

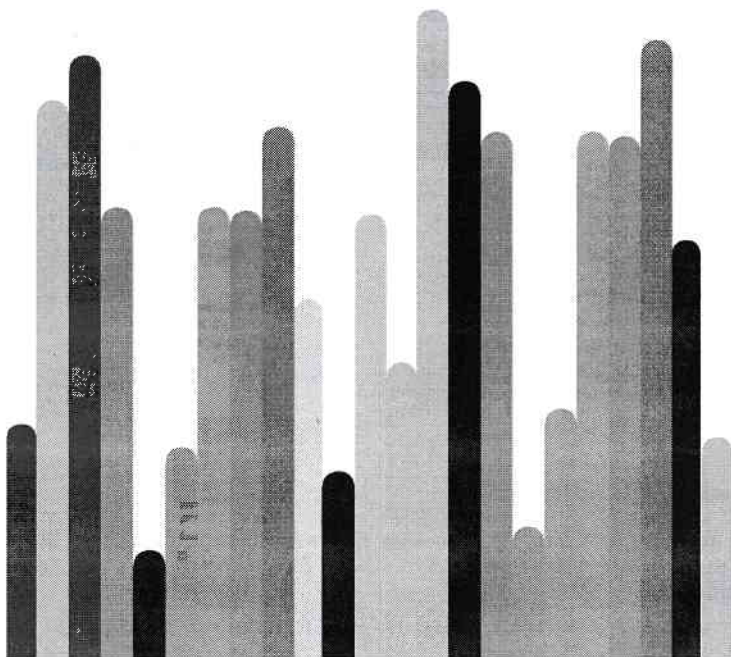
Florin A. Sava

ANALIZA DATELOR

ÎN CERCETAREA

PSIHOLOGICĂ

Ediția a doua, revizuită



REZUMAT ÎN LIMBA ENGLEZĂ	xv
INTRODUCERE LA EDIȚIA A II-A	xxi
PREFAȚA PRIMEI EDIȚII	xxiii
Capitolul 1. ELEMENTE DE STATISTICĂ DESCRIPTIVĂ	1
1.1. Modalități de măsurare a variabilelor studiate	2
1.2. Analiza descriptivă univariată	3
1.3. Analiza descriptivă bivariată	13
1.3.1. Coeficienții de corelație	13
1.3.2. Reprezentarea grafică a asocierii dintre variabile	15
1.3.3. Tabele de contingență – moduri de prezentare a asocierii dintre datele non-numerice	25
Bibliografie	27
Capitolul 2. TESTAREA IPOTEZELOR	29
2.1. Ipoteza nulă – clarificări conceptuale	30
2.1.1. Logica testării ipotezelor	31
2.1.2. Ipoteza nulă și mărimea eșantionului	33
2.1.3. Erorile de tip I și II în testarea ipotezelor	33
2.2. Mărimea efectului	35
2.2.1. Indicatori ai mărimii efectului	36
2.2.2. Interpretarea mărimii efectului	37
2.2.3. Stabilirea mărimii efectului	39
2.2.4. Limitele mărimii efectului	39
2.3. Puterea statistică a unei cercetări	40
2.4. Appendix – Ce înseamnă gradele de libertate?	49
Bibliografie	51
Capitolul 3. TEHNICI INTRODUCATIVE DE TESTARE A IPOTEZELOR	53
3.1. Testarea asocierii dintre două variabile	53
3.1.1. Coeficientul de corelație r	53
3.1.2. Corelația parțială	63
3.2. Tehnici de comparație între grupuri	65
3.2.1. Tehnicile χ^2	65
3.2.2. Testele t și z de comparare a diferențelor dintre două medii	74
3.3. Managementul datelor brute	86
3.3.1. Managementul datelor lipsă	86
3.3.2. Managementul datelor neobișnuite	87
3.3.3. Transformarea datelor brute	88
Bibliografie	91
Capitolul 4. TEHNICILE ANOVA	93
4.1. ANOVA simplă	97

4.1.1. Procedurile post hoc	99
4.1.1.1. Mărima efectului în cadrul procedurilor post hoc	102
4.1.2. Procedurile a priori	104
4.1.2.1. Contraste nestandardizate	104
4.1.2.2. Mărima efectului în cazul contrastelor	109
4.1.2.3. Contrastele standardizate	109
4.1.2.4. Contrastele polinomiale	112
4.2. ANOVA factorială	115
4.2.1. Metoda efectelor simple	120
4.2.2. Mărima efectelor în cazul ANOVA factorială	123
4.2.3. Condiții necesare pentru aplicarea ANOVA simplă și ANOVA factorială	124
4.3. ANOVA cu măsurători repetate	125
4.3.1. Mărima efectelor în ANOVA cu măsurători repetate	130
4.3.2. Condiții necesare pentru aplicarea ANOVA cu măsurători repetate	130
4.4. Alte modalități ANOVA	134
4.5. Exemplu de ANOVA factorială: cazul euristicii de accesibilitate	134
4.6. Tehnicile ANCOVA	138
4.6.1. Mărima efectului în tehnicile ANCOVA	143
4.6.2. Condiții suplimentare aplicării ANCOVA	143
4.7. Tipuri de factori în design-urile de tip ANOVA sau ANCOVA	146
Bibliografie	148
Capitolul 5. ANALIZA FACTORIALĂ	151
5.1. Analiza factorială exploratorie	152
5.1.1. Teoria analizei factoriale	152
5.1.2. Analiza componentelor principale vs. analiza factorială	155
5.1.3. Design-ul și metodologia analizei factoriale	157
5.1.3.1. Tipuri de matrice și metode de extragere a factorilor	159
5.1.3.2. Metode de selectare a numărului adecvat de factori	161
5.1.3.3. Analiza statistică a factorilor extrași	166
5.1.3.4. Rotirea factorilor	167
5.1.3.5. Interpretarea naturii factorilor	173
5.1.3.6. Ameliorarea structurii factoriale	174
5.1.3.7. Aspecte metodologice complementare	175
5.1.4. Exemplu de analiză factorială exploratorie - cazul inteligenței	180
5.2. Analiza factorială confirmatorie	183
5.2.1. Specificarea modelelor concurente	185
5.2.2. Identificarea modelelor	186
5.2.3. Stabilirea gradului de adecvare a modelelor	190
5.2.3.1. Indicatori absoluți	190
5.2.3.2. Indicatori de comparare	191
5.2.4. Modificarea modelului	195
5.2.5. Aspecte metodologice ale aplicării analizei factoriale confirmatorii	197
5.3. Adaptarea culturală a instrumentelor psihometrice	198
5.3.1. Exemplu - adaptarea scalei de măsurare a motivației de a fuma	203
Bibliografie	208
Capitolul 6. REGRESIA LINIARĂ	211
6.1. Regresia liniară simplă	213

6.1.1. Principii ale regresiei liniare	213
6.1.2. Evaluarea eficienței unei ecuații de regresie	218
6.1.3. Testarea individuală a eficienței predictorilor în estimarea evoluției criteriului	221
6.2. Modalități de analiză de regresie multiliniară	224
6.3. Regresia multiliniară în scop predictiv	227
6.3.1. Interpretarea coeficienților de regresie	231
6.3.2. Rolul ecuației de regresie în interpretarea datelor cu scop predictiv	235
6.3.3. Instrumente complementare regresiei în scop predictiv	236
6.3.3.1. <i>Decizia organizațională</i>	236
6.3.3.2. <i>Contravalidarea</i>	239
6.3.3.3. <i>Analiza cazurilor influente</i>	241
6.4. Regresia multiliniară în scop explicativ	243
6.4.1. Logica regresiei utilizate în scop explicativ	244
6.4.2. Testarea eficienței explicative a modelului propus	246
6.4.3. Interpretarea coeficienților de regresie în analizele cu scop explicativ	247
6.4.4. Indicatorii mărimii efectului în cazul regresiei multiple	249
6.5. Condiții necesare pentru aplicarea regresiei liniare	250
6.5.1. Evitarea erorii de specificare	250
6.5.2. Evitarea erorii de măsurare	251
6.5.3. Eroarea reziduală cât mai mică	252
6.5.4. Evitarea multicoliniarității	252
6.5.5. Existența unor relații de tip liniar și aditiv	254
6.6. Exemplu de regresie multiplă liniară – cazul citirii	255
6.7. Direcții de evoluție ale regresiei	260
6.7.1. Alte tipuri de regresie	260
6.7.2. Identificarea efectelor de moderare (interacțiune) în regresia multiplă	260
6.7.3. Analiza relațiilor de mediere prin instrumentul regresiei liniare	271
6.7.4. Analiza de cale	279
Bibliografie	282
Capitolul 7. REGRESIA LOGISTICĂ	285
7.1. Concepte și principii teoretice	285
7.2. Modelul regresiei logistice	287
7.3. Estimarea modelelor logistice	292
7.3.1. Testarea modelelor explicative logistice	292
7.3.2. Compararea modelelor logistice	296
7.3.3. Stabilirea eficienței explicative a modelului ales	298
7.3.4. Testarea și interpretarea coeficienților logit	303
7.3.5. Stabilirea unor profiluri ideale	307
7.3.6. Analiza cazurilor influente asupra modelului	308
7.4. Exemplu de regresie logistică binară – cazul infidelității	309
Bibliografie	311
Capitolul 8. SCALAREA MULTIDIMENSIONALĂ	313
8.1. Modele de scalare multidimensională	314
8.2. Design-ul cercetării în scalarea multidimensională	317
8.2.1. Selectarea obiectelor și a numărului acestora	318
8.2.2. Metode de colectare a datelor	319
8.2.2.1. <i>Compararea fiecărei perechi de stimuli</i>	319

8.2.2.2. <i>Sortarea stimulilor</i>	320
8.2.2.3. <i>Ierarhizarea stimulilor</i>	321
8.2.3. Managementul datelor colectate prin SPSS	322
8.3. Harta perceptivă. Reprezentarea spațială a distanțelor psihologice	325
8.3.1. Selectarea numărului de dimensiuni din harta perceptivă	325
8.3.2. Interpretarea configurației MDS	330
8.3.3. Aplicații INDSCAL	336
8.4. Modele de scalare multidimensională conexe	338
8.4.1. Modele MDS constrânse	338
8.4.2. Modele MDS extinse	339
8.5. Incursiune în opțiunile SPSS	341
Bibliografie	342

CAPITOLUL 1

ELEMENTE DE STATISTICĂ DESCRIPTIVĂ

Analiza datelor poate părea un proces anost, rutinier. Însă importanța acestui proces a fost sesizată de către întreaga comunitate științifică. Până și titulatura de cercetare științifică cantitativă sau calitativă are în vedere tocmai modul de analiză a datelor.

Astăzi, comunitatea psihologilor recunoaște și încurajează ambele tipuri de cercetare, cu toate că, în unele medii conservatoare, cercetarea cantitativă are întâietate. Acest fapt poate fi regăsit și în practica unor cadre didactice universitare care încurajează întotdeauna aplicarea unor metode cantitative complementare unui demers calitativ, în schimb, utilizează cu caracter facultativ, recomandarea unor demersuri calitative într-o cercetare cantitativă.

Cartea de față vizează doar tehnici statistice cantitative, însă nu face acest lucru dintr-o lipsă de respect față de metodele calitative. Tehnicile cantitative alese pentru a fi descrise în prima ediție a acestui volum au urmărit să completeze metodele prezentate în textele introductive din domeniu, precum cele ale lui Clocotici și Stan (2000), Novak (1977), Hohn (2000), Radu și colab. (1993), Rotariu și colab. (1999). Până la apariția celei de-a doua ediții au apărut însă mai multe publicații în spațiul universitar românesc, apropiate ca tematică, unele focalizate în special pe SPSS (Howitt și Cramer, 2006; Labăr, 2008), altele focalizate predominant pe aspecte teoretice (Opariuc-Dan, 2009; Popa, 2008; Popa, 2010). Sper ca tot acest efort publicistic din ultimii cinci ani să îi ajute, deopotrivă pe studenții și pe cercetătorii din psihologie, să realizeze analize statistice adecvate, similare în complexitate cu cele întâlnite în revistele de specialitate incluse în fluxul internațional de cunoaștere.

La baza tuturor analizelor complexe, de tip inferențial, se situează însă un alt gen de abordare, denumită statistică descriptivă. Analiza datelor cu scop descriptiv include un complex de metode și tehnici menite să faciliteze înțelegerea realității prin analiza cantităților observate. Reprezentările grafice, mediile, tabelele de frecvențe sau alți indicatori statistici reprezintă elemente întâlnite des în cotidian sau în textele de specialitate, de aceea este necesară o bună cunoaștere a lor.

Acest minim de cunoștințe poate fi integrat în termenul generic de **statistică descriptivă**. Ea nu-și propune altceva decât să facă o radiografie cât mai corectă a realității (fără a face inferențe de la nivelul datelor colectate prin eșantionul studiat la nivelul populației din care acesta face parte). Dacă analizăm un singur aspect al

realității, vorbim despre o **statistică descriptivă univariată** (pentru o singură variabilă), în situația în care descriem evoluția a două variabile, vorbim de o **statistică descriptivă bivariată** (pentru două variabile). Se vorbește și despre o **statistică multivariată**, în situațiile în care sunt analizate simultan cel puțin trei aspecte ale realității. Prelucrarea statistică ține cont nu doar de numărul variabilelor, ci și de calitatea scalelor de măsurare utilizate.

O altă idee de bază în analiza statistică a datelor are în vedere diferența dintre **analiza la nivelul lotului** testat și **analiza la nivelul populației** din care face parte acel eșantion. De exemplu, putem obține anumite valori descriptive la nivelul lotului testat (ex. nivelul de empatie la o clasă de elevi de a IX-a) și alte valori la nivelul populației (ex. valoarea medie a empatiei la nivelul populației de elevi de clasa a IX-a). De multe ori, analiza datelor se bazează pe rezultatele obținute la nivelul lotului de subiecți testați, însă inferența se face la nivelul populației din care acesta face parte. Pentru a facilita distincția dintre cele două domenii, se utilizează termeni diferiți precum cei de parametri, respectiv indicatori statistici. Parametrii se referă la nivelul populației, iar indicatorii statistici se referă la nivelul lotului analizat. Această diferențiere se regăsește, frecvent, și în modul de notare. Astfel, media și abaterea standard la nivelul populației se notează cu μ , respectiv σ , iar la nivelul eșantionului se utilizează simbolurile m , respectiv s .

Vom reaminti, pe scurt, informațiile de bază din statistica descriptivă, pentru a pătrunde mai bine semnificația informațiilor prezentate în capitolele următoare. Cititorii interesați pot consulta lucrările deja menționate pentru mai multe detalii. Acestea vor fi grupate în trei secțiuni: importanța modului de măsurare a variabilelor în analiza datelor (1); analiza descriptivă univariată (2) și analiza descriptivă bivariată (3).

1.1. Modalități de măsurare a variabilelor studiate

Primul element de bază în analiza datelor este identificarea modului în care au fost măsurate variabilele de interes. Astfel, se pot distinge patru scale de măsurare: nominală, ordinală, de interval și de proporții (raport), fiecare având calități diferite în ceea ce privește posibilitățile pe care le oferă de a analiza datele.

De pildă, variabilele de tip **nominal** presupun răspunsuri care nu sunt numerice și nici nu sunt ierarhizabile. Prin urmare, scalele de tip nominal implică analize de frecvențe. Un exemplu tipic de variabilă măsurată prin scală nominală este „apartenența la religie”, deoarece răspunsurile aferente ortodox, catolic, protestant, musulman etc. nu sunt nici numere, nici ordonabile intrinsec. Variabilele de tip **ordinal**, deși nu sunt de tip numeric, permit ordonarea intrinsecă a răspunsurilor, permițând astfel realizarea unor ierarhii. Acest fapt permite cercetătorilor să apeleze la analize bazate pe ierarhii (ranguri). Un exemplu de variabilă măsurată prin scală

ordinală este întâlnit frecvent în studiile despre atitudini și implică răspunsuri de genul „foarte rar”, „rar”, „mediu”, „des”, „foarte des”. Un alt tip de scală de măsurare este cea de tip **interval**. Aceasta este o scală numerică (ex. un IQ de 120; 7 puncte la o probă de asertivitate etc.). Tot o scală numerică este și scala de **proporții**, avantajul acesteia față de scala de tip interval constând în faptul că valoarea zero nu este aleasă arbitrar. Un exemplu de variabilă măsurată prin scale de proporții este timpul de reacție (ex. 170 ms) sau venitul lunar (ex. 10.000.000 lei). Unii autori, precum Kinnear și Gray (2000), tind să restrângă ultimele două modalități de măsurare la termenul generic de scală numerică, deoarece majoritatea tehnicilor statistice utilizate în cazul scalelor de tip interval sunt valabile și în cazul scalelor de tip proporții. Același poziție a fost adoptată și de consilierii statistici ai SPSS-ului, care au renunțat la a mai diferenția între cele două tipuri de variabile (SPSS 11.0, 2001).

Un alt aspect ce trebuie subliniat cu privire la scalele de măsurare este acela că aproape orice variabilă poate fi măsurată prin mai multe scale de măsurare. Spre exemplu, nivelul de școlarizare poate fi evaluat prin scală nominală în cazul unor răspunsuri de tip „școlarizat” sau „neșcolarizat”, prin scală ordinală, în cazul unor răspunsuri precum – „studii primare”, „studii medii”, „studii postliceale”, „studii universitare” etc. Realizarea unor scoruri vizând gradul de alfabetizare sau nivelul cunoștințelor generale implică obținerea unor indicatori numerici, specifici scalelor de tip interval. În cele din urmă, prin intermediul scalei de proporții, nivelul de școlarizare poate fi reflectat prin numărul de ani de școală absolviți (ex. 5; 12; 16 etc.).

De asemenea, trebuie remarcată distincția calitativă dintre scalele nominale și ordinale, pe de o parte, respectiv scalele de interval și proporții, pe de altă parte. Primele sunt scale non-numerice, presupunând apelul la **tehnici statistice non-parametrice**, mai puțin sofisticate. În schimb, variabilele măsurate prin scale numerice necesită, în cele mai multe cazuri, **tehnici statistice parametrice**, mai complexe. În cartea de față ne vom axa preponderent asupra ultimelor, deoarece în cercetarea psihologică ele constituie o majoritate clară.

1.2. Analiza descriptivă univariată

Lucrările amintite anterior și cele prezentate în nota de subsol 1 oferă o bună introducere în problema analizei statistice descriptive. De aceea, ne vom rezuma să prezentăm doar principalele concluzii ale unui asemenea demers descriptiv, insistând însă asupra coeficientului de corelație r , ca element fundamental în înțelegerea unor tehnici statistice complexe, precum analiza factorială sau regresia liniară.

(1) Descrierea unui set de date necesită cel puțin doi parametri.¹

Unul vizează tendința centrală a datelor, adică găsirea unei valori reprezentative pentru setul respectiv de date, iar celălalt vizează gradul de împrăștiere a datelor, urmărind omogenitatea sau eterogenitatea unui grup. De pildă, dacă am încerca să descriem un grup doar prin prisma primului parametru (ex. ambele clase de elevi obțin o medie generală de 8,25) nu am putea afla dacă clasele sunt similare, deoarece una ar putea fi formată din elevi omogeni, toți având media în jurul valorii de 8,25, în timp ce cealaltă clasă de elevi ar putea fi formată din două categorii diferite, de elevi foarte buni și elevi slabi, dar care împreună ar conduce la o medie egală cu cea a primului grup.

În funcție de scalele de măsurare utilizate, se stabilește și parametrul tendinței centrale, respectiv al împrăștierii datelor care ar trebui calculat.

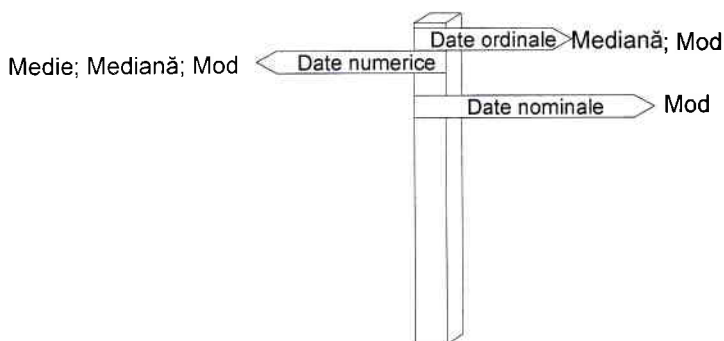


Figura 1.1. Parametri posibili ai tendinței centrale, în funcție de tipul de variabile

Între parametrii împrăștierii, abaterea standard și dispersia sunt cei mai întâlniți în cazul datelor numerice, iar abaterea intercuartilă este mai des utilizată în cazul datelor măsurate prin scale ordinale. În ceea ce privește datele nominale, situația indicatorilor împrăștierii este neclară. Indicatori precum entropia, indicele de fragmentare, indicele de diversitate sau cel al dispersiei calitative sunt aproape necunoscuți și lipsesc cu desăvârșire din programele statistice pe calculator. În consecință, răspândirea lor este minimă, majoritatea cercetătorilor rezumându-se la a urmări distribuția procentajelor pentru fiecare răspuns posibil (Weisberg, 1992).

În general, indicatorii împrăștierii iau valoarea zero dacă toate datele dintr-un set sunt identice. Cu cât datele sunt mai împrăștiate (diferite), cu atât valorile acestor indicatori sunt mai mari.

¹ Mai multe informații despre algoritmul de calcul al parametrilor tendinței centrale și ai împrăștierii pot fi găsite în orice manual introductiv, precum lucrările lui Hohn (2000), Radu și colab. (1993) ori Sava (2002).

În cazul datelor numerice, cea mai frecvent întâlnită pereche de parametri descriptivi este media, ca indicator al tendinței centrale, și abaterea standard, ca indicator al împrăștierii datelor. De altfel, acești parametri au și cele mai multe avantaje tehnice, permițând realizarea mai multor operații, precum standardizarea datelor, clasificarea răspunsurilor etc.

Totuși, în anumite situații, media și abaterea standard pot conduce la rezultate distorsionate, iar pentru a evita acest lucru este esențial să inspectăm grafic situația rezultatelor obținute.

(2) Reprezentarea grafică însoțește analiza descriptivă a datelor.

Reprezentarea grafică este o metodă intuitivă de a observa distribuția datelor, ea fiind cu atât mai utilă, cu cât numărul de date este mai mare.

Există o multitudine de grafice posibil de realizat. Alegerea unui anumit tip de grafic se face în funcție de modul de măsurare a variabilelor, dar și de aspecte preferențiale, ce țin de subiectivitatea fiecăruia. În funcție de tipul de date colectate, cele mai întâlnite grafice sunt:

- diagrama circulară („plăcinta”) pentru variabile de tip nominal;
- histograma modificată pentru variabile de tip nominal sau ordinal;
- histograma sau poligonul frecvențelor pentru variabile numerice.

Iată un exemplu de reprezentare grafică prin metoda poligonului de frecvențe:

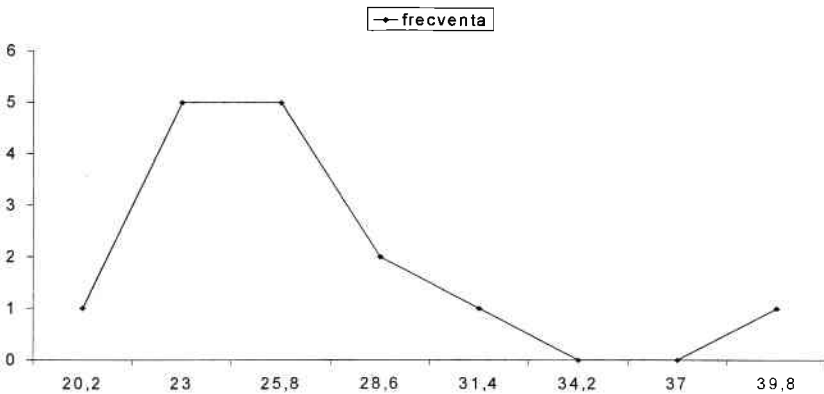


Figura 1.2. Exemplu de poligon al frecvențelor, reprezentare tipică pentru datele numerice

(3) Analiza datelor este influențată de forma distribuției datelor.

Dincolo de aspectul pur intuitiv al reprezentării grafice, foarte importantă este interpretarea formei distribuției datelor. Aceasta ne ajută să înțelegem două aspecte

esențiale: (a) care sunt parametrii tendinței centrale care reprezintă cel mai bine șirul de date?; (b) ce teste statistice pot fi utilizate pentru a testa anumite ipoteze?

Se spune că multe dintre aspectele realității au o distribuție normală. Să luăm cazul inteligenței. Cei mai mulți oameni au o inteligență medie. Întâlnim tot mai puțin oameni, cu cât mergem fie spre o inteligență superioară, fie spre un deficit grav de inteligență. O asemenea situație, denumită distribuție normală, poate fi reprezentată grafic sub forma unui clopot.

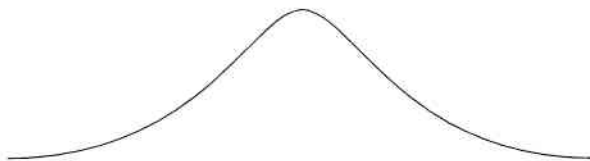


Figura 1.3. Forma unei distribuții normale a datelor

Multe teste statistice parametrice (utilizate în cazul datelor numerice) au drept condiție de aplicare normalitatea distribuției datelor. De aceea este important să caracterizăm forma distribuției obținute prin intermediul a două aspecte: (a) **nivelul de boltire**, respectiv (b) **nivelul de oblicitate** (asimetrie).

(a) O distribuție normală este *mezocurtică* (normală), adică are un nivel mediu de boltire (nici prea joasă, nici prea înaltă). Totodată, ea este simetrică, adică nu este înclinată evident nici spre stânga, nici spre dreapta.

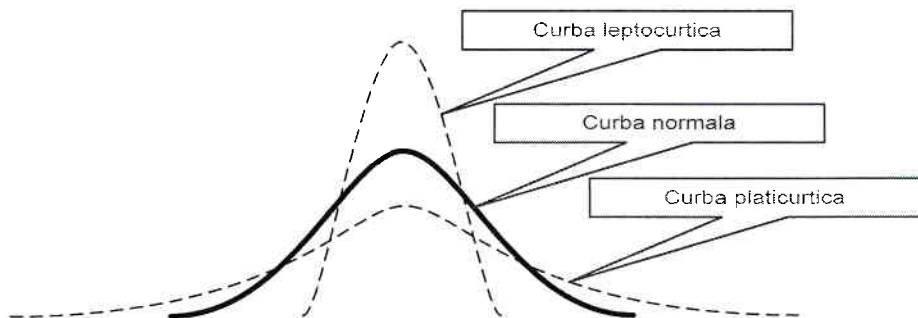


Figura 1.4. Devieri ale unei curbe simetrice din perspectiva boltirii datelor

Cum interpretăm distribuțiile prezentate în figura 1.4? O curbă *platicurtică* se obține atunci când subiecții au rezultate foarte împrăștiate. Nu există o frecvență mare a rezultatelor în jurul mediei, fapt ce ne conduce la ideea unui grup eterogen. Un exemplu ar fi o clasă de elevi, ale căror rezultate sunt foarte dispersate. Câțiva copii

au note cuprinse între 1 și 4, câțiva copii au rezultate între 5 și 7, iar alți elevi au rezultate superioare, cuprinse între 8 și 10. Reversul situației este reprezentat de o distribuție *leptocurtică*. Aceasta indică un grup omogen de date, în care cele mai multe rezultate se găsesc distribuite în jurul mediei. În cazul clasei de elevi, aceasta ar fi alcătuită din copii cu performanțe apropiate. Cei mai mulți ar înregistra rezultate între 5 și 7, în timp ce foarte puțini elevi ar fi incluși în categoria elevilor cu note sub 5 sau a celor cu note peste 7. Într-o asemenea clasă lipsesc notele extreme, fapt ce face ca nu toată gama de la 1 la 10 să fie distribuită. În sfârșit, o curbă mezocurtică are un nivel mediu, normal de boltire, ceea ce înseamnă că mulți copii au rezultate apropiate de valoarea mediei, deși există unii elevi foarte slabi, respectiv foarte buni, toată gama de note fiind acoperită.

(b) Distribuția datelor poate varia și în funcție de nivelul de simetrie a datelor. În cazul în care distribuția este *asimetrică pozitiv*, există mai multe valori mici decât mari. Într-o situație școlară, aceasta ar corespunde unui test dificil, în care majoritatea elevilor obțin note mici, deși sunt și elevi care obțin rezultate școlare bune. În schimb, o distribuție *asimetrică negativ* indică un dezechilibru al valorilor în direcție opusă. Vor apărea mai multe rezultate mari decât mici, fapt care, transpus în note școlare, arată predominanța notelor ridicate. În cazul probelor psihologice, prima distribuție poate fi socotită un indicator al unui grad de dificultate ridicat pentru proba respectivă, în timp ce o distribuție asimetrică negativ semnifică un grad scăzut de dificultate a probei. Bineînțeles că aceste concluzii sunt valide în condițiile în care valorile ridicate indică performanțe superioare, în timp ce valorile mici arată performanțe scăzute.



Figura 1.5. Distribuție asimetrică pozitiv (în stânga) și distribuție asimetrică negativ (în dreapta)

Pentru a evita interpretările subiective se pot calcula doi indicatori care să ajute la interpretarea formei distribuției.

Indicatorul de oblicitate (în engleză „skewness statistic”) evaluează gradul de asimetrie a unei distribuții. Valoarea zero indică o distribuție perfect simetrică. Valorile negative, îndepărtate de zero, indică o distribuție asimetrică negativ, în timp ce valorile pozitive, cu cât se îndepărtează de zero, arată o asimetrie pozitivă. Pentru a diferenția o distribuție ușor asimetrică, ce poate fi tratată ca o distribuție normală, de o distribuție clar asimetrică, Lewis-Beck (1995) ia drept prag valoarea absolută de

0,80. Astfel, o valoare a indicatorului de oblicitate inclusă în intervalul $[-0,80$ și $0,80]$ poate fi tratată ca un semn al unui distribuții simetrice. Valorile mai mici de $-0,80$ indică o asimetrie negativă, în timp ce valorile mai mari de $0,80$ semnifică o asimetrie pozitivă. Această regulă implicită merită a fi aplicată mai ales în condițiile unui eșantion mare de subiecți, să zicem de peste 150 de persoane.

În cazul în care volumul eșantionului este mediu sau mic, putem adopta o altă strategie, bazată pe calcularea cotei z a indicelui de oblicitate. Aceasta presupune împărțirea valorii indicatorului de oblicitate la eroarea standard estimată a acestuia. Ambele elemente sunt oferite de SPSS. Field (2000) propune două praguri ale lui z în funcție de numărul de persoane testate: 2,58 pentru un eșantion mai mic de 30 de subiecți și 1,96 pentru un eșantion mai mare de 30 de subiecți. Același autor recomandă renunțarea la regulă în cazul unui număr mare de subiecți, să zicem peste 150, deoarece z are mari șanse de a deveni semnificativ, datorită scăderii erorii standard a oblicității odată cu mărirea eșantionului.

Tabelul 1.1. Exemplu de valori ale mediei, mediane și oblicității privind vârsta ideală pentru căsătorie

Vârsta ideală pentru căsătorie	
N	15
Media	26,00
Mediana	25,00
Oblicitatea (skewness)	2,240
Eroarea standard a oblicității	0,580

Fiind testați doar 15 subiecți, vom împărți oblicitatea la eroarea sa standard și vom obține 3,86 ($2,24:0,58$). Rezultatul, în valoare absolută, este superior pragului specificat, de 2,58, prin urmare avem de a face cu o distribuție asimetrică pozitivă. Deoarece în astfel de cazuri mediana este un parametru adecvat pentru a descrie șirul de date, am putea spune că subiecții chestionați consideră vârsta de 25 de ani ca vârstă ideală pentru căsătorie. E de remarcat că operația trebuie completată de inspectarea graficului rezultat, deoarece distribuțiile multimodale nu pot fi detectate prin urmărirea indicatorului de oblicitate.

Nu numai oblicitatea poate fi estimată exact. Boltirea beneficiază, de asemenea, de un astfel de coeficient pentru a determina normalitatea unei distribuții.

Indicatorul de boltire (în engleză „kurtosis”) este estimat de SPSS pe baza acelorași principii ca în cazul oblicității. Valoarea zero indică o distribuție perfect mezocurtică, valorile pozitive, cu cât sunt mai mari, arată o tendință spre distribuții leptocurtice, în timp ce valorile mai mici de zero, semnalează tendința spre o distribuție platocurtică. Field (2000) recomandă calcularea lui z kurtosis după aceleași criterii enunțate pentru oblicitate.

Dacă există dubii cu privire la gradul de normalitate a distribuției datelor, se recomandă testarea condiției de normalitate prin testul z al lui Kolmogorov-Smirnov. Dacă valorile sunt semnificative statistic, putem afirma că distribuția nu îndeplinește