cu cutie și mustăți (box-and-whisker plot) și componentele sale

2.4.6. Grafice cu bare de eroare

Graficele cu bare de eroare sunt un instrument frecvent utilizat în cercetarea medicală pentru compararea grupurilor. Acestea reprezintă media cu un cerc și ilustrează variabilitatea folosind fie abaterea standard, fie eroarea standard (*figura 2.7*). Graficele cu bare de eroare oferă perspective asupra asemănărilor în distribuțiile între grupuri.

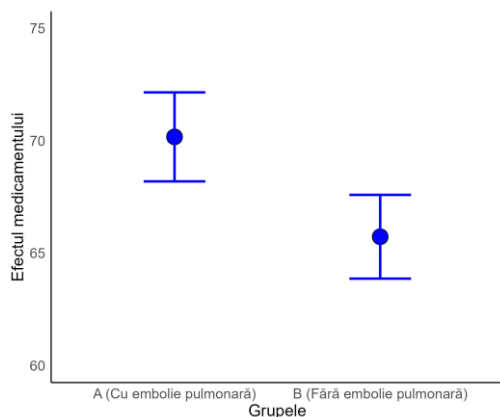


Figura 2.7. Grafice cu bare de eroare privind efectul medicamentului pentru pacienți cu (A) și fără (B) embolie pulmonară

2.4.7. Grafice de dispersie

Un grafic de dispersie este utilizat pentru a ilustra relația dintre două variabile numerice diferite. Fiecare punct de pe grafic reprezintă simultan o pereche de valori (*figura 2.8*).

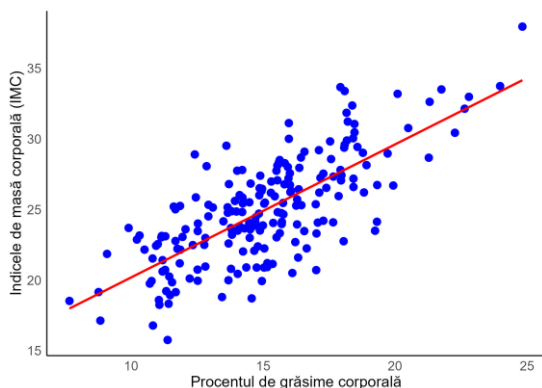


Figura 2.8. Grafic de dispersie care ilustrează relația între IMC și procentul de grăsime corporală

Exerciții de recapitulare

1. Specificați tipul variabilei și scala de măsurare adecvată pentru următoarele seturi de date:
 - a) Salariile a 125 de medici dintr-o clinică.
 - b) Notele la teste ale tuturor mediciniștilor care susțin examenele de iarnă într-un anumit an.
 - c) Nivelul colesterolului seric al persoanelor sănătoase.
 - d) Prezența diareei într-un grup de sugari.
 - e) Vârsta de debut a cancerului de sân la femei.
 - f) Temperatura corporală a pacienților.
 - g) Rezultatul tratamentului pacienților externați.
 - h) Numărul nașterilor într-un an dat.
2. Utilizați datele de mai jos pentru a le reprezenta grafic prin toate graficele adecvate. Justificați decizia dumneavoastră.

- a) Divergența dintre diagnosticele clinice și patologice în spital, 2017-2021

Ani	2017	2018	2019	2020	2021
Divergență, %	11	9,8	8,0	9,2	8,2

- b) Morbidity prin hepatite virale acute în Republica Moldova, 2021

Tip	A	B	C	D	E
%	34,4	41,4	17,6	3,8	2,8

1. Propuneți un set de date care poate fi reprezentat printr-un grafic cu linii. Justificați decizia dumneavoastră.
2. Propuneți un set de date care poate fi reprezentat printr-un grafic cu bare. Justificați decizia dumneavoastră.

Întrebări de recapitulare

1. Ce sunt statisticile descriptive?
2. Care sunt clasificările variabilelor? Prezentați exemple.
3. Cum se deosebește o variabilă alternativă de una non-alternativă? Oferiți exemple.
4. Definiți scalele de măsurare și clasificați-le. Dați câte un exemplu pentru fiecare tip.
5. Cum se diferențiază datele ordinale de datele nominale? Dați exemple.
6. Ce tipuri de metode de prezentare a datelor cunoașteți? Explicați diferențele dintre ele.
7. Când ar putea fi benefic să folosiți frecvențele relative în locul frecvențelor absolute atunci când construiți un tabel?
8. Descrieți prezentarea grafică a datelor: conținut și tipuri. Dați câte un exemplu pentru fiecare tip.
9. Care este graficul adecvat pentru prezentarea variabilelor nominale? Dați exemple.
10. Care este graficul adecvat pentru prezentarea variabilelor ordinare? Dați exemple.
11. Care este graficul adecvat pentru prezentarea variabilelor numerice? Dați exemple.

CAPITOLUL 3. STATISTICA DESCRIPTIVĂ.

REZUMAREA DATELOR NUMERICE

Concepte cheie

❖ *Statistica descriptivă* are rolul de a descrie, organiza și sumariza datele.

❖ *Modul, mediana și media* sunt folosite pentru date numerice (de interval și de raport). Mediana se aplică și pentru date ordinale.

❖ Modul și mediana sunt *insensibile* la valorile extreme dintr-o distribuție, spre deosebire de medie, care este *sensibilă* la acestea.

❖ Dacă valorile originale ale unei distribuții nu sunt disponibile, *media ponderată* poate fi estimată pe baza unui tabel de frecvență.

❖ Într-o distribuție *normală* (în formă de clopot), toate cele trei măsuri de tendință centrală sunt egale.

❖ Într-o distribuție *pozitivă* (asimetrică spre dreapta), modul este mai mic decât mediana.

❖ Într-o distribuție *negativă* (asimetrică spre stânga), mediana este mai mică decât modul.

❖ *Media* se utilizează pentru date numerice și distribuții normale.

❖ *Mediana* se aplică pentru date ordinale sau pentru date numerice în cazul distribuțiilor asimetrice.

❖ *Modul* este utilizat în principal pentru distribuțiile bimodale.

❖ O distribuție *unimodală* are un singur mod, una *bimodală* are două moduri, iar o distribuție *uniformă* nu are mod.

❖ *Amplitudinea* este sensibilă la valorile extreme dintr-o distribuție.

❖ *Intervalul intercuartil (IQR)* este utilizat pentru a descrie centrul de 50% dintr-o distribuție, indiferent de forma acesteia.

❖ IQR reprezintă diferența dintre percentila a 75-a și percentila a 25-a.

❖ *Varianța (S^2)* reflectă măsura variabilității unui set de date în jurul mediei.

❖ *Deviația standard (S)* indică distanța medie între valorile individuale ale unei distribuții și media calculată pentru aceasta.

❖ *Coeficientul de variație (CV)* este o măsură a dispersiei relative care permite compararea observațiilor măsurate pe scale diferite.

❖ *Deviația (abaterea) standard* se utilizează atunci când se folosește media (date numerice simetrice).

❖ *IQR* se utilizează atunci când se folosește mediana (date ordinale sau date numerice asimetrice).

❖ *Regula empirică* (regula 68-95-99,7) poate fi aplicată numai dacă datele urmează o distribuție normală.

❖ *Coeficientul de asimetrie al lui Pearson* este utilizat pentru a determina asimetria într-un eșantion.

Statisticile descriptive sunt utilizate pentru a organiza și descrie caracteristicile unui set de date. Spre deosebire de statisticile inferențiale, statisticile descriptive nu implică testarea ipotezelor sau analiza datelor. Ele ne permit să caracterizăm într-un mod concis distribuția valorilor în ansamblul său.

Statisticile descriptive pentru sumarizarea datelor numerice:

⇒ *Măsurile tendinței centrale:* media, mediana și modul.

⇒ *Măsurile variabilității (dispersiei):* amplitudinea, intervalul intercuartil, varianța, deviația standard, coeficientul de variație.

3.1. Măsuri de tendință centrală

Măsurile de tendință centrală reprezintă indicatori numerici ce caracterizează centrul unui set de date, unde observațiile tind să se aglomereze. Cele trei măsuri utilizate frecvent în sănătate sunt media aritmetică, mediana și modul. Toate trei sunt utilizate pentru însumarea datelor numerice, iar mediana este aplicată și în cazul datelor ordinale.

3.1.1. Media

Cea mai frecvent utilizată măsură a tendinței centrale este media aritmetică. Media este notată cu $\bar{X}$ și se calculează prin împărțirea sumei (Σ) valorilor individuale (X_i) la numărul de observații (n):

a) *Media simplă* – utilizată pentru distribuții de frecvențe negrupate (nestructurate).

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

b) *Media ponderată* – utilizată pentru distribuții de frecvențe grupate (structurate):

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i f$$

unde f reprezintă frecvența valorilor individuale.

Media este adesea utilizată pentru descrierea datelor numerice distribuite normal. Este extrem de sensibilă la valorile extreme din setul de date, cunoscute și sub denumirea de valori aberante. De exemplu, media setului de date (1;2;2;3) este 8/4 sau 2. Dacă numărul 3 este înlocuit cu 19, setul de date devine (1;2;2;19), iar media este 24/4 sau 6. Astfel, media de 2 este mai adecvată pentru setul de date decât media de 6.

3.1.2. Mediana

Mediana (M_d) împarte setul de date ordonat în două părți egale. Mediana este punctul de mijloc în setul ordonat de date de observație, unde jumătate dintre observații sunt mai mici, iar cealaltă jumătate sunt mai mari. Spre deosebire de media aritmetică, mediana este mai puțin influențată de valorile extreme. Este adesea utilizată pentru a estima centrul distribuției unei variabile ordinale sau numerice asimetrice. În cazul în care distribuția datelor nu este simetrică, mediana este considerată cea mai adecvată măsură a tendinței centrale.

Pentru a calcula mediana, se ordonează observațiile în ordine crescătoare:

- ⇒ Dacă numărul de observații este *impar*, mediana M_d va fi valoarea de mijloc din setul de date ordonate sau observația de la poziția $[(n+1)/2]$. De exemplu, mediana unui set de date cu $n=5$ (1; 2; 4; 5; 6) este $M_d = 4$.
- ⇒ Dacă numărul de observații este *par*, mediana M_d va fi media aritmetică a celor două valori centrale. De exemplu, mediana unui set de date cu $n=4$ (1; 2; 4; 5) este $M_d = 3$.

3.1.3. Modul

Modul (M_o) reprezintă valoarea care apare cel mai frecvent într-un set de date. De exemplu, în setul de date (3; 4; 5; 6; 6; 6; 7; 8; 9), $M_o=6$. Dacă toate valorile sunt diferite (distribuție uniformă), modul nu există. Într-o distribuție pot exista mai multe moduri, caz în care distribuția este numită bimodală sau multimodală. Modul este rar utilizat în practică.

3.1.4. Relația empirică între măsurile tendinței centrale

Cea mai adecvată măsură a tendinței centrale pentru un set de date specific depinde de forma distribuției acestora:

⇒ *Distribuție normală.* Într-o distribuție normală, valorile datelor sunt simetrice în jurul centrului și formează o curbă în formă de clopot. Media, mediana și modul ar trebui să fie aproximativ egale. În acest caz, media reprezintă cea mai adecvată măsură a tendinței centrale.

$$\bar{X} = M_d = M_o$$

⇒ *Distribuție asimetrică.* Într-o distribuție asimetrică, valorile extreme apar într-o direcție. Mediana este cea mai precisă măsură a tendinței centrale în astfel de distribuții. Media este sensibilă la valorile extreme și poate fi considerabil supraestimată sau subestimată.

Există două tipuri de distribuții asimetrice:

- *Asimetrie negativă* (asimetrică spre stânga) - valorile extreme sunt reduse.

$$\bar{X} < M_d < M_o$$

- *Asimetrie pozitivă* (asimetrică spre dreapta) - valorile extreme sunt ridicate.

$$M_o < M_d < \bar{X}$$

Când datele sunt asimetrice spre dreapta, media se situează la dreapta medianei, iar când sunt asimetrice spre stânga, media se situează la stânga medianei (*figura 3.1*).

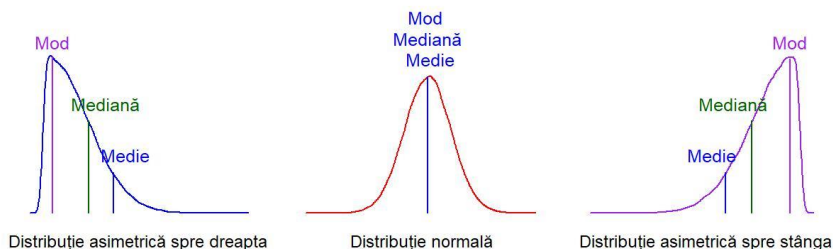


Figura 3.1. Posibile distribuții ale datelor

Orientări pentru alegerea măsurii potrivite de tendință centrală:

Aplicarea corectă a măsurilor de tendință centrală depinde de scala de măsurare și de forma distribuției.

1. Media este utilizată pentru date numerice cu o distribuție simetrică (normală).
2. Mediana este utilizată pentru date ordinale sau date numerice cu o distribuție asimetrică.
3. Modul este folosit aproape întotdeauna pentru distribuțiile bimodale.

3.2. Măsuri ale variabilității

Măsurile tendinței centrale descriu doar centrul unui set de date. Pentru a evalua corect acuratețea acestei măsuri, este esențial să înțelegem variabilitatea datelor. Aceasta presupune cunoașterea faptului dacă observațiile sunt similare și concentrate în jurul centrului sau dacă sunt dispersate pe un interval larg de valori. În exemplul de mai jos, observăm două distribuții foarte diferite de valori ale datelor, unde media, mediana și modul sunt egale:

Setul de date 1: -200; -20; -10; 7; 10; 20; 200 ($n=7$; $\bar{X}=1$; $M_d=7$)

Setul de date 2: -20; -5; -2; 7; 2; 5; 20 ($n=7$; $\bar{X}=1$; $M_d=7$)

În pofida diferențelor semnificative între seturile de date, măsurile tendinței centrale rămân aceleași. Acest exemplu subliniază importanța utilizării măsurilor tendinței centrale în paralel cu măsurile variabilității pentru a descrie adecvat un set de date. Pentru a descrie complet variabilitatea datelor, se folosesc măsuri precum amplitudinea, intervalul intercuartil, varianța, abaterea standard și coeficientul de variație.

3.2.1. Amplitudinea

Amplitudinea reprezintă diferența dintre valorile maximă și minimă dintr-un set de date.

Amplitudinea (R) = $\text{Max}(X_i) - \text{Min}(X_i)$.

Exemplu:

Set de date 1: (-200; -20; -10; 7; 10; 20; 200); $n=7$; $\bar{X}=1$

Set de date 2: (-20; -5; -2; 7; 2; 5; 20); $n=7$; $\bar{X}=1$

$R_1 = 400$

$R_2 = 40$.

Amplitudinea este influențată considerabil de valorile extreme și nu ia în considerare restul distribuției. Acest indicator este extrem de sensibil la valori excepțional de mari sau mici și este utilizat pentru date numerice cu scopul de a evidenția valorile extreme.

3.2.2. Intervalul intercuartil

Intervalul intercuartil (IQR) reduce influența valorilor extreme în setul de date. Cuartilele sunt valorile care împart un set de date ordonat în patru segmente egale. IQR se calculează ca diferența dintre percentila a 75-a ($Q3$) și percentila a 25-a ($Q1$), reprezentând cele 50% observații centrale.

$$IQR = Q3 - Q1$$

Percentila a 50-a (Q_2) este mediana setului de date, indicând punctul median al acestuia. *Figura 3.2* ilustrează cuartilele și intervalul intercuartil utilizând un grafic cu cutie cu mustăți (box-and-whisker plot).

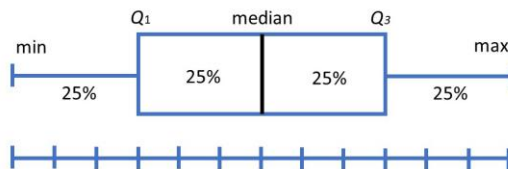


Figura 3.2. Prezentarea grafică a cuartilelor și a intervalului intercuartil

Pentru a calcula cuartilele, datele trebuie ordonate de la cea mai mică la cea mai mare valoare:

$$Q_1 = (n+1) / 4$$

$$Q_2 = (n+1) / 2$$

$$Q_3 = 3(n+1) / 4$$

Prima cuartilă reprezintă valoarea mediană dintre cea mai mică valoare și mediana distribuției, indicând percentila de 25%. A doua cuartilă corespunde mediane distribuției și indică percentila de 50%. A treia cuartilă este valoarea mediană dintre mediană și cea mai mare valoare din distribuție, indicând percentila de 75%.

Intervalul intercuartil este utilizat atunci când se folosește mediana (pentru date ordinale sau date numerice asimetrice). Este potrivit pentru descrierea celor 50% centrale ale distribuției, indiferent de forma acesteia.

3.2.3. *Varianța și abaterea standard*

Varianța (S^2) cuantifică gradul de variabilitate sau dispersie în jurul mediei. Aceasta se calculează ca media deviațiilor pătrate ale fiecărui număr față de medie. Pentru date negrupate, varianța se calculează folosind formula:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Pentru date grupate, varianța se calculează luând în considerare frecvența:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 f_i$$

unde:

X_i – valorile individuale ale setului de date

$\bar{X}$ - media

n – numărul de observații

f_i – frecvența

Deviația standard (S) reprezintă rădăcina pătrată a varianței și măsoară dispersia în jurul mediei. Pentru date negrupate:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

Pentru date grupate:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2 f_i}{n-1}}$$

unde:

S^2 – varianța

Deviația standard se utilizează atunci când se folosește media (date numerice simetrice) și, împreună cu media, rezumă caracteristicile întregii distribuții a valorilor. Deviația standard are aceleași unități de măsură ca și media.

3.2.4. Coeficientul de variație

Coeficientul de variație (CV) reprezintă o măsură relativă a variabilității și este utilizat pentru a compara distribuțiile măsurate pe scale diferite. CV-ul este definit ca raportul dintre deviația standard și media aritmetică, înmulțit cu 100%.

$$CV = \frac{S}{\bar{X}} \times 100\%$$

unde:

S – deviația standard

$\bar{X}$ – media

CV-ul este valoros pentru compararea nivelurilor de variabilitate în diverse seturi de date și se exprimă ca procent. CV-ul poate fi utilizat pentru a estima nivelul de variabilitate într-un set de date conform scalei prezentate în *tabelul 3.1*.

Tabelul 3.1. Scala coeficientului de variație

Coeficientul de variație, %	Nivelul de variabilitate
<10	Scăzut
10-35	Mediu
>35	Ridicat

În cercetările științifice, un set de date ar trebui să prezinte o variabilitate scăzută sau medie pentru a asigura că media aritmetică este reprezentativă pentru setul de date.

Orientări pentru alegerea măsurii de variabilitate adecvate:

1. Deviația standard se folosește atunci când se utilizează media aritmetică (date numerice simetrice).
2. Intervalul intercuartil se folosește atunci când:
 - a) se utilizează mediana (date ordinale sau date numerice asimetrice);
 - b) se compară observații individuale cu un set de norme;

- c) se descriu 50% centrale ale unei distribuții, indiferent de formă.
- 3. Amplitudinea se folosește cu date numerice pentru a sublinia valorile extreme.
- 4. Coeficientul de variație se folosește pentru a compara variabilitatea diferitelor seturi de date.

3.3. Distribuția normală și proprietățile acesteia

Distribuțiile normale prezintă caracteristici esențiale, ușor de identificat pe un grafic, așa cum este ilustrat în *figura 3.3*:

- 1. Media, mediana și modul coincid.
- 2. Distribuția este simetrică față de media aritmetică: jumătate din valori sunt mai mici decât media și jumătate sunt mai mari.
- 3. Distribuția poate fi descrisă prin două valori: media și deviația standard.
- 4. Suprafața totală sub curbă este egală cu 1.

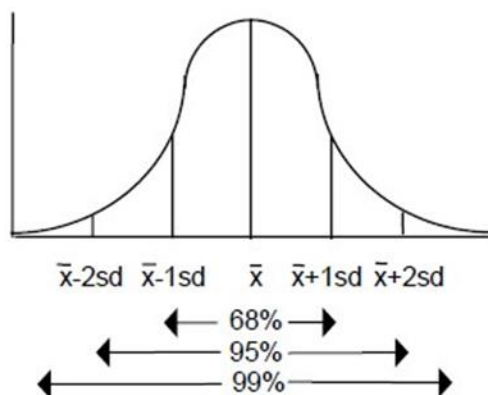


Figura 3.3. Proprietățile distribuției normale

Dacă datele sunt distribuite normal (formând o curbă în formă de clopot), aproximativ 68% din date vor fi situate în intervalul unei deviații

standard, aproximativ 95% în intervalul a două deviații standard și aproximativ 99,7% în intervalul a trei deviații standard.

3.4. Asimetria și kurtosis

Asimetria reprezintă gradul de abatere al unei distribuții în cadrul unui set de date față de valoarea medie, indicând direcția și măsura în care distribuția se abate de la o distribuție normală simetrică (curbă în formă de clopot). Există diverse formule pentru măsurarea asimetriei. Cei mai des utilizați sunt următorii doi coeficienți de asimetrie Pearson (S_k) pentru date numerice (de interval sau de raport).

Primul coeficient de asimetrie Pearson (asimetrie după mod):

$$S_k = \frac{\text{Media} - \text{Modul}}{\text{Deviația standard}} = \frac{\bar{X} - M_o}{S}$$

Atunci când se utilizează primul coeficient de asimetrie Pearson, valorile variază între -1 și +1. Valoarea 0 indică o distribuție perfect simetrică. Valorile apropiate de -1 sau +1 indică distribuții asimetrice din ce în ce mai negative sau pozitive.

Al doilea coeficient de asimetrie Pearson (asimetrie după mediană):

$$S_k = \frac{3 \times (\text{Media} - \text{Mediana})}{\text{Deviația Standard}} = \frac{3 \times (\bar{X} - M_d)}{S}$$

Atunci când se utilizează al doilea coeficient de asimetrie Pearson, valorile variază între -3 și +3. Valoarea 0 indică o distribuție perfect simetrică. Valorile apropiate de -3 sau +3 indică distribuții asimetrice din ce în ce mai negative sau pozitive.

Interpretarea celui de-al doilea coeficient de asimetrie Pearson:

$S_k = 0$: Distribuția este perfect simetrică.

S_k este între -0,5 și 0,5: Distribuția este aproape simetrică.

S_k este între -1 și -0,5: Asimetrie negativă moderată.

S_k este între 0,5 și 1: Asimetrie pozitivă moderată.

S_k este mai mic de -1 sau mai mare de 1: Datele sunt puternic asimetrice.

Kurtosisul (boltirea sau aplatizarea) reprezintă o altă măsură a formei unei distribuții și se referă la forma vârfului curbei (*figura 3.4*). În timp ce asimetria evaluează gradul de nesimetrie, kurtosisul semnifică gradul de înălțime al vârfului unei distribuții de frecvențe. Kurtosisul este utilizat pentru a identifica prezența valorilor aberante în date.

Există trei tipuri de kurtosis:

- ⇒ *Platykurtic* – vârful curbei este plat comparativ cu cel normal, iar cozile sunt lungi (kurtosis negativ).
- ⇒ *Mesokurtic* – vârful curbei este normal, iar cozile de pe ambele părți ale mediei sunt, de asemenea, normale.
- ⇒ *Leptokurtic* – vârful curbei este îngust, iar cozile de pe ambele părți ale mediei sunt scurte (kurtosis pozitiv).

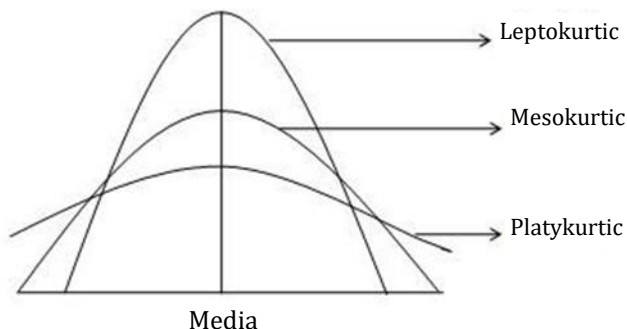


Figura 3.4. Tipuri de kurtosis

3.5. Exemplu de calcul

Să presupunem că vârstele celor 19 pacienți studiați sunt următoarele:

31; 24; 26; 30; 24; 35; 35; 31; 35; 33; 26; 33; 26; 31; 26; 30; 31; 31; 30.

Calculați măsurile de tendință centrală și variabilitate. Formulați decizia dvs.

1. Ordonăți datele:

24; 24; 26; 26; 26; 26; 30; 30; 30; 31; 31; 31; 31; 31; 33; 33; 35; 35; 35.

2. Aranjați datele în tabelul de frecvență:

X_i	Frecvență, f	$x_i f_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})^2 f_i$
24	2	48	-5,8	33,64	67,28
26	4	104	-3,8	14,44	57,76
30	3	90	+0,2	0,04	0,12
31	5	153	+1,2	1,44	7,2
33	2	66	+3,2	10,24	20,48
35	3	105	+5,2	27,04	81,12
Total	n=19	566			233,96

3. Calculați media:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i f_i$$

$$\bar{X} = \frac{24 \times 2 + 26 \times 4 + 30 \times 3 + 31 \times 5 + 33 \times 2 + 35 \times 3}{19} = \frac{566}{19} = 29,8$$

4. Determinați modul:

$$\text{Modul } (M_o) = 31$$

5. Calculați mediana

Deoarece $n = 19$ (impar), poziția mediane este $\frac{n+1}{2} = \frac{19+1}{2} = 10$.

A zecea observație în datele ordonate este 31.

$$\text{Mediana } (M_d) = 31$$

6. Calculați varianța și abaterea standard:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 f_i = \frac{233,96}{18} = 13$$

$$s = \sqrt{S^2} = \sqrt{13} \approx 3,6$$

7. Calculați coeficientul de variație:

$$CV = \frac{s}{\bar{x}} \times 100\%$$

$$CV = \frac{3,6}{29,8} \times 100\% = 12,1\%$$

8. Concluzii

- Vârsta medie a pacienților este de $29,8 \pm 3,6$ ani.
- Nivelul de variabilitate a datelor este mediu ($CV=12,1\%$), adecvat pentru cercetările științifice.
- Distribuția datelor este aproape simetrică, cu valori ale modului, mediei și mediane apropiate, având o ușoară asimetrie spre stânga.
- Media este reprezentativă pentru acest set de date.

Exerciții de recapitulare

1. Folosind următorul set de date privind IMC-ul a 10 pacienți:
Date: 29, 22, 24, 37, 23, 35, 35, 27, 50, 37.
 - Găsiți măsurile tendinței centrale.
 - Analizați relația dintre măsurile tendinței centrale și identificați forma distribuției. Formulați concluzia dumneavoastră.

2. Folosind următorul set de date privind nivelul de glucoză în sânge (mmol/l) a 11 pacienți:
Date: 3,3, 12,0, 9,0, 6,0, 11,0, 11,8, 11,8, 11,0, 5,5, 3,3.
 - Găsiți toate măsurile tendinței centrale.
 - Analizați relația dintre măsurile tendinței centrale și identificați forma distribuției. Formulați concluzia dumneavoastră.

3. Folosind setul de date privind IMC-ul din exercițiul 1:
 - Găsiți toate măsurile variabilității. Formulați concluzia dumneavoastră.
 - Determinați cât de reprezentativă este media pentru acest set de date.
 - În funcție de scala de măsurare și de forma distribuției, decideți care măsuri de tendință centrală sunt mai adecvate.
 - Construiți un grafic cu cutie cu mustăți (box-and-whisker plot).

4. Folosind setul de date privind nivelul de glucoză din sânge din exercițiul 2:

- Găsiți toate măsurile variabilității. Formulați concluzia dumneavoastră.
- Determinați cât de reprezentativă este media pentru acest set de date.
- În funcție de scala de măsurare și de forma distribuției, decideți care măsuri de tendință centrală sunt mai adecvate.

Întrebări de recapitulare

1. Media. Definiție, tipuri și reguli de calcul. Dați un exemplu.
2. Mediana. Definiție și reguli de calcul. Dați un exemplu.
3. Modul. Definiție și reguli de calcul. Dați un exemplu.
4. Comparați media, mediana și modul ca măsuri ale tendinței centrale.
5. În ce condiții este preferată utilizarea mediei?
6. În ce condiții este preferată utilizarea medianei?
7. În ce condiții este preferată utilizarea modului?
8. Măsuri ale variabilității. Motive de aplicare.
9. Amplituda. Definiție, caracteristici și condiții preferate de utilizare. Regula de calcul.
10. Quartilele. Definiție, caracteristici și regula de calcul.
11. Intervalul interquartil. Definiție, caracteristici și condiții preferate de utilizare. Regula de calcul.
12. Deviația standard. Semnificație, caracteristici și condiții preferate de utilizare. Regula de calcul.
13. Coeficientul de variație. Definiție, caracteristici și condiții preferate de utilizare. Regula de calcul.
14. Definiți distribuția normală și proprietățile acesteia.

CAPITOLUL 4. STATISTICA DESCRIPTIVĂ: REZUMAREA DATELOR NOMINALE ȘI ORDINALE

Concepte Cheie

- ❖ *Datele calitative (catoriale)* pot fi măsurate prin diverse metode: rapoarte, proporții și rate.
- ❖ *Un raport* reprezintă numărul de observații care prezintă o caracteristică specifică „a” împărțit la numărul de observații care nu au această caracteristică „b”.
- ❖ *O proporție* reprezintă numărul de observații care prezintă o caracteristică specifică „a” împărțit la numărul total de observații „a+b”.
- ❖ *O rată* este asemănătoare cu proporția, dar utilizează un multiplicator (de exemplu, 1000, 10000 sau 100000) și este calculată pe o perioadă determinată.
- ❖ *Diagramele cu bare/coloane și diagramele liniare* sunt utilizate pentru a reprezenta grafic rapoartele și ratele.
- ❖ *Diagrama circulară* este utilizată pentru a reprezenta grafic proporția.
- ❖ *Proporția* reflectă structura unui fenomen și constituie un indicator static.
- ❖ *Rata* reflectă frecvența (nivelul sau intensitatea) unui fenomen de-a lungul timpului și constituie un indicator dinamic.
- ❖ *Ratele* sunt esențiale în epidemiologie și în medicina bazată pe dovezi; ele formează baza pentru calcularea statisticilor vitale, care descriu starea de sănătate a populațiilor.
- ❖ *Ratele* sunt de obicei calculate pentru un an (rate anuale).
- ❖ *Ratele* pot fi brute, specifice sau standardizate.
- ❖ *Ratele* trebuie standardizate atunci când se compară populații cu diferențe semnificative în structura lor (de exemplu, după vârstă și sex).
- ❖ *Prevalența și incidența* sunt două măsuri importante ale morbidității.

4.1. Metode de descriere a datelor categoriale

În cercetare, informațiile statistice sunt de regulă prezentate sub formă de valori absolute. Aceste valori pot fi dificil de interpretat, deoarece nu permit realizarea unor comparații, sinteze sau corelații semnificative între diferite caracteristici. Pentru a realiza comparații mai relevante între grupuri, se pot utiliza valori relative în locul numerelor absolute. Datele categoriale (calitative) pot fi măsurate prin trei metode:

- proporții și procente;
- raporturi;
- rate.

4.1.1. Proporții și procente

Proporțiile reprezintă indicatori statistici care reflectă structura unui fenomen. Ele sunt definite ca partea fenomenului împărțită la întregul fenomen. Proporția se calculează împărțind numărul de observații cu caracteristica de interes (a) la numărul total de observații ($a+b$). Procentele sunt echivalentul proporțiilor înmulțite cu 100%.

$$\text{Proporția (procentul)} = \left(\frac{a}{a+b} \right) \times 100\%$$

unde:

- a reprezintă numărul de observații cu caracteristica de interes;
- b reprezintă numărul de observații fără caracteristica respectivă.

Exemplu de proporție

Mortalitatea proporțională

$$= \frac{\text{Numărul de decese cauzate de X într – un an}}{\text{Numărul total de decese într – un an}} \times 100$$

Numărul de decese cauzate de cancer în Republica Moldova în 2023 = 5.959.

Numărul total de decese în Republica Moldova în 2023 = 33733.

$$\text{Mortalitatea proporțională prin cancer în 2023} = \frac{5959}{33733} \times 100 = 18\%$$

Proporțiile sunt denumite indicatori extensivi, deoarece ilustrează structura unui fenomen. Ele sunt indicatori statici care oferă o imagine de ansamblu a fenomenului într-un moment dat și nu reflectă schimbările dinamice. Acestea sunt utile pentru date ordinale și numerice, precum și pentru date nominale, în special atunci când observațiile sunt prezentate într-un tabel de frecvență.

Reprezentare grafică: diagramă circulară.

4.1.2. Raporturi

Un raport se definește ca o parte împărțită la altă parte, reprezentând două fenomene independente. Raportul se calculează împărțind numărul observațiilor într-un grup cu o caracteristică specifică (a) la numărul observațiilor fără acea caracteristică (b):

$$\text{Raportul} = \frac{a}{b}$$

Exemple de raporturi :

$$\text{Raportul de sex} = \frac{\text{Numărul de bărbați}}{\text{Numărul de femei}}$$

Numărul de nou-născuți de sex masculin în Republica Moldova în 2023 = 12239

Numărul de nou-născuți de sex feminin în Republica Moldova în 2023 = 11794

$$\text{Raportul de sex la naștere în 2023} = \frac{12239}{11794} = 1,04$$

În acest exemplu, raportul nu are unități. Totuși, raportul poate fi redimensionat prin înmulțirea sa cu un multiplicator, cum ar fi 100 sau 1.000. Raportul de sex la naștere în Republica Moldova în 2023 a fost de 1,04 sau 104% (1,04 x 100%), ceea ce înseamnă 104 nou-născuți de sex masculin pentru fiecare 100 de nou-născuți de sex feminin.

$$\text{Asigurarea cu medici} = \frac{\text{Numărul de medici}}{\text{Populația totală}} \times \text{multiplicator}$$

Numărul de medici în Republica Moldova în 2022 = 12.600

Numărul total al populației în Republica Moldova în 2022 = 2.565.030

$$\begin{aligned} \text{Asigurarea cu medici în 2022} &= \frac{12600}{2565030} \times 1000 \\ &= 5 \text{ medici la } 1000 \text{ locuitori} \end{aligned}$$

Prezentare grafică: diagramă liniară, diagramă cu bare/coloane.

4.1.3. Rate

O rată este similară cu o proporție, dar include un multiplicator și o dimensiune temporală. Rata este întotdeauna calculată pentru o perioadă specifică, de obicei un an (rată anuală). Ratele sunt indicatori statistici intensivi care exprimă frecvența sau nivelul unui fenomen pe o perioadă determinată. Acestea sunt calculate după formula:

$$\text{Rata} = \left(\frac{a}{a+b} \right) \times \text{multiplicator}$$

unde:

- *a* reprezintă numărul de observații cu o caracteristică specifică (de exemplu, numărul de decese sau nașteri într-un an și loc);
- *a+b* reprezintă numărul total de observații (de exemplu, populația totală într-un an dat);
- multiplicatorul este un factor de scalare (de exemplu, 100; 1000; 10000; 100000).

Exemplu de rată

Într-un studiu care a durat un an, proporția pacienților care au dezvoltat o boală a fost de 0,02. Rata la 1.000 de pacienți ar fi $(0,02) \times (1000)$, adică 20 de cazuri de boală la 1.000 de pacienți pe parcursul unui an.

Prezentare grafică: diagramă liniară, diagramă cu bare/coloane.

4.2. Indicatorii stării de sănătate în statistica descriptivă

Indicatorii stării de sănătate evaluează sănătatea unei populații folosind statistici vitale. Există trei tipuri principale de indicatori:

- ⇒ Indicatori de mortalitate;
- ⇒ Indicatori de morbiditate;
- ⇒ Indicatori de dizabilitate.

Majoritatea indicatorilor stării de sănătate sunt reprezentați prin rate și, uneori, prin proporții și raporturi. Ratele de mortalitate și ratele de morbiditate sunt cele mai frecvent utilizate.

4.2.1. Ratele de mortalitate

Rata de mortalitate sau rata de deces este definită ca numărul de decese care au loc într-o perioadă specificată, împărțit la numărul total de persoane expuse riscului de deces în aceeași perioadă.

- ⇒ *Rata brută (generală)* - se calculează pentru toate persoanele dintr-o populație dată, fără a ține cont de diferențele cauzate de vârstă, sex, rasă etc.
- ⇒ *Rata specifică* - se calculează în cadrul subgrupurilor populaționale relativ mici și bine definite. De exemplu:
 - rata de deces specifică pe vârstă;
 - rata de deces specifică pe sex;
 - rata de deces specifică cauzei.

4.2.2. Ratele de morbiditate

Rata de morbiditate este definită ca numărul de persoane care au dezvoltat o boală într-o perioadă specificată, împărțit la numărul total de persoane expuse riscului în aceeași perioadă.

Incidența și prevalența sunt principalele măsuri ale morbidității și sunt frecvent utilizate pentru evaluarea stării de sănătate a populației în cadrul cercetărilor medicale și epidemiologice.

Incidența - numărul de cazuri noi care apar într-o perioadă determinată, împărțit la numărul total de persoane expuse riscului în aceeași perioadă.

Prevalența - numărul de persoane cu o anumită afecțiune la un moment dat, împărțit la numărul total al populației expuse riscului pentru acea afecțiune la acel moment.

Ratele de morbiditate oferă un mod standardizat de evaluare atât a ratelor generale, cât și a celor specifice.

Exemplu de calcul

Într-un an și într-o localitate specifică, populația era de 75.000 de persoane. În acel an, au fost înregistrate 897 de decese. În acea localitate erau 40 de medici: 20 de medici de familie, 10 chirurgi și 10 medici specialiști în alte domenii. Calculați indicatorii statistici care sintetizează datele nominale.

1. Proporție și procent

$$\text{Proporția (procentul) medicilor de familie} = \frac{20}{40} \times 100\% = 50\%$$

2. Raport

$$\begin{aligned} \text{Asigurarea cu medici} &= \frac{40}{75000} \times 10000 \\ &= 5,3 \text{ medici la } 10000 \text{ de locuitori} \end{aligned}$$

3. Rată

$$\begin{aligned} \text{Rata brută de mortalitate} &= \frac{897}{75000} \times 1000 \\ &= 11,9 \text{ decese la } 1000 \text{ de locuitori} \end{aligned}$$

4.3. Rate standardizate. Metoda directă de standardizare

Ratele brute pot fi utilizate pentru compararea a două populații diferite doar în cazul în care aceste populații sunt omogene în toate caracteristicile relevante. Totuși, dacă populațiile prezintă diferențe în caracteristici precum sexul sau vârsta, utilizarea ratelor brute poate conduce la rezultate inexacte. Aceasta se datorează influenței puternice a structurii pe vârste asupra ratelor brute. De exemplu, o populație poate avea o proporție mai mare de persoane vârstnice comparativ cu alta. În aceste situații, ratele brute trebuie ajustate sau standardizate pentru a permite comparații corecte.

Standardizarea ajustează ratele brute pentru a elimina efectele diferențelor în compoziția populației, în principal pe baza vârstei și sexului. Există două metode principale de standardizare: metoda directă și metoda indirectă. Metoda directă de standardizare calculează ratele care ar rezulta dacă toate grupurile comparate ar avea aceeași compoziție standard. Ratele standardizate sunt, așadar, valori convenționale create exclusiv pentru scopuri comparative și nu pot fi utilizate în mod independent.

Metoda directă de standardizare a ratelor include patru etape:

1. Calcularea ratelor specifice pentru fiecare grup.
2. Alegerea populației standard.
3. Calcularea numărului așteptat de evenimente pentru fiecare grup.
4. Calcularea ratelor standardizate.

Mai jos este prezentat un exemplu detaliat al calculului ratelor de deces standardizate pentru cele două regiuni.

Exemplu de calcul. Metoda directă de standardizare a ratelor de mortalitate.

Etapa 1. Calcularea ratelor specifice pentru fiecare grupă

Sexul	Regiunea A		Regiunea B		Rata de mortalitate specifică după sex (la 1.000)	
	Persoane	Decese	Persoane	Decese	Regiunea A	Regiunea B
Bărbați	50	1	170	4	20	24
Femei	200	10	30	3	50	100
Total	250	11	200	7	44	35

De exemplu, rata de mortalitate specifică după sex pentru bărbați în Regiunea A = $\frac{1}{50} \times 1000 = 20$ decese la 1000 bărbați.

Etapa 2. Selectarea populației standard

Sex	Regiunea A		Regiunea B		Populația standard
	Persoane	Decese	Persoane	Decese	Persoane din A + Persoane din B
Bărbați	50	1	170	4	220
Femei	200	10	30	3	230
Total	250	11	200	7	450

În acest exemplu, populația standard este suma populațiilor din ambele regiuni. Cu alte cuvinte, se presupune că structura populației pe sexe este identică în ambele regiuni.

Etapa 3. Calcularea numărului așteptat de evenimente pentru fiecare grup

Sex	Rata de mortalitate specifică după sex (la 1.000)		Populația standard	Evenimente așteptate (decese)	
	Regiunea A	Regiunea B	Persoane din A + Persoane din B	Regiunea A	Regiunea B
Bărbați	20	24	220	4	5
Femei	50	100	230	12	23
Total				16	28

$$\text{Numărul de evenimente așteptate} = \frac{\text{Rata specifică} \times \text{populație standard}}{1000}$$

De exemplu, numărul de decese așteptate pentru regiunea A la bărbați = $\frac{20 \times 220}{1000} = 4$ decese. Cu alte cuvinte, dacă numărul populației masculine din regiunea A ar fi populația standard (220 persoane), ne-am aștepta la 4 decese aici.

Etapa 4. Calcularea ratei standardizate pentru fiecare grup

Sex	Populația standard	Evenimente așteptate (decese)		Rata standardizată (la 1000)	
		Regiunea A	Regiunea B	Regiunea A	Regiunea B
Bărbați	220	4	5		
Femei	230	12	23		
Total	450	16	28	36	62

$$\text{Rata standardizată} = \frac{\text{Numărul total de evenimente așteptate}}{\text{Populația totală standard}} \times 1000$$

De exemplu, rata de mortalitate standardizată pentru regiunea A = $\frac{16}{450} \times 1000 = 36$ decese la 1000 de populație standard.

Concluzie. Comparația dintre ratele brute și standardizate

Rata brută (la 1.000)		Rata standardizată (la 1.000)	
Regiunea A	Regiunea B	Regiunea A	Regiunea B
44	35	35	63

Intensitatea (nivelul) mortalității este mai mare în regiunea B. Nivelul mortalității în regiunea B este de 1,8 ori mai mare decât în regiunea A ($63 \div 35 = 1,8$).

Exerciții de recapitulare

- În localitatea A, într-un anumit an, au fost înregistrate 2500 de cazuri de îmbolnăvire: 800 dintre acestea au fost boli cardiovasculare; 500 boli pulmonare; 450 traumatisme și 750 alte boli. Populația totală a fost de 900.000 de locuitori.
 - Calculați toate statisticile vitale posibile.
 - Realizați o prezentare grafică adecvată pentru aceste date.
- În localitatea B, într-un anumit an, populația a fost de 78000 de locuitori. În acest an, 110 persoane au decedat și 400 de persoane au dezvoltat pentru prima dată boli cardiovasculare.
 - Calculați toate statisticile vitale posibile.
 - Realizați o prezentare grafică adecvată pentru aceste date.
- Analizați următoarele date comparând letalitatea în abdomenul acut în spitalele „A” și „B”:

Durata spitalizării, ore	Spitalul „A”		Spitalul „B”	
	Numărul de pacienți	Numărul de cazuri letale	Numărul de pacienți	Numărul de cazuri letale
< 6	650	72	490	34
6-12	450	83	380	66
> 24	131	23	736	206
Total:	1231	178	1606	306

- Calculați ratele brute și comparați aceste rate.
- Cum diferă ratele standardizate de ratele brute în fiecare dintre aceste spitale? Explicați rezultatele (interpretare și concluzii).

4. Analizați următoarele date comparând mortalitatea spitalicească în spitalele „A” și „B”:

Boală	Spital „A”		Spital „B”	
	Numărul de pacienți	Numărul de decese	Numărul de pacienți	Numărul de decese
Gastrointestinală	1200	24	1700	40
Tumoră malignă	190	55	100	30
Cardiovasculară	160	100	1100	72
Total:	1650	179	2900	142

- Calculați ratele brute și comparați aceste rate.
- Cum diferă ratele standardizate de ratele brute în fiecare dintre aceste spitale? Explicați rezultatele obținute (interpretare și concluzii).

Întrebări de recapitulare

1. Valori absolute și relative.:Semnificația și aplicarea lor în biostatistică. În ce condiții este preferabilă utilizarea valorilor relative? Dați un exemplu.
2. Tipuri de mărimi relative. Care sunt diferențele și asemănările între acestea? Dați câte un exemplu pentru fiecare tip.
3. Proporții, raporturi și rate. Particularități, reguli de calcul, condiții de aplicare și prezentare grafică adecvată. Dați câte un exemplu.
4. Statistica vitală. Definiție și principalele rate utilizate.
5. Care este diferența între ratele brute și cele specifice?
6. Care este diferența între ratele de mortalitate și cele de morbiditate?
7. Care sunt diferențele și asemănările între prevalență și incidență?
8. În ce condiții este preferabilă utilizarea ratei standardizate?
9. Metoda directă de standardizare. Definiție și etapele procesului.
10. În ce circumstanțe ar trebui utilizate ratele brute, specifice și standardizate?

CAPITOLUL 5. CORELAȚIE ȘI REGRESIE

Concepte cheie

- ❖ *Coeficientul de corelație (r)* măsoară intensitatea și direcția relației liniare dintre două variabile.
- ❖ *Variabila X* - adesea denumită variabilă independentă sau explicativă.
- ❖ *Variabila Y* este cunoscută ca variabilă dependentă sau rezultativă.
- ❖ Valoarea coeficientului de corelație variază între -1 și +1.
- ❖ *Direcția corelației* este indicată de semnul (+ sau -) coeficientului.
- ❖ *Intensitatea corelației* este indicată de magnitudinea coeficientului.
- ❖ *Corelație pozitivă (+)* - valorile mari ale unei variabile sunt asociate cu valorile mari ale celeilalte variabile.
- ❖ *Corelație negativă (-)* - valorile mari ale unei variabile sunt asociate cu valorile mici ale celeilalte variabile.
- ❖ *Coeficientul de corelație Pearson (r)* evaluează relația liniară între două variabile numerice.
- ❖ *Coeficientul de corelație al rangurilor Spearman (r_s)* evaluează relația dintre două variabile folosind ordonarea rangurilor.
- ❖ *Regresia* prezice valoarea unei variabile dependente (Y) pe baza uneia sau mai multor variabile independente (X).
- ❖ *Regresia liniară* poate fi simplă sau multiplă.

Cercetările biomedicale explorează adesea relațiile dintre două sau mai multe variabile. De exemplu, există o corelație între consumul de sare și tensiunea arterială sau între fumat și speranța de viață? Înțelegerea acestor relații este esențială pentru identificarea asociațiilor și realizarea predicțiilor. Pentru analiza acestor relații, se aplică două

tehnici statistice esențiale: corelația și regresia.

1. *Corelația* - metodă utilizată pentru a determina și cuantifica intensitatea și direcția relației dintre două variabile.
2. *Regresia* - metodă utilizată pentru a exprima relația funcțională dintre două sau mai multe variabile. Analiza de regresie permite cercetătorilor să prezică valoarea unei variabile pe baza valorilor altor variabile.

5.1. Corelație

5.1.1. Tipuri de coeficienți de corelație

Coeficientul de corelație Pearson

Coeficientul de corelație Pearson este o măsură parametrică a relației dintre două variabile numerice cu distribuție normală. Variabila independentă sau explicativă este denumită „X”, iar variabila dependentă sau de rezultat este „Y”.

Coeficientul de corelație, simbolizat prin „r”, se calculează folosind formula:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

unde:

X_i - valoarea individuală a variabilei independente

Y_i - valoarea individuală a variabilei dependente

$\bar{X}$ - media variabilei independente

$\bar{Y}$ - media variabilei dependente.

Coeficientul de corelație este un număr fără dimensiune, adică nu are unități de măsură. Valoarea sa variază între -1 și 1. Pentru orice set de date, coeficientul de corelație (r) respectă inegalitatea:

$$-1 \leq r \leq 1$$

Direcția relației este indicată de semnul coeficientului, în timp ce intensitatea relației este indicată de magnitudinea coeficientului.

Direcția corelației:

- ⇒ *Corelația pozitivă* ($0 < r \leq 1$) indică faptul că o creștere a unei variabile tinde să fie asociată cu o creștere a celeilalte variabile. De exemplu, un consum crescut de sare corelează cu tensiunea arterială mai mare.
- ⇒ *Corelația negativă* ($-1 \leq r < 0$) semnalează că o creștere a unei variabile este asociată cu o scădere a celeilalte variabile. De pildă, un consum crescut de țigări corelează cu reducerea speranței de viață.

Valori specifice:

- $r = 1$ - indică o relație liniară pozitivă perfectă.
- $r = -1$ - indică o relație liniară negativă perfectă.
- $r = 0$ - indică absența unei relații liniare (fără corelație) între variabile.

Interpretarea intensității legăturii de corelație:

- ⇒ 0 până la 0,24 ($\pm$) - indică lipsa corelației sau corelație slabă;
- ⇒ 0,25 până la 0,49 ($\pm$) indică corelație moderată;
- ⇒ 0,50 până la 0,74 ($\pm$) indică corelație puternică;
- ⇒ 0,75 până la 1 ($\pm$) indică corelație foarte puternică;
- ⇒ ± 1 indică corelație perfectă.

Această scală oferă o înțelegere a puterii legăturii de corelație dintre cele două variabile. Valorile absolute mai mari ale lui r indică o legătură mai puternică, în timp ce valorile apropiate de zero sugerează o legătură mai slabă.

Diagrame de dispersie

Diagramele de dispersie reprezintă un instrument vizual valoros pentru reprezentarea relației dintre două variabile numerice. Acestea facilitează evaluarea linearității relației și identificarea valorilor extreme.

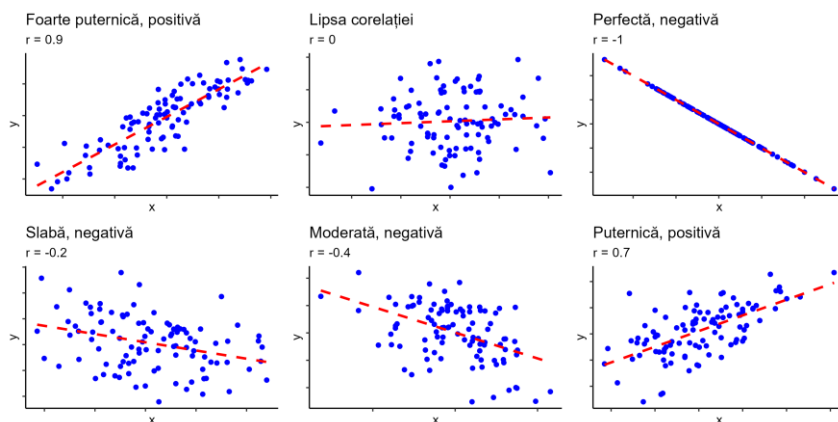


Figura 5.1. Diagrame de dispersie care ilustrează diferite niveluri de corelație

Exemplu de calcul al coeficientului de corelație Pearson

Să considerăm relația dintre talie (lungime) (variabilă independentă) și greutate (variabilă dependentă) la nouă nou-născuți (tabelul 5.1). Pentru a calcula coeficientul de corelație Pearson:

1. Determinați media variabilelor X și Y.

$$\bar{X} = \frac{\sum X_i}{n} = \frac{437}{9} = 48,6$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{25.7}{9} = 2,9$$

2. Calculați abaterile pentru fiecare variabilă X și Y.

$$(X_i - \bar{X}) \quad \text{și} \quad (Y_i - \bar{Y})$$

3. Ridicați la pătrat abaterile variabilelor X și Y:

$$(X_i - \bar{X})^2 \quad \text{și} \quad (Y_i - \bar{Y})^2$$

4. Calculați coeficientul utilizând formula:

$$r = \frac{\Sigma (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2 \Sigma(Y_i - \bar{Y})^2}} = \frac{17,5}{\sqrt{106,2 \times 3,5}} = \frac{17,5}{19,2} = 0,91$$

Tabelul 5.1. Date privind talia (lungimea) (cm) (X_i) și greutatea (kg) (Y_i) colectate pentru nouă nou-născuți.

Unitate de observație	X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X}) \times (Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
1	43,0	2,2	30,9	0,4	3,6	30,9	0,4
2	45,0	2,1	12,6	0,6	2,7	12,6	0,6
3	47,0	2,4	2,4	0,2	0,7	2,4	0,2
4	48,0	2,4	0,3	0,2	0,3	0,3	0,2
5	49,0	3,0	0,2	0,0	0,1	0,2	0,0
6	49,0	2,7	0,2	0,0	-0,1	0,2	0,0
7	50,0	3,5	2,1	0,4	0,9	2,1	0,4
8	50,0	3,4	2,1	0,3	0,8	2,1	0,3
9	56,0	4,0	55,4	1,3	8,5	55,4	1,3
Total	437	25,7	106,2	3,5	17,5	106,2	3,5

Concluzie: coeficientul de corelație r este 0,91. Relația dintre lungimea și greutatea nou-născuților este pozitivă și foarte puternică.

Coeficientul de corelație al rangurilor Spearman

Corelația rangurilor Spearman este o măsură neparametrică utilizată pentru următoarele tipuri de date:

- două variabile ordinale;
- o variabilă ordinală și una numerică;
- două variabile numerice, în cazul când una dintre ele nu este distribuită normal.

Acest coeficient este simbolizat ca r_s și se calculează prin ordonarea valorilor fiecărei variabile și aplicarea formulei Pearson asupra rangurilor. Formula pentru corelația rangurilor Spearman este:

$$r_s = \frac{\Sigma (R_x - \bar{R}_x)(R_y - \bar{R}_y)}{\sqrt{\Sigma(R_x - \bar{R}_x)^2 \Sigma(R_y - \bar{R}_y)^2}}$$

unde:

R_x și R_y sunt rangurile variabilelor X și Y

$\bar{R}_x$ și $\bar{R}_y$ sunt rangurile medii ale variabilelor X și Y.

Asemănător coeficientului de corelație Pearson, coeficientul de corelație al rangurilor Spearman variază în intervalul de la -1 la 1. Valorile r_s apropiate de limitele extreme indică un grad ridicat de corelație între variabilele X și Y, în timp ce valorile apropiate de 0 denotă o relație mai slabă.

Alte tipuri de corelații

- Coeficientul de corelație punct-biserial - utilizat atunci când una dintre variabile este numerică, iar cealaltă este dihotomică.
- Coeficientul Phi - utilizat pentru două variabile dihotomice.

Considerații importante

- ⇒ *Corelația nu implică cauzalitate.* O corelație semnificativă între două variabile nu indică faptul că modificările într-o variabilă cauzează modificări în cealaltă. Coeficientul de corelație este exclusiv o măsură a relației existente între două variabile. A trasa o relație cauzală pe baza unei corelații reprezintă o eroare frecventă și fundamentală.
- ⇒ Prezența unei corelații între două variabile într-un eșantion nu garantează că aceeași corelație există în întreaga populație. Este necesar să se efectueze teste statistice, cum ar fi testele t, pentru a evalua semnificația corelației (vezi capitolul 6).

5.1.2 Coeficientul de determinare

Coeficientul de determinare, notat R^2 , este pătratul coeficientului de corelație. Acesta reflectă proporția varianței unei variabile explicată de varianța celeilalte variabile. Atunci când variabilele sunt corelate, există o anumită cantitate de varianță comună între ele. Cu cât corelația este mai mare, cu atât varianța comună este mai semnificativă și

coeficientul de determinare este mai ridicat. Valoarea R^2 variază între 0 și 100% (figura 5.2).

Exemplu de calcul al coeficientului de determinare

Să luăm în considerare un studiu care examinează relația dintre orele de studiu săptămânale (variabila X) și notele la examene (variabila Y) în rândul mediciiștilor. Să presupunem că studiul identifică o corelație r de 0,70. Pentru a calcula coeficientul de determinare:

$$R^2 = (0,70)^2 = 0,49 \text{ sau } 49\%$$

Acest rezultat arată că 49% din varianța notelor la examene este explicată de varianța orelor de studiu. Astfel, aproape jumătate din variația notelor la examene între studenți poate fi atribuită diferențelor în obiceiurile lor de studiu.

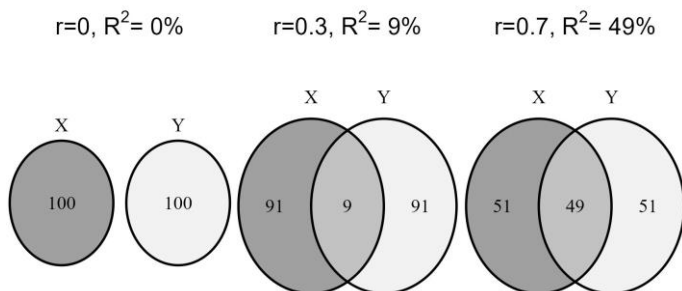


Figura 5.2. Explicația grafică a coeficientului de determinare (R^2)

5.2. Regresie. Abordări generale

5.2.1. Regresie liniară simplă

În cazul în care două variabile sunt foarte corelate, este posibil să se prezică valoarea variabilei dependente pe baza valorii variabilei independente utilizând metode de regresie. Spre deosebire de analizele de corelație care tratează variabilele simetric, regresia liniară simplă

este o tehnică utilizată pentru a explora natura relației dintre două variabile continue. Diferența esențială dintre aceste două metode statistice este că regresia permite examinarea schimbărilor în variabila Y (variabila dependentă sau de rezultat) în raport cu modificările în variabila X (variabila independentă sau explicativă), conform *ecuației de regresie*, care descrie relația liniară dintre cele două variabile. *Linia dreaptă de regresie* este echivalentă cu „linia de cel mai bun fit” pentru graficul de dispersie, utilizată și în calcularea coeficientului de corelație.

În cazul ecuației liniare simple, valoarea unei variabile (X) este utilizată pentru a prezice valoarea celeilalte variabile (Y) prin intermediul ecuației de regresie. Ecuația pentru regresia liniară simplă este:

$$\text{Valoarea estimată } Y = \alpha + \beta X$$

unde:

Y – valoarea estimată a lui Y (variabila dependentă)

α – constantă cunoscută sub denumirea de „constantă de interceptare” (intercept)

β – „constantă de pantă” a liniei de regresie (slope)

X – valoarea variabilei X (variabila independentă)

Ecuația regresiei liniare simple este cunoscută sub denumirea de forma pantă-interceptare. Panta (β) reprezintă numărul care se înmulțește cu valoarea lui X, iar interceptarea (α) este numărul adăugat sau scăzut. Constanta pantei indică modificarea lui Y atunci când X crește cu 1 unitate. Constanta de interceptare reprezintă valoarea lui Y când X este 0 (punctul unde linia de regresie intersectează axa Y).

Interpretarea coeficientului de regresie (β) este următoarea:

Dacă $\beta = 0$, variabila Y nu depinde de variabila X

Dacă $\beta \neq 0$, variabila Y depinde de variabila X astfel:

$\beta > 0$, indică o relație pozitivă între Y și X

$\beta < 0$, indică o relație negativă între Y și X.

Odată ce valorile a și b au fost stabilite, valoarea estimată a lui Y poate fi prezisă pentru orice valoare a lui X . De exemplu, s-a demonstrat că rata de clearance hepatic a lidocainei (variabila Y , ml/min/kg) poate fi prezisă pe baza ratei de clearance hepatic al colorantului verde de indocianină (variabila X , ml/min/kg), utilizând ecuația $Y = 0,30 + 1,07X$. Această metodă permite anesteziologilor să minimizeze riscul de supradozaj de lidocaină prin evaluarea clearance-ului colorantului (Pagano M., Gauvreau K., 2000).

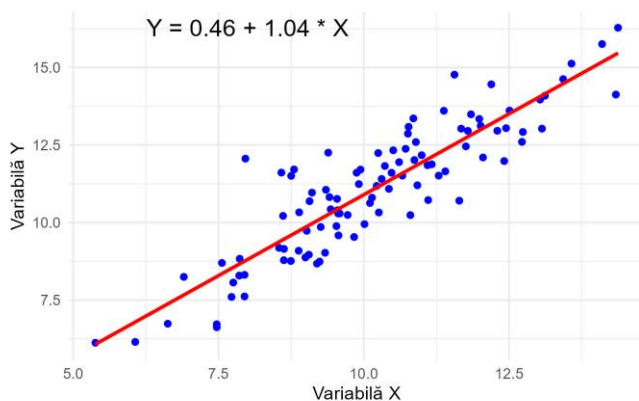


Figura 5.3. Diagrama de dispersie cu linia de regresie pentru un set de date ipotetice

5.2.2. Regresie liniară multiplă

Tehnicile de *regresie liniară multiplă* sunt aplicate atunci când se utilizează mai multe variabile continue X pentru a prezice valoarea așteptată a lui Y . Ecuația regresiei multiple este formulată astfel:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

5.2.3. Regresie logistică

În cazul regresiei liniare, variabila de răspuns Y este numerică și se consideră că urmează o distribuție normală. Regresia logistică se ocupă de situațiile în care variabilele predictor (independente X) sunt numerice și variabilele prezise (dependente Y) sunt nominale (de exemplu, supraviețuire versus deces, vindecat versus nevindecat). În loc să prezică o valoare medie, regresia logistică estimează probabilitatea asociată cu răspunsul dihotomic pentru diferite valori ale unei variabile explicative.

Exerciții de recapitulare

1. Un grup de 10 studenți a fost monitorizat în ceea ce privește numărul de ore de studiu și punctajul lor la examen.

Set de date 1. Ore de studiu versus Punctajul la examen

Unitate de observație	1	2	3	4	5	6	7	8	9	10
Ore studiate	2	3	4	5	6	7	8	9	10	11
Punctajul la examen	56	61	68	70	72	75	78	80	85	88

- Calculați coeficientul de corelație adecvat. Explicați alegerea dumneavoastră.
 - Interpretați coeficientul de corelație calculat.
 - Realizați o diagramă de dispersie pentru aceste date.
2. Un grup de 10 persoane și-a înregistrat timpul zilnic de exerciții (în minute) și pierderea în greutate corespunzătoare (în kg) pe parcursul unei luni.

Set de date 2. Timpul zilnic de exerciții versus Pierderea în greutate

Unitate de observație	1	2	3	4	5	6	7	8	9	10
Timpul de exerciții (minute)	15	20	25	30	35	40	45	50	55	60
Pierderea în greutate (kg)	0,5	0,8	1,0	1,3	1,5	1,8	2,0	2,3	2,5	2,8

- Calculați coeficientul de corelație adecvat. Explicați alegerea dumneavoastră.
- Interpretați coeficientul de corelație calculat.
- Realizați o diagramă de dispersie pentru aceste date.

3. Dacă relația dintre două măsuri este liniară și coeficientul de corelație este apropiat de 1, diagrama de dispersie a observațiilor:

- a) Este o linie dreaptă orizontală
- b) Este o linie dreaptă verticală
- c) Este o linie dreaptă care nu este nici orizontală, nici verticală
- d) Are o pantă negativă
- e) Are o pantă pozitivă

4. Dacă relația dintre două măsuri este liniară și coeficientul de corelație este apropiat de -1, diagrama de dispersie a observațiilor:

- a) Este o linie dreaptă orizontală
- b) Este o linie dreaptă verticală
- c) Este o linie dreaptă care nu este nici orizontală, nici verticală
- d) Are o pantă negativă
- e) Are o pantă pozitivă

Întrebări de recapitulare

1. În ce situații utilizarea corelației este considerată preferabilă?
2. Care sunt avantajele și limitările coeficientului de corelație Pearson?
3. Care este diferența între corelația rangurilor Spearman de corelația Pearson?
4. De ce este esențial să se elaboreze un grafic de dispersie atunci când se analizează relația dintre două variabile continue?
5. Dacă un test de ipoteză sugerează că corelația dintre două variabile nu este semnificativ diferită de 0, acest lucru implică neapărat că variabilele sunt independente? Explicați.
6. Care este diferența principală între analizele de corelație și cele de regresie?
7. În ce condiții este recomandată utilizarea regresiei liniare simple?
8. În ce circumstanțe este preferabilă utilizarea regresiei liniare multiple?
9. În ce situații se justifică utilizarea regresiei logistice?

CAPITOLUL 6. STATISTICA INFERENȚIALĂ. TEORIA PROBABILITĂȚII ȘI TESTAREA IPOTEZELOR

Concepte cheie

- ❖ *Statistica inferențială* se bazează pe datele colectate din eșantioane și este utilizată pentru a trasa inferențe (concluzii) despre parametrii unei populații.
- ❖ Statistica inferențială se aplică exclusiv eșantioanelor *probabilistice*. În cazul eșantioanelor *non-probabilistice*, se folosesc doar statistici descriptive.
- ❖ *Probabilitatea* joacă un rol esențial în statistica inferențială. Decizia privind semnificația statistică a unui rezultat se bazează pe evaluarea probabilității.
- ❖ *Distribuția de eșantionare a mediilor* va tinde întotdeauna spre normalitate, indiferent de forma distribuției populației din care sunt extrase eșantioanele (*Teorema Limitei Centrale*).
- ❖ *Eroarea standard* reprezintă diferența medie între media populației și media unui eșantion individual.
- ❖ *Eroarea standard* este invers proporțională cu rădăcina pătrată a măririi eșantionului (n). Eșantioanele mici generează erori standard mari.
- ❖ *Ipoteza nulă* (H_0) indică întotdeauna absența unui efect în populație (de exemplu, mediile a două populații nu diferă).
- ❖ Ipoteza alternativă (H_1) preconizează existența diferențelor între grupuri (de exemplu, mediile a două populații vor diferi).
- ❖ O ipoteză alternativă *bilaterală* sau *nedirecțională* nu face speculații cu privire la care dintre valori va fi mai mare.
- ❖ O ipoteză alternativă *unilaterală* sau *direcțională* specifică care valoare va fi mai mare.
- ❖ Regula „AAA” pentru eroarea de tip I - Eroarea Alpha Acceptă Ipoteza Alternativă falsă (*Alpha error Accepts the Alternative*).
- ❖ Regula „BEAN” pentru eroarea de tip II - Eroarea Beta Acceptă Ipoteza Nulă falsă (*Beta Error Accepts the Null*).

- ❖ *Nivelul de încredere* reprezintă probabilitatea de a accepta ipoteza nulă atunci când aceasta este adevărată.
- ❖ *Puterea studiului* reprezintă probabilitatea de a respinge ipoteza nulă atunci când aceasta este falsă.
- ❖ *Nivelul de semnificație (α -nivel)* este probabilitatea de a face o eroare de tip I, stabilită înainte de calcularea testului statistic.
- ❖ Nivelul α este de obicei setat la 0,05 sau mai mic.
- ❖ *Valoarea p* reprezintă probabilitatea de a face o eroare de tip I, calculată după efectuarea testului statistic.
- ❖ Dacă valoarea $p < \text{nivelul } \alpha$ (valoarea $p < 0,05$), se poate respinge ipoteza nulă. Astfel, se poate concluziona că diferența dintre cele două medii este semnificativă din punct de vedere statistic.
- ❖ *Un interval de încredere* este un interval calculat folosind statistice de eșantion care conțin parametrul populației (de exemplu, media) cu un anumit grad de încredere (de exemplu, 95% sau 99% încredere).

Statistica inferențială folosește datele eșantionului pentru a face concluzii referitoare la parametrii populației. Aceste statistice ne ajută să stabilim dacă un fenomen observat într-un eșantion există într-adevăr în populația mai largă din care a fost extras eșantionul.

6.1. Teoria probabilităților

6.1.1. Concepte generale

Probabilitatea este esențială în statistica inferențială, în special atunci când se evaluează dacă rezultatul unui studiu este semnificativ din punct de vedere statistic, adică dacă rezultatul nu este rezultatul întâmplării. Teoria probabilităților este fundamentală pentru analiza seturilor mari de date și a experimentelor repetate, denumite trialuri, care generează rezultate variate.

Definiția clasică a probabilității (p) afirmă că probabilitatea unui eveniment este numărul de rezultate favorabile (m) împărțit la numărul total de rezultate posibile (n):

$$p = \frac{m}{n}$$

Probabilitatea ca un eveniment să nu aibă loc (q) este:

$$q = \frac{n - m}{n} = 1 - \frac{m}{n} = 1 - p$$

Astfel, suma probabilităților ca un eveniment să aibă loc și să nu aibă loc este egală cu 1 (sau 100%). Valoarea lui p variază de la 0 la 1 (0% la 100%), o valoare mai mare a lui p indicând o probabilitate mai mare a producerii evenimentului.

Două concepte matematice importante în teoria probabilităților sunt legea numerelor mari și teorema limitei centrale.

6.1.2. Legea numerelor mari

Legea numerelor mari afirmă că, pe măsură ce numărul de încercări într-un experiment crește, media rezultatelor se apropie tot mai mult de valoarea așteptată. Cu cât mai mult se apropie numărul de încercări de un anumit prag, cu atât mai mult media rezultatelor experimentale se apropie de valoarea teoretică așteptată.

6.1.3. Teorema limitei centrale

Teorema limitei centrale descrie caracteristicile distribuției de eșantionare a mediilor:

1. Media distribuției de eșantionare este egală cu media populației (μ).

2. Deviația standard a distribuției de eșantionare a mediilor este cunoscută sub denumirea de *eroare standard a mediei* ($SE_{\bar{x}}$). Eroarea standard a mediei se calculează prin împărțirea deviației standard a populației (σ) la rădăcina pătrată a mărării eșantionului (n):

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

unde:

$SE_{\bar{x}}$ - eroarea standard a mediei

σ - deviația standard a populației

n – mărimea eșantionului

Cu alte cuvinte, eroarea standard reprezintă diferența medie între media populației și media unui eșantion individual.

3. Atunci când n este suficient de mare, forma distribuției de eșantionare este aproximativ normală, indiferent de forma distribuției populației din care au fost extrase eșantioanele.

6.1.4. Utilizarea erorii standard

Eroarea standard este influențată de mărimea eșantionului; cu cât mărimea eșantionului este mai mare, cu atât media eșantionului ($\bar{X}$) se apropie mai mult de media populației (μ). Eroarea standard reflectă incertitudinea estimării mediei, indicând gradul de variație al mediei eșantionului în cazul unor eșantionări repetate. Erorile standard mari sunt caracteristice eșantioanelor mici cu deviații standard mari. Eroarea standard este utilizată pentru calcularea intervalelor de încredere, care facilitează inferența parametrilor populației.

6.2. Eșantionare

6.2.1. Definiția eșantionării

Teoria probabilităților ne permite să facem inferențe despre caracteristicile populației folosind datele eșantionului. Eșantionarea implică selectarea unui grup din populație pentru a colecta date pentru cercetare. Motivele eșantionării includ economisirea timpului și a banilor, obținerea unor rezultate mai precise, reducerea eterogenității și estimarea erorilor.

6.2.2. Metode de eșantionare

Mai multe metode de eșantionare sunt utilizate în cercetarea medicală, fiecare dintre ele necesitând selecție aleatorie pentru utilizarea eficientă a statisticii inferențiale. Eșantioanele probabilistice asigură că probabilitatea de includere a fiecărui subiect este cunoscută, ceea ce permite inferențe valide. Eșantioanele non-probabilistice permit doar statistici descriptive.

A. Eșantionare probabilistică:

- ⇒ Eșantionare aleatorie simplă. Fiecare membru al populației are șanse egale de selecție. Unitățile sunt selectate aleatoriu până când se atinge mărimea dorită a eșantionului.
- ⇒ Eșantionare sistematică. Fiecare al k -lea element este selectat dintr-o listă a populației, cu k determinat prin împărțirea mărimii populației la mărimea dorită a eșantionului.
- ⇒ Eșantionare stratificată. Populația este împărțită în grupuri mutual exclusive (straturi), apoi, din fiecare strat sunt extrase eșantioane aleatorii.
- ⇒ Eșantionare în cuiburi (clustere). Inițial, se selectează aleatoriu un set de grupuri sau „cuiburi” din populație, iar ulterior cazurile

sunt selectate aleatoriu din fiecare cuib. Cuiburile sunt adesea bazate pe zone geografice.

B. Eșantionare non-probabilistică:

Eșantioanele non-probabilistice sunt acelea pentru care probabilitatea de selecție a unei unități este necunoscută.

- ⇒ Eșantionare de comoditate. Cercetătorul selectează cei mai accesibili membri ai populației pentru colectarea informațiilor.
- ⇒ Eșantionare pe cote. Cercetătorul interviează un număr prestabilit de persoane în fiecare dintre mai multe categorii.

6.3. Estimarea și testarea ipotezelor

Estimarea și testarea ipotezelor sunt componente fundamentale ale statisticii inferențiale, care permit cercetătorilor să formuleze concluzii bazate pe date și relațiile dintre variabile.

Estimarea:

- *Estimarea punctuală* utilizează datele eșantionului pentru a calcula un singur număr care estimează parametrul de interes, precum media populației, fără a oferi informații despre variabilitatea estimării. Nu se cunoaște cât de aproape este media eșantionului ($\bar{X}$) de media populației (μ).
- *Estimarea prin interval* folosește un interval de valori pentru a estima parametrul, oferind un interval de încredere (ÎÎ) care conține media populației (μ) cu un anumit grad de încredere.

Testarea ipotezelor implică formularea unei ipoteze nule (H_0) și a unei ipoteze alternative (H_1) și efectuarea unui test statistic pentru a determina care ipoteză să fie acceptată. Scopul este de a respinge ipoteza nulă și de a accepta ipoteza alternativă.

6.4. Intervalele de încredere

Un interval de încredere este o estimare prin interval a unui parametru al populației, reprezentat prin medie, proporție, coeficient de corelație sau diferențe între două medii sau proporții. Extremitățile intervalului se numesc *limitele intervalului de încredere*. Intervalele de încredere ($\hat{I}$) sunt calculate folosind media eșantionului și eroarea standard:

$$\hat{I} = \bar{X} \pm z \times SE_{\bar{x}}$$

unde:

$\bar{X}$ - media eșantionului

$SE_{\bar{x}}$ - eroarea standard

z - valoarea z egală cu 1,96 pentru $\hat{I}_{95}$ și 2,56 pentru $\hat{I}_{99}$.

Exemplu de calculare a intervalului de încredere

Acest exemplu ilustrează metoda de calcul a intervalului de încredere cu un grad de încredere de 95% ($\hat{I}_{95}$) pentru următorul set de date.

Tabelul 6.1. Date privind nivelul colesterolului colectate pentru 10 pacienți

Unitate de observație	1	2	3	4	5	6	7	8	9	10
Nivelul colesterolului în sânge (mg/dl), x_i	168	258	228	230	247	156	172	165	210	264

1. Determinați media eșantionului.

$$\bar{X} = \frac{\sum x_i}{n} \quad \bar{X} = 210 \text{ mg/dl}$$

2. Determinați deviația standard.

$$s = \sqrt{\frac{\sum (x_i - \bar{X})^2}{n-1}} \quad s = 41,1 \text{ mg/dl}$$

3. Determinați eroarea standard (ES).

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

$$ES = \frac{41,1}{3,16} = 13,0 \text{ mg/dl}$$

4. Stabiliți limitele inferioară și superioară ale intervalului de încredere $\hat{I}_{95}$

$$\hat{I}_{95} = \bar{X} \pm z \times ES$$

$$\hat{I}_{95} = 210 \pm 1,96 \times 13,1$$

$\hat{I}_{95}$: de la 183,5 (limita inferioară) la 236 (limita superioară)

Interpretarea rezultatelor: Dacă experimentul se repetă de 100 de ori, în doar 5 cazuri din 100, media populației (μ) se va situa în afara limitelor intervalului de încredere. În 95 din 100 de cazuri, media populației se va afla între limitele inferioară și superioară ale intervalului de încredere. Pentru $\hat{I}_{95}$, probabilitatea de a găsi media populației în interiorul limitelor sale este de 0,95. Probabilitatea ca media populației să fie în afara acestor limite este de 0,05.

Compararea a două medii folosind intervalele de încredere

În cercetarea medicală, compararea mediilor diferitelor grupuri este o sarcină frecventă pentru a stabili dacă un tratament sau o condiție are un efect semnificativ. Prin compararea intervalelor de încredere ($\hat{I}$) ale diferitelor grupuri, putem concluziona dacă diferența dintre cele două medii este semnificativă din punct de vedere statistic sau nu.

Figura 6.1 compară ritmul cardiac în trei grupuri de pacienți: grupul B și grupul C sunt comparate cu grupul A, care este grupul de referință. Ritmul cardiac este prezentat prin medii și intervalele de încredere de 95% (grafice cu bare de eroare). Ritmul cardiac pentru grupul B este semnificativ mai mare decât cel al grupului A, deoarece intervalul de încredere de 95% pentru grupul B nu se suprapune cu intervalul de încredere de 95% pentru grupul A (valoarea p pentru testul $t < 0,05$). Diferența în ritmurile cardiace între grupul A și grupul C nu este

semnificativă din punct de vedere statistic; suprapunerea intervalelor de încredere sugerează că diferența observată ar putea fi cauzată de variația aleatorie (valoarea p pentru testul $t > 0,05$).

Tabelul 6.2 oferă un cadru pentru interpretarea suprapunerii intervalelor de încredere și implicațiile acesteia pentru semnificația statistică a diferențelor dintre grupuri.

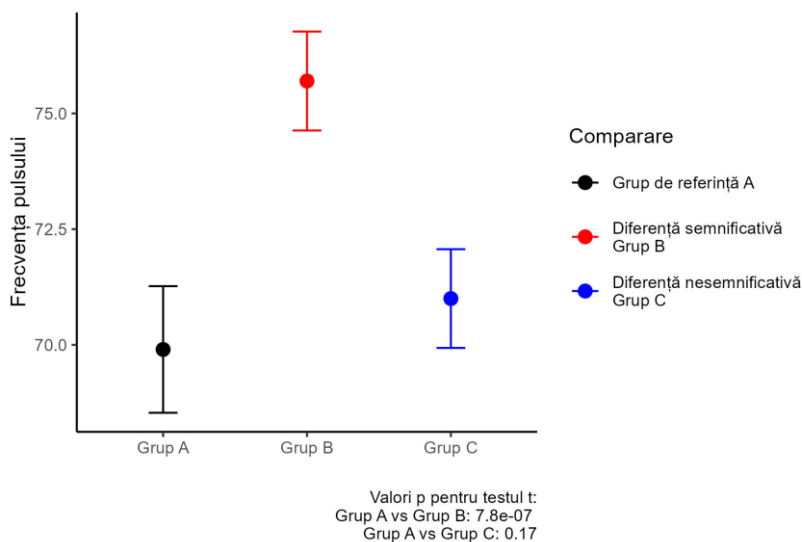


Figura 6.1. Compararea ritmului cardiac între trei grupuri folosind intervale de încredere de 95%

Notă: Comparativ cu grupul A (grupul de referință), diferența în media ritmului cardiac este semnificativă din punct de vedere statistic doar pentru grupul B (fără suprapunere între intervalele de încredere), dar nu și pentru grupul C (suprapunere între intervalele de încredere).

Tabelul 6.2. Compararea a două medii folosind intervalele de încredere (Î)

Suprapunerea Î	Semnificație statistică între grupurile comparate
Lipsă	Diferență foarte semnificativă
Ușoară	Posibil semnificativă, dar nu foarte semnificativă
Extinsă	Cu siguranță nesemnificativă

6.5. Testarea ipotezelor: concepte teoretice de bază

6.5.1. Definirea ipotezei

O ipoteză reprezintă o presupunere bine argumentată despre un fenomen specific. Aceasta are rolul de a ghida cercetarea, oferind claritate, specificitate și obiectivitate.

6.5.2. Tipuri de ipoteze

⇒ *Ipoteza nulă* (H_0) susține că nu există nicio diferență semnificativă sau relație între variabile (de exemplu, mediile a două populații sunt egale).

$$H_0: \mu_1 = \mu_2$$

⇒ *Ipoteza alternativă* (H_1) se opune ipotezei nule, afirmând că există diferențe între grupuri (de exemplu, mediile a două populații sunt diferite).

- *Ipoteza alternativă nedirecțională* sau *bilaterală* (cu două cozi) afirma existența unei diferențe fără a specifica care valoare este mai mare; testată printr-un test bilateral (test cu două cozi).

$$H_1: \mu_1 \neq \mu_2$$

- *Ipoteza alternativă direcțională* sau *unilaterală* (cu o coadă) indică direcția așteptată a diferenței, specificând care

valoare este mai mare; testată printr-un test unilateral (cu o coadă).

$$H_1: \mu_1 > \mu_2 \quad \text{or} \quad H_1: \mu_1 < \mu_2$$

Împreună, ipotezele nulă și alternativă acoperă întreg spectrul valorilor posibile ale mediei populației (μ). În consecință, una dintre cele două afirmații trebuie să fie adevărată.

6.5.3. Erori de tip I și de tip II

În testarea ipotezelor pot apărea două tipuri de erori:

- ⇒ *Eroarea de tip I* - respingerea ipotezei nule când aceasta este adevărată (fals pozitiv). Probabilitatea producerii unei erori de tip I (α) este, de asemenea, cunoscută sub numele de *eroare de respingere*.
- ⇒ *Eroarea de tip II* - acceptarea ipotezei nule când aceasta este falsă (fals negativ). Probabilitatea producerii unei erori de tip II (β) este, de asemenea, cunoscută sub denumirea de *eroare de acceptare*.

6.5.4. Puterea studiului

Puterea unui studiu se referă la capacitatea acestuia de a identifica o diferență reală atunci când aceasta este prezentă. În mod specific, aceasta reprezintă probabilitatea de a respinge ipoteza nulă atunci când aceasta este falsă, concluzionând astfel că ipoteza alternativă este adevărată atunci când este într-adevăr validă. Cu alte cuvinte, puterea studiului este probabilitatea de a evita o eroare de tip II (β). Matematic, se exprimă astfel:

$$\text{Puterea studiului} = 1 - \beta,$$

unde β este probabilitatea de a comite o eroare de tip II.

Pentru ca un studiu să fie considerat acceptabil, este de obicei necesar să aibă o putere de cel puțin 0,8 (sau un β de 0,2). Cu alte

cuvinte, un studiu ar trebui să aibă cel puțin 80% șanse de a detecta o diferență reală, dacă aceasta există. Un studiu cu o putere mai mică de 80% este în general inacceptabil.

Una dintre cele mai eficiente și importante metode de a crește puterea unui studiu este prin mărirea volumului eșantionului. Volume mai mari ale eșantionului reduc eroarea standard, facilitând identificarea diferențelor sau efectelor reale.

6.5.5. Nivelul de încredere

Nivelul de încredere se referă la capacitatea unui studiu de a nu detecta o diferență falsă. Este probabilitatea de a accepta ipoteza nulă atunci când aceasta este, într-adevăr, adevărată. Nivelul de încredere se definește astfel:

$$\text{Nivelul de încredere} = 1 - \alpha$$

unde α este probabilitatea de a comite o eroare de tip I.

Tabelul 6.3 rezumă cele patru rezultate posibile ale testării ipotezei. Este esențial să ne amintim că testăm întotdeauna ipoteza nulă, care, într-o realitate necunoscută, poate fi fie adevărată, fie falsă.

Tabelul 6.3. Patru posibile rezultate ale testării ipotezei nule (H_0)

	H₀ ADEVĂRATĂ	H₀ FALSĂ
ACCEPTARE H₀	Decizie corectă <i>Nivelul de încredere</i> $1-\alpha$	Eroare de tip II β
RESPINGERE H₀	Eroare de tip I α	Decizie corectă <i>Puterea studiului</i> $1-\beta$

6.5.6. Nivelul de semnificație

Pentru a respinge ipoteza nulă, trebuie să avem o certitudine rezonabilă că orice diferență observată între statistica eșantionului ($\bar{X}$) și parametrul populației (μ) nu se datorează întâmplării. Cât de diferită trebuie să fie media eșantionului față de media populației pentru a considera diferența *semnificativă* din punct de vedere statistic? Acest criteriu este cunoscut sub denumirea de nivelul de semnificație sau nivelul α , care reprezintă probabilitatea de a comite o eroare de tip I, adică probabilitatea de a respinge ipoteza nulă atunci când aceasta este, de fapt, adevărată.

Nivelul de semnificație (nivelul α) reprezintă probabilitatea de a comite o eroare de tip I, stabilită înainte de calcularea testului statistic. Pentru a minimiza riscul de a respinge incorect ipoteza nulă, nivelul de semnificație ar trebui să fie suficient de mic. Este de obicei stabilit la 0,05 sau mai mic (0,01 sau 0,001). Cu cât nivelul α este mai mic, cu atât este mai puțin probabilă comiterea erorii de tip I.

6.5.7. Valoarea p

Valoarea p reprezintă un concept strâns asociat cu nivelul de semnificație α . Aceasta reprezintă probabilitatea de a obține rezultatele observate dacă ipoteza nulă este adevărată. Este probabilitatea de a face o eroare de tip I după calcularea testului statistic.

În urma efectuării unui test statistic, valoarea p se compară cu nivelul α . Dacă valoarea p este inferioară nivelului α (de exemplu, $p < 0,05$), ipoteza nulă poate fi respinsă, permițând acceptarea ipotezei alternative. Acest lucru sugerează că diferența observată între cele două medii este semnificativă statistic și este improbabil să fie rezultatul întâmplării.

Pe de altă parte, dacă valoarea p depășește nivelul α (de exemplu, $p > 0,05$), ipoteza nulă nu poate fi respinsă, conducând la concluzia că diferența dintre cele două medii nu este semnificativă statistic și este, cel mai probabil, atribuibilă întâmplării.

6.6. Procesul de testare a ipotezelor. Abordări generale

În cadrul testării ipotezelor, scopul este de a formula concluzii cu privire la un parametru al populației utilizând datele din eșantion. O metodă constă în construirea unui interval de încredere pentru media populației (μ); o altă metodă este efectuarea unui test statistic. Pașii esențiali în procesul de testare a ipotezelor sunt:

1. Formularea ipotezelor - ipoteza nulă (H_0) și ipoteza alternativă (H_1).
2. Selectarea testului statistic adecvat.
3. Determinarea nivelului de semnificație (nivelul α).
4. Identificarea valorii critice necesare pentru ca testul statistic să fie considerat semnificativ.
5. Realizarea calculelor necesare.
6. Formularea concluziilor.

Etapele detaliate ale testării ipotezelor, alături de exemple, vor fi prezentate în capitolul 7.

Exerciții de recapitulare

1. Care este scopul testării ipotezelor?
2. Explicați pe scurt relația dintre intervalul de încredere și testarea ipotezelor.
3. În ce circumstanțe se folosește testul unilateral în locul testului bilateral?
4. Descrieți cele două tipuri de erori care pot apărea în timpul testării ipotezelor.
5. Distribuția nivelului de colesterol seric într-o eșantion de 400 de bărbați adulți este asimetrică spre stânga. Distribuția de eșantionare a mediilor colesterolului seric este:
 - a) asimetrică spre stânga
 - b) asimetrică spre dreapta
 - c) normală
 - d) nu este posibil de determinat.
6. Eroarea standard a unei statistici este:
 - a) media distribuției eșantionului
 - b) deviația standard a distribuției eșantionului
 - c) media împărțită la rădăcina pătrată din mărimea eșantionului (n).
7. Media tensiunii arteriale sistolice într-un grup de 500 de persoane selectate dintr-un total de 100.000 de persoane din localitatea A este de 130 mmHg, iar deviația standard este de 15 mmHg.

Calculați: eroarea standard și intervalul de încredere de 95% pentru media populației. Interpretați rezultatele.
8. A fost realizat un studiu privind tensiunea arterială sistolică la bărbați de 60 de ani cu diabet zaharat. În acest studiu, a fost selectat un eșantion aleatoriu de 300 de bărbați de 60 de ani cu

diabet zaharat, media tensiunii arteriale sistolice fiind de 160 mmHg, iar deviația standard a eșantionului de 25 mmHg.

- a) Calculați intervalul de încredere de 95% pentru media tensiunii arteriale sistolice în rândul populației de bărbați de 60 de ani cu diabet zaharat.
- b) Să presupunem că mărimea eșantionului ar fi fost de 150 în loc de 300, iar media eșantionului și deviația standard ale eșantionului rămân aceleași. Intervalul de încredere s-ar lărgi sau s-ar micșora? De ce?

Întrebări de recapitulare

1. Ce este statistica inferențială?
2. Care este rolul și importanța teoriei probabilităților în statistică?
3. Descrieți legea numerelor mari și relevanța sa pentru cercetare.
4. Explicați teorema limitei centrale.
5. Ce reprezintă eșantionarea și de ce este esențială în analiza statistică?
6. De ce este crucială eșantionarea aleatorie pentru validitatea rezultatelor cercetărilor?
7. În ce scenarii este preferabilă eșantionarea sistematică în locul eșantionării simple aleatorii?
8. Când este metoda eșantionării în straturi cea mai potrivită?
9. În ce condiții eșantionarea în cuiburi reprezintă abordarea optimă?
10. Care este scopul principal al testării ipotezelor?
11. Care sunt diferențele dintre estimarea punctuală și estimarea pe interval?

12. Definiți o ipoteză și explicați diferitele tipuri de ipoteze.
13. Ce sunt intervalele de încredere și limitele de încredere și cum sunt ele utilizate în analiza statistică?
14. Ce factori influențează lărgimea unui interval de încredere pentru o medie?
15. Descrieți cele două tipuri principale de erori care pot apărea în testarea ipotezelor.
16. Explicați diferența dintre eroarea de tip I și eroarea de tip II în testarea ipotezelor.
17. Ce reprezintă puterea unui studiu și cum poate fi descrisă?
18. Ce este nivelul de încredere și cum se interpretează?
19. Ce reprezintă nivelul de semnificație și cum poate fi explicat?
20. Ce este valoarea p și ce semnifică aceasta în contextul testării statistice?

CAPITOLUL 7. TESTAREA IPOTEZELOR. METODE PARAMETRICE ȘI NEPARAMETRICE

Concepte cheie

- ❖ *Testele statistice parametrice* sunt aplicate pentru analiza datelor numerice (pe scala de interval sau de raport), presupunând că datele din populație prezintă o distribuție normală.
- ❖ *Testele t* sunt exemple de teste statistice parametrice.
- ❖ *Testele statistice neparametrice* sunt adecvate pentru datele categoriale (nominale sau ordinale).
- ❖ *Testul chi-pătrat (χ^2)* constituie un exemplu de test statistic neparametric, utilizat pentru evaluarea diferențelor în proporții și a asociațiilor dintre două variabile.
- ❖ *Testele statistice neparametrice* sunt mai puțin puternice decât testele parametrice.
- ❖ Tipul de testare a ipotezei, test bilateral, unilateral stânga sau unilateral dreapta, depinde de ipoteza alternativă.
- ❖ Dacă valoarea absolută a statisticii testului depășește valoarea critică, ipoteza nulă se respinge. Aceasta indică faptul că valoarea p este mai mică decât nivelul α .
- ❖ Dacă valoarea absolută a statisticii testului este mai mică decât valoarea critică, ipoteza nulă nu se respinge, sugerând că valoarea p este mai mare decât nivelul α .

7.1. Teste parametrice și neparametrice

Metodele de testare a ipotezelor se clasifică în parametrice și neparametrice, în funcție de tipul datelor analizate. Metodele parametrice sunt adecvate pentru datele pe scala de interval și de raport, presupunând o distribuție normală a datelor. În schimb, metodele neparametrice sunt utilizate pentru datele pe scala ordinală

sau nominală și nu impun presupuneri referitoare la distribuția normală a datelor într-o populație.

Alegerea metodei statistice corespunzătoare se bazează în principal pe mai mulți factori esențiali:

- ⇒ Tipul de date (numerice, nominale sau ordinale);
- ⇒ Natura eșantioanelor (independente sau dependente (legate));
- ⇒ Mărimea eșantionului ($n > 30$ sau $n < 30$);
- ⇒ Numărul de grupuri (unul, două sau mai multe);
- ⇒ Tipul de ipoteză alternativă (direcțională sau nedirecțională);
- ⇒ Tipul distribuției datelor (normală sau asimetrică);
- ⇒ Omogenitatea varianțelor.

Figura 7.1 ilustrează exemple de teste statistice parametrice și neparametrice.

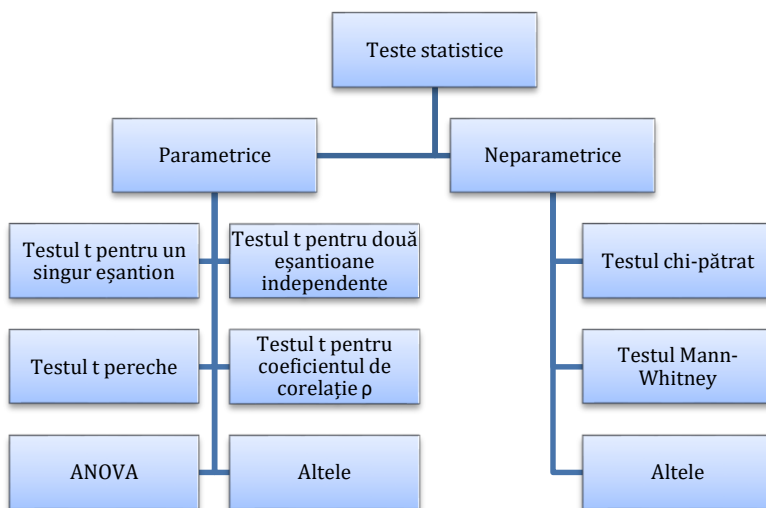


Figura 7.1. Teste statistice parametrice vs neparametrice

7.2. Abordarea generală a testării ipotezelor

7.2.1. Etapele testării ipotezelor

În pofida diversității testelor statistice disponibile, etapele testării ipotezelor rămân constante.

Etapa 1. Formularea ipotezelor:

- Ipoteza nulă (H_0) - aceasta presupune că nu există niciun efect sau diferență. Este ipoteza pe care cercetătorul urmărește să o respingă.
- Ipoteza alternativă (H_1): Aceasta este ipoteza conform căreia există un efect sau o diferență. Este ipoteza pe care cercetătorul urmărește să o dovedească.

Etapa 2. Selectarea testului statistic adecvat:

- Alegerea testului statistic depinde de natura datelor și de designul studiului. Printre testele frecvent utilizate se numără testele t , testul chi-pătrat, ANOVA, etc.
 - Testul t pentru un singur eșantion- se folosește pentru compararea mediei unui eșantion unic cu o medie cunoscută.
 - Testul t pentru două eșantioane independente - Se folosește pentru compararea mediilor a două grupuri independente.
 - Testul t pereche - Se folosește pentru compararea mediilor aceleiași grupe la momente diferite.
 - Testul t pentru coeficientul de corelație ρ (rho) - se folosește pentru a evalua intensitatea și direcția relației dintre două variabile continue.
 - Testul chi-pătrat - se folosește pentru date categoriale pentru a evalua probabilitatea ca distribuția observată să fie rezultatul întâmplării.

- ANOVA (Analiza variației, *eng. Analysis of Variance*)- se folosește pentru compararea mediilor între trei sau mai multe grupuri.

Etapa 3. Stabilirea nivelului de semnificație α :

- Nivelul de semnificație α reprezintă probabilitatea de a respinge ipoteza nulă atunci când aceasta este adevărată (eroare de tip I).
- Valorile α utilizate frecvent sunt 0,05, 0,01 și 0,001.

Etapa 4. Determinarea gradelor de libertate (GL) și a valorii critice:

- *Gradele de libertate* (GL, *eng. df, degrees of freedom*). Acestea sunt determinate de tipul testului statistic și de mărimea eșantionului (n):
 - Pentru testul t cu un singur eșantion: $GL = n - 1$.
 - Pentru testul t cu două eșantioane independente:
 $GL = n_1 + n_2 - 2$.
 - Pentru testul t pentru coeficientul de corelație ρ :
 $GL = n - 2$.
 - Pentru testul chi-pătrat, GL se calculează în mod obișnuit pe baza numărului de categorii.
- *Valoarea critică a testului* (t -critic) - această valoare se obține din tabele statistice (distribuția t , distribuția chi-pătrat etc.), în funcție de nivelul α și de gradele de libertate. Tabelele statistice sunt disponibile în anexa A (tabelul distribuției t) și anexa B (tabelul chi-pătrat).

Etapa 5. Efectuarea calculului:

- Se calculează statistica testului (t -calculat) utilizând formula adecvată pentru testul selectat.

Etapa 6. Formularea concluziilor:

- Se compară valoarea absolută a statisticii testului (t -calculat) cu valoarea absolută a valorii critice (t -critic).
 - Dacă valoarea absolută a statisticii testului ($|t\text{-calculat}|$) depășește valoarea critică ($|t\text{-critic}|$), *ipoteza nulă se respinge*. Aceasta sugerează că valoarea p este mai mică decât nivelul α , indicând semnificație statistică (valoarea $p < \text{nivelul } \alpha$).
 - Dacă valoarea absolută a statisticii testului ($|t\text{-calculat}|$) este inferioară valorii critice ($|t\text{-critic}|$), *ipoteza nulă nu se respinge*. Aceasta sugerează că valoarea p este mai mare decât nivelul α , indicând absența semnificației statistice (valoarea $p > \text{nivelul } \alpha$).

7.2.2. Testarea ipotezelor. Teste bilaterale, unilaterale stânga și unilaterale dreapta

Testarea ipotezelor ne ajută să determinăm dacă o afirmație despre un parametru al populației este adevărată. Testarea ipotezelor implică formularea unei ipoteze nule (H_0) și a unei ipoteze alternative (H_1). Există trei tipuri de teste de ipoteză, în funcție de natura ipotezei alternative:

- ⇒ Test bilateral - ipoteza alternativă conține semnul " $\neq$ ".
- ⇒ Test unilateral stânga - ipoteza alternativă conține semnul " $<$ ".
- ⇒ Test unilateral dreapta - ipoteza alternativă conține semnul " $>$ ".

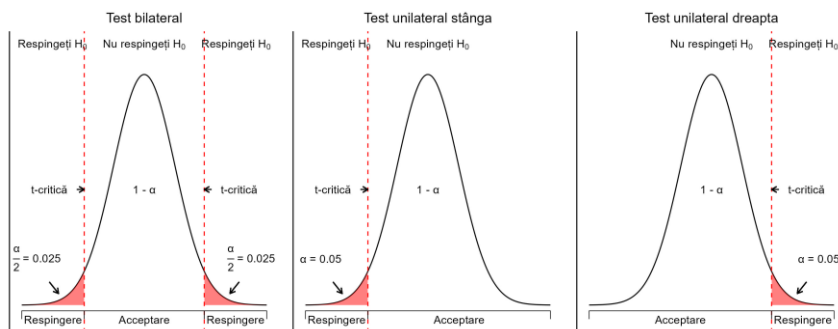


Figura 7.2. Ilustrarea valorilor critice, a regiunilor de respingere și de acceptare în testarea ipotezelor

Figura 7.2 ilustrează teste bilaterale, unilaterale stânga și unilaterale dreapta folosind o diagramă de distribuție a probabilității reprezentată de o curbă normală în formă de clopot. Suprafața de sub curbă este egală cu 1 sau 100%. O valoare critică t (linia întreruptă în figura 7.2) împarte zona de sub curbă în regiuni de respingere și de acceptare a ipotezei nule.

Pentru un *test bilateral*, ipoteza nulă este respinsă dacă statistica testului (t -calculat) este prea mică sau prea mare, creând două regiuni de respingere: una în stânga și una în dreapta. Ipoteza nulă este respinsă dacă statistica testului este fie mai mică, fie mai mare decât valorile critice.

Pentru un *test unilateral stânga*, ipoteza nulă este respinsă dacă statistica testului este prea mică, plasând regiunea de respingere în partea stângă a distribuției. Ipoteza nulă este respinsă dacă statistica testului este sub valoarea critică din stânga.

Pentru un *test unilateral dreapta*, ipoteza nulă este respinsă dacă statistica testului este prea mare, plasând regiunea de respingere în partea dreaptă a distribuției. Ipoteza nulă este respinsă dacă statistica testului depășește valoarea critică din dreapta.

Folosind abordarea valorii critice, determinăm dacă statistica testului calculat este mai mare (extremă) decât valoarea critică. Statistica testului calculat este comparată cu valoarea critică, care reprezintă un prag decisiv. Dacă statistica testului depășește această valoare critică, ipoteza nulă este respinsă. Dacă statistica testului nu atinge această valoare critică, ipoteza nulă nu este respinsă.

7.3. Teste parametrice

7.3.1. Testul t pentru un singur eșantion

Definiție. Testul t pentru un singur eșantion este o metodă statistică utilizată pentru a evalua dacă media unui singur eșantion este semnificativ diferită de o medie cunoscută sau ipotetică a populației. Acest test parametric este deosebit de util atunci când se dorește compararea mediei eșantionului cu o valoare standard sau anticipat.

Condiții pentru aplicarea testului t cu un singur eșantion:

- Datele trebuie să fie continue (scala de interval sau de raport).
- Eșantionul trebuie să fie selectat aleatoriu.
- Datele din eșantion trebuie să fie distribuite aproximativ normal.
- Deviația standard a populației este necunoscută.

Exemplu:

Să presupunem că un eșantion de 15 medici raportează numărul orelor de somn în timpul săptămânii de examene finale. Datele colectate sunt următoarele:

7, 6, 5, 6, 7, 8, 6, 7, 5, 6, 7, 8, 5, 6, 7

Scopul cercetării este de a verifica dacă media orelor de somn din această perioadă ($\bar{X}$) este semnificativ diferită de cele 8 ore de somn recomandate (media ipotetică a populației, μ).

Soluție:

1. Formularea ipotezelor:

- Ipoteza nulă (H_0): $\mu = 8$ (Durata medie a somnului este de 8 ore).
- Ipoteza alternativă (H_1): $\mu \neq 8$ (Durata medie a somnului nu este de 8 ore). În acest caz, ipoteza alternativă este bilaterală, indicând faptul că testul statistic va fi bilateral.

2. Selectarea testului statistic adecvat:

- Utilizarea testului t pentru un singur eșantion, deoarece se compară media eșantionului cu o medie cunoscută a populației.

3. Stabilirea nivelului de semnificație α :

- Alegem $\alpha = 0,05$.

4. Determinarea gradelor de libertate (GL) și a valorii critice:

- $GL = 15 - 1 = 14$
- Valoarea t critică pentru un test bilateral cu $GL = 14$ și $\alpha = 0,05$ poate fi identificată utilizând tabelul de distribuție t (vezi anexa A): t -critică = $\pm 2,145$.

5. Efectuarea calculului:

- Calcularea mediei eșantionului ($\bar{X}$):

$$\bar{X} = \frac{\sum X_i}{n} = \frac{96}{15} = 6,4$$

- Calcularea deviației standard a eșantionului (s):

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}} = \sqrt{\frac{\sum (7 - 6,4)^2 + (6 - 6,4)^2 + \dots + (7 - 6,4)^2}{15 - 1}} \approx 1,12$$

- Calcularea statisticii testului t folosind formula corespunzătoare testului t pentru un singur eșantion:

$$t = \frac{\bar{X} - \mu}{SE_{\bar{X}}} = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{6,4 - 8}{1,12/\sqrt{15}} = \frac{-1,6}{0,289} \approx -5,54$$

6. Formularea concluziilor:

- Compararea valorii t calculate cu valoarea t critică:
 - valoarea t calculată este $-5,54$
 - valoarea t critică este $\pm 2,145$.
- Concluzie: valoarea absolută a statisticii testului ($5,54$) depășește valoarea absolută a valorii critice ($2,145$). Astfel, respingem ipoteza nulă și concluzionăm că media orelor de somn ale mediciniștilor în săptămâna de examene finale este semnificativ diferită de cele 8 ore recomandate (valoarea $p < 0,05$).

Ilustrarea valorilor critice și a regiunilor de respingere:

În exemplul nostru, testul t pentru un singur eșantion este bilateral, fapt indicat de cele două valori critice $\pm 2,145$, corespunzătoare unui nivel de semnificație (α) de $0,05$ și GL de 14 (*figura 7.3*). Zonele umbrite indică regiunile de respingere, unde ipoteza nulă ar fi respinsă. Zona centrală neumbrită reprezintă regiunea unde ipoteza nulă nu ar fi respinsă. Valoarea t calculată ($-5,54$) se află în regiunea de respingere (dincolo de valoarea critică de $-2,145$), ceea ce duce la respingerea ipotezei nule.

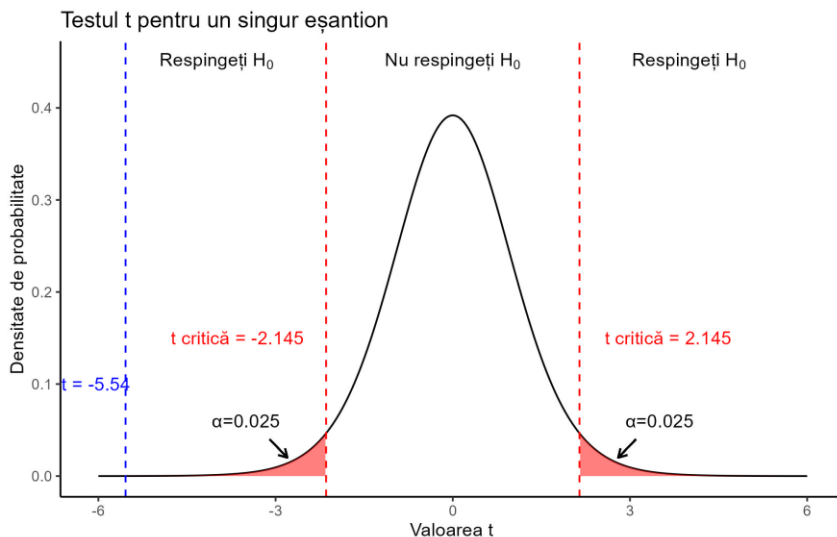


Figura 7.3. Testul t pentru un singur eșantion. Vizualizarea valorilor critice și a regiunilor de respingere pentru un test bilateral

7.3.4. Testul t pentru două eșantioane independente

Definiție. Testul t pentru două eșantioane independente se utilizează pentru a compara mediile a două grupuri independente, cu scopul de a stabili dacă există dovezi statistice că mediile populațiilor corespunzătoare sunt semnificativ diferite.

Condiții pentru utilizarea testului t pentru două eșantioane independente:

- Cele două eșantioane trebuie să fie selectate aleatoriu.
- Eșantioanele trebuie să fie independente unul de celălalt.
- Variabila dependentă trebuie măsurată pe o scală de interval sau de raport.
- Datele din fiecare grup trebuie să fie distribuite aproximativ normal.

- Testul poate fi aplicat în cazul varianțelor egale sau inegale în cele două grupuri. În funcție de această condiție, se folosesc formule diferite.

Exemplu:

Un cercetător dorește să determine dacă există o diferență semnificativă în timpul mediu de recuperare (în zile) între pacienții tratați cu medicamentul A și cei tratați cu medicamentul B pentru o anumită boală. Cercetătorul colectează date privind timpul de recuperare de la două grupuri independente de pacienți.

Date:

- Grupul A (medicamentul A):
 - media eșantionului ($\bar{X}_A$) - 8 zile
 - deviația standard a eșantionului (s_A) - 2 zile
 - mărimea eșantionului (n_A) - 31
- Grupul B (medicamentul B):
 - media eșantionului ($\bar{X}_B$) - 6 zile
 - deviația standard a eșantionului (s_B) - 1,5 zile
 - mărimea eșantionului (n_B) - 31

Soluție:

Vom aplica un test cu ipoteză unilaterală, presupunând varianțe inegale.

1. Formularea ipotezelor:

- Ipoteza nulă (H_0) - $\mu_A = \mu_B$. Nu există nicio diferență în timpul mediu de recuperare între pacienții tratați cu medicamentul A și cei tratați cu medicamentul B. H_0 poate fi formulată și ca $\mu_A - \mu_B = 0$.
- Ipoteza alternativă (H_1) - $\mu_A > \mu_B$. Durata medie de recuperare a pacienților tratați cu medicamentul A este mai mare decât cea

a pacienților tratați cu medicamentul B. H_1 poate fi formulată și ca $\mu_A - \mu_B > 0$.

2. Selectarea testului statistic adecvat:

- Utilizăm testul t pentru două eșantioane independente cu varianțe inegale (testul Welch).

3. Stabilirea nivelului de semnificație α :

- Alegem $\alpha = 0,05$.

4. Determinarea gradelor de libertate (GL) și a valorii critice:

- $GL = 31 + 31 - 2 = 60$
- Valoarea t critică pentru un test cu ipoteză unilaterală și $GL = 60$, la nivelul de semnificație $\alpha = 0,05$, poate fi determinată din tabelul de distribuție t (vezi anexa A). t-critică = 1,671.

5. Efectuarea calculului:

- Calcularea statisticii testului t folosind formula corespunzătoare testului t pentru două eșantioane independente:

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{8 - 6}{\sqrt{\frac{2^2}{31} + \frac{1,5^2}{31}}} = \frac{2}{\sqrt{0,129 + 0,072}} \approx 4,46$$

6. Formularea concluziilor:

- Compararea valorii t calculate cu valoarea t critică:
 - valoarea t calculată este 4,46
 - valoarea t critică este 1,671.
- Concluzie. În urma calculelor efectuate, statistica testului a fost determinată ca fiind 4,46. Valoarea absolută a statisticii testului depășește valoarea critică de 1,671. Prin urmare, respingem ipoteza nulă și concluzionăm că există dovezi semnificative care sugerează că timpul mediu de recuperare pentru pacienții tratați cu medicamentul A este mai mare comparativ cu cel pentru pacienții tratați cu medicamentul B (valoarea $p < 0,05$).

Ilustrarea valorilor critice și a regiunilor de respingere

În exemplul nostru, valoarea t calculată (4,46) se încadrează în regiunea de respingere (depășind valoarea critică de 1,671). Aceasta conduce la respingerea ipotezei nule. Testul t pentru două eșantioane independente utilizat este un test unilateral dreapta. Zona umbrită din partea dreaptă reprezintă regiunea de respingere, unde ipoteza nulă ar fi respinsă. Restul zonei neumbrite indică regiunea în care ipoteza nulă nu ar fi respinsă (figura 7.4).

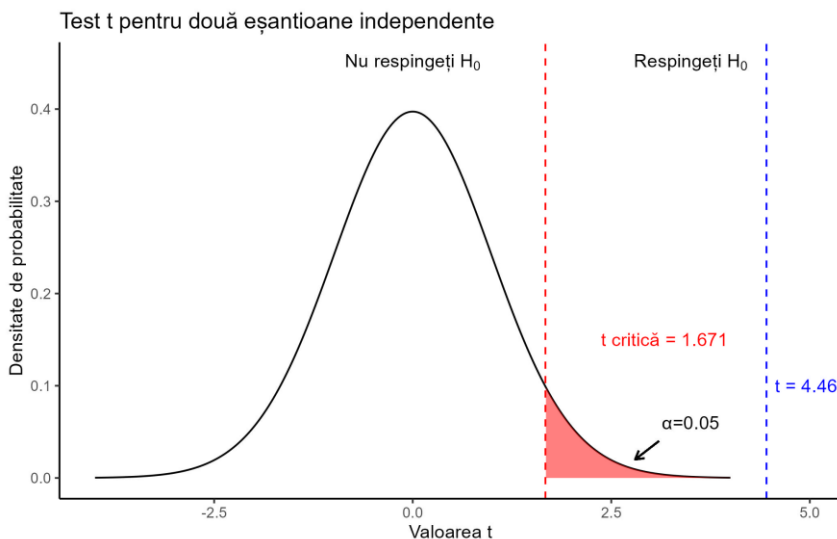


Figura 7.4. Testul t pentru două eșantioane independente. Vizualizarea valorilor critice și a regiunilor de respingere pentru un test unilateral dreapta

7.3.5. Testul t pentru coeficientul de corelație

Definiție. Testul t pentru coeficientul de corelație ρ (rho) determină dacă există o relație liniară semnificativă din punct de vedere statistic între două variabile continue.

Condiții pentru utilizarea testului coeficientului de corelație:

- Ambele variabile trebuie să fie continue.
- Datele trebuie să urmeze o distribuție bivariată.
- Eșantionul trebuie să fie selectat aleatoriu.
- Relația dintre variabile trebuie să fie liniară.

Exemplu:

Un cercetător studiază relația dintre numărul de ore de activitate fizică săptămânală a pacienților și nivelurile de colesterol din sânge. Au fost colectate date de la 30 de pacienți. Pentru fiecare pacient, s-a înregistrat numărul de ore de exerciții pe săptămână și nivelul de colesterol (mg/dl). Coeficientul de corelație calculat (r) este de $-0,15$. Scopul este de a determina dacă există o corelație negativă semnificativă din punct de vedere statistic între durata exercițiilor fizice și nivelul de colesterol.

Soluție:

1. Formularea ipotezelor:

- Ipoteza nulă (H_0): $\rho = 0$. Nu există nicio corelație între numărul de ore de exerciții fizice pe săptămână și nivelurile de colesterol.
- Ipoteza alternativă (H_1): $\rho < 0$. Există o corelație negativă între numărul de ore de exerciții fizice pe săptămână și nivelurile de colesterol.

2. Selectarea testului statistic adecvat:

- Se utilizează testul de corelație Pearson.

3. Stabilirea nivelului de semnificație α :

- Alegem $\alpha = 0,05$.

4. Determinarea gradelor de libertate (GL) și a valorii critice:

- $GL = 30 - 2 = 28$

- Valoarea critică t pentru un test unilateral cu $GL = 28$ și $\alpha = 0,05$ poate fi găsită în tabelul de distribuție t (vezi anexa A). t -critică = 1,701.

5. Efectuarea calculului:

- Având $r = -0,15$, se calculează statistica testului t utilizând formula:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$t = \frac{-0,15\sqrt{30-2}}{\sqrt{1-(-0,15)^2}} = \frac{-0,15\sqrt{28}}{\sqrt{1-0,0225}} = \frac{-0,15 \times 5,2915}{0,9887} = \frac{-0,7937}{0,9887} = -0,803$$

6. Formularea concluziilor:

- Compararea valorii t calculate cu valoarea t critică:
 - valoarea t calculată este -0,803
 - valoarea t critică este 1,701.
- Concluzie: Pe baza calculelor efectuate, statistica testului a fost determinată la -0,803. Valoarea absolută a statisticii testului (0,803) este inferioară valorii critice de 1,701. Astfel, *nu respingem ipoteza nulă* și concluzionăm că nu există o corelație negativă semnificativă statistic între durata exercițiilor fizice și nivelurile de colesterol.

Ilustrarea valorilor critice și a regiunilor de respingere

În *figura 7.5* este reprezentat testul unilateral stânga pentru coeficientul de corelație. Valoarea critică t este indicată printr-o linie punctată la -1,701. Dacă valoarea t calculată se situează în zona umbrită, ipoteza nulă se respinge. În acest exemplu, valoarea t calculată de -0,803 se situează în regiunea de acceptare, ceea ce sugerează că nu există o corelație negativă semnificativă din punct de vedere statistic între orele de exerciții și nivelurile de colesterol.

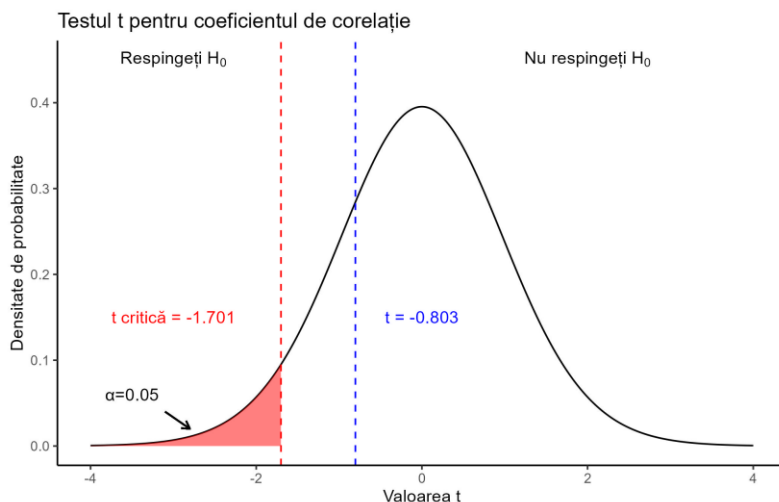


Figura 7.5. Testul t pentru coeficientul de corelație. Vizualizarea valorilor critice și a regiunilor de respingere pentru un test unilateral stânga

7.4. Testele neparametrice

Testele neparametrice nu presupun că datele dintr-o populație sunt distribuite normal, motiv pentru care sunt denumite teste fără presupuneri de distribuție. Aceste teste sunt utilizate pentru analiza datelor nominale, ordinale sau numerice asimetrice.

7.4.1. Testul chi-pătrat

Definiție. Testul chi-pătrat este o metodă statistică utilizată pentru a stabili dacă există o asociere semnificativă între două variabile categoricale.

Condiții pentru utilizarea testului chi-pătrat:

- Datele trebuie să fie prezentate sub formă de numărători sau frecvențe într-un tabel de contingență.

- Categoriile trebuie să fie mutual exclusive.
- Mărimea eșantionului trebuie să fie suficient de mare (frecvențele așteptate în fiecare celulă trebuie să fie de cel puțin 5).

Exemplu:

Un cercetător analizează relația între statutul de fumător (fumător vs. nefumător) și incidența cancerului pulmonar (prezent vs. absent). Au fost colectate date de la 100 de pacienți, așa cum este prezentat în *tabelul 7.1*. Este necesar să se stabilească dacă există o asociere statistic semnificativă între statutul de fumător și cancerul pulmonar.

Tabelul 7.1. Statutul de fumător și cancerul pulmonar într-un tabel 2x2 (frecvențe înregistrate)

	Cancer pulmonar prezent	Cancer pulmonar absent	Total
Fumător	30	20	50
Nefumător	10	40	50
Total	40	60	100

Soluție:

1. Formularea ipotezelor:

- Ipoteza nulă (H_0) - Nu există o asociere între statutul de fumător și cancerul pulmonar.
- Ipoteza alternativă (H_1) - Există o asociere între statutul de fumător și cancerul pulmonar.

2. Selectarea testului statistic adecvat:

- Se utilizează testul chi-pătrat.

3. Stabilirea nivelului de semnificație α :

- Alegem $\alpha = 0,05$.

4. Determinarea gradelor de libertate (GL) și a valorii critice:

- $GL = (\text{numărul de rânduri} - 1) \times (\text{numărul de coloane} - 1)$
- $GL = (2 - 1) \times (2 - 1) = (2 - 1) \times (2 - 1) = 1$
- Valoarea critică pentru $GL = 1$ și $\alpha = 0,05$ poate fi găsită utilizând un tabel al distribuției chi-pătrat (vezi anexa B). Aceasta este 3,841.

5. Efectuarea calculului:

- Se calculează frecvențele așteptate pentru fiecare celulă:

$$E_{ij} = \frac{(\text{Total rând}) \times (\text{Total coloană})}{\text{Total general}}$$

De exemplu, pentru celula „Fumător și cancer pulmonar prezent”:

$$E_{1,1} = \frac{50 \times 40}{100} = 20$$

În același mod, se calculează valorile așteptate în celelalte celule, așa cum sunt prezentate în tabelul 7.2.

Tabelul 7.2. Statutul de fumător și cancerul pulmonar într-un tabel 2x2 (frecvențele așteptate)

	Cancer pulmonar prezent	Cancer pulmonar absent	Total
Fumător	20	30	50
Nefumător	20	30	50
Total	40	60	100

- Calcularea testului chi-pătrat (χ^2) folosind frecvențele observate (O_{ij}) și cele așteptate (E_{ij}):

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2 = \frac{(30 - 20)^2}{20} + \frac{(20 - 30)^2}{30} + \frac{(10 - 20)^2}{20} + \frac{(40 - 30)^2}{30} = 16,66$$

6. Formularea concluziilor:

- Se compară valoarea calculată a lui chi-pătrat cu valoarea critică:
 - valoarea calculată a lui chi-pătrat este 16,66.
 - valoarea critică este 3,841.
- Concluzie. Valoarea calculată a testului chi-pătrat (16,66) este mai mare decât valoarea critică de 3,841. Prin urmare, *ipoteza nulă se respinge* și se concluzionează că există o asocierie statistic semnificativă între statutul de fumător și cancerul pulmonar ($p < 0,05$).

Exerciții de recapitulare

Folosiți următoarele seturi de date și:

1. Calculați media pentru ambele grupuri.
2. Calculați măsurile de variație și verificați dacă mediile sunt reprezentative.
3. Calculați intervalul de încredere pentru $\alpha = 0,05$.
4. Comparați mediile folosind un test t și formulați concluzii despre semnificația statistică a diferenței dintre medii.

Ipoteze:

H_0 : Diferența dintre mediile eșantioanelor nu este semnificativă statistic.

H_1 : Diferența dintre mediile eșantioanelor este semnificativă statistic.

$p > 0,05 \Rightarrow H_0$ acceptată

$p < 0,05 \Rightarrow H_0$ respinsă

Setul de date 1. Rezultatele nivelurilor de colesterol din sânge pentru două eșantioane independente ($n = 15$ fiecare):

Nr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Gr. 1	168	258	228	247	156	172	165	210	264	220	258	200	195	245	189
Gr. 2	136	148	125	121	157	148	116	140	161	122	128	122	137	139	128

Setul de date 2. Rezultatele tensiunii arteriale sistolice (TAS) pentru două eșantioane independente ($n = 12$ fiecare):

Nr.	1	2	3	5	6	7	8	9	10	11	12
Gr.1	130	130	120	110	90	120	125	115	135	140	120
Gr.2	170	175	160	170	170	190	185	185	170	160	190

Întrebări de recapitulare

1. Care sunt metodele parametrice de testare a ipotezelor?
2. Care sunt metodele neparametrice de testare a ipotezelor?
3. Când se aplică metodele parametrice de testare a ipotezelor?
4. Când se aplică metodele neparametrice de testare a ipotezelor?
5. Care sunt principalele diferențe între metodele parametrice și cele neparametrice?
6. Când se folosește testul t pentru un singur eșantion?
7. Când se folosește testul t pentru două eșantioane independente?
8. Când se folosește testul t pentru coeficientul de corelație?
9. Când se folosește testul chi-pătrat?

CAPITOLUL 8. INTRODUCERE ÎN METODOLOGIA CERCETĂRII

Concepte cheie

- ❖ *Din punct de vedere al aplicării*, cercetările se clasifică în studii fundamentale și aplicate.
- ❖ *Din punct de vedere al metodologiei utilizate* pentru a răspunde la o întrebare de cercetare, studiile se împart în cercetări cantitative și calitative.
- ❖ În domeniul sănătății, designul studiilor este împărțit în două categorii majore: studii *observaționale*, în care subiecții sunt observați fără intervenție, și studii *experimentale*, în care este evaluat efectul unei intervenții.
- ❖ Designul studiilor *descriptive* poate doar să formuleze ipoteze privind o posibilă legătură între un factor de risc și un rezultat.
- ❖ Designul studiilor *analitice* are capacitatea de a evalua și analiza relația dintre un factor de risc și o boală, oferind posibilitatea testării ipotezelor de cauzalitate.
- ❖ *Sinteza sistematică* este un studiu secundar care rezumă literatura științifică existentă.
- ❖ *Eroarea sistematică* apare atunci când modul în care un studiu este conceput sau realizat duce la erori în rezultate și concluzii. Aceasta poate apărea din cauza unei selecții necorespunzătoare a subiecților sau a metodelor inadecvate de colectare și analiză a datelor.
- ❖ *Eroarea de evocare* (memorie) apare atunci când participanții la un studiu au o probabilitate sistematic diferită de a-și aminti informațiile legate de expunere, în funcție de starea lor de sănătate.

8.1. Definiția, caracteristicile și tipurile de cercetări

8.1.1. Definiția și caracteristicile cercetării

Cercetarea reprezintă o activitate sistematică ce utilizează metodologii științifice riguroase pentru a aborda probleme complexe și a genera noi cunoștințe cu aplicabilitate extinsă.

Cercetarea implică un proces structurat de colectare, analiză și interpretare a datelor pentru a răspunde la anumite întrebări fundamentale. Aceasta se caracterizează prin următoarele aspecte:

1. **Obiectivitate.** Cercetarea trebuie să fie lipsită de orice formă de prejudecată personală sau de eroare sistematică. Acestea din urmă pot apărea dacă studiul este proiectat sau realizat într-un mod care conduce la denaturarea rezultatelor și concluziilor. Erorile sistematice pot fi cauzate de o selecție incorectă a subiecților sau de utilizarea unor metode necorespunzătoare de colectare și analiză a datelor.
2. **Validitate.** Validitatea este sinonimă cu acuratețea cercetării (acuratețea procedurilor, a instrumentelor de cercetare, a testelor etc.). Există două tipuri de validitate:
 - *Validitatea internă* - rezultatele sunt valabile pentru eșantionul specific de subiecți care au fost efectiv studiați.
 - *Validitatea externă* - rezultatele sunt aplicabile populației din care a fost extras eșantionul. Dacă rezultatele pot fi extinse și la alte populații, studiul este considerat *generalizabil*.
3. **Fiabilitate.** Fiabilitatea, sinonimă cu repetabilitatea sau reproductibilitatea, implică faptul că rezultatele pot fi reproduse și verificate atât de cercetătorul inițial, cât și de alți cercetători.
4. **Comparabilitate.** Procesul de investigare trebuie să fie minuțios și lipsit de erori, astfel încât concluziile și rezultatele cercetării să reziste unei analize critice și să fie comparabile cu alte studii.

5. *Abordare sistematică.* Toate procedurile de investigare trebuie să urmeze o secvență logică riguroasă. Fiecare etapă trebuie realizată într-un mod ordonat, asigurând coerența și consistența pe tot parcursul procesului de cercetare.
6. *Relevanță.* Există două tipuri esențiale de relevanță:
 - ⇒ *Relevanța științifică* - studiul contribuie la avansarea cunoștințelor noastre într-un anumit domeniu științific, fie că este vorba de un concept, un proces sau o boală, consolidând baza de cunoștințe științifice;
 - ⇒ *Relevanța socială* - studiul oferă beneficii directe societății, cum ar fi îmbunătățirea sănătății publice, informarea deciziilor politice sau creșterea calității vieții prin aplicații practice ale rezultatelor cercetării.

8.1.2. Erori aleatorii și erori sistematice în cercetare

În cercetare, erorile sistematice pot avea un impact profund asupra validității și fiabilității rezultatelor obținute. Diferențierea între erorile aleatorii și erorile sistematice este esențială pentru proiectarea riguroasă a studiilor și pentru interpretarea precisă a rezultatelor.

⇒ Erori aleatorii

Erorile aleatorii reprezintă variații imprevizibile care apar în procesul de măsurare. Aceste fluctuații pot proveni din variațiile instrumentelor de măsurare sau din alți factori imprevizibili. Deși erorile aleatorii pot influența precizia rezultatelor, ele nu generează, în mod obișnuit, erori sistematice într-o anumită direcție. Astfel, erorile aleatorii au un impact redus asupra rezultatelor finale ale studiului.

Exemplu: Variabilitatea în măsurătorile tensiunii arteriale poate fi cauzată de diferențe subtile în poziționarea manșetei sau în postura subiectului în timpul măsurătorilor. Fiecare măsurătoare poate varia ușor, dar în ansamblu, ar trebui să se concentreze în jurul valorii reale.

⇒ **Erori sistematice**

Erorile sistematice (*bias*) sunt inexactități constante și repetabile care rezultă din erorile în proiectarea studiului, colectarea datelor sau metodele de analiză. Aceste erori pot conduce la concluzii eronate, printr-o distorsiune consistentă a rezultatelor într-o anumită direcție. Erorile sistematice au un impact semnificativ asupra rezultatelor studiului.

În general, se disting trei tipuri majore de erori sistematice:

1. **Confuzia (*confounding*)**. Aceasta apare atunci când factorul principal de risc investigat este confundat cu o altă variabilă, ceea ce face dificilă izolarea efectului real al factorului principal de risc.

Exemplu: Într-un studiu care examinează relația dintre consumul de cafea și bolile de inimă, dacă consumatorii de cafea tind să fumeze mai mult, devine dificil să se discearnă impactul specific al consumului de cafea asupra riscului de boli de inimă.

2. **Eroare de selecție**. Aceasta este o eroare sistematică care apare atunci când participanții sau subiecții incluși într-un studiu nu reflectă în mod fidel populația țintă, ceea ce duce la rezultate distorsionate.

Exemple:

- ✓ *Eroare de non-răspuns*. Dacă persoanele care nu răspund la un sondaj diferă semnificativ de cei care răspund, rezultatele obținute pot să nu fie reprezentative pentru întreaga populație.
- ✓ *Eroare de excludere*. Dacă anumite grupuri sunt excluse sistematic din studiu, concluziile obținute pot să nu fie generalizabile. De exemplu, excluderea pacienților vârstnici dintr-un studiu clinic poate conduce la rezultate care nu sunt aplicabile populației în vârstă.

- ✓ *Eroare de volum al eșantionului.* Dacă mărimea eșantionului este prea mică sau nu este randomizată adecvat, rezultatele pot să nu reflecte corect populația generală.

3. Eroare de informație. Acest tip de distorsionare apare în timpul colectării datelor, adesea ca rezultat al erorilor sistematice de măsurare sau al clasificării incorecte a subiecților.

Exemple:

- ✓ *Eroarea de intervievare.* În cazul în care interviuatorii nu sunt suficient de pregătiți sau dacă influențează neintenționat răspunsurile, datele colectate pot fi afectate de erori (bias).
- ✓ *Eroarea de evocare (memorie) (recall bias).* În studiile retrospective, participanții pot avea dificultăți în a-și aminti cu precizie evenimentele sau expunerile anterioare, ceea ce poate conduce la date distorsionate. De exemplu, pacienții cu o anumită afecțiune pot rememora expunerea la factori de risc diferit față de persoanele sănătoase.
- ✓ *Eroarea de raportare.* Dacă participanții dezvăluie selectiv informații, precum subestimarea comportamentelor socialmente nedorite (de exemplu, consumul de tutun sau de alcool), rezultatele studiului pot fi influențate negativ.

8.1.3. Tipuri de cercetări

Cercetările pot fi clasificate din trei perspective principale, așa cum este detaliat în *tabelul 8.1*:

- aplicabilitatea cercetării;
- metodologia cercetării utilizată;
- obiectivele cercetării.

Tabelul 8.1. Tipuri de cercetări după perspective

<i>Aplicarea cercetării</i>	<i>Metodologia cercetării</i>	<i>Obiectivele cercetării</i>
1. Fundamentală	1. Cantitativă	1. Istorică
2. Aplicată	2. Calitativă	2. Descriptivă
		3. Corelațională
		4. Experimentală
		5. Exploratorie

Tipuri de cercetări în funcție de aplicare:

- ⇒ *Cercetare fundamentală* - acest tip de cercetare se concentrează pe dezvoltarea și testarea teoriei și ipotezelor care sunt intelectual provocatoare. Deși nu poate avea aplicații practice imediate, contribuie la avansarea cunoștințelor teoretice.
- ⇒ *Cercetare aplicată* - această cercetare este orientată spre rezolvarea problemelor practice specifice. Scopul este de a aborda probleme reale.

Tipuri de cercetări în funcție de metodologia utilizată:

- ⇒ *Cercetare cantitativă* - această abordare permite cuantificarea amplitudinii unei probleme, chestiuni sau fenomen, prin măsurarea variabilelor și analiza statistică. Întrebarea centrală este: „Câte sunt?”.
 - ⇒ *Cercetare calitativă* - această abordare explorează natura unei probleme, chestiuni sau fenomen fără a-l cuantifica, concentrându-se pe înțelegerea „cum” și „de ce” prin date descriptive.
- Ambele abordări, cantitativă și calitativă, au caracteristici unice și sunt adesea complementare.

Tipuri de cercetări în funcție de obiectivele urmărite:

- ⇒ *Cercetare istorică* - se axează pe formularea de concluzii referitoare la evenimente, tendințe, cauze sau efecte din trecut. Aceasta implică adesea analiza surselor primare sau interviuarea martorilor oculari. Acest tip de cercetare contribuie la înțelegerea evenimentelor curente și la prognozarea tendințelor viitoare.
- ⇒ *Cercetare descriptivă* - oferă o descriere sistematică a unei situații, probleme sau fenomene. Se concentrează pe colectarea de date pentru a răspunde la întrebări referitoare la starea actuală a subiectului. Metodele utilizate includ chestionare, interviuri sau observații pentru a aduna informații despre condițiile actuale.
- ⇒ *Cercetare corelațională* - vizează identificarea și măsurarea relațiilor dintre două sau mai multe variabile. Aceasta merge dincolo de descriere, explorând modul în care variabilele sunt interconectate și poate fi folosită pentru a face predicții sau a testa ipoteze. Cercetarea corelațională este frecvent utilizată pentru validarea instrumentelor și metodologiilor predictive.
- ⇒ *Cercetare experimentală* - se concentrează pe stabilirea relațiilor cauză-efect prin manipularea activă a variabilelor și observarea efectelor acestora. Spre deosebire de cercetarea corelațională, cercetarea experimentală implică intervenția directă a cercetătorului pentru a determina aceste relații cauzale.
- ⇒ *Cercetare exploratorie*. - se realizează în condițiile în care cunoștințele despre un subiect sunt limitate și are ca obiectiv investigarea fezabilității unui studiu sau identificarea domeniilor pentru cercetări viitoare. Este adesea utilizată ca studiu preliminar sau pilot pentru a explora idei noi și a colecta date preliminare.

8.2. Etapele procesului de cercetare

În cadrul desfășurării cercetării, sunt necesare două tipuri de decizii esențiale:

1. Ce anume se dorește să se descopere.
2. Cum se va proceda pentru a realiza această descoperire.

Pentru a obține răspunsuri la întrebările de cercetare, este necesar să se urmeze o serie de pași practici. Abordarea aleasă pentru a răspunde la aceste întrebări este cunoscută sub denumirea de metodologie de cercetare. La fiecare etapă a procesului de cercetare, se alege dintr-o varietate de metode și tehnici concepute pentru a atinge în mod optim obiectivele cercetării. Pentru a asigura un proces de cercetare sistematic și eficient, trebuie urmați următorii pași:

1. **Formularea problemei de cercetare** - se formulează clar problema sau întrebarea care se dorește a fi explorată. Aceasta implică identificarea lacunelor în cercetare și specificarea aspectelor care se intenționează să se investigheze.
2. **Efectuarea sintezei literaturii** - se examinează cercetările existente și literatura relevantă pentru subiectul în cauză. Aceasta ajută la înțelegerea stării actuale a cunoștințelor, rafinarea problemei de cercetare și identificarea lacunelor care ar putea fi abordate de studiul preconizat spre realizare.
3. **Formularea scopului și obiectivelor** - se definește clar scopul principal al cercetării și se detaliază în obiective specifice și măsurabile. Aceste obiective vor ghida dezvoltarea designului și metodologiei cercetării.
4. **Dezvoltarea designului cercetării** - se elaborează un plan detaliat care include metodologia cercetării, designul studiului, metodele de eșantionare, procedurile de colectare a datelor și tehnicile de analiză.

5. **Colectarea datelor** - se adună datele necesare utilizând metodele selectate. Acest lucru poate include studii, interviuri, experimente sau analiza datelor secundare, în funcție de designul cercetării.
6. **Analiza datelor** - se procesează și se analizează datele colectate. Se folosesc tehnici statistice adecvate, bazate pe obiectivele cercetării și tipul de date.
7. **Generalizarea și interpretarea rezultatelor** - se formulează concluzii pe baza analizei datelor. Se interpretează rezultatele în raport cu obiectivele cercetării și se elaborează recomandări bazate pe aceste rezultate.
8. **Prezentarea rezultatelor** - se comunică constatările cercetării printr-un raport bine structurat și, dacă este necesar, printr-o prezentare orală.

8.3. Formularea problemei de cercetare

Formularea problemei de cercetare constituie primul și cel mai esențial pas în procesul de cercetare. O problemă de cercetare bine conturată oferă direcție și concentrare întregului demers. În selectarea problemei de cercetare, se iau în considerare următoarele aspecte:

- ⇒ *Interesul* - se optează pentru o temă care stârnește un interes autentic. Pasiunea pentru subiect va sprijini menținerea motivației pe parcursul întregului proces de cercetare.
- ⇒ *Magnitudinea* - se va asigura că tema este gestionabilă în contextul timpului și resurselor disponibile.
- ⇒ *Relevanța* - cercetarea ar trebui să aducă o contribuție semnificativă la corpul de cunoștințe existente.
- ⇒ *Disponibilitatea datelor* - se va verifica accesibilitatea și posibilitatea colectării datelor necesare.
- ⇒ *Considerentele etice* - se va asigura că cercetarea respectă standardele etice relevante.

Formularea problemei de cercetare implică mai mulți pași fundamentali. O cunoaștere adecvată a domeniului general este crucială pentru a defini clar și eficient problema de cercetare. Etapele de urmat pentru a contura o problemă de cercetare solidă:

1. *Identificarea domeniului general de interes* - se începe prin selectarea unei arii generale care corespunde intereselor și expertizei cercetătorului.
2. *Descompunerea domeniului general în subdomenii*- se segmentează domeniul general în subdomenii mai specifice pentru a restrânge aria de investigare.
3. *Selectarea subdomeniilor de interes*se aleg subdomeniile cele mai relevante și atrăgătoare pentru obiectivele cercetării.
4. *Dezvoltarea întrebărilor de cercetare* - se formulează întrebări precise la care urmează să răspundă cercetarea. Aceste întrebări vor ghida direcția studiului.
5. *Stabilirea ipotezei* - pe baza întrebărilor de cercetare, se elaborează ipoteze care propun răspunsuri sau explicații posibile.
6. *Verificarea repetată* - se va asigura că problema cercetării este bine definită, realizabilă și aliniată cu obiectivele de cercetare.

8.4. Sinteza literaturii

Sinteza literaturii constituie o etapă fundamentală în procesul de cercetare, esențial pentru înțelegerea corpusului existent de cunoștințe în domeniul de interes. Acest pas nu doar că oferă context pentru cercetarea efectuată, dar contribuie în mod substanțial la rafinarea și orientarea fiecărei etape a studiului.

Funcțiile sintezei literaturii:

1. Clarifică și focalizează problema de cercetare.
2. Îmbunătățește metodologia studiului.
3. Extinde domeniul de cunoștințe.
4. Contextualizează rezultatele obținute.

Proceduri pentru realizarea sintezei literaturii:

- ⇒ **Identificarea literaturii existente** - se începe prin identificarea și colectarea literaturii relevante pentru domeniul de cercetare. Se utilizează baze de date academice și biblioteci pentru a găsi sursele necesare.
- ⇒ **Analiza literaturii selectate** - se efectuează o analiză critică și o sinteză a literaturii pentru a evalua starea actuală a cercetărilor pe subiectul în cauză.
- ⇒ **Dezvoltarea unui cadru teoretic** - se elaborează un cadru teoretic care conturează teoriile și conceptele ce vor ghida cercetarea.
- ⇒ **Elaborarea unui cadru practic** - se proiectează un cadru practic care integrează metodologii, tehnici și abordări relevante pentru studiul în cauză.
- ⇒ **Organizarea și documentarea surselor** - se va asigura catalogarea sistematică a surselor de literatură analizate.

Instrumente și resurse

Pentru a căuta literatura în mod eficient, este necesară înțelegerea clară a temei de cercetare și stabilirea parametrilor de căutare. Pot fi folosite următoarele baze de date și resurse online:

- MedLine
https://www.nlm.nih.gov/medline/medline_overview.html
- Research4Life <https://portal.research4life.org/>
- HINARI
<https://www.emro.who.int/information-resources/hinari/hinari.html>
- PubMed <https://pubmed.ncbi.nlm.nih.gov/>
- și multe altele...

Pentru un articol academic, trebuie să se folosească cărți, articole și site-uri web care adună informații relevante despre subiectul cercetat. În scrierea academică, este esențial să se recurgă la o varietate de surse, inclusiv cărți, articole științifice și site-uri web de încredere. În timpul sintezei literaturii, se iau notițe detaliate despre metodologiile și instrumentele utilizate în cercetările anterioare. Acest lucru va facilita alegerea abordărilor adecvate pentru propriul studiu, inclusiv tehnicile de eșantionare, metodele de colectare a datelor și procedurile analitice. Se va asigura citarea surselor conform unui format standardizat. Bibliografia sintezei literaturii trebuie să constituie o listă completă și clară a tuturor surselor menționate, organizată alfabetic după numele autorilor. Cele trei stiluri de citare cel mai frecvent utilizate sunt:

⇒ ***Sistemul Harvard*** - stilul de citare „autor-data”.

- Citarea în text: (Smith, 2020).
- Lista de referințe: Smith, J. (2020). *Understanding Modern Research*. Oxford University Press.

⇒ ***Sistemul Vancouver*** - stilul de citare numeric introdus în Canada în 1978.

- Citarea în text: (1).
- Lista de referințe: Smith, J. *Understanding Modern Research*. Oxford University Press; 2020.

⇒ **Sistemele literă-număr** - un sistem de citare hibrid care combină elemente ale sistemelor numerice și alfabetice.

- Citarea în text: (Smith 2020a).
- Lista de referințe: Smith J. *Understanding Modern Research*. Oxford University Press; 2020a.

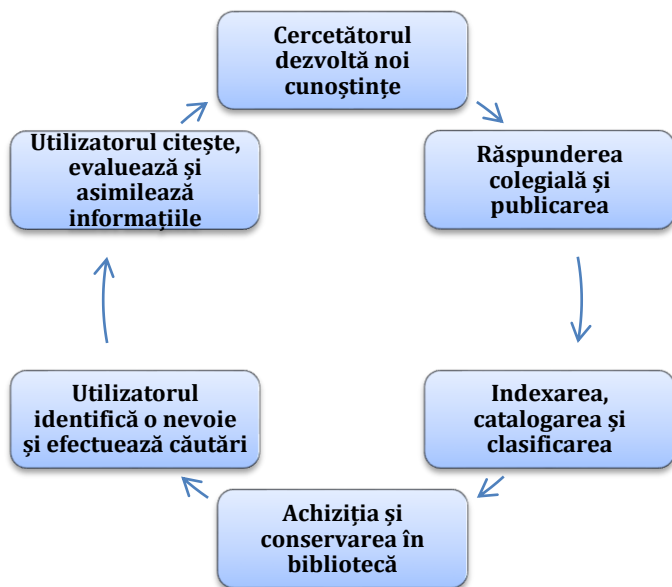


Figura 8.1. Ciclul informațional

Diagrama 8.1 ilustrează modul în care informația trece prin diferite etape pentru a deveni parte a cunoștințelor publicate. Aceste cunoștințe sunt ulterior disponibile cercetătorilor pentru a fi utilizate în continuarea studiilor, pentru a inspira noi domenii de cercetare și pentru a crea noi cunoștințe.

8.5. Formularea scopului și obiectivelor studiului

Scopul reprezintă obiectivul general pe care studiul își propune să-l atingă, servind ca declarație generală a finalității cercetării. Obiectivele sunt sarcinile specifice necesare pentru realizarea acestui scop.

Este esențial ca aceste obiective să fie formulate într-un mod clar și specific. Obiectivele trebuie enumerate numeric, fiecare având în vedere un singur aspect al studiului. În formularea obiectivelor, este recomandat să se folosească verbe orientate spre acțiune. Prin urmare, obiectivele ar trebui să înceapă cu termeni precum: a determina, a descoperi, a stabili, a măsura, a explora etc.

8.6. Pregătirea designului de cercetare și colectarea datelor

8.6.1. Definirea designului de cercetare și pașii necesari

Designul de cercetare reprezintă cadrul conceptual în care se desfășoară cercetarea.

Funcția sa este de a asigura colectarea de informații relevante cu un minim de efort, timp și resurse financiare. Pregătirea unui design de cercetare adecvat pentru o problemă de cercetare specifică implică următorii pași:

- ⇒ Determinarea *designului eșantionului*.
- ⇒ Elaborarea instrumentelor pentru colectarea datelor.
- ⇒ Adoptarea unui *design de cercetare*.

8.6.2 Determinarea designului eșantionului

Cercetătorii de obicei trag concluzii despre o populație studiind un eșantion. Proiectarea eșantionului implică trei decizii fundamentale:

1. Cine va fi supus cercetării? (*Eșantionul*)

2. Câte persoane vor fi incluse în cercetare? (*Mărimea eșantionului*)
3. Cum ar trebui selectat eșantionul? (*Tipul de eșantionare*)

8.6.3. Instrumente de colectare a datelor

Dezvoltarea unui instrument de cercetare pentru colectarea datelor reprezintă un aspect crucial al protocolului de cercetare. Rezultatele și concluziile se bazează pe datele colectate, care sunt influențate de întrebările incluse în chestionar.

Orientări pentru elaborarea unui chestionar:

Pasul I - definirea și listarea clară a tuturor obiectivelor specifice sau întrebărilor de cercetare ale studiului.

Pasul II - pentru fiecare obiectiv sau întrebare de cercetare, se enumeră toate întrebările asociate la care se dorește să se obțină răspunsuri în cadrul studiului.

Pasul III - pentru fiecare întrebare de cercetare listată la pasul I și fiecare obiectiv enunțat la pasul II, se identifică informațiile necesare pentru a răspunde la acestea.

Pasul IV - formularea întrebărilor pentru a obține aceste informații.

Un chestionar constă dintr-un set de întrebări adresate respondenților. Există numeroase modalități de a formula întrebările, dar este esențial să se asigure claritatea și să se obțină informațiile dorite. Chestionarul trebuie elaborat și testat cu rigurozitate înainte de a fi utilizat pe scară largă.

Tipuri de structură a întrebărilor:

- ⇒ Întrebări închise;
- ⇒ Întrebări deschise;
- ⇒ Combinație între cele două.

Întrebări închise - aceste întrebări includ toate răspunsurile posibile într-o serie de categorii predefinite, dintre care respondenții aleg (de exemplu, întrebări cu opțiuni multiple, întrebări cu scară). Întrebările închise sunt folosite pentru generarea de statistici în cercetările cantitative. Principalul avantaj al acestora este ușurința cu care răspunsurile pot fi analizate și raportate.

Întrebări deschise - aceste întrebări permit respondenților să răspundă cu propriile lor cuvinte. Avantajul principal constă în capacitatea de a captura gândurile respondenților în propriile lor cuvinte. Totuși, analiza acestor date poate fi mai complexă.

Întrebări combinate - această structură începe cu o serie de întrebări închise și se finalizează cu întrebări deschise pentru obținerea unor răspunsuri mai detaliate.

Tabelul 8.2. Întrebări deschise versus întrebări închise

Criterii	Întrebări deschise	Întrebări închise
Scopul	Captarea cuvintelor exacte sau a citatelor	Obținerea celor mai frecvente răspunsuri
Respondenții	Produc răspunsuri detaliate și aprofundate	Preferă răspunsuri rapide și simplificate
Contextul întrebării	Opțiunile sunt necunoscute	Opțiunile pot fi anticipate
Analiza	Analiza conținutului; consumatoare de timp	Numărare sau evaluare
Raportarea	Răspunsuri individuale sau grupate	Date statistice

Sursa: Conform Dawson and Trapp, 2004

Cele mai multe studii utilizează chestionare auto-administrate, fie personal, fie prin poștă ori e-mail sau interviuri, fie față-în-față, fie telefonic. Fiecare metodă prezintă avantaje și dezavantaje, unele dintre acestea fiind ilustrate în *tabelul 8.3*.

Tabelul 8.3. Avantajele și dezavantajele diferitelor metode de studii

	Autoadministrat poștă/e-mail	Autoadministrat în persoană	Interviu telefonic	Interviu față-în- față
Costul	++	+	-	-
Timpul	++	+	-	-
Standardizarea	+	+	+/-	+/-
Profundimea / detalierea	-	-	+	++
Rata de răspuns	-	++	+	++
Răspunsuri lipsă	-	+	++	++
+ Avantaje; - Dezavantaje; +/- Neutru				

Sursa: Dawson and Trapp, 2004.

Structura generală a chestionarului:

- ⇒ Titlu
- ⇒ Instrucțiuni
- ⇒ Informații generale despre respondent
- ⇒ Întrebări
- ⇒ Mesaj de mulțumire

8.6.5. Clasificarea designului de cercetare

Există diverse scheme pentru clasificarea metodelor de proiectare a studiilor. Una dintre aceste clasificări este ilustrată în tabelul 8.4. Această clasificare împarte designurile de cercetare în trei categorii principale: studii observaționale (în care participanții sunt monitorizați fără a se interveni asupra lor), studii experimentale (în care se implementează o anumită intervenție) și studii secundare (în care studiile primare sunt supuse unei pre-evaluări sau filtrări).

Tabelul 8.4. Clasificarea designului de cercetare

DESIGNUL	METODA	TIPURILE
I. Studii primare:		
– Studii observaționale	Descriptive	– Serie de cazuri / raportare de caz – Studii transversale – Studii ecologice (la nivel de populație)
	Analitice	– Studii caz-control – Studii de cohortă
– Studii experimentale	Analitice	– Studii clinice – Studii clinice la nivel de populație
II. Studii secundare:		
– Studii pre-evaluate sau filtrate	Cantitative Calitative	– Sinteză sistematică – Sinteză narativă – Meta-analiză

Prezentare generală a designului de cercetare

Designurile de cercetare reprezintă metode esențiale pentru colectarea datelor și evaluarea relației dintre factorii de risc (expuneri sau cauze) și boli (rezultate sau efecte). Obiectivul central constă în stabilirea *cauzalității*, adică a relației dintre expunere și rezultat sau între cauză și efect. Alegerea designului de cercetare este determinată de întrebările de cercetare, de preocupările privind validitatea și de considerentele practice sau etice.

⇒ **Studii observaționale** - aceste studii oferă perspective asupra expunerilor în contexte naturale, evitând astfel problemele etice care pot apărea în cadrul studiilor experimentale. Ele includ:

- **Studii serie de cazuri** - documentează o serie de cazuri cu o condiție similară, fără a include un grup de comparație.

- **Studii transversale**- evaluează datele la un anumit moment, oferind o imagine de ansamblu a relației dintre factorii de risc și rezultate.
- **Studii caz-control**- compară persoanele cu o boală (cazuri) cu cei care nu prezintă această boală (controale), pentru a identifica diferențele în ceea ce privește expunerea.
- **Studii de cohortă** monitorizează un grup expus unui factor de risc de-a lungul timpului pentru a observa rezultatele asupra sănătății.

Studiile serie de cazuri și studiile transversale sunt studii *descriptive*; acestea formulează ipoteze privind relațiile dintre factorii de risc și rezultate, dar nu stabilesc cauzalitatea. În contrast, studiile de tip caz-control și studiile de cohortă sunt *analitice*; ele testează ipoteze pentru a stabili cauzalitatea.

Studiile de cohortă sunt *longitudinale*, colectând date despre factorii de risc și rezultate în mai multe momente temporale, în timp ce studiile *transversale* adună date la un anumit moment de timp.

Studiile de cohortă sunt *prospective*, urmărind tranziția de la factorii de risc la rezultate, în timp ce studiile de tip caz-control sunt *retrospective*, urmărind traseul de la rezultate la factorii de risc. Studiile transversale nu au o direcție temporală, întrucât datele sunt colectate într-un anumit moment de timp (*figura 8.2*).

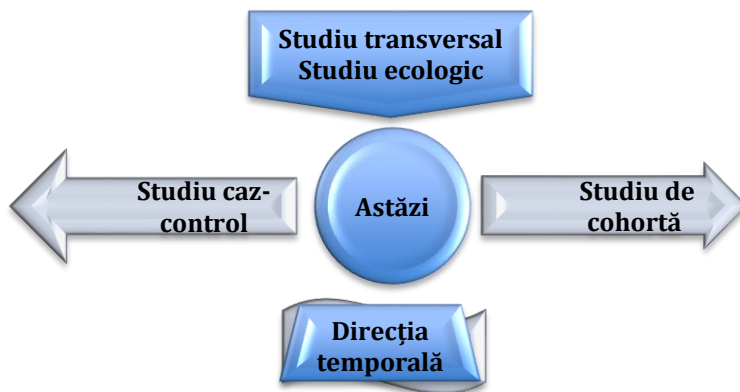


Figura 8.2. Relația temporală între diferite tipuri de studii observaționale

⇒ **Studiile experimentale (trialuri clinice)** - aceste studii implică intervenții, cum ar fi medicamentele sau tratamentele, și sunt concepute pentru a evalua efectele acestora.

⇒ **Sintezele sistematice** - acest studiu secundar evaluează critic și sintetizează literatura clinică pe un subiect specific, utilizând o metodologie structurată. Cercetătorii localizează, adună și evaluează sistematic studiile primare relevante conform unor criterii predefinite. Scopul sintezei este de a răspunde unei întrebări de cercetare bine definite, aplicând atât metode calitative, cât și cantitative.

⇒ **Sintezele narrative** aceste sinteze au un domeniu mai larg și pot să nu urmeze criteriile riguroase pentru selecția sau evaluarea articolelor. Ele adesea nu dispun de criterii explicite pentru includere și pot să nu evalueze validitatea studiilor sintetizate, ceea ce poate introduce erori (bias).

⇒ **Meta-analiza** - un tip de studiu secundar care combină cantitativ rezultatele din mai multe studii primare pentru a oferi un rezumat

cuprinzător și concluzii. Meta-analizele se axează frecvent pe evaluarea eficacității terapeutice sau pe planificarea cercetărilor ulterioare.

8.7.3. Puterea dovezilor în designul studiilor

Piramida dovezilor clasifică designurile studiilor de la cele mai puternice la cele mai slabe, cu studiile de cea mai înaltă calitate situate în partea superioară și cele care furnizează dovezi mai slabe în partea inferioară. Așa cum este ilustrat în *figura 8.3*, studiile care oferă dovezi puternice sunt relativ rare.

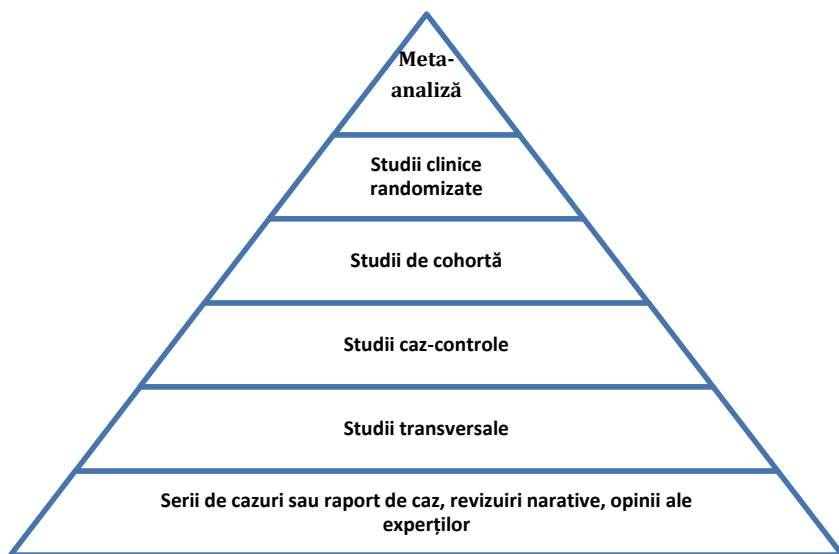


Figura 8.3. Piramida puterii dovezilor pentru designul studiilor de cercetare

Studiile secundare, precum *sintezele sistematice* și *meta-analizele*, oferă cea mai mare validitate internă. Aceste studii sintetizează rezultatele obținute din multiple studii primare pentru a furniza dovezi robuste privind cauzalitatea.

Studiile experimentale, inclusiv studiile clinice randomizate (RCT), sunt eficace în stabilirea relațiilor cauzale prin intervenții directe. Deși sunt mai controlate comparativ cu studiile observaționale, ele sunt mai puțin frecvente din cauza provocărilor practice și etice.

Studiile de cohortă urmăresc grupuri expuse la factori de risc pe o perioadă extinsă pentru a observa rezultatele de sănătate, furnizând dovezi puternice ale cauzalității.

Studiile caz-control compară persoanele care suferă de o boală cu cei care nu au această boală pentru a identifica diferențele de expunere. În timp ce sunt valoroase, studiile caz-control sunt considerate mai slabe decât studiile de cohortă deoarece sunt retrospective.

Studiile transversale analizează datele obținute de la un grup de subiecți într-un anumit moment de timp, oferind o imagine a asociațiilor fără a stabili cauzalitatea.

Studiile serii de cazuri, recenziile narative și opiniile experților sunt plasate la baza piramidei. Aceste studii sunt mai puțin controlate, oferind dovezi mai slabe de cauzalitate.

Exerciții de recapitulare

1. Un manager de clinică intenționează să realizeze un studiu pe un eșantion aleatoriu de pacienți pentru a evalua percepția acestora asupra schimbărilor recente implementate în operațiunile clinicii. Managerul a elaborat un chestionar și solicită o revizuire a acestuia. Una dintre întrebări este: „Sunteți de acord că noul program de lucru al clinicii reprezintă o îmbunătățire față de programul anterior?”

- Ce recomandări ați oferi managerului în legătură cu formularea acestei întrebări? Argumentați alegerea dumneavoastră.

2. Să presupunem că doriți să determinați distanța până la care sunt dispuși să călătorească medicii pentru a participa la cursuri de educație continuă, având în vedere că este necesar un anumit număr de ore pe an. În plus, doriți să aflați ce subiecte ar dori să fie incluse în viitoarele programe. Cum ați selecta eșantionul de medici pentru a-l include în studiul dumneavoastră? Justificați alegerea dumneavoastră.

- a) Toți medicii care au participat la programele de anul trecut.
- b) Toți medicii care vor participa la cele două programe viitoare.
- c) Un eșantion aleatoriu de medici care au participat la programele din anul precedent.
- d) Un eșantion aleatoriu de medici obținut dintr-o listă întreținută de asociația profesională a medicilor.
- e) Un eșantion aleatoriu de medici din fiecare district, obținut dintr-o listă întreținută de societățile medicale de district.

Întrebări de recapitulare

1. Definiția și caracteristicile cercetării.
2. Rolul validității în procesul de cercetare.
3. Caracteristicile cercetării și semnificația lor. Exemplificați fiecare dintre acestea.
4. Clasificarea tipurilor de cercetări.
5. Clasificarea tipurilor de cercetări în funcție de aplicabilitate. Tipuri și semnificațiile lor. Dați exemple.
6. Clasificarea tipurilor de cercetări din perspectiva obiectivelor. Tipuri și semnificațiile lor. Ilustrați cu exemple.
7. Etapele procesului de cercetare. Conținutul și specificitatea acestora.
8. Formularea problemei de cercetare. Funcția principală și criteriile de selecție.
9. Etapele formulării unei probleme de cercetare. Conținutul și particularitățile acestora. Dați un exemplu.
10. Sinteza literaturii. Funcțiile, procedurile și sistemele de citare a referințelor.
11. Scopul și obiectivele cercetării. definiția și regulile de formulare.
12. Definiția designului de cercetare și etapele acestuia.
13. Pașii pentru construirea unui chestionar.
14. Structura de bază a întrebărilor. Conținutul acestora.
15. Tipuri de metode de studiu. Conținutul acestora.
16. Clasificarea designului studiului.
17. Designul studiilor observaționale versus experimentale. Semnificația și specificitatea. Avantaje și dezavantaje.
18. Ce tip de design de studiu este cel mai adecvat în funcție de întrebările de cercetare?
 - a) Întrebări legate de terapie

- b) Diagnosticare / screening
 - c) Prognoză
 - d) Apariție
 - e) Cauzalitate.
19. Specificați diferența principală între designul studiului observațional descriptiv și cel al studiului observațional analitic.
 20. Specificați diferența principală între designul studiului caz-control și cel al studiului de cohortă.
 21. Specificați diferența principală între designul studiului serii de cazuri și cel al studiului caz-control.
 22. Clasificați designurile studiilor în funcție de puterea dovezilor.

CAPITOLUL 9. STUDII OBSERVAȚIONALE DESCRIPTIVE

Concepte cheie

- ❖ Studiile de tip serie de cazuri și cele transversale fac parte din categoria studiilor observaționale și descriptive.
- ❖ *Studiul transversal* este denumit și studiu de prevalență sau studiu comunitar.
- ❖ *Studiul transversal* colectează informații despre expunere și rezultat o singură dată, într-un moment specific.
- ❖ Fiecare tip de design de studiu are avantaje și dezavantaje specifice.

9.1. Studiul serie de cazuri / raportare de caz

Un studiu de tip serie de cazuri sau raportare de caz reprezintă cea mai simplă metodologie de cercetare, în cadrul căreia autorul descrie observații deosebite sau neobișnuite observate la un număr redus de pacienți (studiu serie de cazuri) sau chiar la un singur pacient cu o afecțiune rară (raportare de caz). Când anumite caracteristici ale unui grup (sau serie) de pacienți (sau cazuri) sunt prezentate într-un raport publicat, acesta poartă denumirea de studiu serie de cazuri. Acest tip de studiu generează frecvent ipoteze care pot fi ulterior investigate în studii de tip caz-control sau de cohortă.

Utilizări:

- ⇒ Identificarea unor boli sau rezultate noi.
- ⇒ Formularea ipotezelor.

Avantaje:

1. Redactare facilă.
2. Observațiile pot fi de mare valoare pentru alți cercetători.

Dezavantaje:

1. Vulnerabil la multiple erori sistematice.

2. Nu este adecvat pentru luarea unor decizii definitive.

9.2. Studiul transversal

Studiul transversal analizează datele colectate de la un grup de subiecți într-un anumit moment, fără a-i urmări pe o perioadă extinsă. Este proiectat pentru a răspunde la întrebarea „Ce se întâmplă în prezent?”. Subiecții sunt selectați, iar informațiile despre expunerea la factorii de risc și rezultat (boală) sunt adunate într-un interval scurt, așa cum este prezentat în *figura 9.1*.

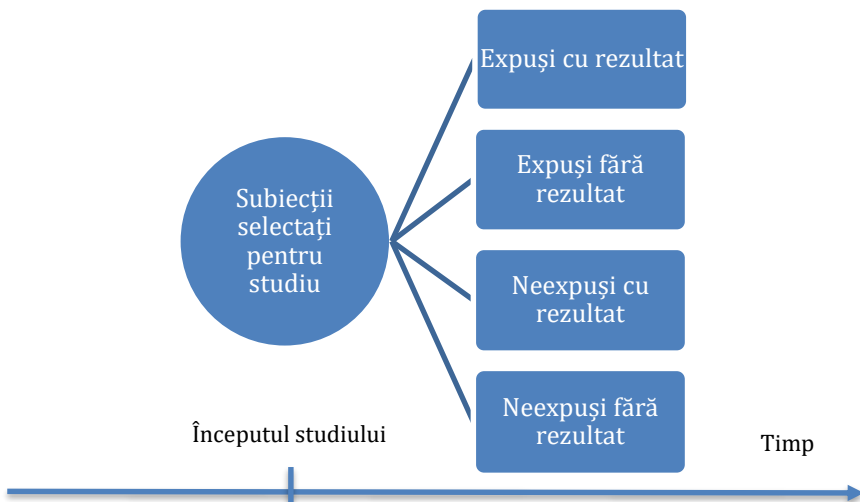


Figura 9.1. Diagrama de flux a designului studiului transversal

Studiul transversal este utilizat pentru a măsura prevalența unei boli și pentru a analiza potențialii factori de risc sau cauze. Deoarece studiile transversale examinează relația dintre expunere și prevalența bolii într-o populație definită la un anumit moment, ele sunt de asemenea denumite *studii de prevalență*. *Studiile comunitare* sunt în

mod obișnuit studii transversale, deși sondajele pot face parte și din studii de cohortă sau studii caz-control.

Designul studiului transversal este cel mai adecvat pentru scopuri precum diagnostic / screening, evaluarea prevalenței unei boli sau a unui factor de risc, realizarea studiilor comunitare sau formularea întrebărilor de cercetare.

Proceduri statistice pentru analiza datelor studiului transversal:

- Calcularea proporțiilor, ratelor (inclusiv standardizate) și raporturilor.
- Calculul intervalelor de încredere pentru proporții sau medii.
- Analize de corelație și regresie, inclusiv regresie logistică.
- Aplicarea testelor parametrice, cum ar fi testul t , analiza varianței, testul chi-pătrat și alte teste neparametrice.

Avantaje:

1. Permite estimarea poverii unei boli într-o populație – rata prevalenței poate fi determinată.
2. Ieftin și rapid de realizat.
3. Valoros pentru evaluarea procedurilor de diagnostic.
4. Util în analiza factorilor de risc comuni.
5. Eficient în studierea bolilor (rezultatelor) frecvente.

Dezavantaje:

1. Participanții pot fi mai puțin dispuși să colaboreze.
2. Nu oferă informații despre fluxul evenimentelor.
3. Arată asocierea dintre factorii de risc și boala studiată, dar nu și relația de cauzalitate.
4. Nu este eficient pentru identificarea cauzelor maladiilor.
5. În general, este util pentru studiul bolilor cronice.

6. Factorii de confuzie pot fi distribuiți inegal.
7. Mărimile grupurilor pot fi inegale.
8. Eroarea de evocare (amintire) este o problemă.

Întrebări de recapitulare

1. Definiți studiul de tip serie de cazuri și conținutul acestuia.
2. Care sunt avantajele studiului de tip serie de cazuri?
3. Care sunt dezavantajele studiului de tip serie de cazuri?
4. Care analiză statistică este adecvată pentru designul studiului de tip serie de cazuri?
5. Definiți studiul transversal și menționați sinonimele acestuia.
6. Descrieți conținutul diagramei pentru studiul transversal.
7. Care sunt avantajele studiului transversal?
8. Care sunt dezavantajele studiului transversal?
9. Care analiză statistică este adecvată pentru studiul transversal?

CAPITOLUL 10. STUDII ANALITICE OBSERVAȚIONALE

Concepte cheie

- ❖ Un studiu de tip *caz-control* include un grup de cazuri (persoane care prezintă un anumit rezultat) și un grup de controale (persoane fără acel rezultat).
- ❖ În cadrul unui studiu *caz-control*, informațiile referitoare la expunere sunt colectate *retrospectiv*.
- ❖ Deși studiul de tip *caz-control* oferă date referitoare la relațiile de cauzalitate, eroarea de evocare reprezintă o problemă frecventă.
- ❖ *Formarea de perechi* (potrivirea) într-un studiu de tip *caz-control* reduce impactul factorilor de confuzie.
- ❖ *Raportul șanselor* indică de câte ori este mai probabil ca un caz să fi fost expus unui factor de risc în comparație cu un control.
- ❖ *O cohortă* este un grup de persoane care nu prezintă boala de interes la începutul studiului și care sunt selectate pentru a fi observate pe o perioadă îndelungată.
- ❖ Într-un *studiu de cohortă*, cercetătorul va colecta date despre noile cazuri de boală (*incidență*) pe parcursul examinărilor regulate viitoare (*prospectiv*).
- ❖ Într-un studiu *longitudinal*, sunt colectate două sau mai multe seturi de observații pentru fiecare subiect de-a lungul timpului.
- ❖ Un studiu de cohortă este cunoscut și sub denumirile de studiu *prospectiv*, *longitudinal* sau de *incidență*.
- ❖ *Riscul relativ* indică de câte ori este mai mare probabilitatea ca o persoană expusă să dezvolte boala în comparație cu o persoană neexpusă.
- ❖ *Riscul atribuibil* indică proporția de incidență a bolii care poate fi atribuită (sau explicată prin) unei anumite expuneri (în rândul celor expuși).

- ❖ Într-un studiu caz-control, rezultatul este cunoscut, iar cercetarea se concentrează pe identificarea expunerii în trecut (*retrospectiv*).
- ❖ Într-un studiu de cohortă, expunerea este cunoscută, iar subiecții sunt monitorizați de-a lungul timpului pentru a identifica rezultatul în cauză (*prospectiv*).

10.1 Studiul caz-control

Studiul caz-control reprezintă un design de cercetare *retrospectiv*, în care informațiile despre factorii de risc sunt obținute printr-o analiză retrospectivă a istoricului participanților. Deși acest tip de studiu poate oferi perspective valoroase asupra relațiilor cauzale, eroarea de evocare (amintire) reprezintă o problemă comună.

Designul studiului

Cazuri - persoane care prezintă o anumită boală sau un anumit rezultat.

Controale - persoane care nu prezintă boala sau rezultatul în cauză.

Pentru a minimiza influența variabilelor de confuzie, cercetătorii trebuie să utilizeze potrivirea pentru a asocia grupul de controale cu cel de cazuri, pe baza unor caracteristici precum vârsta și sexul. Cazurile și controalele trebuie să fie similare din toate punctele de vedere, cu excepția expunerii la factorul de risc studiat.

Participanții sunt selectați pe baza stării lor de sănătate (variabilă dependentă), iar ambele grupuri sunt investigate cu privire la expunerea lor la potențiali factori de risc (variabilă independentă). Studiile de tip caz-control sunt esențiale pentru investigarea posibilelor cauze ale bolilor (întrebarea de cercetare cauzală).

Metode de colectare a datelor

⇒ Înregistrări disponibile din spitale, statistica vitală și alte registre relevante.

- ⇒ Interviu.
- ⇒ Chestionare auto-administrate.
- ⇒ Măsurători directe.

Ulterior, se compară istoricul expunerii la factorii de risc între cazuri și controale. Studiile de tip caz-control urmăresc să răspundă la întrebarea „Ce s-a întâmplat?”. Cunoscând rezultatul final, cercetătorii examinează retrospectiv expunerile anterioare, așa cum este ilustrat în *figura 10.1*.

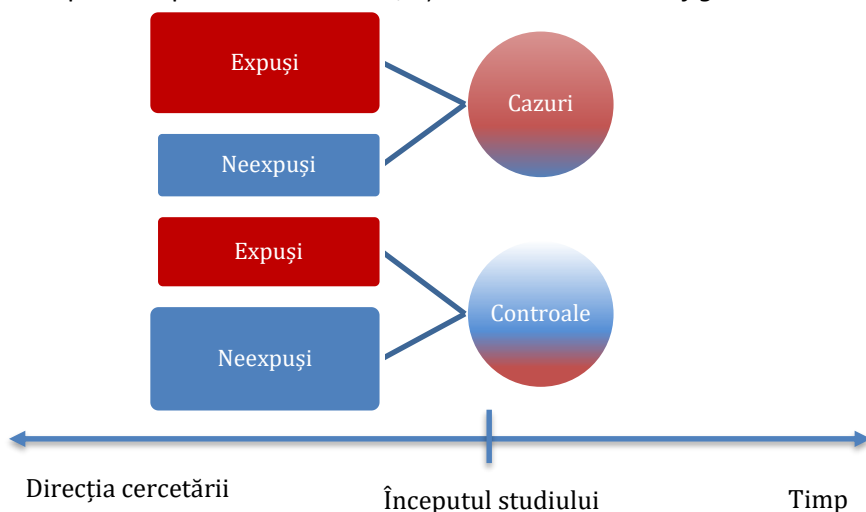


Figura 10.1. Diagramă de flux a designului studiului caz-control

Analiza datelor

Analiza unui studiu de tip caz-control implică calcularea unei măsuri de asociere denumite raportul șanselor (Odds Ratio). Șansele sunt definite ca fiind raportul dintre probabilitatea ca un eveniment să se întâmple și probabilitatea ca același eveniment să nu se întâmple: $p / (1-p)$. Raportul șanselor se calculează astfel:

Raportul șanselor

$$= \frac{\text{Șansele ca un caz să fi fost expus la factorul de risc}}{\text{Șansele ca un control să fi fost expus la factorul de risc}}$$

Raportul șanselor permite evaluarea riscului în studiile de tip caz-control, indicând de câte ori este mai mare probabilitatea ca un caz să fi fost expus la un factor de risc comparativ cu un control. Calculul este simplu atunci când datele sunt prezentate într-un tabel de contingență 2x2. Un tabel de contingență este un tip special de tabel de distribuție a frecvențelor, care prezintă simultan două variabile (*tabelul 10.1*).

Tabelul 10.1. Tabel de contingență 2x2 pentru studiul caz-control

Factor de expunere	Rezultatul (boala)		Total
	Cazuri	Controale	
Expuși	a	b	a + b
Neexpuși	c	d	c + d
Total	a + c	b + d	a + b + c + d

$$\text{Raportul șanselor (OR)} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Raportul șanselor este, de asemenea, cunoscut sub denumirea de raportul produselor încrucișate, deoarece poate fi definit ca raportul produsului diagonalelor într-un tabel 2x2.

Interpretarea raportului șanselor

- ⇒ OR = 1 - nu există asocierie (fără diferență în expunerea între cazuri și controale).
- ⇒ OR > 1 - expunere periculoasă.
- ⇒ OR < 1 - expunere benefică.

Testarea semnificației

Raportul șanselor poate fi testat pentru semnificație prin calculul unui interval de încredere, care reprezintă un interval de valori ce este

probabil să conțină raportul șanselor al populației cu un anumit nivel de încredere. Dacă intervalul de încredere de 95% pentru OR include valoarea 1, rezultatele nu sunt statistic semnificative.

Exemplu de interpretare a OR

Interpretarea unui raport al șanselor (OR) statistic semnificativ se face astfel:

Dacă $OR = 3,23$, persoanele afectate de boală au de 3,23 ori mai multe șanse să fi fost expuse față de persoanele neafectate de boală.

Avantaje

1. Permite analiza simultană a mai multor factori de risc.
2. Oferă posibilitatea studierii efectelor pe termen lung ale expunerii într-o perioadă relativ scurtă de timp.
3. Necesită un număr mai mic de subiecți comparativ cu alte metode.
4. Procesul de studiu este relativ rapid și mai ieftin.
5. Este adecvat pentru studiul bolilor rare.

Dezavantaje:

1. Risc crescut de erori (bias) și influență a variabilelor de confuzie din cauza colectării retrospective a datelor.
2. Dificultăți în selecția unui grup de control adecvat.
3. Posibilitatea apariției erorii (bias) de evocare datorită naturii retrospective a studiului.
4. Imposibilitatea de a determina incidența sau prevalența bolii.
5. Dificultăți în stabilirea unei relații temporale între expunere și rezultat.

10.2. Studiul de cohortă

Studiul de cohortă, cunoscut și sub denumirile de studiu de incidență, studiu longitudinal sau studiu prospectiv, urmărește o grupă

de persoane pe o perioadă extinsă pentru a observa dezvoltarea de noi cazuri de boală.

Designul studiului

O cohortă reprezintă un grup de persoane care împărtășesc o trăsătură comună, dar nu prezintă boala de interes la începutul studiului. Aceste persoane sunt monitorizate pe parcursul timpului pentru a observa incidența noilor cazuri de boală. Cercetătorii colectează date privind expunerile și urmăresc cohorta în timp pentru a observa noi cazuri de boală (incidență) în cadrul unor examinări regulate viitoare (natura prospectivă). Într-un studiu de cohortă, întrebarea formulată este „Ce se va întâmpla?”. Cercetătorii cunosc statutul expunerii la începutul studiului și urmăresc subiecții de-a lungul timpului pentru a observa rezultatele, conform diagramei din *figura 10.2*.

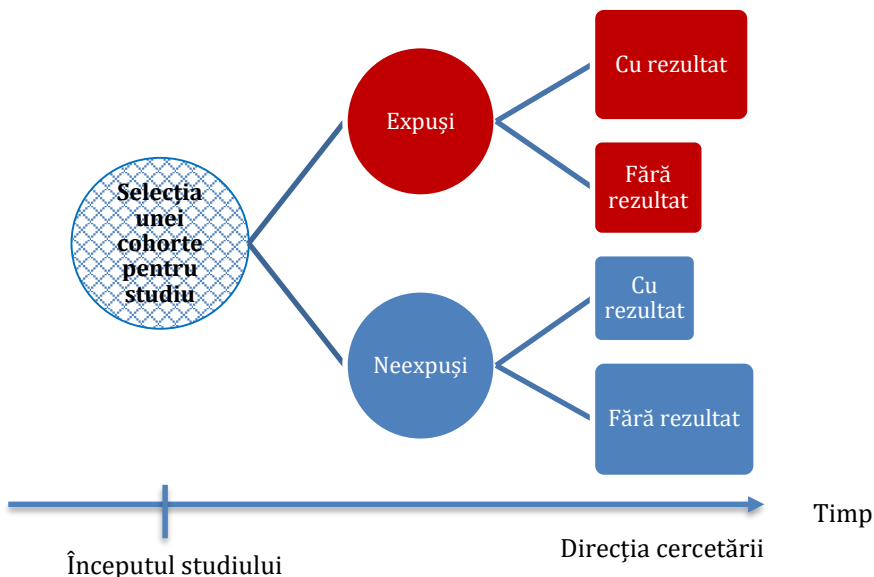


Figura 10.2. Diagramă de flux a designului studiului de cohortă

Studiile de cohortă tipice sunt, de obicei, prospective, deoarece rezultatele privind starea de sănătate sunt observate după începutul studiului. Studiile de cohortă utilizează grupuri care sunt similare în toate aspectele, cu excepția expunerii.

Studiile de cohortă sunt utilizate pentru:

- ⇒ Măsurarea incidenței bolii.
- ⇒ Investigarea cauzelor bolilor.
- ⇒ Stabilirea prognosticului.
- ⇒ Determinarea calendarului și direcționalității evenimentelor.

Metode de colectare a datelor

Obținerea datelor într-un studiu de cohortă se realizează prin interviuri personale, examinări medicale și studii de mediu.

Analiza datelor

Scopul principal al analizei datelor într-un studiu de cohortă este de a compara apariția rezultatelor în grupurile expuse și neexpuse.

Pentru a estima relația dintre un factor de risc și apariția unui anumit rezultat, se utilizează următoarele măsuri de asociere:

- ⇒ Riscul relativ (RR);
- ⇒ Riscul atribuibil (AR).

Riscul relativ (RR) reprezintă raportul dintre rata incidenței în grupul expus și rata incidenței în grupul neexpus. RR este simplu de calculat pentru un studiu de cohortă atunci când datele sunt prezentate în tabelul 2x2 (*tabelul 10.2*).

Riscul relativ indică de câte ori o persoană expusă este mai predispusă să dezvolte un rezultat comparativ cu o persoană neexpusă.

Interpretarea riscului relativ

- ⇒ $RR = 1$ - fără asociere (fără diferențe în incidența bolii între grupurile expuse și neexpuse).
- ⇒ $RR > 1$ - expunere periculoasă.
- ⇒ $RR < 1$ - expunere benefică.

Tabelul 10.2. Tabel de contingență 2x2 pentru studiul de cohortă

Factorul de expunere	Rezultatul (boala)		Total
	Yes	No	
Expuși	a	b	a + b
Neexpuși	c	d	c + d
Total	a + c	b + d	a + b + c + d

$$Riscul\ relativ\ (RR) = \frac{\text{Incidența la expuși}}{\text{Incidența la neexpuși}} = \frac{a / (a + b)}{c / (c + d)}$$

Testarea semnificației

Riscul relativ (RR) poate fi evaluat pentru semnificație prin utilizarea intervalelor de încredere. Dacă intervalul de încredere de 95% pentru RR include valoarea 1, rezultatele nu sunt considerate statistic semnificative.

Exemplu de interpretare a RR

Dacă $RR = 3,23$, persoanele expuse au de 3,23 ori mai multe șanse să dezvolte boala comparativ cu cele neexpuse.

Riscul atribuibil (AR) evaluează proporția incidenței bolii în rândul expușilor care poate fi atribuită expunerii. Acesta este raportul dintre diferența incidenței între expuși și neexpuși și incidența printre expuși, exprimat în procente.

$$\begin{aligned} \text{Riscul atribuibil (AR)} &= \frac{\text{Incidența la expuși} - \text{incidența la neexpuși}}{\text{Incidența la expuși}} \times 100\% \\ &= \frac{(a / (a + b)) - (c / (c + d))}{a / (a + b)} \times 100\% \end{aligned}$$

Exemplu de interpretare a AR

Dacă AR = 80%, înseamnă că 80% din incidența bolii în rândul expușilor poate fi atribuită expunerii.

Avantaje:

1. Permite măsurarea factorilor de risc înainte de apariția bolii, oferind dovezi privind cauzalitatea.
2. Permite studiul unui număr variat de rezultate ale bolii.
3. Oferă rate de incidență și estimări ale riscului relativ.
4. Potrivit pentru studiul expunerilor rare.
5. Reduce eroarea de selecție și eroarea de informație.

Dezavantaje:

1. Costuri ridicate.
2. Ineficiență în studiul rezultatelor (bolilor) rare.
3. Necesită o perioadă extinsă de urmărire și/sau o populație mare.
4. Pierderile în procesul de urmărire pot afecta validitatea rezultatelor.
5. Ineficiență în cazul bolilor rare.
6. Comportă probleme etice.

Exerciții de recapitulare

Scenariul 1. Studiu de cohortă privind radiațiile solare și cancerul de piele, tabel 2x2

Expunerea la radiații solare intense	Rezultatul (cancer de piele)		Total
	Da	Nu	
Da	39	12	51
Nu	10	64	74
Total	49	76	125

1. În baza acestor rezultate, încercați să restabiliți scenariile studiului în cuvinte.
2. Calculați toate măsurile de asociere posibile.
3. Interpretați rezultatele obținute.

Scenariul 2. Studiu de cohortă privind activitatea fizică și boala ischemică a inimii, tabel 2x2

Expunerea la activități fizice	Rezultatul (boala ischemică a inimii)		Total
	Da	Nu	
Da	1024	2376	3400
Nu	1205	604	1809
Total	2229	2980	5209

1. În baza acestor rezultate, încercați să restabiliți scenariile studiului în cuvinte.
2. Calculați toate măsurile de asociere posibile.
3. Interpretați rezultatele obținute.

Scenariul 3. Studiu de tip caz-control privind diabetul și infarctul miocardic, tabel 2x2

Expunerea la diabet	Rezultatul (infarct miocardic)		Total
	Da	Nu	
Da	60	40	100
Nu	340	360	700
Total	400	400	800

1. În baza acestor rezultate, încercați să restabiliți scenariile studiului în cuvinte.
2. Calculați toate măsurile de asociere posibile.
3. Interpretați rezultatele obținute.

Întrebări de recapitulare

1. Definiți designul studiului de tip caz-control și oferiți sinonimele acestuia.
2. Care sunt principalele măsuri de asociere utilizate într-un studiu de tip caz-control? Definiți-le și explicați interpretarea lor.
3. Care sunt avantajele unui studiu de tip caz-control?
4. Care sunt dezavantajele unui studiu de tip caz-control?
5. Definiți designul studiului de cohortă și oferiți sinonimele acestuia.
6. Care sunt principalele măsuri de asociere utilizate într-un studiu de cohortă? Definiți-le și explicați interpretarea lor.
7. Care sunt avantajele unui studiu de cohortă?
8. Care sunt dezavantajele unui studiu de cohortă?
9. Care este diferența principală între un studiu de tip caz-control și un studiu de cohortă? Dați un exemplu.

CAPITOLUL 11. STUDII EXPERIMENTALE

Concepte cheie

- ❖ *Studiile clinice* (trialuri) sunt clasificate în două categorii principale: studii controlate și studii necontrolate.
- ❖ *Studiile controlate* compară un tratament experimental cu un control (tratament standard, tratament anterior sau placebo). Aceste studii sunt considerate mai valide datorită capacității lor de a minimiza erorile și de a determina clar eficacitatea intervenției.
- ❖ *Studiile fără control* nu dispun de un grup de control, ceea ce le face mai puțin robuste din punct de vedere al validității, întrucât nu oferă o bază de comparație.
- ❖ *Studiile clinice cu grupuri de control paralele* implică testarea simultană a grupurilor experimental și de control în cadrul aceleiași cercetări, asigurând că diferențele observate se datorează intervenției. Studiile oarbe (simple sau duble) sunt utilizate pentru a minimiza erorile.
- ❖ *Studiile clinice randomizate* (RCT) reprezintă standardul de aur pentru cercetările clinice, în care participanții sunt alocați aleatoriu în grupuri, reducând erorile de selecție și sporind fiabilitatea rezultatelor.
- ❖ *Studiile clinice nerandomizate* nu utilizează alocarea aleatorie, ceea ce complică asigurarea că diferențele dintre grupuri se datorează exclusiv intervenției, conducând astfel la dovezi mai slabe.
- ❖ În studiile *crossover* (*modelul de tratament încrucișat*), pacienții primesc ambele tratamente (experimental și control) în momente diferite, servind drept propriile lor controale, ceea ce reduce variabilitatea.
- ❖ *Studiile cu grupuri de control externe* compară rezultatele experimentale cu date istorice din alte studii. Deși oferă perspective

valoroase, sunt considerate mai puțin fiabile din cauza diferențelor în condițiile studiului și în populații.

- ❖ *Măsurile statistice* utilizate în studiile clinice includ rata evenimentelor experimentale (EER), rata evenimentelor de control (CER), riscul relativ (RR), reducerea absolută a riscului (ARR), reducerea relativă a riscului (RRR) și numărul necesar pentru tratament (NNT).

s11.1. Clasificarea studiilor clinice

Studiile clinice reprezintă cercetări experimentale care implică subiecți umani, având ca obiectiv evaluarea eficacității procedurilor medicale sau tratamentelor. Acestea sunt esențiale pentru a răspunde la întrebări legate de terapie și pot fi clasificate în două categorii principale: studii controlate și studii fără control.

Clasificarea studiilor clinice:

I. Studii controlate:

1.1 Studii cu grupuri de control paralele:

- a) randomizate*
- b) nerandomizate*

1.2 Studii cu grupuri de control succesive:

- a) autocontrol*
- b) crossover*

1.3 Studii cu grupuri de control externe

II. Studii fără control.

Studiile controlate compară medicamentul sau procedura experimentală cu un alt tratament, care poate fi un tratament standard, un tratament anterior acceptat sau un placebo.

Studiile fără control descriu rezultatele unui medicament sau ale unei proceduri experimentale fără a le compara cu un alt tratament. Aceste studii nu dispun de un grup de control și sunt, în general, considerate mai puțin riguroase din perspectiva validității.

Studiile controlate sunt mai valide în cercetarea medicală, deoarece sunt concepute pentru a izola efectele intervenției, minimizând erorile și permițând o evaluare mai clară a eficacității reale a intervenției. Studiile fără control, deși relevante în anumite contexte, nu oferă același nivel de dovezi din cauza lipsei unei baze comparative.

11.2. Studii clinice cu grupuri de control paralele

Una dintre metodele frecvent utilizate pentru desfășurarea unui studiu clinic controlat constă în crearea a două grupuri de subiecți: grupul experimental, care primește procedura experimentală, și grupul de control, care primește tratamentul standard sau un placebo, așa cum este prezentat în *figura 11.1*.

Pentru a asigura validitatea științifică a rezultatelor, este imperativ ca grupurile experimental și de control să fie cât mai similare posibil, astfel încât orice diferență observată să poată fi atribuită în mod exclusiv intervenției. Controlul paralel presupune că intervențiile în ambele grupuri sunt realizate simultan în cadrul aceluiași studiu.

Pentru a reduce la minimum riscul de erori, cercetătorii pot implementa studii „orbite”:

- *Studii orbite simple* - subiecții nu cunosc intervenția pe care o primesc.
- *Studii orbite duble* - nici subiecții și nici cercetătorii nu cunosc despre alocarea subiecților în grupurile experimentale sau de control.

Din rațiuni etice, studiile clinice sunt permise doar atunci când intervențiile sunt anticipate a fi benefice pentru subiecți.

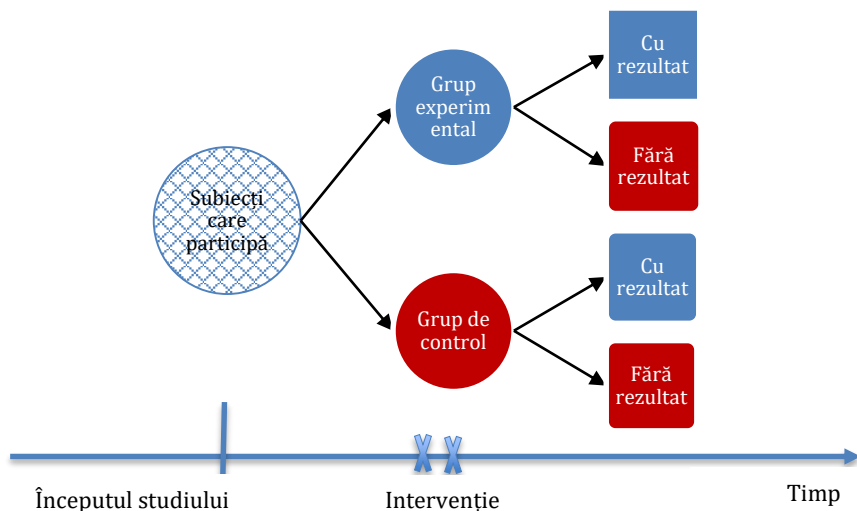


Figura 11.1. Diagramă de flux a unui studiu clinic cu modelul de tratament în paralel

⇒ **Studii clinice randomizate (RCT-uri)**

RCT-urile sunt considerate standardul de referință pentru stabilirea relațiilor cauzale, deoarece asigură cu cea mai mare certitudine că rezultatele observate sunt exclusiv efectul intervenției aplicate.

Într-un studiu clinic randomizat controlat (RCT), participanții sunt repartizați în mod aleatoriu fie în grupul experimental, fie în grupul de control, garantând astfel că fiecare participant are șanse egale de a primi oricare dintre intervențiile posibile. Acest proces de randomizare contribuie la eliminarea erorii de selecție, asigurând un echilibru între factorii cunoscuți și necunoscuți în ambele grupuri. În plus, în cadrul RCT-urilor orbite, nici participanții, nici cercetătorii nu sunt informați despre intervenția administrată, ceea ce reduce suplimentar riscul de eroare și îmbunătățește fiabilitatea rezultatelor obținute.

⇒ ***Studii clinice nerandomizate***

Studiile clinice nerandomizate sunt cercetări în care alocarea subiecților în grupurile experimental sau de control nu se realizează prin randomizare. Aceste studii, cunoscute și sub denumirea de studii clinice sau studii comparative fără randomizare, sunt considerate mai puțin robuste deoarece nu reduc eroarea care rezultă din alocarea pacienților, ceea ce complică stabilirea faptului că diferențele dintre grupuri sunt cauzate exclusiv de intervenție.

11.3. Studii clinice cu grupuri de control succesive

⇒ ***Autocontrol***

Studiile cu autocontrol sunt cercetări în care același grup de subiecți funcționează atât ca grup experimental, cât și ca grup de control. Acest tip de design permite un control moderat prin compararea rezultatelor în cadrul aceluiași grup de persoane, dar în condiții diferite.

⇒ ***Studii crossover***

Studiile crossover (modelul de tratament încrucișat) implică două grupuri de pacienți: un grup primește inițial tratamentul experimental, în timp ce celălalt grup beneficiază de placebo sau de tratamentul de control. După o perioadă prestabilită, ambele tratamente sunt suspendate pentru o perioadă de „spălare”, în care nu se administrează tratamente pentru a asigura dispariția efectelor tratamentelor anterioare. După perioada de spălare, grupurile își schimbă tratamentele: primul grup primește placebo, iar al doilea grup primește tratamentul experimental (*figura 11.2*). Studiul crossover, cunoscut sub denumirea de design „între-subiecți” („within-subjects”), se dovedește a fi deosebit de eficient atunci când este aplicat corect. În acest design, fiecare pacient primește atât tratamentul activ, cât și placebo la

momente diferite, permițând comparații directe între aceleași persoane. Astfel, fiecare pacient acționează ca propriul său control, ceea ce îmbunătățește fiabilitatea și validitatea rezultatelor prin reducerea variabilității între diferiți subiecți.

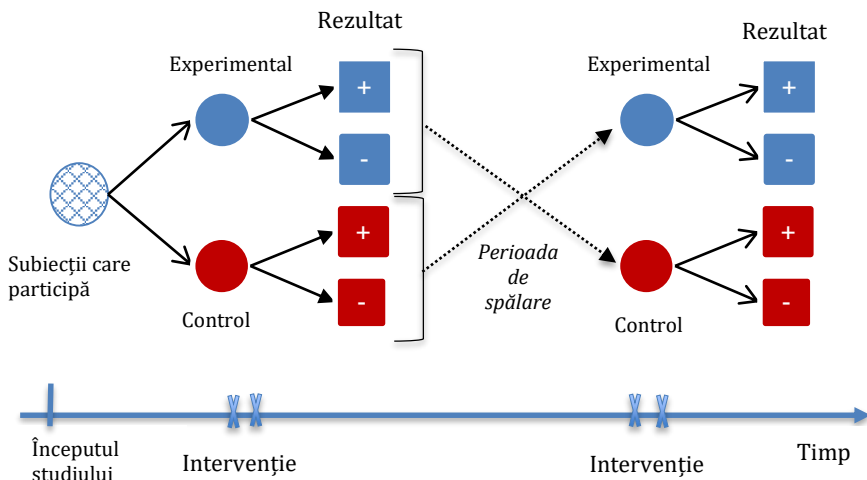


Figura 11.2. Diagramă de flux a unui studiu clinic cu controale secvențiale

11.4. Studii clinice cu grupuri de control externe

Studiile clinice cu controale externe implică compararea rezultatelor unui tratament experimental cu date provenite din alte studii sau cu date colectate anterior, așa cum este prezentat în *figura 11.3*. Aceste comparații, cunoscute și sub denumirea de controale istorice, permit cercetătorilor să evalueze eficacitatea unei intervenții prin utilizarea datelor existente ca referință. Deși această metodă poate oferi informații valoroase, este în general considerată mai puțin fiabilă comparativ cu studiile cu controale simultane, din cauza posibilelor

diferențe în condițiile de studiu și în caracteristicile populațiilor investigate.

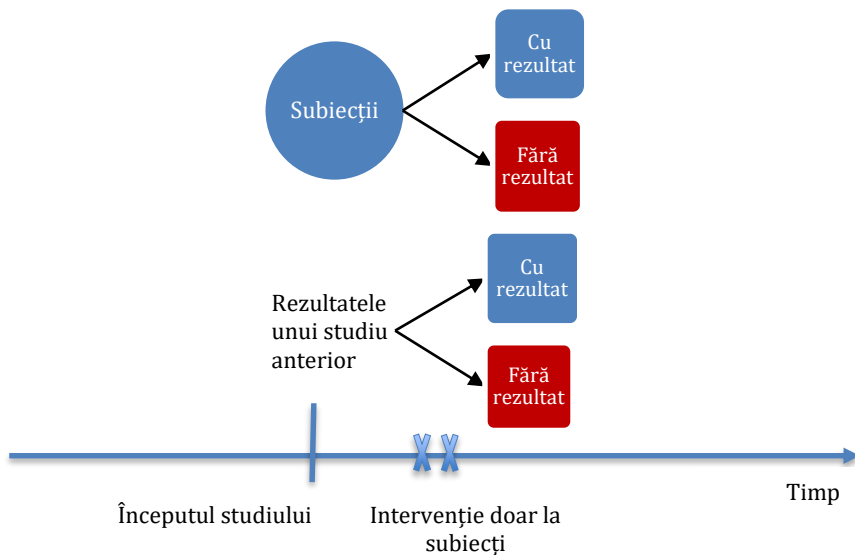


Figura 11.3. Diagramă de flux a unui studiu clinic cu un grup de control extern

11.5. Analiza statistică a studiilor clinice

Analiza statistică în cadrul studiilor clinice implică utilizarea unor măsuri specifice pentru a evalua eficacitatea intervențiilor. Aceste măsuri includ:

- Rata evenimentului experimental (*Experimental Event Rate, EER*);
- Rata evenimentului de control (*Control Event Rate, CER*);
- Riscul relativ (*Relative Risk, RR*);
- Reducerea absolută a riscului (*Absolute Risk Reduction, ARR*);
- Reducerea relativă a riscului (*Relative Risk Reduction, RRR*);
- Numărul pacienților care necesită tratament (*Number Needed to Treat, NNT*).

Pentru a determina aceste măsuri de asociere, rezultatele studiilor clinice sunt adesea organizate într-un tabel 2x2 (*tabelul 11.2*).

Tabelul 11.1. Tabel 2x2 pentru studiul clinic

Factorul de expunere (intervenție)	Rezultatul (<i>nefavorabil</i>)		Total
	Da	Nu	
<i>Expuși:</i> Tratament experimental	a	b	a + b
<i>Neexpuși:</i> Placebo	c	d	c + d
Total	a + c	b + d	a + b + c + d

⇒ **Rata evenimentului experimental (EER)** - rata evenimentului (risc de rezultat advers, cum ar fi o boală sau decesul) în grupul care primește tratamentul experimental:

$$EER = \frac{a}{a + b}$$

- ⇒ **Rata evenimentului de control (CER)** - rata evenimentului în grupul care primește placebo sau tratamentul de control:

$$CER = \frac{c}{c + d}$$

- ⇒ **Riscul relativ (RR)** - raportul dintre riscul de boală în grupul experimental și riscul de boală în grupul de control. Acest indicator arată cât de probabil este ca grupul expus să experimenteze rezultatul comparativ cu grupul neexpus.

$$RR = \frac{EER}{CER}$$

Interpretarea RR:

RR = 1 - intervenția nu are niciun efect.

RR > 1 - intervenția reprezintă un factor de risc.

RR < 1 - intervenția constituie un factor benefic/protector.

- ⇒ **Reducerea absolută a riscului (ARR)** - indicator care măsoară scăderea riscului datorită intervenției, comparativ cu riscul inițial, și arată câți subiecți evită evenimentul pentru fiecare 100 de subiecți tratați. ARR se calculează ca diferența dintre rata evenimentului de control (CER) și rata evenimentului experimental (EER):

$$ARR = CER - EER$$

Exemplu: Într-un studiu, dacă EER pentru boala cardiovasculară (BCV) în grupul care ia aspirină este de 0,14 și CER în grupul de control este de 0,30, atunci:

$$ARR = 0,30 - 0,14 = 0,16$$

Interpretarea ARR:

Riscul de BCV este de 14 subiecți la 100 în grupul care ia aspirină și de 30 la 100 în grupul care ia placebo. Prin urmare, 16 persoane din 100 evită BCV datorită administrării de aspirină.

⇒ **Reducerea relativă a riscului (RRR)** - proporția reducerii riscului în raport cu riscul inițial:

$$RRR = \frac{CER - EER}{CER} = \frac{ARR}{CER}$$

Exemplu: Folosind exemplul anterior cu aspirină,

$$RRR = \frac{0,16}{0,30} = 0,53 \text{ or } 53\%$$

Interpretarea RRR:

Reducerea relativă a riscului este de 53%. Acest lucru indică faptul că, în raport cu riscul inițial de 30 de cazuri de BCV la 100 de persoane, aspirina reduce riscul cu 53%.

⇒ **Numărul pacienților care necesită tratament (NNT)** - indicator care arată numărul de pacienți ce trebuie tratați pentru a preveni un eveniment advers.

$$NNT = \frac{1}{ARR}$$

Exemplu: Pentru studiul cu aspirină, NNT se calculează astfel:

$$NNT = \frac{1}{0,16} = 6,25$$

Interpretarea NNT:

Astfel, aproximativ 6,25 pacienți trebuie tratați cu aspirină pentru a preveni un caz de BCV. Această informație ajută clinicienii să cântărească beneficiile și riscurile unui tratament.

Avantaje:

1. Oferă dovezi puternice privind cauzalitatea.
2. Reduce erorile sistematice.
3. Permite utilizarea controalelor istorice pentru studiile preliminare.

Dezavantaje:

1. Costisitoare.
2. Implică probleme etice.
3. Necesită timp.
4. Depinde de respectarea recomandărilor de către participanții la studiu.

Exerciții de recapitulare

1. Rezultatele unui studiu clinic randomizat controlat sunt prezentate în următorul tabel 2x2:

Factor de expunere (Intervenție)	Rezultat		Total
	Fără scădere a colesterolului	Scădere a colesterolului cu 20 mg/dl	
<i>Expuși:</i> Tratament experimental: Un inhibitor al enzimei 3-hidroxi-3-metilglutaril coenzimă A reductază	3	43	46
<i>Neexpuși:</i> Placebo	8	39	47
Total	11	82	93

- 1.1.Descrieți în cuvinte scenariul studiului conform acestor rezultate.
- 1.2.Calculați toate măsurile de asociere posibile.
- 1.3.Interpretați rezultatele obținute.
2. Alegeți un studiu recent dintr-o revistă științifică recenzată care v-a captat interesul.
 - 2.1.Identificați întrebarea principală de cercetare pe care o abordează.
 - 2.2.Determinați ce design de studiu (de exemplu, studiu clinic randomizat, studiu de cohortă) ar fi cel mai adecvat pentru a răspunde la întrebarea de cercetare.
 - 2.3.Verificați dacă studiul a folosit designul pe care l-ați considerat optim. Dacă nu, analizați motivele pentru care autorii au ales alt design.

Întrebări de recapitulare

1. Care este definiția și clasificarea studiilor clinice?
2. Prin ce se distinge un studiu clinic controlat de unul necontrolat?
3. Ce definește un studiu clinic randomizat?
4. Ce reprezintă studiile clinice cu grupuri de control paralele? Prezentați definiții, tipuri și o schemă a designului corespunzător.
5. Ce sunt studiile clinice cu grupuri de control succesive? Prezentați definiții, tipuri și o schemă a designului corespunzător.
6. Ce sunt studiile clinice cu grupuri de control istorice? Prezentați o definiție și o schemă a designului corespunzător.
7. Care sunt principalele măsuri de asociere în analiza studiilor clinice? Definiți fiecare măsură și explicați interpretarea acesteia.

CAPITOLUL 12. REDACTAREA REZULTATELOR CERCETĂRII. METODE GENERALE

12.1. Redactarea unei lucrări de cercetare

Redactarea lucrării de cercetare constituie etapa finală a procesului de cercetare. Această lucrare servește la comunicarea obiectivelor studiului, a metodologiei utilizate, a rezultatelor obținute și a concluziilor formulate. Redactarea trebuie realizată într-un stil academic riguros, evitându-se expresiile colocviale sau jurnalistice.

Formatul tipic al unei lucrări de cercetare este următorul:

A. *Pagina de titlu*

- Titlul proiectului de cercetare
- Numele cercetătorului
- Numele instituției
- Data publicării

B. *Conținutul lucrării*

- **Introducere** - oferă o contextualizare amplă și definește problema de cercetare sau ipoteza.
- **Sinteza literaturii** - sintetizează cercetările existente relevante pentru tematica abordată.
- **Material și metode** - detaliază designul studiului, procedurile aplicate și materialele utilizate.
- **Rezultate** - prezintă analiza statistică a datelor și interpretarea rezultatelor obținute.
- **Discuții** - oferă o sinteză a constatărilor, discută implicațiile acestora și le raportează la literatura de specialitate.
- **Concluzii** - trage concluziile esențiale pe baza analizei și discuțiilor prezentate.

- **Recomandări** - propun direcții de aplicabilitate sau sugestii pentru cercetări viitoare.
- **Referințe / Bibliografie** - includ toate sursele citate în cadrul lucrării.
- **Anexe** - conțin materialele suplimentare, precum datele brute, chestionare sau calcule detaliate.

12.2. Prezentarea publică a rezultatelor cercetărilor medicale

Scopul unei prezentări orale este de a comunica eficient descoperirile științifice ale cercetării medicale către publicul țintă.

Designul general al unei prezentări orale:

- **Primul slide** - titlul prezentării și numele autorilor.
- **Introducere** - 1-2 slide-uri care oferă context și informații relevante.
- **Scop și obiective** - 1-2 slide-uri care detaliază obiectivele cercetării.
- **Materiale și metode** - 1-2 slide-uri care descriu designul și metodologia utilizate în studiu.
- **Rezultate** - 2-3 slide-uri care evidențiază cele mai semnificative descoperiri.
- **Discuții** - 1 slide care rezumă principalele interpretări și implicații ale rezultatelor.
- **Concluzii** - 1 slide ce prezintă concluziile esențiale extrase din studiul realizat.
- **Încheiere** - 1 slide pentru observații finale și recunoștință.

Sfaturi pentru prezentare:

- **Durată.** Se recomandă o prezentare de 10 minute, cu 8-10 slide-uri. În general, se alocă aproximativ un minut pentru fiecare slide.
- **Prezentare grafică.** Se vor utiliza grafice și tabele pentru a ilustra datele, deoarece acestea sunt de obicei mai eficiente decât textul.

Se va asigura că toate elementele vizuale sunt clare și relevante pentru mesajul transmis.

Recomandări practice:

- **Primul slide** - se va folosi o linie distinctă cu font îngroșat sau culori contrastante pentru a spori vizibilitatea.
- **Text** - se va asigura că textul este lizibil de la distanță. Textul se va limita la 5-7 rânduri per slide, conform regulii 7x7 - nu mai mult de 7 rânduri și 7 cuvinte per rând.
- **Grafice și tabele** se va păstra claritatea și simplitatea. Tabelele nu ar trebui să conțină mai mult de 3-4 coloane și 5-7 rânduri pentru a evita supraaglomerarea.

12.3. Structura și principiile de elaborare a lucrării de absolvire / tezei de licență la USMF „Nicolae Testemitanu”

Lucrarea de absolvire / teza de licență trebuie să demonstreze capacitatea studentului de a interacționa eficient cu literatura relevantă pentru subiectul abordat. Aceasta trebuie să fie metodologic solidă, cu o analiză și interpretare riguroasă a datelor, și să urmeze o structură logică. De asemenea, lucrarea trebuie să fie redactată într-un limbaj științific, conform standardelor academice și ghidurilor de scriere științifică stabilite de Universitatea de Stat de Medicină și Farmacie „Nicolae Testemitanu”. Respectarea acestor standarde garantează că lucrarea îndeplinește cerințele universitare pentru elaborarea și susținerea lucrărilor de licență. Pentru orientări detaliate se vor consulta documentele oficiale ale universității la www.usmf.md.

CAPITOLUL 13. INTRODUCERE ÎN ETICA CERCETĂRII

13.1. Definirea și obiectivele eticii în cercetare

Etica reprezintă un set de reguli care reglementează așteptările noastre privind comportamentul propriu și al altora.

Etica în cercetare constituie un set de linii directoare etice ce stabilesc modul în care cercetarea științifică trebuie să fie realizată și comunicată.

Obiectivele eticii în cercetare sunt următoarele:

1. Protejarea participanților, asigurându-le demnitatea, drepturile și bunăstarea.
2. Asigurarea desfășurării cercetărilor într-un mod care contribuie la bunăstarea persoanelor, grupurilor și societății în ansamblu.
3. Evaluarea integrității etice a activităților de cercetare specifice.

13.2. Principiile eticii în cercetare

Etica în cercetare se fundamentează pe trei abordări esențiale:

- ⇒ **Respectul pentru persoane.** Recunoașterea autonomiei și drepturilor persoanelor.
- ⇒ **Beneficența (bunăvoința).** Maximizarea beneficiilor și minimizarea daunelor pentru participanți.
- ⇒ **Justiția.** Asigurarea unei distribuții echitabile a beneficiilor și sarcinilor cercetării.

Sunt cinci principii fundamentale ale eticii în cercetare, bazate pe aceste abordări:

1. **Minimizarea riscurilor de vătămare** - implementarea măsurilor necesare pentru prevenirea daunelor aduse participanților.

2. **Obținerea consimțământului informat** - asigurarea faptului că participanții sunt pe deplin informați despre cercetare și că își dau consimțământul în mod voluntar pentru a participa.
3. **Protecția anonimatului și confidențialității** - garantarea siguranței informațiilor personale ale participanților.
4. **Evitarea practicilor înșelătoare** - menținerea onestității și transparenței față de participanți.
5. **Asigurarea dreptului de retragere** - permisiunea ca participanții să părăsească studiul în orice moment, fără consecințe.

Sfaturi esențiale pentru asigurarea unei cercetări etice:

- Colectați informațiile necesare și discutați deschis despre proprietatea intelectuală.
- Identificați și abordați problemele etice.
- Recunoașteți părțile interesate și luați în considerare interesele acestora.
- Determinați posibilele sacrificii și riscuri.
- Recunoașteți responsabilitățile (principii, drepturi, justiție).
- Reflectați asupra integrității și onestității personale.
- Gândiți creativ în privința acțiunilor posibile.
- Respectați intimitatea și confidențialitatea.
- Decideți asupra acțiunii etice corespunzătoare și fiți pregătiți să gestionați punctele de vedere divergente.

CAPITOLUL 14. PREVENIREA PLAGIATULUI. PRINCIPII DE BAZĂ

14.1. Definiția și clasificarea plagiatului

Plagiatul reprezintă utilizarea lucrărilor sau ideilor altora fără a le acorda recunoașterea academică deplină prin citări adecvate, inclusiv în listele de referințe, mulțumiri etc.

Clasificarea plagiatului:

⇒ **Plagiat direct** - transcrierea literală a lucrării altcuiva fără indicarea sursei și fără utilizarea ghilimelelor.

Plagiatul direct poate fi de mai multe tipuri:

- *Plagiat global* - asumarea integrală a unui text, prezentându-l ca fiind creația proprie.
- *Plagiat prin parafrizare* - reformularea unei idei aparținând altcuiva, cu scopul de a o prezenta ca fiind originală.
- *Plagiat mozaic* - construcția unui text prin combinarea fragmentelor din diverse lucrări, pentru a crea impresia unei opere originale.

⇒ **Auto-plagiat** - reutilizarea unei lucrări deja publicate sau prezentate anterior într-un context academic. În cazul în care vreți să integrați orice text, idei sau date deja utilizate, asigurați-vă că menționați sursa, citându-vă pe voi înșivă.

⇒ **Plagiat accidental** - plagiatul neintenționat se referă la apropierea accidentală a ideilor și materialelor altora, cauzată de o înțelegere insuficientă a normelor de citare și documentare. Chiar dacă este neintenționat, acest tip de plagiat este inacceptabil și reprezintă o încălcare a normelor academice.

14.2. Metode de prevenire a plagiatului

Există câteva tehnici esențiale care ar trebui aplicate pentru a preveni plagiatul în lucrările științifice:

- ⇒ Asigurați-vă că înțelegeți pe deplin ce înseamnă plagiatul.
- ⇒ Formulați idei originale, identificând noi perspective și subiecte de discutat.
- ⇒ Utilizați ghilimele pentru a marca clar utilizarea ideilor altora.
- ⇒ Oferiți recunoaștere academică completă tuturor surselor folosite prin citare corectă și includerea acestora în lista de referințe.
- ⇒ Fiți foarte atenți la parafrizare- textul parafrizat trebuie să fie redactat în cuvintele voastre și chiar și în aceste condiții trebuie să atribuiți sursa corectă.
- ⇒ Nu uitați să vă citați și lucrările anterioare.
- ⇒ Utilizați programe pentru detectarea plagiatului, pentru a vă asigura că textul este original și respectă normele academice.

BIBLIOGRAFIE

- BERRY G., MATTHEWS JNS, ARMITAGE P. *Statistical Methods in Medical Research*, 4th Edition, Blackwell Scientific, 2001.
- COLTON T. *Statistics in Medicine*, Little, Brown, 1974.
- COMSTOCK G. *Research Ethics: A Philosophical Guide to the Responsible Conduct of Research*, 1st Edition. Cambridge University Press, 2013.
- DANIEL W.W. *Biostatistics: A Foundation for Analysis in the Health Sciences*, 7th ed. Wiley, 1998
- DAWSON B., TRAPP G. R. *Basic and Clinical Biostatistics*, Fourth Edition, McGraw-Hill Companies, Inc., USA, 2004.
- FEINSTEIN A.R. *Clinical Epidemiology: The Architecture of Research*, WB Saunders, 1985.
- FISHER LD, VAN BELLE G. *Biostatistics: A Methodology for Health Sciences*, Wiley, 1996.
- FLEISS JL. *Design and Analysis of Clinical Experiments*, Wiley, 1999.
- FLEISS JL. *Statistical Methods for Rates and Proportion*, 2nd Edition, Wiley, 1981.
- GLANTZ, STANTON A. *Primer of Biostatistics*, University of California. 4th Edition, McGraw-Hill, Inc, 1994: перевод на русский язык, Издательский дом «Практика», 1999.
- GLASER, ANTONY N. *High-Yield Biostatistics*, Medical University of South Carolina. 4th Edition, Lippincott Williams & Wilkins, a Wolters Kluwer, Philadelphia, 2014.
- GREENBERG RS. *Prospective studies*. In Kotz S, Johnson NL (editors): *Encyclopedia of Statistics Sciences*, Vol. 7, pp.315-319. Wiley, 1986.
- GREENBERG RS. *Retrospective studies*. In Kotz S, Johnson NL (editors): *Encyclopedia of Statistics Sciences*, Vol. 8, pp.120-124. Wiley, 1988.
- HENNESEY DESENA L. *Preventing plagiarism. Tips and Techniques*, National Council of Teachers of English, 2007

- HULLEY SB (ED), CUMMINGS SR, BROWNER WS ET AL. *Designing Clinical Research*, 2nd Edition Lippincott Williams and Wilkins, 2001.
- INGELFINGER JA, WARE JH, THIBODEAU LA. *Biostatistics in Clinical Medicine*, 3rd Edition, Macmillan, 1994.
- KANE RL. *Understanding Health Care Outcomes Research*, Aspen Publishers, 1997.
- KRUGER RA, CASEY MA. *Focus Groups: A Practical Guide for Applied Research*. Sage, 2000.
- LANDRIVON G., DELAHAYE F. *La Recherche Clinique. De l'idée a la publication*. RECIF. Masson, Paris, 1995: traducere limba română Edit DAN, 2002.
- NAGESVARO RAO G. *Biostatistics and Research Methodology*, PharmaMed Press, 2018.
- PAGANO M., GAUVREAU K., *Principles of Biostatistics*, Second Edition, Belmont, CA, USA, 2000.
- RAEVSCHI E., TINTIUC D., *Biostatistics & Research Methodology*, Nicolae Testemitanu SUMPh, CEP Medicina, Chisinau, 2012.
- REA LM, PARKER RA: *Designing and Conducting Survey Research: A comprehensive Guide*, 2nd Edition Jossey-Bass, 1997
- SCHLESSELMAN JJ: *Case-Control Studies: Design, Conduct, Analysis*. Oxford, 1982.
- TAYLOR B. R. Medical Writing: A Guide for Clinicians, Educators, and Researchers, 3rd Edition, Springer, 2018.
- TINTIUC D., BADAN V., RAEVSCHI E., GROSSU IU., GREJDEANU T., ET AL. *Biostatistica si Metodologia Cercetarii Stiintifice*, USMF „Nicolae Testemitanu”, CEP Medicina, Chisinau, 2011.
- WEINSTEIN MC, FINEBERG HV: *Clinical Decision Analysis*, WB Saunders, 1998.

ANEXA A. Valori critice pentru distribuția „t”

Grade de libertate (GL)	Test unilateral				
	0.05	0.025	0.01	0.005	0.0005
	Test bilateral				
	0.10	0.05	0.02	0.01	0.001
1	6.314	12.706	31.821	63.657	636.62
2	2.920	4.303	6.965	9.925	31.598
3	2.353	3.182	4.541	5.841	12.924
4	2.132	2.776	3.747	4.604	8.610
5	2.015	2.571	3.365	4.032	6.869
6	1.943	2.447	3.143	3.707	5.959
7	1.895	2.365	2.998	3.499	5.408
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
16	1.746	2.120	2.583	2.921	4.015
17	1.740	2.110	2.567	2.898	3.965
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.903	2.539	2.861	3.883
20	1.725	2.086	2.528	2.865	3.850
21	1.721	2.080	2.518	2.831	3.819
22	1.717	2.074	2.508	2.819	3.792
23	1.714	2.069	2.500	2.807	3.767
24	1.711	2.064	2.492	2.797	3.745
25	1.708	2.060	2.485	2.787	3.725
26	1.706	2.056	2.479	2.779	3.707
27	1.703	2.052	2.473	2.771	3.690
28	1.701	2.048	2.467	2.763	3.674
29	1.699	2.045	2.462	2.756	3.659
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.291

ANEXA B. Valori critice pentru distribuția chi pătrat

Grade de libertate (GL)	Nivelul de semnificație (α)								
	0.995	0.00	0.975	0.95	0.9	0.1	0.05	0.025	0.01
1	0	0	0	0	0.02	2.71	3.84	5.02	6.63
2	0.01	0.02	0.05	0.1	0.21	4.61	5.99	7.38	9.21
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34
4	0.21	0.3	0.48	0.71	1.06	7.78	9.49	11.14	13.28
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09
6	0.68	0.87	1.24	1.64	2.2	10.64	12.59	14.45	16.81
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09
9	1.73	2.09	2.7	3.33	4.17	14.68	16.92	19.02	21.67
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21
11	2.6	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72
12	3.07	3.57	4.4	5.23	6.3	18.55	21.03	23.34	26.22
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14
15	4.6	5.23	6.26	7.26	8.55	22.31	25	27.49	30.58
16	5.14	5.81	6.91	7.96	9.31	23.54	26.3	28.85	32
17	5.7	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81
19	6.84	7.63	8.91	10.12	11.65	27.2	30.14	32.85	36.19
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29
24	9.89	10.86	12.4	13.85	15.66	33.2	36.42	39.36	42.98
26	11.16	12.2	13.84	15.38	17.29	35.56	38.89	41.92	45.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28
30	13.79	14.95	16.79	18.49	20.6	40.26	43.77	46.98	50.89
32	15.13	16.36	18.29	20.07	22.27	42.58	46.19	49.48	53.49
34	16.5	17.79	19.81	21.66	23.95	44.9	48.6	51.97	56.06
38	19.29	20.69	22.88	24.88	27.34	49.51	53.38	56.9	61.16
42	22.14	23.65	26	28.14	30.77	54.09	58.12	61.78	66.21
46	25.04	26.66	29.16	31.44	34.22	58.64	62.83	66.62	71.2
50	27.99	29.71	32.36	34.76	37.69	63.17	67.5	71.42	76.15
55	31.73	33.57	36.4	38.96	42.06	68.8	73.31	77.38	82.29
60	35.53	37.48	40.48	43.19	46.46	74.4	79.08	83.3	88.38
65	39.38	41.44	44.6	47.45	50.88	79.97	84.82	89.18	94.42
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.43
75	47.21	49.48	52.94	56.05	59.79	91.06	96.22	100.84	106.39
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33
85	55.17	57.63	61.39	64.75	68.78	102.08	107.52	112.39	118.24
90	59.2	61.75	65.65	69.13	73.29	107.57	113.15	118.14	124.12
95	63.25	65.9	69.92	73.52	77.82	113.04	118.75	123.86	129.97
100	67.33	70.09	74.22	77.93	82.36	118.5	124.34	129.56	135.81

USMF „Nicolae Testemițanu”

Centrul Editorial-Poligrafic *Medicina*

Formatul hârtiei 60x84 ¹/₁₆ Tiraj: 500 ex.

Coli de autor 5,2 Comanda nr. 30

Chișinău, bd. Ștefan cel Mare și Sfânt, 165