

UNIVERSITATEA DE STAT DE MEDICINĂ ȘI FARMACIE  
*NICOLAE TESTEMIȚANU*

ȘCOALA DE MANAGEMENT ÎN SĂNĂTATE PUBLICĂ

**Larisa SPINEI   Oleg LOZAN   Vladislav BADAN**

# **BIOSTATISTICA**

Chișinău • 2009

CZU: 311.2  
S 73

Aprobat de Consiliul de Experti al Ministerului Sănătății Republicii Moldova  
din 27.06.08, proces-verbal nr. 6

**Autori:**

*Larisa Spinei*, doctor habilitat în medicină, profesor universitar  
*Lozan Oleg*, doctor în medicină, conferențiar universitar  
*Badan Vladislav*, lector superior

**Recenzenți:**

*Maria Sagaidac*, doctor în matematică, conferențiar universitar, Universitatea  
de Stat RM  
*Ețco Constantin*, doctor habilitat în medicină, profesor universitar, șef ca-  
tedra Economie, Management și Psihopedagogie USMF  
„N.Testemițanu”

Lucrarea dată este elaborată în conformitate cu cerințele didactice actuale și în corespundere cu planul de studii și programa analitică la *Biostatistică* pentru audienți Școlii de Management în Sănătate Publică.

*Editarea manualului „Biostatistica” a fost posibilă datorită suportului financiar  
al Fundației SOROS-Moldova.*

**Redactor:** Lidia Ciobanu

**Machetare computerizată:** Natalia Dorogan

**Corectori:** Nadejda Koporskaia, Tatiana Colin

© Larisa Spinei, Oleg Lozan, Vladislav Badan  
© Școala de Management în Sănătate Publică, 2009

**Descrierea CIP a Camerei Naționale a Cărții**

**Spinei, Larisa**

Biostatistica / Larisa Spinei, Oleg Lozan, Vladislav Badan; Univ.  
de Stat de Medicină și Farmacie „Nicolae Testemițanu”; Școala de Mana-  
gement în Sănătate Publică. – Ch. : Școala de Management în Sănătate  
Publică, 2009 (Î.S. F.E.-P. „Tipografia Centrală”). - 186 p.  
250 ex.  
ISBN 978-9975-78-743-7

311.2

## CUPRINS

<b>CAPITOLUL I. SCURT ISTORIC. GENERALITĂȚI</b> .....	4
1.1. Apariția și dezvoltarea statisticii .....	4
1.2. Obiectul și metoda statisticii .....	10
<b>CAPITOLUL II. NOȚIUNI DE BAZĂ. METODE STATISTICE</b> .....	18
2.1. Definiția, obiectivele, compartimentele .....	18
2.2. Teoria probabilităților .....	19
2.3. Legea cifrelor mari .....	20
2.4. Metode de cercetare utilizate în biostatistică .....	21
2.5. Noțiuni de bază în biostatistică .....	26
2.5.1. Totalitatea statistică .....	27
2.5.2. Unitatea de observație .....	30
2.5.3. Indicatori statistici .....	30
2.5.4. Date statistice .....	32
2.6. Gruparea materialului statistic. Seria statistică de variație. Indicatorii tendinței centrale .....	33
2.7. Seriile cronologice .....	48
2.8. Standardizarea .....	62
2.9. Corelația și regresia .....	66
2.9.1. Corelația .....	66
2.9.2. Regresia .....	77
2.10. Analiza statistică a variabilității .....	80
2.11. Aprecierea veridicității valorilor medii și relative .....	88
2.12. Compararea valorilor absolute sau a distribuțiilor de frecvență: testul $X^2$ .....	107
<b>CAPITOLUL III. METODOLOGIA GENERALĂ A CERCETĂRII STATISTICE</b> .....	116
3.1. Colectarea sau observarea materialului statistic .....	116
3.2. Prelucrarea materialului statistic .....	126
3.3. Eșantionarea .....	132
3.4. Prezentarea materialului statistic .....	145
3.4.1. Prezentarea prin tabele .....	145
3.4.2. Reprezentarea grafică a materialului statistic .....	149
3.5. Analiza și interpretarea materialului statistic .....	176
<b>BIBLIOGRAFIE</b> .....	179

### 1.1. Apariția și dezvoltarea statisticii

Statistica, prin obiectul și metoda sa, face parte din rândul științelor ce studiază aspectele cantitative ale fenomenelor și proceselor din cadrul naturii, tehnologiei și societății. Ea s-a dezvoltat ca știință cu obiect și metodă proprii, într-un proces istoric îndelungat și s-a diferențiat în funcție de particularitățile domeniului studiat.

Pentru a cuprinde conținutul, categoriile și principiile cu care operează statistica este necesar a se avea în vedere, în primul rând, rolul pe care îl are această disciplină în procesul cunoașterii societății.

Cunoașterea și transformarea societății nu pot fi realizate decât folosind combinat principiile, metodele și tehnicile tuturor științelor care studiază fenomenele și procesele ce se manifestă în cadrul societății în scopul descoperirii legilor ce acționează în mod obiectiv în diferitele etape de dezvoltare. În acest sens, în dezvoltarea istorică a științelor social-economice există, pe de o parte, un proces de diferențiere în funcție de particularitățile ei de studiu și, pe de altă parte, apariția unor discipline cu caracter multidisciplinar sau de graniță impuse de multiplele conexiuni din cadrul societății. În cadrul studiilor social-economice, statistica a fost și este folosită întotdeauna ca un instrument indispensabil pentru cunoașterea vieții social-economice.

Procesul de conturare a statisticii a fost marcat de mai multe momente semnificative care au asigurat trecerea de la primele înregistrări izolate la statistica contemporană. *Statistica face parte din cadrul disciplinelor care studiază fenomenele într-o viziune sistematică, la nivel micro- și macroeconomic* și, în consecință, ea ține seama de dinamismul structurilor existente și de factorii care acționează variabil în timp, în spațiu și în funcție de modul de organizare socială și economică.

Rădăcinile istorice ale statisticii moderne sunt:

- *statistica practică* în sensul unor înregistrări sistematice sau izolate, ce pot fi asimilate unor observări statistice utilizate și azi;
- *statistica descriptivă* ca disciplină de învățământ și de fundamentare a statisticii practice;
- *aritmetica politică și calculul probabilităților*, ca bază conceptuală și ca mod de interpretare a fenomenelor în statistica modernă.

Forma cea mai veche – *statistica practică* – datează de peste patru milenii. Ea a servit unor scopuri fiscale, demografice și administrative. De exemplu: inventarierea (din doi în doi ani) a aurului și pământului în Egipt (2650-2190 î.e.n.). Din documentele descoperite rezultă că și chinezii dispuneau încă din milenii IV și III î.e.n. de date cu privire la numărul populației, structura terenurilor și că utilizau diferite tabele statistice referitor la unele aspecte ale activității agricole.

Ca știință statistica s-a dezvoltat în două direcții. Una dintre ele a fost *statistica descriptivă*, apărută în Germania, și care s-a dezvoltat în paralel cu practica statistică în cadrul universităților. Numele acestei școli provine de la faptul că, potrivit concepției sale, statistica se ocupă de descrierea situației geografice, economice și politice a unui stat. Printre primele descrieri sistematice, folosind statistica descriptivă, se numără și lucrarea lui Francisco Sansovino (1521-1586). Ca reprezentanți de seamă ai acestei școli se mai amintesc: Giovanni Botero (1540-1617); Hermann Conring (1606-1681), care a elaborat și primul curs de statistică; Gottfried Achenwall (1719-1772), considerat părintele statisticii pentru faptul că a dat noii discipline numele de STATISTICA, argumentând că etimologic provine de la „status” în sens de „stare”.

Trăsătura caracteristică a statisticii descriptive este faptul că ea se rezumă la descrierea fenomenelor sau a unor părți ale acestora fără să se ocupe și de cunoașterea regularităților care se manifestau în interiorul lor, a legilor care guvernau aceste fenomene. Dar limitarea ei numai la descrierea fenomenelor fără o analiză cauzală, bazată pe identificarea factorilor determinanți cu acțiune variabilă asupra lor, a dus la declinul acestei școli.

În plus, practic, continuitatea folosirii statisticii în cunoașterea fenomenelor sociale a fost asigurată prin instituționalizarea statisticilor naționale într-un cadru oficial organizat (în 1796 – în Suedia; în 1797 – în Norvegia; în 1800 – în Franța).

Treptat interesul pentru studierea statisticii descriptive a scăzut. Acest declin se datorează orientării cunoașterii științifice spre un mod de cercetare specific științelor naturii, fizicii și chimiei, caracterizat prin descoperirea legilor care guvernează fenomenele respective, prin folosirea unor procedee bazate pe experimente, ce conduc la formularea unor concluzii exacte. Astfel a apărut a doua direcție în Anglia, școala aritmeticii politice, care a reprezentat în același timp și un moment im-

portant în conturarea statisticii ca știință. Numele școlii provine de la lucrarea lui William Petty (1623 – 1687) „Aritmetica politică” (1690), în care sunt utilizate o serie de noțiuni de bază cu care operează și azi statistica (medie, proporție etc.).

Reprezentanții acestei școli se deosebesc esențial de statistica descriptivă. Prin analogie cu științele naturii, *aritmetica politică* tinde spre exactitate și în cunoașterea socială, obiectivul de bază fiind găsirea regularităților ce se produc în manifestările sociale și economice. Ea depășește faza de descriere și își fundamentează concluziile pe un număr mare de cazuri individuale cu care, prin generalizare, se pot interpreta tendințele de producere a fenomenelor economico-sociale. Cu alte cuvinte statistica începe să-și contureze etapele procesului său de cunoaștere: înregistrarea empirică a fenomenelor, sistematizarea și prelucrarea datelor individuale în vederea generalizării rezultatelor sintetice, pentru ca în final să efectueze analiza și interpretarea statistică a fenomenelor medicale.

Fondatorul acestei școli este considerat John Graunt (1620 – 1674) care, pe baza listelor privind natalitatea și mortalitatea din Londra (întocmite începând din anul 1600), a formulat în 1662 o serie de regularități privind mișcarea naturală a populației (proporția pe sexe, fertilitatea, mortalitatea etc.).

Pe linia utilizării, în mai mare măsură, a metodelor cantitativ-numerice în studiul fenomenelor economice și sociale se înscriu lucrările lui L. A. I. Quételet (1796 – 1874), considerat de unii autori ca fondator al statisticii moderne. El a folosit în cercetările sale, în domeniul demografiei și criminalității, calculul probabilităților și legea numerelor mari, fără să țină seama de particularitățile fenomenelor social-economice, de variabilitatea lor în timp și în spațiu. Neținând seama de tipurile calitative existente obiectiv în cadrul societății a fost „normal” să ajungă la concluzia că legile care guvernează aceste fenomene au caracter imuabil. Absolutizând și extinzând la societatea omenească noțiunile cu care se operează în statistică – repartiții, frecvențe, medii, dispersii etc. – el a ajuns să formuleze așa denumita teorie a „omului mediu” fără să țină seamă că nu pentru orice categorie de fenomene pot fi folosite noțiuni abstracto-convenționale cu care să se sintetizeze bogăția manifestărilor individuale ale fenomenelor studiate. Ceea ce are sens pentru un anu-

mit domeniu al statisticii poate să rămână o „abstracție pură” fără nici o reprezentare reală pentru un alt domeniu. După părerea unor autori „Lăsând la o parte latura sa fabuloasă, conceptul de *om mediu* se cuvine evocat pentru ideile implicate: repartiție; medie; dispersie; observare de masă; regularitate; noțiuni esențiale în cercetările statistice”.

Pentru statistica din mijlocul secolului al XVIII-lea și secolul al XIX-lea este specifică folosirea tot mai frecventă a metodelor matematice și, în special, a *calculului probabilităților* în investigarea și interpretarea rezultatelor privind fenomenele și procesele din societate, utilizându-le ca instrumente de cercetare în producerea fenomenelor, la cunoașterea legilor care le determină și pe această bază, la efectuarea de previziuni științifice.

Combinând metoda inductivă și deductivă, folosind din ce în ce mai frecvent rezultatele bazate pe experimentări succesive, s-au formulat, pe de o parte, principiile teoriei selecției și a extinderii rezultatelor acesteia pentru caracterizarea întregului ansamblu și, pe de altă parte, necesitatea de a privi fenomenele din cadrul societății ca interdependente. La naștere astfel *statistica inductivă*, la care și-au adus contribuții importante R. Fisher, G. Yule, K. Pearson, E. Pearson, P. Cebășev, A. Ciuprov, A. Liapunov, A. Marcov și alții.

Paralel cu dezvoltarea pe plan științific are loc și o extindere a ariei de aplicabilitate a statisticii oficiale pentru studierea nu numai a unor fenomene izolate, ci a întregului ansamblu, fiind folosită atât pentru studierea prezentului și a trecutului, cât și pentru elaborarea unor ipoteze asupra modului de comportare a aceluiași fenomen în viitor. Statistica folosind metode cantitative studiază fenomenele în continua lor dezvoltare istorică în condiții date de timp, spațiu, în deplină concordanță cu aspectul calitativ al acestora. De aici se conturează, cu mai multă claritate, cele două aspecte: dezvoltarea statisticii ca activitate practică de cunoaștere nemijlocită a fenomenelor din cadrul societății omenești, activitate considerată ca instrument indispensabil procesului de conducere și decizie; statistica ca disciplină științifică și de învățământ.

De asemenea, trebuie menționat că în literatura de specialitate, statisticii i se atribuie frecvent și unele sensuri cu conținut mai restrâns ca,

\* G. Moineagu, I. Negură, A. Urseanu, Statistica, Ed. Științifică și Enciclopedică, București, 1976, p. 21.

de exemplu: mulțimea datelor statistice obținute prin activitatea practică; metodologia statistică formată din totalitatea metodelor de culegere, prelucrare și interpretare a datelor folosite într-o cercetare concretă; metoda statistică de cercetare a fenomenelor de masă, indiferent de domeniul în care ele se produc, ceea ce îi conferă un aspect de generalitate.

Dezvoltarea statisticii ca activitate practică trebuie urmărită în raport cu transformarea actului de conducere, din activitatea intuitivă în știință a conducerii, precum și în interdependență cu progresele înregistrate în domeniul tehnicii de prelucrare a datelor pe plan național și internațional.

Evoluția statisticii ca disciplină de învățământ trebuie urmărită având în vedere dezvoltarea generală a științei și a creșterii rolului său în procesul dezvoltării societății omenești.

Apariția și dezvoltarea tehnicii electronice de calcul a impus regândirea locului și rolului statisticii în procesul de cunoaștere și predicție a fenomenelor atât în plan teoretic cât și practic. Aceasta cu atât mai mult cu cât au apărut în vecinătatea statisticii o serie de științe cum sunt toate „metriile”, ca: econometrie, biometrie, psihometrie, sociometrie etc., care aplică sistematic statistica. La aceasta se adaugă cibernetica, informatica, managementul, marketingul, cercetările operaționale și economiile de ramură. Tuturor acestor discipline, statistica le oferă din plin serviciile devenind un fel de „știință metodologică, în gen de omnibus al cunoașterii empirice”. Toate aceste discipline care se folosesc de statistică în procesul de cunoaștere au o serie de puncte comune cu statistica în ceea ce privește obiectul și metodele lor de cercetare, dar nici una nu este în măsură să o suplinească într-o manieră satisfăcătoare pe cealaltă. Tocmai acest aspect obliga la precizarea și delimitarea statisticii, ca știință independentă.

#### ARGUMENTE ÎN FAVOAREA CUNOAȘTERII STATISTICE:

- suntem în mod curent utilizatori și furnizori de informație, fie în viața particulară fie în cea profesională;
- suntem adesea decidenți și calitatea deciziilor noastre depinde de o bună informare;

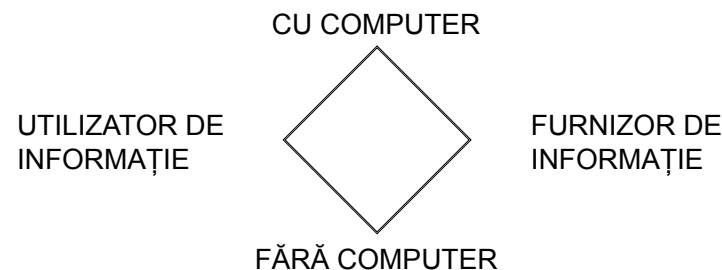
\* G. Menges, Grundriss der Statistik, Teii I. Theorie, „Westdeutscher Verlag, Kohl und Opladen, 1968, p. 16.

- factorii decizionali au nevoie să știe cum să descrie și să prezinte în modul cel mai potrivit informațiile;
- factorii decizionali au nevoie să știe cum să obțină previziuni credibile privind variabilele de interes;
- factorii decizionali au nevoie să știe cum să îmbunătățească desfășurarea activităților de care sunt răspunzători;
- factorii decizionali au nevoie să știe cum să tragă concluzii despre colectivități numeroase, doar pe baza informațiilor obținute din eșantioane
- suntem adesea interpreți și calitatea înțelegerii noastre depinde de o bună cunoaștere.

#### ”CE ÎNSEAMNĂ O GÂNDIRE STATISTICĂ?”:

- **Înseamnă recunoașterea variației în orice proces și fenomen și mai înseamnă că studiind această variație și cauzele ei vom găsi noi cunoștințe și vom putea lua decizii mai bune.**
- Înseamnă o înțelegere concretă, rapidă, în context și în corelație a realității economice și sociale.

#### ”DE CE TREBUIE SĂ CUNOAȘTEM TEORIA STATISTICĂ DACĂ COMPUTERUL OFERĂ FACILITĂȚI ÎN DOMENIU?”



Să reținem că față de statistică, teorie sau practică, ne găsim permanent în una din situațiile de mai jos:

- suntem utilizator de informație statistică având sau nu la îndemână un computer;
- suntem furnizor de informație statistică având sau nu la îndemână o rețea de calculatoare;

- suntem concomitent utilizator și furnizor de informație statistică căci potrivit teoriei sistemelor suntem sistem și subsistem în același timp.

Pentru a fi riguroși și eficienți în oricare din situațiile prezentate mai sus se impune deopotrivă cunoașterea teoriei statistice și a calculatoarelor.

## 1.2. Obiectul și metoda statisticii

În toate momentele dezvoltării sale, statistica s-a ocupat cu acele fenomene și procese care se produceau într-un număr mare de cazuri, prezentau în reproducerea lor anumite regularități și care pot fi denumite *fenomene de masă* sau *fenomene de tip colectiv*. Pentru a înțelege caracterul și particularitățile acestor fenomene trebuie pornit de la natura raporturilor de cauzalitate a acestora. În general, în cadrul societății, ca și în natură și în tehnologie, fenomenele pot să apară ca rezultat al unei singure cauze, sau ca rezultat al mai multor cauze care se manifestă izolat sau în interdependență între ele. În primul caz, sunt fenomene univoc determinate și, de regulă, se prezintă ca fenomene simple, identice între ele, denumite și fenomene tipice, în cel de-al doilea caz, apar ca fenomene multicauzale al căror proces de formare poate să prezinte grade diferite de complexitate, cu relații multiple de interdependență, formând împreună un ansamblu a cărui dimensiune și structură pot fi delimitate în timp, în spațiu și organizatoric.

Spre deosebire de fenomenele tipice care apar ca rezultat al unei singure cauze, fenomenele de masă apar ca rezultat al influenței comune a unui număr mare de cauze și condiții variabile, cu grade și sensuri diferite de influență, ceea ce face ca ele să se prezinte ca o masă compactă de fenomene atipice, aparent independente între ele. Fenomenele de masă, aparținând unei forme superioare de mișcare a materiei se produc sub acțiunea unor factori cu caracter sistematic, asociați cu cei întâmplători și ca atare fenomenologia cauzalității lor este complexă. Astfel în cadrul fenomenelor simple univoc determinate, pe măsură ce se produce cauza se produce și efectul, dacă condițiile rămân neschimbate. Ele apar deci, ca fenomene deterministe, certe, ca rezultat al acțiunii unor legi ale dinamicii, ce pot fi cunoscute cu ajutorul metodei experimentale și pentru care legea odată descoperită poate fi verificată pentru fiecare caz în parte.

Fenomenele de masă, în general, apar ca o mulțime de forme individuale diferite, cu existență distinctă, aparent fără nici o legătură de la o formă la alta, dar care analizate comparativ se constată că au o aceeași esență. Aceasta se explică, în principal, prin faptul că ele sunt generate de o serie de cauze comune care se manifestă, de regula, în condiții diferite, în consecință, la fenomenele de masă din societate, relațiile de multicauzalitate directă sau indirectă care le determină fac imposibilă cunoașterea legilor care le produc și guvernează, dacă se iau în studiu izolat doar câteva forme de manifestare a lor, ignorând ansamblul din care fac parte. Explicația constă în faptul că în practică formele individuale de manifestare diferă de la o unitate la alta în funcție de modul în care se asociază și se combină factorii sistematici cu cei întâmplători, cei esențiali cu cei neesențiali, cei obiectivi cu cei subiectivi, lăsând impresia că fiecare formă individuală din cadrul ansamblului se produce la întâmplare, fără să existe o cauzalitate certă. Dar, analizându-le la nivel de ansamblu, ele par asemănătoare între ele, fiind generate de cauze esențiale comune, supunându-se aceleiași legi de apariție și dezvoltare. La aceasta trebuie adăugat și un alt aspect și anume, relațiile de cauzalitate dintre fenomenele sociale au caracter dinamic, ca urmare a modificării permanente a condițiilor în care se manifestă. De aici, se poate trage concluzia logică că în astfel de cazuri, legea acționează atât static, când fenomenele de masă sunt circumscrise în aceleași condiții de timp, cât și dinamic, când fenomenele sunt delimitate în spațiu și organizatoric, dar înregistrate în unități de timp diferite.

Ca atare, pentru a descoperi legea de apariție a unor astfel de fenomene nu sunt suficiente numai metodele experimentale, ca în cazul fenomenelor de tip determinist ci și altele, bazate pe o serie de abstracțiuni succesive, prin care să se rețină numai ceea ce este esențial, tipic în forma lor de manifestare, prin eliminarea aspectelor întâmplătoare și neesențiale. Aceasta presupune că se iau în studiu toate cazurile individuale deoarece, spre deosebire de fenomenele care se produc pe baza legilor dinamicii, care apar ca fenomene identice, fenomenele de masă având cauze comune de apariție dar și unele cauze particulare care diferă de la un caz la altul, sunt numai asemănătoare între ele. Principala lor proprietate este *variabilitatea în timp și în spațiu și legea de apariție a acestora se manifestă ca tendință ce nu poate fi cunoscută și verificată decât la nivelul ansamblului și nu în fiecare caz*

*în parte.* Prezența printre factorii determinanți și a factorilor aleatori face ca astfel de fenomene să nu poată fi interpretate decât recurgând la principiile teoriei probabilităților. Ca atare, ele pot fi considerate ca fiind fenomene de tip nedeterminist sau stochastic. În același timp sub raport statistic, interpretarea acțiunii factorilor întâmplători nu se poate realiza decât folosind proprietățile legii numerelor mari, potrivit cărora variațiile întâmplătoare de la tendința generală se compensează reciproc dacă există un număr mare de cazuri individuale de aceeași esență, luate în studiu.

Fenomenele de masă sunt legile statistice specifice – legi care se manifestă sub formă de tendință, față de care abaterile întâmplătoare într-un sens sau altul se compensează reciproc, în consecință, este necesar ca în cercetările statistice să se ia toate cazurile individuale sau un număr suficient de mare, dar reprezentativ pentru întregul ansamblu, pentru ca ele să poată intra sub acțiunea legii numerelor mari.

Sintetizând cele relatate se poate conchide că o primă particularitate a statisticii economico-sociale constă în faptul că: *statistica studiază fenomenele social-economice de masă* în cadrul cărora acționează legile statistice și care prezintă proprietatea de a fi variabile în timp și spațiu, în literatura de specialitate ele se întâlnesc sub denumirea de fenomene de masă, fenomene de tip colectiv, fenomene stochastice sau fenomene atipice.

Studierea fenomenelor sociale cu folosirea mai multor discipline, printre care și statistica, în aceste condiții, este necesar, să se stabilească obiectul din care studiază statistica aceste fenomene de masă, ajungând astfel la cea, de-a doua particularitate, și anume: *caracterizarea laturii cantitative a fenomenelor social-economice*, stabilindu-le dimensiunea, dinamica, intensitatea, structura, raporturile de interdependență și celelalte aspecte ce pot fi caracterizate numeric în funcție de locul și timpul în care s-au produs în strânsă interdependență cu latura lor calitativă.

Deci, operând cu fenomene care apar în cadrul societății, statistica contribuie la procesul de cunoaștere cantitativă pornind de la conținutul calitativ al acestora. Desigur, în analiza statistică accentul cade pe latura cantitativă, dar cercetarea trebuie să pornească de la înțelegerea trăsăturilor calitative ale fenomenelor. Abstractizările și generalizările realizate prin intermediul statisticii în vederea evidențierii a ceea ce este specific, tipic în manifestările unui fenomen sunt corecte numai în condițiile înțe-

legerii și respectării integrității calitative a acestuia. Aceste considerații privitoare la particularitățile statisticii conduc la definirea *obiectului său de studiu*: statistica este știința care studiază aspectele cantitative ale determinărilor calitative ale fenomenelor de masă, fenomene care sunt supuse acțiunii legilor statistice ce se manifestă în condiții concrete, variabile în timp, spațiu și de organizare socio-economică.

Deci, rațiunea de a fi a statisticii sociale și economice constă în necesitatea cunoașterii aspectelor cantitative ale fenomenelor de masă cu largă răspândire în viața social-economică. Datorită întinderii, diversității și variabilității de manifestare, aspectele esențiale ale acestora nu pot fi cunoscute decât la nivelul întregului ansamblu, prin eliminarea a tot ceea ce este întâmplător și neesențial în producerea lor. De aici rezultă și modul de efectuare a studiului statistic al fenomenelor, care trebuie să țină seama de faptul că astfel de fenomene colective sunt generate de legi și cauzalități ce se manifestă sub formă de tendință, valabilă pentru întregul ansamblu. Cu alte cuvinte, metodologia de investigație statistică a fenomenelor trebuie să pornească de la trăsăturile caracteristice acestor fenomene de masă, astfel încât să se realizeze atât cunoașterea individuală a lor cât și ansamblul și structurile corespunzătoare acestora. Aceste probleme apar numai pentru statisticile care se referă la un anumit domeniu de manifestare concretă a fenomenelor de masă, unde, deci, este obligatoriu pentru statistică să țină seama, în primul rând, de specificitatea atât a ansamblului cât și a structurilor componente. Această precizare trebuie făcută deoarece în procesul cunoașterii și statistica socială și economică folosește și instrumentele formale furnizate de matematică, fapt ce a făcut uneori să se afirme că orice statistică este o ramură a matematicii.

Pentru a face o delimitare mai clară se acceptă, în general, că partea din statistică care poate fi considerată ramură a matematicii este denumită *statistică matematică*, iar celor care au un domeniu concret de investigare să li se atașeze un atribut corespunzător ariei de aplicabilitate.

Deci, de fiecare dată vor apare două științe complementare: statistica matematică și statistica domeniului de investigare. Aceste două științe însă se diferențiază din multe puncte de vedere, dintre care esențial este acela al metodei de cercetare. Se știe că, de regulă, câmpul de activitate al matematicianului este domeniul obiectelor ideale, ipotetice

și o mare parte din cunoașterea matematică se realizează în mod deductiv, ceea ce înseamnă rezolvarea problemelor cunoașterii pornind de la general la particular. Aceasta nu înseamnă că nu se procedează uneori și inductiv. Dar deosebirea constă în faptul că statistica care are ca obiect de studiu un anumit domeniu din natură, tehnologie sau societate, pornește de la particularul obținut prin observarea (înregistrarea) valorilor empirice spre ceea ce este general, valabil pentru întregul ansamblu pe cale inductivă. Orice cunoaștere științifică, inclusiv cea statistică, se realizează prin alternanța „inductiv-deductiv”, dar calea tipic statistică este cea inductivă. Aceste deosebiri apar și mai evident când fenomenele studiate au conținut social-economic.

Considerând societatea omenească ca fiind de natură obiectivă ce se dezvoltă pe baza unor legi care pot fi cunoscute și verificate în condiții specifice de timp și spațiu, statisticii, ca principal instrument de cunoaștere, îi revine o serie de sarcini ce pot fi rezolvate cu ajutorul unor metode, procedee, tehnici de calcul și interpretare statistică. Totalitatea operațiilor, tehnicilor, procedeele și metodelor de investigare statistică a fenomenelor formează *metodologia statistică*.

Datorită faptului că metoda statistică poate să fie folosită în studiile concrete efectuate de celelalte științe, care studiază dintr-un anumit unghi întreaga societate sau numai o parte a acesteia, a făcut ca ea să fie uneori considerată ca o disciplină metodologică ce poate fi aplicată de fiecare dată când apare necesitatea caracterizării cantitativ-numerice a fenomenelor și proceselor economice și sociale. De fapt, se și apreciază că metoda statistică a devenit un instrument eficient și indispensabil tuturor științelor empirice. Aceasta apreciere este nemijlocit legată de faptul că statistica vine în întâmpinarea particularizării tot mai accentuate a științei, precum și dorința de cuantificare specifică pentru aproape toate aceste științe. Pentru a putea deveni un instrument indispensabil cunoașterii actuale a fenomenelor din cadrul diferitelor domenii particulare este necesar ca ea însăși să se preocupe în permanență de elaborarea și perfecționarea unei metodologii științifice, unitare care să-i confere statutul de disciplină științifică autonomă. Ca atare, definirea metodei statisticii social-economice trebuie să se facă potrivit cu natura fenomenelor cercetate și cu scopul cercetării. Procedeele și metodele se aplică urmând etapele oricărui proces de cunoaștere, conceput ca un proces dialectic. Specificitatea metodologiei sale

poate fi urmărită și prin folosirea unor noțiuni și categorii proprii legate de obiectul său (colectivitate statistică, unitate statistică, caracteristică statistică) sau de metodologia sa (date statistice, indicator statistic, indici, ecuații de estimare, teste de semnificație etc.). Astfel se începe cu observarea fiecărui caz în parte pentru a cunoaște mai întâi manifestările cantitative individuale. Observarea se realizează concret, sub diverse forme, ca: raportări statistice (dări de seamă), recensământuri, anchete etc. În continuare, datele individuale se prelucrează prin metode specifice statisticii, în vederea eliminării abaterilor individuale din manifestările cantitative. Astfel se obțin indicatori statistici cu grade diferite de generalizare, în funcție de metodele folosite care pot fi: metoda grupărilor, metoda mediilor, analiza variațională, metoda indicilor, metoda corelației și altele. Pentru aceasta este necesară analiza prealabilă a domeniului respectiv, realizată pe căi statistice și nestatistice, ceea ce conferă statisticii caracter interdisciplinar, care se realizează în tot procesul de cunoaștere.

Procesul de cunoaștere statistică se încheie cu analiza și interpretarea rezultatelor și cu formularea concluziilor statistice, care obligatoriu trebuie să cuprindă variante probabile de apariție a acelorași fenomene în viitor.

În literatura didactică de specialitate, elaborarea principiilor, tehnicilor și metodelor folosite în cadrul culegerii (observării), prelucrării și analizei datelor statistice formează conținutul cursului ce poate fi intitulat: *Teoria statisticii*, *Statistica generală* sau *Statistica teoretică*. Această parte a statisticii formează fundamentul metodologic pentru toate disciplinele statistice care studiază fenomene și procese din diferite ramuri ale statisticii social-economice, având același obiect și aceeași metodă, diferențindu-se numai în ceea ce privește conținutul particular al fenomenelor și proceselor specifice fiecarei ramuri. Datorită acestui fapt ele apar ca discipline relativ independente, deși au un caracter unitar prin folosirea acelorași noțiuni și metode elaborate de *Teoria statisticii*, în cadrul unor facultăți economice ea se întâlnește și sub denumirea de *Bazele statisticii social-economice*.

Adâncirea diviziunii sociale a muncii a impus lărgirea ariei de investigație statistică, ducând în consecință la apariția de noi ramuri ale statisticii social-economice: *statisticii teritoriale naționale și internaționale*, *statistica indicatorilor sociali*, *statistica mediului înconjurător* și altele.



În concepția de astăzi a științei, statistica poate fi considerată ca o disciplină de graniță care, prin metodele sale, asigură cercetării științifice un caracter interdisciplinar. Apariția unor științe noi: cibernetică, informatică, teoria sistemelor și altele, nu numai că nu se suprapun ca obiect și metodă statisticii, dar au impulsivat folosirea din ce în ce mai mult a metodelor sale în lărgirea și adâncirea cunoașterii fenomenelor de masă, cu deosebire a celor complexe din cadrul societății. Considerarea societății, ca un sistem de relații cantitative și calitative au impus găsirea unor noi metode de reflectare și analiză numerică a acesteia, modele care nu pot fi elaborate și aplicate decât dacă se dispune de o informație statistică corespunzătoare, în același timp se poate aprecia efectul pozitiv pe care l-a avut asupra statisticii procesul de modernizare a științelor sociale, apariția unor științe care se ocupă cu studiul societății, sănătății populației și care influențează asupra perfecționării metodelor sale de investigație, ca, de exemplu, sociologia, antropologia și altele.

Rezultă că aceste relații ale statisticii cu celelalte științe trebuie înțelese ca un proces de reciprocitate, care are ca rezultat benefic atât dezvoltarea științelor generale, cât și a celor particulare și care face posibil ca statistica să se plaseze la granița dintre empiric și logic, dintre particular și general, dintre obiectiv și subiectiv. În plan practic, pentru statistică, relațiile de interdependență se realizează în analiza teoretică ce precede orice studiu statistic, dar mai ales în etapa finală la formularea concluziei statistice privind fenomenele investigate prin metode statistice.

În concluzie, se poate afirma că statistica se aplică în toate domeniile în care fenomenele se pot exprima cantitativ și prezintă proprietatea de a fi variabile între ele ca formă concretă de manifestare. Astfel de fenomene, întâlnindu-se în natură, tehnologie și în societate au impus o diferențiere a statisticii atât în plan teoretic, cât și practic. Statistica socială și-a dezvoltat o metodologie proprie cu aplicabilitate în toate compartimentele vieții sociale pornind de la două premise esențiale:

- studiile sale au întotdeauna caracter concret, deci trebuie să pornească de la specificitatea fenomenelor abordate, ce trebuie delimitate în timp, spațiu și structura organizatorică, ceea ce înseamnă că ea se bazează întotdeauna pe o cunoaștere empirică;
- variabilitatea fenomenelor sociale are, de regulă, un grad mult mai mare de amplitudine fiind rezultatul combinării a numeroși

factori, care fac ca în cadrul ansamblelor studiate să apară structuri variate, ce pot să se modifice esențial în dinamică.

De regulă nu există probe martor cu care să se facă direct verificarea ipotezelor utilizate, ceea ce presupune că accentul trebuie să cadă, în special, pe analiza logică cantitativă și calitativă. Aceasta nu înseamnă, însă, că statistica socială nu poate studia și unilateral unele fenomene care interesează la un moment dat, de exemplu, prețurile, consumurile de materii prime, consumurile energetice, resursele materiale etc. Specificitatea acestor studii constă în faptul că și în acest caz, este necesar să se stabilească dimensiunea și dinamica fenomenului, factorii care îl influențează, implicațiile pe care acest fenomen le are asupra altor fenomene coexistente în spații diferite sau în dinamică. De aici și necesitatea ca statistica să studieze static și dinamic fenomenele și să elaboreze variante de predicții a fenomenelor pentru perioada viitoare.

### 2.1. Definiția, obiectivele, compartimentele

Fondatorul biostatisticii este John Graunt.

**STATISTICA** – este o știință socială, care studiază partea cantitativă a fenomenelor sociale de masă în legătură strânsă cu particularitățile lor calitative în condiții concrete de spațiu și timp.

Reieșind din definiție putem să evidențiem două particularități ale ei:

- nu se ocupă cu cazuri individuale ci numai cu fenomenele de masă;
- studiază fenomenele de masă în condiții concrete de timp și spațiu.

**BIOSTATISTICA** – este statistica, ce studiază aspectele, legate de medicină și ocrotirea sănătății.

**Bazele teoretice ale biostatisticii sunt:**

- 1) dialectica materiei;
- 2) teoria generală a statisticii;
- 3) legăturile economice;
- 4) bazele medicinei sau unor ramuri ale ei.

**Obiectivele biostatisticii.** Biostatistica studiază:

1. Starea sănătății populației:
  - Reproducerea și mortalitatea populației;
  - Morbiditatea și invaliditatea;
  - Parametrii dezvoltării fizice.
2. Legătura dintre influența mediului ambiant și factorilor sociali asupra sănătății populației.
3. Acumularea și analiza datelor referitor la activitatea instituțiilor și cadrelor medicale.
4. Evaluarea eficacității metodelor de profilaxie și tratament.
5. Planificarea, economia și finanțarea ocrotirii sănătății.

Pornind de la obiective deosebit de importante următoarele **compartimente ale biostatisticii**

- Statistica sănătății populației;
- Statistica ocrotirii sănătății;
- Statistica managementului de profilaxie, tratament și altor activități în medicină.

### 2.2. Teoria probabilităților

**Teoria probabilităților** este o teorie matematică ce se ocupă cu studiul fenomenelor întâmplătoare ce pot apărea în gruparea unei caracteristici studiate pe baza unui experiment probabilist aleator. Adică, atunci când rezultatele nu pot fi prevăzute cu exactitate, dar pot avea o oarecare frecvență mai mult sau mai puțin legitimă. De exemplu, adresarea fiecărui cetățean la serviciul de urgență nu poate fi programată de nimeni, însă în totalitatea lor integrală sau parțială aceste adresări pot fi repartizate cu o anumită legitime în fiecare zi, pe ore. Studiind această legitime, putem face prognoze privind adresările din anumite zile și ore. Iar acest lucru ne permite să planificăm repartiția mijloacelor serviciului pentru satisfacerea necesităților populației.

Deci, **probabilitate** se numește acea măsură de posibilitate a apariției unor fenomene întâmplătoare în condițiile concrete date. Probabilitatea, de obicei, se înseamnă prin litera “P”.

Probabilitatea de apariție a unui fenomen poate fi estimată pe două căi:

- clasică – dacă fenomenul se produce în “h” posibilități ale “n” experimente, atunci probabilitatea acestui fenomen va fi  $h/n$ ;
- frecvența empirică – când experimentul se repetă de mai multe ori (f), iar fenomenul se produce cu o frecvență de acum fixată “h” (cum e cazul cu adresările la serviciul de urgență). Aici probabilitatea apariției fenomenului va fi  $h/f$ .

Probabilitatea apariției în totalitatea parțială a unui fenomen “P” se determină prin raportul fenomenelor deja apărute (m) la numărul tuturor cazurilor posibile (n) – aceasta e definiția clasică a probabilității lui Laplace. Atunci probabilitatea fenomenului “P” va fi:

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

Drept contraprobabilitate determinăm alternativa – probabilitatea lipsei fenomenului, însemnată prin litera „q”.

De aici:

$$q = \frac{n - m}{n} = 1 - \frac{m}{n} = 1 - P \quad \text{sau} \quad q = 1 - P; P + q = 1$$

Așadar, suma probabilității prezenței fenomenului cu contraprobabilitatea lui este egală cu unitatea, iar în procente cu 100,0%.

Rezultatele de mai sus pot fi exprimate și altfel, că probabilitatea apariției fenomenului „P” se află în hotarele dintre 0 și 1 sau 0 -100,0%. Cu cât probabilitatea e mai aproape de 1 sau 100,0% cu atât prezența fenomenului în totalitate e mai reală. Această teorie argumentează legea cifrelor mari.

### 2.3. Legea cifrelor mari

**Legea cifrelor mari** (teorema lui P. Cebâșev) are două aplicații importante pentru determinarea totalităților selective:

1. Pe măsura majorării numărului cazurilor de observare rezultatele cercetării căpătate pe baza totalității selective tind să reproducă datele totalității integrale;
2. La atingerea unui anumit număr de cazuri de observare în totalitatea selectivă rezultatele cercetării vor fi maximal apropiate de cele posibile pe baza totalității integrale.

Deci, la un număr suficient de mare de cazuri de observare se manifestă acea legitate ce stă la baza totalității integrale și care nu poate fi observată în cazul unui număr mic de cazuri cercetate.

Pentru o înțelegere mai bună a teoriei probabilității și regulilor legii cifrelor mari descriem exemplul cu urna lui Galton, în care era introdus un număr egal de bile negre și albe (câte 500), deci, în proporție de 1:1 sau 50% și 50%. Se scoate câte o bilă, se înseamnă culoarea și apoi ea este pusă înapoi în urnă. Operația se repetă de 10, 20, 50 ori și mai mult. La început proporția bilor albe față de cele negre era 1:4 sau 20% și 80%. La un număr mai mare de extrageri (100-200-250) proporția se apropie de cea reală 0,5/0,5 sau 50% și 50%. Din acest moment oricât s-ar prelungi numărul de extrageri proporția nu se va schimba. Deci, limita maximă corespunzătoare a totalității selective poate fi calculată cu o oarecare probabilitate.

Să presupunem că ponderea fenomenului în totalitatea selectivă ( $P_1$ ) e mai mică decât cea din totalitatea integrală (P). Diferența dintre aceste mărimi exprimă eroarea reprezentativă (m) și se explică prin faptul că nu toate unitățile de observare au fost studiate. Comparând aceste ponderi, dintre care una e căpătată pe baza unei totalități selective de un volum mai mare, vedem că valorile lor s-au apropiat. Deci, în

caz de un eșantion de volum mai mare probabilitatea de necoincidență a ponderilor e atât de mică că ea poate fi neglijată.

Teoria statistică a demonstrat că în caz de totalitate selectivă mare ( $n > 30$ ) ponderea fenomenului ( $P_1$ ) diferă față de cea din totalitatea integrală (P) cu probabilitatea de 95% și este egală cu  $2m$ ; cu probabilitatea de 99,7% diferența ponderilor ( $P_1 - P$ ) nu va depăși  $3m$ . Cifrele 1, 2, 3 n, cu care se înmulțește valoarea erorii (m) poartă denumirea de coeficienți de exactitate sau factori de probabilitate și se înseamnă prin simbolul “t”. Cu majorarea coeficientului “t” crește probabilitatea cu care vom putea spune că diferența ponderilor căpătate din ambele totalități este situată în intervalul:  $\Delta = tm$ , unde  $\Delta$  este eroarea limită admisă pentru studiul dat.

$$\text{Deci, } P = P_1 \pm \Delta$$

În cazul experimentului cu bilele s-a observat o legitație, și anume că la un anumit număr de extrageri se înregistrează raportul real al bilor albe și negre.

Bazându-se pe legea cifrelor mari și majorând volumul totalității selective, se poate influența asupra mărimii maxime a erorii, aducând-o la valori minime. Pentru a determina numărul de cazuri de observație (n) al totalității trebuie să cunoaștem eroarea limită admisă ( $\Delta$ ) pentru studiul dat.

### 2.4. Metode de cercetare utilizate în biostatistică

Metodele de cercetare aplicate în biostatistică nu diferă cu mult de cele folosite în alte domenii, unele din ele au fost chiar împrumutate de la ele (cele matematice, economice). Utilizarea lor se va face însă ținând cont de specificul acesteia. De exemplu, specificul statisticii sanitare în cadrul studiului sănătății publice constă nu numai în depistarea fenomenelor legate de ea, ci și a dinamicii lor, în evidențierea tendinței acestor fenomene, corelației lor cu factorii ce le provoacă etc. Cunoașterea acestor lucruri se poate efectua prin intermediul următoarelor metode:

1. **Metoda observării** – constă în urmărirea desfășurării unor fenomene pentru a le putea analiza în dinamică, sau prin compararea lor, ca apoi să se realizeze sinteza caracteristicilor lor esențiale.

Cercetările efectuate cu folosirea acestei metode în studiul sănătății publice sau unor cazuri clinice se realizează pe diferite căi:

a) *statistică* – când se acumulează informația sub formă de valori numerice despre schimbările fenomenului studiat sau activitatea organelor, instituțiilor sanitare și altor servicii pentru sănătate. Pe parcurs, observarea statistică după volum poate fi efectuată prin:

- *cercetarea integrală* (adică a întregului volum numeric al totalității) când avem de a face cu un fenomen de apariție rară, cu un număr mic de cazuri în legătură cu care se cere de luat măsuri urgente, sau când ne interesează determinarea stării întregii populații prin recensăminte;
- *cercetarea selectivă* – față de cea integrală se folosește mai des și urmărește studiul unui fenomen cu o frecvență mai mare pe baza unei totalități selective. Dacă este nevoie de a studia aprofundat un aspect, un fenomen răspândit într-o localitate, un grup de populație, va fi folosită cercetarea monografică, prin cohortă, prin sondaj etc.

b) *aprecierea prin expertiză* – prezintă un supliment, când fenomenul general e studiat pe fiecare aspect al lui în particular prin intermediul experților. De exemplu, dacă ne interesează morbiditatea spitalizată, atunci fiecare caz de boală tratat în staționar fie că va fi paralel studiat în decursul spitalizării, fie că după externare pe baza istoriei de boală. Experții se pronunță cu privire la termenele spitalizării bolnavului, dată fiind starea lui, apreciază calitatea investigațiilor, corectitudinea diagnosticului și tratamentului în fiecare caz concret de patologie. Același lucru se poate efectua în cadrul serviciului de ambulator, de urgență etc. Datele generalizate servesc la căpătarea unor coeficienți de corectare a indicilor obținuți pe cale statistică. Expertiza este o metodă mai mult calitativă de investigație, permițând elaborarea de măsuri concrete de ajustare a fenomenului studiat.

În funcție de  **timp**  când se efectuează observarea delimităm:

- *cercetarea curentă* (permanentă, continuă) – se utilizează atunci când variabilitatea fenomenului studiat se poate schimba destul de des, iar aceste schimbări importante pot surveni zi de zi și chiar oră de oră. Astfel de cercetări se efectuează de regulă în clinică, mai ales în secțiile de reanimare sau în maternitate, secțiile de boli infecțioase, unde observarea trebuie să se facă de câteva ori pe zi sau peste un anumit interval de timp.

- *cercetarea periodică* (la anumite intervale de timp) – se utilizează atunci când fenomenul studiat are o variabilitate mult mai mică în timp, sau o apariție și evoluție periodică. De exemplu, se știe că o parte din boli pot da acutizări în anumite perioade ale anului (boala ulcerosă, reumatismul) sau evoluează în condiții atmosferice specifice ale anului (gripa, virozele organelor respiratorii, unele boli infecțioase etc.). Ca urmare, cercetarea poate fi efectuată numai pe parcursul acestor perioade de timp sau la sfârșitul lor. În alte cazuri se recurge la cercetări periodice o dată la 5-10 ani. Astfel de cercetare se impune în cazul unui volum mare al totalității (structura populației după sex, vârstă, ocupație, locul de trai, starea civilă etc.).
- *cercetare la un moment critic* – atunci când se fixează data și chiar ora de înregistrare a fenomenului studiat (recensământul populației, înregistrarea personalului medico-sanitar etc.).

După  **frecvența**  observărilor cercetările pot fi:

- de o singură dată;
- repetate.

2. **Metoda epidemiologică** sintetizează cele evidențiate prin metoda observării și prezintă un studiu corelativ al fenomenelor din cadrul sănătății publice cu factorii (cunoscuți sau presupuși) de risc. Această metodă, ca și cea precedentă, folosește pe larg o serie de metode matematice pentru a găsi legitatea schimbărilor variabile, dinamice, structurii fenomenelor studiate în legătură cu factorii ce le determină (valorile medii, dispersia, corelația, regresia etc.).

3. **Metoda istorică** e strâns legată cu primele două metode fundamentând studiul sanitar al sănătății publice sau al activității serviciilor pentru sănătate în cadrul dezvoltării istorice a societății.

Este un fapt dovedit că gradul de dezvoltare a societății, orânduirea socială în trecut și în prezent determină nivelul de trai al populației, calitatea vieții, deci și sănătatea ei.

Aici mai des se folosește metoda comparării aceluiași fenomen raportat la diferitele categorii de populație (diferențiate după semne – sex, grupuri sociale, vârstă, ocupații etc.), teritorii (continente, țări, regiuni, localități) studiate acum și în trecut.

4. **Metoda economică** precizează starea de sănătate publică, determinată de prezența surselor bănești, materiale, economice etc.

Această metodă se folosește la aprecierea eficacității activității serviciilor pentru sănătatea populației (din punct de vedere social și economic) în cadrul medicinei de stat, prin asigurare, precum și celei private.

O varietate a acestei metode e *metoda economico-matematică* când se utilizează în combinație cu o serie de metode matematice ce ne permit să optimizăm acțiunile sanitare, rețelei sanitare legate de probleme de sănătate a populației cu analiza de sistem, de prognozare etc.

5. **Metoda experimentală** urmărește scopul de a elabora și aproba noi forme și metode de organizare a muncii, a asistenței medicale, aplicarea în practică a experienței înaintate, verificarea reciprocă a diferitelor proiecte, ipoteze, a noilor metode de diagnosticare și tratament etc. Specificul acestei metode constă în aceea că cercetătorul singur își „crează” obiectul și metoda de cercetare, reproducând astfel fenomenele sau aspectele ce-l interesează în condiții de laborator, iar mai apoi analizând totul în detaliu conform scopului stabilit.

Orice experiment nu depinde numai de dorința și năzuința savantului, ci e o problemă de ordin social. Rezolvarea lui va cere din partea acestuia îmbinarea mai multor metode de cercetare.

În genere, în orice studiu statistic cu scop de cunoaștere nu se folosește numai o metodă izolată de cercetare. Pentru o investigație multilaterală a fenomenului și ținând cont de scopul propus cercetătorul trebuie să decidă ce metode se impun și în ce etapă.

**Etapa I** este legată direct de metoda de observare epidemiologică, economică și de acumulare (culegere) a datelor informative. În **etapa a II-a** se recurge la o serie de metode statistico-matematice de prelucrare a informației acumulate, de asemenea și la metode economice, economico-matematice.

În **etapa a III-a** vom folosi și metoda istorică comparând rezultatele curente ale studiului cu cele ce au fost înregistrate în trecut. Astfel cercetătorul poate să-și atingă scopul studiului și să analizeze în ansamblu toate caracteristicile de bază ale fenomenului trăgând concluzii corecte și elaborând măsurile practice necesare.

Metoda observării după modalitatea îndeplinirii poate fi: statistică și de expertiză. Iar după timpul îndeplinirii: curentă (pe parcursul anului), periodică (o dată la doi ani), la un moment critic (la finele anului calendaristic).

**Metodele de colectare a datelor** pot fi divizate în două grupe mari: directă și indirectă:

La metoda **directă** se referă:

- Metoda observării
- Anchetare
- Interviu
- Monitorizare

Cea **indirectă** reprezintă extragerea datelor din documentația medicală.

Cel mai des în studiile statistice medicale sunt folosite următoarele metode de colectare a informației.

1. **Metoda observării (directă)** se utilizează de obicei în clinică, când medicul, după ce precizează anamneză bolnavului, face studiul obiectiv și investigațiile de laborator ale pacientului, înregistrează în fișă aceste date. Analogic se procedează și în policlinică, și la deservirea chemărilor la domiciliu. Dacă cercetătorul statistic vrea să folosească această metodă, el trebuie să fie prezent alături de medic și să înregistreze datele de care are nevoie. Sunt cazuri când pe cercetător îl interesează nu numai date legate de boală, dar și probleme de ordin personal și social, ca urmare înregistrarea directă poate fi făcută sub formă de anchetă. În acest caz cercetătorul trebuie să beneficieze de încrederea respondentului.

2. **Metoda extragerii informației dintr-un formular statistic (indirectă)** – fișa medicală a bolnavului de staționar, fișa medicală a bolnavului de ambulator, tichetul statistic de evidență a diagnosticului definitiv (precizat), epicriză, raport statistic privind numărul de maladii înregistrate la bolnavii domiciliați în teritoriul de deservire a instituției curative etc. Când se efectuează acest lucru se ține cont de programul de culegere a informației, unde trebuia să fie indicat concret care este unitatea de evidență, caracteristicile ei atributive și cantitative. Datele privind aceste caracteristici se extrag din documentele enumerate mai sus și se notează într-un registru special sau în mai multe fișe, pe fiecare semn aparte. Această metodă se folosește frecvent în studiile statistice și poate fi utilizată de însuși cercetătorul sau de persoane special instruite, în acest caz fiind necesar controlul logic al materialului cules.

3. **Metoda de anchetare** utilizează informația prin intermediul unor anchete, anterior pregătite, care includ întrebări speciale, la care respondenții trebuie să dea răspuns. Această metodă se folosește în studiile sociologice, dar poate fi aplicată și în orice studiu statistic sanitar

În îmbinare cu alte metode de culegere a informației. Este foarte important ca întrebările incluse în anchetă să fie formulate clar și concret, în caz contrar respondenții nu vor ști ce să răspundă la ele. Totodată ancheta trebuie să fie anonimă, adică să nu cerem de la respondenți să ne comunice numele, adresa domiciliului ș.a., deoarece aceasta poate să-l facă pe cel anchetat să nu răspundă în genere la nici o întrebare.

**Anchetele de informare** se utilizează mai ales în cercetările stării de sănătate a populației și în cercetările epidemiologice. Întrebările din anchetă pot fi de tip:

- **deschis** – când la întrebare nu se dau variante de răspuns și respondentul trebuie să le formuleze singur;
- **semideschis** – se dau variante de răspuns, dar se lasă un rând liber pentru expunerea unor opinii personale ale respondentului;
- **închis** – când la întrebare sunt date 2 și mai multe variante de răspuns, iar respondentului i se propune să aleagă unul din ele.

Datele acumulate pot fi analizate prin următoarele metode:

- Metoda istorică
- Metode epidemiologice
- Metode economice
- Metoda experimentală
- Calcularea indicatorilor relativi și medii
- Metode de corelație și regresie
- Determinarea testului de semnificație (t) și pragului de semnificație (p)

## 2.5. Noțiuni de bază în biostatistică

În biostatistică sunt cunoscute următoarele noțiuni de bază:

- totalitatea statistică
- unitatea de observare
- caracteristică statistică
- indicatori statistici
- date statistice

### 2.5.1. Totalitatea (colectivitatea) statistică

Definiția: **totalitatea statistică** reprezintă un număr de elemente (unități de observare) omogene, luate împreună în baza unui factor comun în anumită perioadă de timp și spațiu.

Numărul de unități de observare determină volumul totalității supuse studiului și se notează prin litera „n”. După volum deosebim două tipuri de totalități statistice:

- Totalitatea integrală (generală, „univers statistic”)
- Totalitatea selectivă

**Totalitatea selectivă** are următoarele caracteristici de bază:

- trebuie să dețină caracteristici de bază de care dispune cea integrală;
- trebuie să dispună de un volum.

Metodele de selectare a totalității selective:

- cercetarea prin sondaj
- cercetarea monografică
- cercetarea selectivă:
  - aleatorie
  - mecanică
  - tipică.

**Cercetarea prin sondaj** se bazează pe o metodologie precisă, unitară, valoarea rezultatelor depinzând de efectuarea ei corectă științific. Această cercetare nu permite să se tragă concluzii generale, valabile pentru totalitatea integrală, fiind mai mult o metodă de investigație preventivă unui studiu de mare volum. De obicei se efectuează pe baza unui eșantion mic.

**Cercetarea monografică** e un studiu selectiv în care limitarea volumului eșantionului e completată cu o aprofundare a cercetării caracteristicilor esențiale. Acest tip de eșantion poate fi limitat în cazul unui examen medical clinic cu un scop bine determinat (stabilirea gradului de răspândire al unei boli sau al unui grup de boli din aceeași clasă). Tot astfel de cercetări pot fi aplicate pe un eșantion mic după spațiu, dar majorat în volum după timp, adică în dinamică (studiul natalității, mortalității pe o perioadă de mai mulți ani într-o circumscripție rurală, dar într-o strânsă legătură cu factorii social-economici, sanitaro-igienici etc.).

**Cercetarea selectivă** – studiul selectiv reprezentativ pentru totalitatea integrală și care poate fi efectuat prin selecția:

- **aleatorie** (întâmplătoare, randomizată simplă) se efectuează prin extrageri din liste în care sunt înregistrate toate cazurile individuale fără nici o grupare sistemică prealabilă. O metodă frecventă de selecție aleatorie este tragerea la sorți;

- *mecanică* (sistemică) – este o metodă superioară celei aleatorii, deoarece fiecare unitate de observație are șanse egale să fie aleasă. Selecția eșantionului se face după modelul de șah sau cazurile de evidență sunt ordonate în ordine alfabetică sau localitățile sunt aranjate după hartă și se selectează fiecare a 4-a, a 6-a ori a 10-a, în funcție de pasul de numărare. În acest mod se obține o selecție teritorială uniformă. Intervalul se calculează astfel ca eșantionul să cuprindă de la 5 până la 10% din totalitatea integrală.

De exemplu, 10% – atunci:  $\frac{100\%}{10} = 10$ , fiecare al 10-lea;  $\frac{100\%}{5} = 20$  etc.

Calea aceasta de selectare, deși răspândită, este anevoioasă de înfăptuit. Avantajul constă numai în simplitatea selectării eșantionului, în timp ce exactitatea rezultatelor cercetării poate avea erori mari, fiindcă nu se ține cont de frecvența reală de răspândire a fenomenului, de dispersia lui în spațiu;

- *tipică (stratificată)* (proporțională cu mărimea eșantionului) – urmărește scopul selecției unităților de observație din grupurile tipice ale “universului statistic”. Pentru început, în cadrul “universului statistic” toate unitățile de observare se grupează după anumite caracteristici în grupuri tipice (de exemplu, vârstă, sex sau după intensitatea frecvenței fenomenului). Din fiecare grup, pe cale aleatorie sau mecanică, este selectat un anumit număr de unități astfel ca raportul după caracteristici în eșantion să fie același ca și în totalitatea integrală.

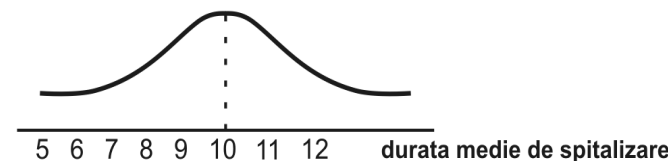
Dacă e nevoie să se facă o nouă stratificare cu scopul obținerii unui grup și mai omogen, metoda în cauză se numește *tipică stratificată cu mai multe trepte*. Avantajul ei este că fenomenele studiate pot fi mai uniform reprezentate și deci și eșantionul va fi reprezentativ;

- *în cuiburi (în serii, în clastere)* – aici din totalitatea integrală se selectează nu unități individuale, ci serii (microzone), localități care sunt ulterior examinate în întregime.

Distribuția datelor în totalitatea selectivă poate fi *alternativă* (Da/Nu), *simetrică* (normală) și *asimetrică*.

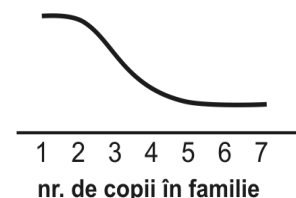
## DISTRIBUȚIA DATELOR ÎN TOTALITĂȚEA SELECTIVĂ

- **ALTERNATIVĂ**   Da/Nu
- **SIMETRICĂ (NORMALĂ)**

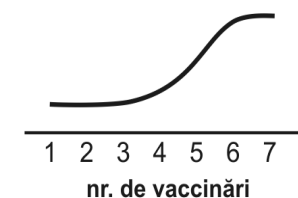


### • ASIMETRICĂ

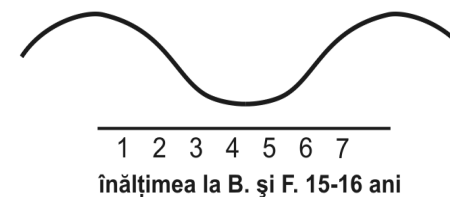
\* pe dreapta



\* pe stânga



\* bimodală



### 2.5.2. Unitatea de observare

Reprezintă fiecare element component al colectivității statistice, care este purtătorul tuturor trăsăturilor comune ale colectivității supuse studiului.

Unitățile statistice pot fi simple, care nu mai suportă diviziune (persoana) și complexe, rezultate ale organizării sociale (familia).

### 2.5.3. Caracteristica sau variabila statistică

Reprezintă trăsătura, proprietatea, însușirea comună unităților de observație, reținută în studiul statistic pentru a fi înregistrată și care variază ca valoare de la o unitate la alta.

Caracteristicile statistice pot fi clasificate după cum urmează:

- **în funcție de modul de exprimare:**
  - caracteristici **calitative (nominative)**, exprimate în cuvinte: profesie, localitatea de domiciliu, culoarea părului, culoarea tegumentelor, sexul, starea de nutriție sau de boală, starea la externare a unui bolnav, etc.;
  - caracteristici **cantitative (numerice)**, exprimate în cifre: salariu, vechimea în muncă, înălțime, greutate, perimetru, tensiune arterială, puls, temperatură, etc. Sunt caracteristici măsurabile.
- **în funcție de numărul variantelor/valorilor de răspuns pe care le pot lua:**
  - caracteristici **alternative (binare sau dihotomice)**, acelea care pot lua doar două variante de răspuns: sex (M/F), starea civilă (căsătorit/necăsătorit), familie cu copii sau fără copii, etc;
  - caracteristici **nealternative** – cele care pot lua mai multe valori/ variante de răspuns: salariu, profesie, localitate de domiciliu, starea la externare a unui bolnav, etc.
- **în funcție de natura variației caracteristicilor cantitative:**
  - caracteristici **continue**, care pot lua orice valoare din scara lor de variație: greutatea unei persoane, înălțimea, temperatura, etc.;
  - caracteristici **discrete** sau **discontinue**, care pot lua numai valori întregi: numărul de copii pe care îi are o familie, numărul de persoane dintr-o familie, număr de medici, număr de paturi, număr de vizite, etc.
- **în funcție de conținutul caracteristicii:**
  - caracteristici **de timp** (anul nașterii);

- caracteristici **de spațiu** (localitatea de domiciliu);
- caracteristici **atributive**, în care variabila reprezintă un atribut, altul decât spațiul ori timpul – cele calitative și cantitative.
- **în funcție de modul de obținere și caracterizare a fenomenului:**
  - caracteristici **primare** (obținute, de regulă, în etapa de colectare a datelor statistice prin măsurare sau numărare);
  - caracteristici **derivate**, obținute în procesul prelucrării datelor statistice.
- **în funcție de modul de influență asupra fenomenului:**
  - caracteristici **factoriale**;
  - caracteristici **rezultative**.

### 2.5.4. Indicatori statistici

Indicatorul statistic este expresia numerică a unor fenomene, procese, activități sau categorii economice și sociale, definite în timp, spațiu și structură organizatorică.

Deosebim următoarele funcții ale indicatorilor statistici:

- funcția de măsurare;
- funcția de comparare;
- funcția de analiză;
- funcția de sinteză;
- funcția de estimare;
- funcția de verificare a ipotezelor și de testare a semnificației unor indicatori statistici calculați.

Indicatorii statistici pot fi prezentați prin valori absolute, relative și medii.

Către valorile relative pot fi atribuite *rata*, *raportul și proporția*.

**Rata** ne arată cât de repede evenimentul (nașteri, îmbolnăviri, decese) apare în populație.

Componentele ratei sunt:

- **Numărător:** numărul de evenimente observate;
- **Numitor:** populația în care evenimentele au loc;
- **Timpul** specificat când au loc evenimentele;
- De obicei un **multiplator** transformă rata dintr-o fracție incomodă sau decimală într-un număr întreg.

Tipurile de rate sunt:

- brute;



- speciale (specifice);
- standardizate.

**Exemplu:** rata brută a mortalității

$$\frac{\text{Numărul deceselor, indiferent de cauză, într-o populație, într-o perioadă specifică}}{\text{Numărul de persoane cu risc de a muri de-a lungul perioadei}} \times 10^n$$

**Exemplu:** rata mortalității din cauze specifice

$$\frac{\text{Numărul de decedați, dintr-o cauză anume într-o populație, într-o perioadă specifică de timp}}{\text{Numărul de persoane cu risc de a muri din acea cauză de-a lungul perioadei}} \times 10^n$$

**Exemplu** "Rata" caz-fatalitate (RCF) – proporția cazurilor unei condiții specifice fatale într-o perioadă specifică de timp.

Sinonimele pentru rată sunt: *frecvență, nivel, răspândire, intensitate*.

**Raportul** permite compararea unei populații cu alta.

Raportul este un număr împărțit la altul.

Exemplu: raport dintre bărbați și femei, nașteri și avorturi, nr. de medici și asistente medicale. În aceste cazuri nu este necesară o relație specifică între numărător și numitor. Raportul poate fi prezentat

astfel  $x : y$  sau  $\frac{x}{y} \times 10^k$ ,

unde: „x” – numărul de paturi, nr. de medici,

„y” – numărul de populație.

**Proporția** ne arată ce fracțiune a populației este afectată. Caracteristicile lor de bază sunt:

- coeficientul a 2 numere;
- numărătorul **este inclus** în numitor;
- proporția întotdeauna deviază între 0 și 1 sau între 0 și 100%.

Exemplu de proporție:

În anul 2006 în Republica Moldova au fost recunoscuți ca invalizi 7695 bărbați și 6005 femei. Care este procentul femeilor din numărul total?

$$\frac{Nr \text{ femei}}{Nr \text{ total}} \times 100 = \frac{6005}{6005 + 7695} \times 100 = 43,8\%$$

Indicatorii de proporție nu pot fi comparați.

**Exemplu:**

Tabelul 1.

Numărul de cazuri de îmbolnăviri în funcție de vârstă și localitatea (abs., %).

	Vârsta	Nr. cazuri	Proporția
<b>Nord</b>	0-14	10	50%
	15-64	6	30%
	65->	4	20%
	Total	20	100%
<b>Sud</b>	0-14	15	50%
	15-64	9	30%
	65->	6	20%
	Total	30	100%

Pentru comparație pot fi utilizate ratele.

**Exemplu:**

Tabelul 2.

Numărul de cazuri de îmbolnăviri în funcție de vârstă și localitatea (abs., %).

	Vârsta	Nr. cazuri	Proporția	Nr. populației	Rata
<b>Nord</b>	0-14	10	50%	100	10%
	15-64	6	30%	200	3%
	65->	4	20%	100	4%
	Total	20	100%	400	
<b>Sud</b>	0-14	15	50%	100	15%
	15-64	9	30%	200	4,5%
	65->	6	20%	100	6%
	Total	30	100%	400	

Sinonimele indicatorului de proporție sunt *cota, ponderea, structura*.

### 2.5.5. Date statistice

Analiza datelor statistice se face în funcție de:

- timp;
- loc;
- persoană.

Caracteristica de **timp** ne permite să stabilim:

- modificări pe termen scurt;
- modificări ciclice;
- modificări seculare (pe termen lung);
- poate fi aranjată în tabele și grafice.

Prin caracteristica de **loc** se descrie unitatea geografică. Pentru prima dată această descriere a fost efectuată de către John Snow în an. 1854 când în Londra a avut loc erupția de holeră.

Caracteristica de **persoană** ne permite să facem analiza datelor acumulate după vârstă, sex, etnie, etc.

## 2.6. Gruparea materialului statistic. Seria statistică de variație.

### Indicatorii tendinței centrale.

**Gruparea materialului statistic** constă în aranjarea unităților statistice, în funcție de diferitele caracteristici sau variabile, în grupe cât mai omogene, pentru a le putea scoate cât mai ușor în evidență. Gruparea este operația statistică ce permite trecerea de la aspectele particulare, individuale ale unităților statistice la aspecte generale, comune pentru colectivitatea supusă studiului.

**În funcție de numărul** caracteristicilor de grupare se realizează **grupări simple**, când la baza repartizării stă un singur criteriu, **și complexe**, când la baza repartizării stau două sau mai multe criterii.

**După natura** caracteristicii gruparea poate fi:

- **în timp**, când repartizarea materialului statistic are la bază anumite intervale de timp;
- **în spațiu**, când repartizarea materialului se face în raport cu locul sau teritoriul unde s-a examinat colectivitatea statistică, de exemplu mediu rural și mediu urban.

**În funcție de variația** caracteristicii de grupare se realizează grupări pe **variante** sau pe **intervale**.

**Limitele grupării** nu se stabilesc mecanic ci depind de natura fenomenului studiat și scopul cercetării. În principiu o grupare prea amănunțită a materialului statistic duce la o fărâmițare ce ne împiedică să observăm ce este caracteristic, esențial. De asemenea dacă o caracteristică are prea multe valori vom repartiza materialul statistic pe grupe de valori. Fiecare grupă de valori sau „clasă” are la rândul ei o limită inferioară, o limită superioară și un centru al grupei egal cu semisuma celor două limite.

În cadrul grupării trebuie să cunoaștem exact valorile extreme – maximă și minimă ale caracteristicii și trebuie să stabilim intervale de grupă egale, iar limitele claselor să fie distincte pentru a nu crea confuzii cu ocazia repartizării unităților de observare.

Prelucrarea materialului statistic nu se oprește la gruparea materialului deoarece datele brute obținute, de cele mai multe ori, nu permit aprecieri comparative. De aceea materialul statistic este supus în continuare unei prelucrări cu ajutorul metodelor de statistică matematică în vederea obținerii unor indicatori statistici ca: valori relative, valori medii, criterii de variație, corelație, veridicitate, ce vor permite aprecieri comparative, corecte și concluzii semnificative.

Urmare grupării datelor statistice rezultă serii statistice.

**Seria statistică de variație** este șirul de valori numerice ale caracteristicii, ordonate crescător sau descrescător în funcție de mărimea acestora.

Seria statistică reprezintă corespondența a două șiruri, cel al valorilor variantelor ( $x$ ) și cel al frecvențelor ( $f$ ), motiv pentru care se mai numește și serie de distribuție/serie de frecvențe. Suma frecvențelor variantelor corespunzătoare corespunde cu numărul de cazuri cercetate ( $\sum f = n$ ).

Seriile statistice de variație pot fi de două categorii: **simple** și **grupate**. Când fiecărei valori a caracteristicii îi corespunde o singură frecvență vorbim de o serie statistică simplă, iar când fiecărei valori îi corespund mai multe frecvențe vorbim de o serie statistică grupată.

Unii autori definesc seria statistică în modul următor: **simplă** – de regulă se formează în cazul unui număr mic de cazuri cercetate –  $\leq 30$  și **grupată** – care se formează în cazul unui număr mare de cazuri cercetate –  $> 30$ .

Cerințele de bază pentru formarea seriei de variație:

1. Ordonarea valorilor variantelor;
2. Numărarea frecvențelor fiecărei variante;
3. Determinarea numărului de grupe și valorii intervalului;
4. Gruparea seriei de variație, utilizând intervalul cu respectarea continuității seriei;
5. Reprezentarea grafică a seriei de variație.

Toate 5 cerințe sînt obligatorii pentru formarea seriei de variație grupate, cerințele 1,2 și 5 – pentru formarea seriei de variație simplă.

Numărul de grupe în serie se determină în dependență de numărul cazurilor cercetate (Tab.3).

Tabelul 3

Numărul de grupe în funcție de numărul de cazuri

Număr cazuri	31-45	46-100	101-200	201-500
Număr grupe	6-7	8-10	11-12	13-17

În cazul seriei de variație în care variantele sînt exprimate în intervale de grupă, pentru a le putea introduce în calculul statistic stabilim mijlocul fiecărei grupe, reprezentat printr-o singură valoare, care este centrul intervalului grupei obținut prin semisuma valorilor variantelor extreme ale fiecărui interval de grupă (Tab.4).

Tabelul 4.

Gruparea seriei statistice în funcție de dinți afectați

Grupe de valori ale dinților afectați	Mijlocul intervalului de grupă	Frecvențe
1-3	2	106
4-6	5	252
7-9	8	75
10-12	11	50
13-15	14	25
Total		508

Cele trei proprietăți majore ale seriilor de variație, pe care le putem analiza folosind indicatorii statistici sunt cele privitoare **la tendința centrală, la variabilitatea și la forma distribuțiilor.**

O clasificare a indicatorilor tendinței centrale se poate face, în funcție de modul de determinare a lor, în:

- **indicatori (mărimi) medii de calcul:** media aritmetică, armonică, cronologică, pătratică, geometrică etc.;
- **indicatori medii de poziție:** modul, mediana.

Indicatorii **fundamentali** ai tendinței centrale sunt: **media aritmetică, modul și mediana**, dar în anumite cazuri speciale putem apela și la alte tipuri de medii.

**Mărimile medii sunt mărimi tipice, caracteristice, ce definesc un fenomen variabil.**

În general, în toate cercetările, dar în special cele care privesc starea de sănătate a populației, de un interes deosebit este cunoașterea

comparativă a fenomenelor studiate față de un etalon, mărimea medie. Fără cunoașterea mărimilor medii, comparația nu este posibilă decât în mod imperfect: compararea dintre două sau mai multe fenomene pentru a constata diferența dintre ele, neputând concluda dacă aceste fenomene sînt sau nu apropiate de o valoare etalon.

Mărimea medie are aceleași dimensiuni concrete cu ale variabilei a cărei repartiție de frecvențe o caracterizează. Astfel dacă variabila privește înălțimea în cm a nou născutului și mărimea medie va fi redată în cm.

După Yule și Kendall condițiile pe care trebuie să le îndeplinească o mărime medie ar fi:

- Să fie definită în mod precis, independent de dorința utilizatorului;
- Să fie expresia, sinteza tuturor observațiilor înregistrate;
- Să posede proprietăți simple, evidente, clare chiar și pentru ne-specialiști;
- Să fie simplu și rapid de calculat;
- Să fie puțin sensibilă la fluctuațiile de selecție;
- Să poată fi studiată rapid cu ajutorul calcului algebric.

În general, indicatorii tendinței centrale, calculați ca mărimi medii realmente nu satisfac toate condițiile prevăzute mai sus. Utilizatorul trebuie însă să fie interesat să cunoască condițiile nesatisfăcute și implicațiile acestora pentru fundamentarea deciziilor.

## MEDIA

**Media** este expresia care sintetizează într-un singur nivel reprezentativ tot ceea ce este esențial, tipic, comun, obiectiv în apariția, manifestarea și dezvoltarea unui fenomen.

Exemple: durata medie de utilizare a patului pe an; durata medie de spitalizare a bolnavului la pat; nota medie la examen a unei grupe de studenți; numărul mediu de vizite la medicul de familie în an la 1 locuitor, care locuiește în teritoriul deservit.

Când este o medie reprezentativă?

- Când calculul mediei se bazează pe folosirea unui număr mare de cazuri individuale:

*Nu putem spune că durata medie de spitalizare a bolnavului la pat în RM a fost în anul 2006 9,1 zile luând în calcul doar trei spitale.*

- Când valorile din care se calculează media sunt omogene:  
*Aceeași afirmație de mai sus poate fi greșită, adică media nu este reprezentativă și în consecință nu este credibilă, dacă durata de tratament a variat de la 7 zile la 77 zile.*
- Când se alege forma de medie care corespunde cel mai bine variației caracteristicii și volumului de date de care se dispune.  
Valorile medii pot fi exprimate prin cifre absolute sau prin indici. De exemplu, în cazul cunoașterii tensiunii arteriale sistolice sau diastolice la un grup de populație, sau în cazul înălțimii sau greutateii medii la diferite vârste și sexe, valorile medii se exprimă în valori absolute (mmHg, cm, kg). În cazul morbidității sau mortalității generale sau specifice pe vârste sau sexe, cauze de boală, etc., valorile medii se exprimă prin cifre relative (la 1000 sau 100000 de locuitori). Fac excepție cazurile foarte rare de boală sau deces care se exprimă mult mai corect prin cifra absolută (de exemplu un caz de difterie sau poliomielită).

Ce tipuri de mărimi medii se pot calcula?

- Media aritmetică
- Media armonică
- Media cronologică
- Media pătratică
- Media geometrică

*Media aritmetică simplă*

Media aritmetică simplă este valoarea medie care se obține din suma valorilor individuale dintr-o colectivitate omogenă, divizată la numărul total al cazurilor studiate.

Media aritmetică simplă se utilizează pentru a stabili valoarea medie în seriile statistice în care fiecărei valori a variantei îi corespunde o singură frecvență.

Formula de calcul:

$$\bar{X}_{as} = \frac{\sum x}{n}$$

**Exemple de calcul:**

Greutatea la naștere la un număr de 6 copii de sex masculin este (în grame): 3000; 2600; 2800; 3100; 3200; 2700.

Aplicând formula de calcul vom obține media greutății, la naștere, la lotul de 6 copii, de 2900 gr:

$$\bar{X}_{as} = \frac{3000 + 2600 + 2800 + 3100 + 3200 + 2700}{6} = 2900$$

Durata de tratament la 8 bolnavi cu aceeași maladie a fost (zile): 10; 12; 15; 18; 20; 22; 24; 27.

Aplicând formula de calcul vom obține media duratei de tratament la lotul de 8 bolnavi de 18,5 zile:

$$\bar{X}_{as} = \frac{10 + 12 + 15 + 18 + 20 + 22 + 24 + 27}{8} = 18,5$$

Cele expuse anterior, cu privire la valorile medii, se pot vedea din formula mediei aritmetice simple: nici un copil nou-născut nu are greutatea medie de 2900 gr.; nici un bolnav nu are durata medie de tratament de 18,5 zile; valorile individuale fiind situate fie sub, fie peste medie și doar pe întreg lotul studiat se obține valoarea medie.

*Media aritmetică ponderată*

Media aritmetică ponderată este valoarea medie care se obține din suma produsului valorilor dintr-o colectivitate omogenă, cu frecvențele corespunzătoare, divizată la numărul total al cazurilor studiate.

Media aritmetică ponderată se utilizează în calcularea valorii medii în cazul în care valorile individuale au frecvențe diferite. În acest caz media aritmetică simplă nu se poate utiliza.

Formula de calcul:

$$\bar{X}_{ap} = \frac{\sum xf}{\sum f(n)}$$

**Exemple de calcul:**

Într-un spital, medicii prezintă următoarea distribuție conform vârstei: 40 persoane de 37 ani; 40 persoane de 38 ani; 10 persoane de 41 ani și 10 persoane de 44 ani.

Aplicând formula de calcul, vom obține vârsta medie a personalului, de 38,5 ani:

$$\bar{X}_{ap} = \frac{40 \times 37 + 40 \times 38 + 10 \times 41 + 10 \times 44}{40 + 40 + 10 + 10} = \frac{3850}{100} = 38,5 \text{ ani}$$

Tabelul 5

Durata de spitalizare a bolnavului la pat (zile):

Zile (x)	10	11	12	13	14	18	21
Bolnavi (f)	2	6	8	12	17	14	6

Aplicând formula de calcul, vom obține durata medie de spitalizare a bolnavului la pat de 14,7 zile:

$$\bar{X}_{ap} = \frac{10 \times 2 + 11 \times 6 + 12 \times 8 + 13 \times 12 + 14 \times 17 + 18 \times 14 + 21 \times 6}{2 + 6 + 8 + 12 + 17 + 14 + 6} = \frac{954}{65} = 14,7$$

Din exemple se poate constata că valoarea mediei aritmetice ponderate se situează în plus sau în minus de partea unde frecvențele sînt mai mari.

În cazul calculării mediei aritmetice ponderate în serii statistice în care valorile variantelor sînt mari (de ordinul sutelor, miilor sau zecilor de mii) și reprezentate de cifre zecimale, iar frecvențele corespunzătoare fiecărei variante sunt de asemenea numeroase, operația devine deosebit de dificilă. În acest caz recurgem la o metodă simplificată a calculării mediei aritmetice ponderate, și anume metoda momentelor.

Formula de calcul:

$$\bar{X}_{ap} = X_o \pm M_1$$

unde:

$\bar{X}$  – media aritmetică ponderată

$X_o$  – media ipotetică

$M_1$  – momentul 1

Momentul 1:

$$M_1 = \frac{\sum df}{\sum f}$$

unde:

d – deviațiile (abaterile) de la medie

f – frecvența

$\Sigma f$  – suma frecvențelor (numărul cazurilor)

Luăm ca exemplu seria de variație a înălțimii nou-născuților de sex masculin, expusă în tab. 6.

Tabelul 6

Înălțimea nou-născuților de sex masculin (cm)

Valorile variantelor înălțimii în cm (x)	Frecvențele (numărul cazurilor)	Produsul valorilor cu frecvențele (xf)	Deviația de la medie (d)	Produsul deviației de la medie cu frecvențele (df)
1	2	3	4	5
46	2	92	-4	-8
47	5	235	-3	-15
48	13	624	-2	-26
49	30	1470	-1	-30 -79
50	89	4450	0	0
51	49	2499	1	49
52	40	2080	2	80
53	20	1060	3	60
54	7	378	4	28
55	3	165	5	15
56	1	56	6	6
57	1	57	7	7 +245
Total:	260 = $\Sigma f$ (n)	13166 = $\Sigma xf$	----	166 = $\Sigma df$

Prima coloană cuprinde seria de variație a valorilor de înălțime (x), iar a doua coloană frecvențele (f) ce corespund fiecăreia din valori. Coloana a 3-a corespunde produsului valorilor cu frecvențele respective, adică produsul datelor corespunzătoare din coloana 1 și a 2-a (xf).

Dacă adunăm coloana a 3-a și împărțim la numărul total de cazuri ( $\Sigma f$ ) vom obține media ponderată, egală cu 50,64 cm. Pentru a substitui acest calcul, care poate deveni destul de dificil în cazul valorilor mari sau cu zecimale, precum și în cazul frecvențelor mari, recurgem la calculul momentului 1.

Acesta se efectuează în felul următor: alegem, în mod arbitrar, ca medie ipotetică, una dintre valorile înălțimii, și anume valoarea varianței cu cele mai multe frecvențe (modulul) – în cazul nostru vom selecta ca medie ipotetică înălțimea de 50 cm. În dreptul acestei valori notăm, în coloana a 4-a, zero corespunzând originii și apoi completăm coloana, în sus, spre valorile mai mici de înălțime, notând cu minus, iar în jos notând cu plus, în progresie aritmetică. Această coloană a deviației de la medie, sau mai bine-zis de la origine o notăm cu d. În coloana a 5-a

trecem produsul dintre coloana a 2-a și a 4-a, adică produsul deviației de la medie cu frecvențele corespunzătoare (df).

Vom obține în coloana a 5-a, în partea de sus, rezultatele cu semnul algebric minus (-), iar în parte de jos sub origine, spre valorile mari ale seriei de variație, rezultatele cu semnul algebric plus (+).

Adunăm separat cele două grupe de produse și apoi scădem suma valorilor cu minus din suma valorilor cu plus, rezultatul împărțindu-l la suma frecvențelor (numărul de cazuri).

Înlocuind simbolurile formulei momentului 1 cu datele rezultate din tab.6 vom obține:

$$M_1 = \frac{-79 + 245}{260} = \frac{166}{260} = 0,64 \text{ cm}$$

Adunăm această valoare a momentului 1 cu media ipotetică aleasă în mod arbitrar și vom obține media aritmetică ponderată:

$$\bar{X}_{ap} = 50,0 \text{ cm} + 0,64 = 50,64 \text{ cm.}$$

Din tabel se poate constata felul în care utilizarea momentului 1 ușurează calcularea mediei. Astfel, în loc de a înmulți coloana 1 cu coloana 2, obținând astfel coloana a 3-a cu cifre mari, care însumate trebuie împărțite la numărul total de cazuri, vom înmulți doar coloana a 4-a cu coloana a 2-a. Înmulțirea se va face cu 4, 3, 2, 1 etc. în loc de 46, 47, 48, 49 etc. Împărțirea la numărul total de cazuri este, de asemenea, ușurată prin metoda momentului 1, căci în loc de a împărți 13166 la 260, ca în calculul simplu al mediei aritmetice ponderate, în cazul momentului 1 împărțim 166 la 260.

Ca medie ipotetică se poate alege oricare din valorile variantelor de înălțime, rezultatul final fiind identic, dar pentru a ușura cât mai mult calculul se alege valoarea care are o frecvență mai mare.

Astfel dacă în exemplul anterior am lua ca medie ipotetică nu 50 cm ci 51 cm am ajunge la aceeași valoare medie, cu singura diferență că în acest caz momentul 1 are o valoare negativă, ce trebuie scăzută din media ipotetică (Tab. 7).

Tabelul 7

Înălțimea nou-născuților de sex masculin (cm).

Valorile variantelor înălțimii în cm (x)	Frecvențele (numărul cazurilor)	Deviația de la medie (d)	Produsul deviației de la medie cu frecvențele (df)
1	2	3	4
46	2	-5	-10
47	5	-4	-20
48	13	-3	-39
49	30	-2	-60
50	89	-1	-89
51	49	0	0
52	40	1	40
53	20	2	40
54	7	3	21
55	3	4	12
56	1	5	5
57	1	6	6
Total:	260 = Σf (n)	-----	-94 = Σdf

Formula momentului 1 se calculează ca și în exemplul anterior:

$$M_1 = \frac{-218 + 124}{260} = \frac{-94}{260} = 0,36 \text{ cm}$$

$$\bar{X}_{ap} = 51,0 \text{ cm} - 0,36 = 50,64 \text{ cm.}$$

Deci rezultatul este identic în toate împrejurările, indiferent de media ipotetică aleasă.

Pentru ușurarea calculului se ia, însă, ca medie ipotetică, totdeauna, valoarea variantei cu cele mai multe frecvențe.

Există cazuri când intervalele valorilor seriei nu sînt de 1, ci de 2 sau 3 unități, sau din contra – de 0,5 sau 0,3.

În aceste cazuri, înainte de a trece la calculul mediei mediei aritmetice ponderate trebuie să înmulțim momentul 1 cu valoarea corespunzătoare a intervalului.

Asupra mediei aritmetice sunt de făcut câteva observații și de subliniat câteva proprietăți:

- 1) Definiția dată mediei aritmetice este valabilă numai dacă valorile individuale înregistrate sunt numerice. Pentru o serie cu valori ne-numerice nu se poate calcula media aritmetică;
- 2) Mărimea mediei aritmetice calculate este unică; o serie nu posedă mai multe medii aritmetice distincte;
- 3) Mărimea mediei aritmetice poate sau nu să coincidă cu vreo valoare individuală înregistrată;
- 4) Media are întotdeauna valoarea cuprinsă între valoarea minimă din serie ( $X_{\min}$ ) și valoarea maximă ( $X_{\max}$ );
- 5) Suma abaterilor valorilor individuale de la media lor este întotdeauna egală cu zero (adică distanțele față de centru se balansează, se compensează reciproc);
- 6) Media aritmetică este legată de toate valorile numerice înregistrate și, în consecință, este sensibilă la prezența valorilor aberante.
- 7) Dacă o serie este alcătuită din mai multe serii componente, pentru care s-au calculat medii parțiale, atunci media întregii serii poate fi calculată ca o medie ponderată din mediile parțiale.

### Media armonică

Media armonică este o medie cu aplicație specială care se determină ca valoarea inversă a mediei aritmetice calculată din inversele valorilor seriei.

*Media armonică simplă*

$$\bar{X}_{arm} = \frac{n}{\sum \frac{1}{X_i}}$$

1. Fie valorile: 1, 5, 10. Care este valoarea medie?

$$\bar{X} = \frac{3}{\frac{1}{1} + \frac{1}{5} + \frac{1}{10}} = \frac{3}{\frac{13}{10}} = 2,3$$

2. Trei investiții produc același venit. Valorile lor sunt următoarele: 5%, 10% și 15% și media de plasare a capitalului pentru toate trei.

$$\bar{X} = \frac{3}{\frac{1}{5} + \frac{1}{10} + \frac{1}{15}} = \frac{3}{\frac{11}{30}} = 8.18\%$$

### Media armonică ponderată

$$\bar{X} = \frac{\sum n}{\sum \frac{1}{X_i} n}$$

Dacă, de exemplu, mortalitatea generală din 3 localități rurale este de 8,0, 9,0 și 10‰, iar numărul de locuitori 4000, 4500 și 7000 și nu cunoaștem cifrele absolute de decedați, media aritmetică simplă este de 9‰ în timp ce media armonică, ponderând valorile relative cu populația la care se referă, va fi 9,12‰:

$$\bar{X} = \frac{4000 + 4500 + 7000}{\frac{1}{8} \times 4000 + \frac{1}{9} \times 4500 + \frac{1}{10} \times 7000} = \frac{15500}{500 + 500 + 700} = 9,12$$

$\Sigma$  – suma

$X_i$  – valorile individuale

$\frac{1}{X_i}$  – inversul valorilor individuale

$n$  – numărul de valori.

În realitate, media armonică se utilizează rar, în special la stabilirea prețurilor medii.

### Media cronologică

Media cronologică se calculează în cazul unei serii de valori, care prezintă variații în timp (lunare, anuale).

Formula de calcul

$$\bar{X} = \frac{\frac{X_1}{2} + X_2 + X_3 + X_4 + \dots + \frac{X_n}{2}}{N-1}$$

unde:

$X(1, 2, 3, 4, n)$  – mărimea valorilor la începutul primei perioade și sfârșitul celor patru perioade;

$N$  – numărul de perioade la care ne referim

Exemplu: numărul de gravide în evidență la medicul de familie la început de an a fost 120, iar la sfârșitul celor patru trimestre – 130, 144, 152, 100. Conform mediei aritmetice simple, ar fi fost 129 gravide, în timp ce media cronologică este de 134, conform formulei:

$$\bar{X} = \frac{\frac{120}{2} + 130 + 144 + 152 + \frac{100}{2}}{5-1} = \frac{536}{4} = 134$$

### Media pătratică

Este tot o medie de calcul cu aplicații speciale și reprezintă valoarea care, înlocuind termenii seriei, nu modifică suma pătratelor lor. Se folosește când fenomenul supus cercetării înregistrează modificări aproximativ în progresie geometrică.

*Media pătratică simplă:*

$$\bar{X} = \sqrt{\frac{\sum x^2}{n}}$$

*Media pătratică ponderată:*

$$\bar{X} = \sqrt{\frac{\sum x^2 f}{\sum f(n)}}$$

### Modulul

Este mărimea medie care corespunde valorii cu cele mai multe frecvențe (în seriile de variație simple). Calculul se efectuează deci simplu, luând valoarea cu frecvență maximă drept valoare medie. Rapiditatea cu care se stabilește modulul este singurul avantaj, deoarece el nu prezintă un etalon precis al valorii medii decât în cazul distribuțiilor normale (simetrice) de frecvențe.

Modulul nu se determină în funcție de toate mărimile valorilor variabilelor, ci de una singură, cea cu frecvență maximă (în repartițiile

unimodale). Totuși el exprimă mărimea cu cea mai mare pondere, deci caracteristica determinantă.

În cazul seriilor de variație grupate modulul se calculează după formula:

$$M_0 = X_{M_0} + h \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

unde:

$X_{M_0}$  – limita inferioară a intervalului modal;

$h$  – mărimea intervalului modal (cu frecvența cea mai mare);

$\Delta_1$  – diferența dintre frecvența intervalului modal și a intervalului precedent;

$\Delta_2$  – diferența dintre frecvența intervalului modal și a intervalului următor;

Se consideră următoarea distribuție (Tab.8).

Tabelul 8

Distribuția persoanelor în funcție de salariu

Salariu	Persoane
400-600	5
600-800	10
800-1000	15
1000-1200	30
1200-1400	20

Intervalul modal este (1.000 – 1.200)

$$M_0 = 1000 + 200 \frac{(30-15)}{(30-15) + (30-20)} = 1000 + 200 \frac{15}{25} = 1120$$

În cazul distribuțiilor de frecvențe moderat asimetrice există următoarele relații matematice aproximative între modul, mediană și medie aritmetică:

$$M_0 = \bar{X} - 3(\bar{X} - Me)$$

$$M_0 = 3Me - 2\bar{X}$$

$$M_0 - Me = 2(Me - \bar{X})$$

În cazul distribuției simetrice de frecvențe cele trei valori medii se suprapun:

$$X = Me = M_0$$



Observații:

- 1) Pe graficul repartiției statistice valoarea modală corespunde punctului în care graficul își atinge maximumul;
- 2) Are avantajul principal față de medie și cuantile că se determină rapid și are o semnificație simplă;
- 3) Există în practică serii cu distribuții multimodale. În astfel de situații se determină mai multe valori modale;

### Mediana (cuantila de ordinul 2)

Ca definiție, mediana, în serii statistice simple, este valoarea acelei variante care împarte în două jumătăți egale numărul variantelor, așezate în ordine crescândă sau descrescândă. În cazul unui număr impar de variante, mediana va corespunde exact valorii de la mijlocul seriei. În seriile cu număr par de variante mediana va corespunde mediei aritmetice simple a celor două valori de la mijlocul seriei.

Exemplu de calcul:

Înălțimea la naștere la un număr de 5 copii a fost: 47, 48, 49, 51, 52 cm. Mediana, corespunzând valorii de la mijlocul seriei va fi 49 cm.

În cazul unei serii de 6 valori, cum ar fi: 47, 48, 49, 51, 52, 53 cm, mediana va fi 50cm (49 + 51/2).

În serii statistice grupate, formula de calcul a medianei este mai complicată, valoarea medianei aflându-se în interiorul intervalului valoric, în care se găsește valoarea frecvenței ce împarte seria în două jumătăți egale.

Formula de calcul:

$$M_e = x_{M_e} + k \frac{\sum f_i - f_{cm}}{f_{M_e}}$$

Unde:

- $x_{M_e}$  – limita inferioară a intervalului median;
- $\sum f$  – frecvențele valorilor variantelor;
- $f_{cm}$  – frecvențele cumulate până la intervalul median;
- $h$  – mărimea intervalului median;
- $f_{M_e}$  – frecvența intervalului median;

### Cum se stabilește intervalul median?

Pe șirul frecvențelor cumulate crescător, intervalul care corespunde primei frecvențe cumulate mai mare decât este intervalul  $\frac{\sum f_i}{2}$  median.

**Cuantilele** sunt indicatorii care descriu anumite poziții particulare din cadrul seriilor de distribuție. Conceptul de cuantilă indică o divizare a distribuției observațiilor într-un număr oarecare de părți. Frecvent se utilizează următoarele cuantile:

- mediana sau cuantila de ordin 2 ( $r = 2$ );
- cuartilele sau cuantilele de ordinul 4 ( $r = 4$ );
- decilele sau cuantilele de ordinul 10 ( $r = 10$ );
- centilele sau cuantilele de ordinul 100 ( $r = 100$ ).

Cuantilele de ordin superior se calculează în cadrul distribuțiilor cu număr mare de grupe sau clase de valori individuale.

## 2.7. Serii cronologice

Un obiectiv important al medicinei și ocrotirii sănătății este studierea sănătății publice, analiza informației privind caracterul și volumul activității instituțiilor medico-sanitare sub aspectul modificărilor lor dinamice. Studiarea acestor modificări este deosebit de importantă pentru prognozarea și planificarea măsurilor cu caracter organizațional, curativ-profilactic etc.

Pentru a analiza modificările dinamice a fenomenelor medico-sociale este necesar de a forma serii cronologice (dinamice), a cunoaște metodele de ajustare și analiză a lor.

### Seria cronologică (serie de timp sau serie dinamică):

- este seria formată din valori omogene comparabile, care caracterizează modificările unui anumit fenomen într-o perioadă de timp.
- vizează măsurarea creșterilor sau descreșterilor de nivel în evoluția unui fenomen.
- fiecare valoare numerică a seriei se numește nivel.
- nivelurile seriei cronologice pot fi prezentate prin valori absolute, relative și medii.

Seriile cronologice se disting printr-o serie de **particularități**, trăsături specifice ei, între care menționăm:

a) **variabilitatea** termenilor SCR – Între valorile individuale care compun seria cronologică există diferențe de mărime explicate prin

acțiunea comună a factorilor esențiali și întâmplători. Gradul de variabilitate a termenilor seriei cronologice depinde de forța cu care factorii aleatori produc abateri, dar și de tendința de variație impusă de factorii cu acțiune sistematică.

b) **omogenitatea** termenilor unei SCR – seriile cronologice sunt omogene deoarece termenii seriei au în comun categoria economică sau socială pe care o reprezintă în momente sau intervale succesive de timp. **Omogenitatea** valorilor seriei este dată de faptul că acestea sunt supuse acțiunii sistematice a acelorași factori esențiali, iar termenii seriei cronologice sunt obținuți prin aceeași metodologie de calcul și folosesc aceeași unitate de măsură.

c) **periodicitatea** termenilor unei SCR – o caracteristică specifică seriilor cronologice. Această trăsătură exprimă continuitatea datelor din punct de vedere al variației timpului. Termenii seriei reprezintă valori ale unui fenomen dinamic, înregistrate la momente sau intervale de timp de regulă egale, astfel încât să se asigure continuitatea seriei. În funcție de scopul concret al analizei efectuate, de natura fenomenului înregistrat și de posibilitățile de obținere a datelor, unitățile de timp pot fi mai mici sau mai mari: minut, oră, zi, săptămână, decadă, lună, trimestru, semestru, an, deceniu, secol.

d) **interdependența** în timp a termenilor unei SCR – este determinată de modalitatea de construire a acestora prin înregistrarea nivelurilor succesive ale unui fenomen pentru aceeași unitate statistică precizată. Din această cauză, orice termen al seriei depinde de nivelurile precedente și influențează mărimile următoare ale termenilor seriei.

Având în vedere aceste particularități ale seriilor cronologice, analiza lor trebuie precedată de verificarea **comparabilității** valorilor individuale înregistrate pentru fenomenul analizat. Pentru a asigura comparabilitatea termenilor seriei cronologice este necesar ca componența seriei să fie identică pentru întreaga perioadă de timp, valorile seriei să fie exprimate în aceleași unități de măsură, iar intervalele de timp între valori să fie egale.

#### Tipuri de serii cronologice:

1. În funcție de modul de definire a timpului deosebim SCR de moment și SCR de interval.

#### Seriile cronologice de momente:

- sunt formate din mărimi care se referă la anumite momente de timp (sfârșitul sau începutul anului, trimestrului, lunii etc.).

- fiecare valoare individuală caracterizează numeric nivelul la care a ajuns fenomenul analizat într-un moment dat.

*De exemplu:* numărul de gravide aflate sub supraveghere la începutul anului; numărul de născuți vii la sfârșitul fiecărui trimestru; numărul populației la o anumită dată, numărul copiilor bolnavi aflați în evidență la finele anului (Tab.9).

Tabelul 9

Număr copii bolnavi aflați în evidență la finele anului, RM

2001	2002	2003	2004	2005	2006
155204	157257	155036	128595	147247	143146

#### Observație:

- **nu permit cumularea valorilor termenilor, deoarece acestea reflectă, în mod repetat, elementele care coexistă în momente diferite de timp.**

*Exemplu:* numărul salariaților unui spital înregistrat la sfârșitul fiecărei luni se regăsește în numărul salariaților înregistrat la sfârșitul anului.

Când intervalele dintre două momente succesive au lungime egală, atunci vom avea o SCR de momente cu intervale egale între momente, iar atunci când intervalele dintre două momente vecine au lungime neegală avem o SCR de momente, cu intervale neegale între momente.

#### Serii cronologice de intervale:

- sunt formate din mărimi care caracterizează fenomenul într-un interval de timp (zi, săptămână, lună, trimestru, an etc.);
- fiecare valoare individuală reprezintă rezultatul unui proces care se desfășoară pe un interval de timp.

*De exemplu:* dinamica anuală a numărului de medici; evoluția lunară a numărului de cazuri îmbolnăvire IRVA, dinamica anuală a natalității (Tab.10).

Tabelul 10

Număr copii născuți vii, mun. Chișinău, aa. 2002-2006.

2002	2003	2004	2005	2006
6924	7143	8296	8364	8626

**Observație:**

- **permite însumarea valorilor, obținându-se astfel un indicator totalizator pentru întreaga perioadă de analiză.**

**Exemplu:** prin însumarea numărului de internări zilnice se obține cifra de internări lunară;

Alegerea perioadei de timp pentru seria cronologică de interval este determinată într-o oarecare măsură de variabilitatea fenomenului. Cu cât mai lent se modifică fenomenul în timp, cu atât mai mari pot fi perioadele de supraveghere.

2. În funcție de modul de exprimare a termenilor seriei deosebit seriei cronologice formate din valori absolute, relative sau medii.

**Seriile cronologice formate din valori absolute** reprezintă situația cea mai frecvent întâlnită. Fiecare termen al seriei este în acest caz o mărime absolută exprimată în unități concrete de măsură. De exemplu: număr de populație, număr de paturi, număr de medici, număr de nou-născuți, număr de decedați, număr de nașteri, număr avorturi, nr. anomalii fetale, depistate ecografic, cheltuielile anuale în IMSP etc.

**Seriile cronologice formate din valori relative.** Termenii acestor serii pot fi reprezentați prin rate, proporții și raport. De exemplu: natalitatea, mortalitatea, morbiditatea, invaliditatea primară, asigurarea populației cu medici, paturi, ponderea populației în vârstă de peste 55 de ani. Baza de raportare trebuie să fie întotdeauna precizată.

**Seriile cronologice formate din valori medii.** De exemplu: numărul mediu de paturi, durata medie de spitalizare, durata medie de utilizare a patului pe an, salariul mediu al medicilor.

3. În funcție de numărul termenilor seriei deosebit seriei cronologice de lungime mică, medie, mare.

**Reprezentări grafice ale seriilor cronologice**

- a) **CRONOGRAMA (historiograma)** – este, așa cum îi arată și numele, reprezentarea grafică tipică, specifică a SCR. Ea se trasează într-un sistem de axe rectangulare, de obicei în cadranul întâi al acestuia. Pe cele două axe se vor reprezenta: timpul – pe abscisă (se marchează momentele sau intervalele), iar termenii SCR – pe ordonată (Fig.1).

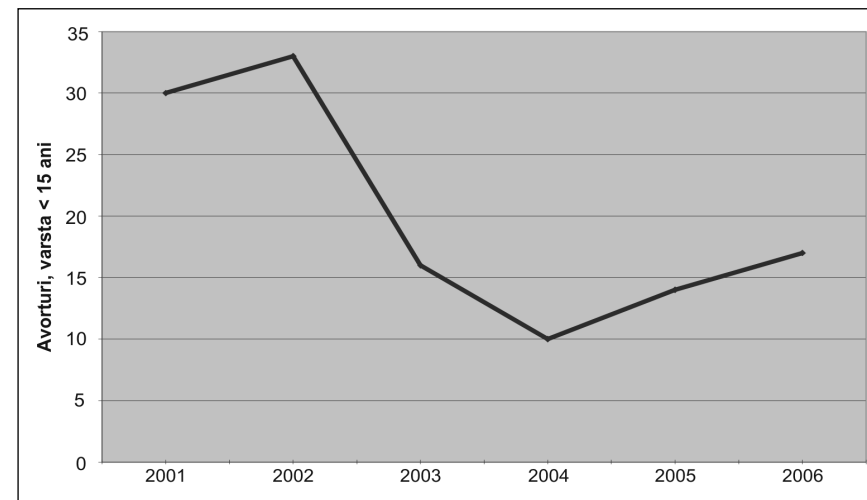


Fig.1. Numărul de avorturi la femeile în vârstă până la 15 ani în RM, aa.2001-2006 (abs.)

- b) **DIAGRAMA PRIN COLOANE** – în care timpul se reprezintă pe abscisă, iar termenii SCR pe ordonată (Fig.2).

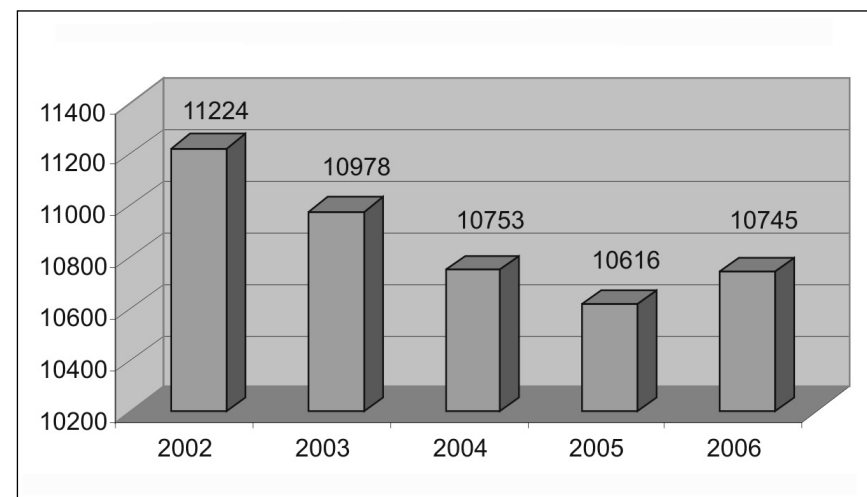


Fig.2. Numărul de medici în Republica Moldova, aa. 2002-2006 (abs.)

- c) **DIAGRAMA PRIN BENZI** – este recomandată a se folosi atunci când se reprezintă (simultan) termenii unor SCR, termeni care constituie niște indicatori strâns legați între ei.
- d) **DIAGrame POLARE** (numite și diagrame radiale sau diagrame în spirală) se construiesc cu ajutorul rețelelor radiale și se utilizează în special în reprezentarea SCR afectate de fluctuații sezoniere.

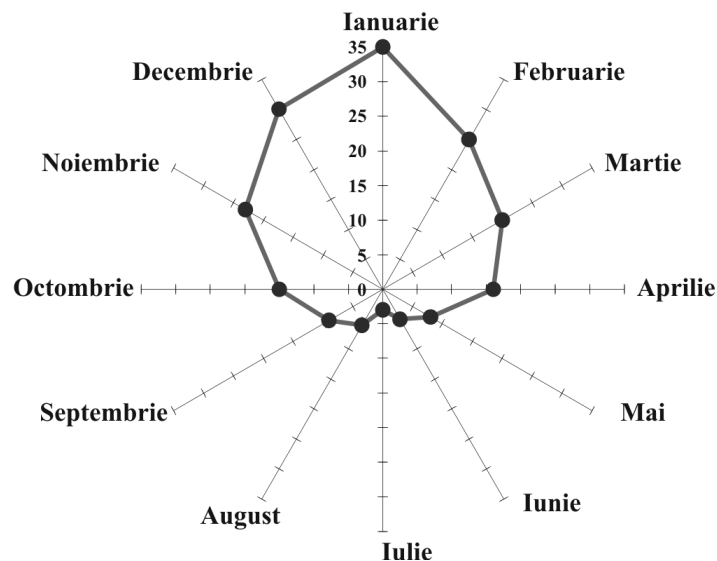


Fig.3. Numărul cazuri IRVA în localitatea N pe parcursul a. 2006

### Indicatorii seriilor cronologice

Pentru caracterizarea unei SCR, se calculează, pe baza termenilor acesteia, un sistem de indicatori statistici, analitici și sintetici care, după modul de calcul și exprimare, pot fi structurați astfel:

- indicatori absoluți;
- indicatori relativi;
- indicatori medii.

Atunci când compararea se face cu primul termen al seriei ( $y_1$ ) vom vorbi de indicatori cu bază fixă, iar atunci când compararea unui termen ( $y_t$ ) se face cu termenul imediat anterior ( $y_{t-1}$ ), vom vorbi de indicatori cu bază în lanț (mobilă).

### Indicatori absoluți

Indicatorii absoluți ai unei serii cronologice de intervale exprimă nivelul, volumul agregat și modificările (în mărime absolută) fenomenului analizat în perioade diferite de timp. Indicatorii absoluți se exprimă în unitatea de măsură a caracteristicii analizate (în unități fizice, valorice, procente etc.).

- 1) Valorile individuale absolute ale caracteristicii redau nivelul fenomenului analizat în fiecare interval de timp.
- 2) Volumul agregat (nivelul totalizat) reprezintă suma termenilor seriei cronologice de intervale.
- 3) Modificarea absolută (sporul sau scăderea absolută) –  $\Delta$ , reflectă creșterea sau descreșterea absolută (în unități concrete de măsură) a valorilor individuale ale fenomenului analizat, de la o perioadă de timp la alta. Se calculează ca diferență între doi termeni ai seriei. În funcție de perioada aleasă ca bază de comparație (constantă sau variabilă), există două forme ale acestui indicator:
  - modificarea absolută cu bază fixă reprezintă distanța (diferența) fiecărui termen al seriei față de o perioadă fixă de referință.
  - modificarea absolută cu bază mobilă se calculează ca diferență între doi termeni succesivi ai seriei cronologice.

Se exprimă în unitățile de măsură ale caracteristicii. Valorile pozitive ale acestor indicatori semnifică sporuri (creșteri, față de perioada aleasă ca bază de comparație), iar valorile negative – scăderi (deficit).

În cazul modificării absolute cu bază fixă este importantă alegerea unei baze de comparație convenabile, reprezentative pentru fenomenul dat și care să nu fie influențată de variații conjuncturale majore. Frecvent, se alege ca bază de comparație primul termen al seriei (începutul perioadei de timp analizate) sau ultimul termen al perioadei anterioare.

### Indicatori relativi

- 1) Indicele de dinamică (indice de modificare, ritm de creștere sau de scădere) – I. Este indicele care arată de câte ori (de cât la sută) s-a modificat mărimea unui fenomen în timp. Se calculează prin raportarea termenului comparat la termenul bază de comparație.
  - indicele cu bază fixă se calculează ca raportul simplu sau procentual al nivelului curent la nivelul ales bază de comparație (nivelul inițial).
  - indicele cu bază mobilă se calculează ca raportul simplu sau procentual al nivelului curent la nivelul precedent.

Se exprimă în unități sau procente. Valori mai mari de 1 sau 100% ale acestui indicator arată creșteri față de perioada bază de comparație. Valorile sub 1 sau 100% semnifică scădere, reducere.

2) Ritmul de dinamică (ritmul sporului) – R. Arată cu cât s-a modificat procentual (a crescut sau a scăzut) mărimea fenomenului într-o anumită perioadă de timp față de o perioadă de referință fixă sau mobilă. Se determină scăzând 100% din indicele de dinamică corespunzător (cu bază fixă sau mobilă). Poate fi calculat și prin următoarele modalități:

- cu bază fixă se calculează ca raportul procentual al sporului absolut cu bază fixă la nivelul ales bază de comparație (nivelul inițial).
- cu bază mobilă se calculează ca raportul procentual al sporului absolut cu bază mobilă la nivelul precedent.

3) Valoarea absolută a unui procent din ritmul de dinamică (de spor) – A. Arată mărimea absolută a modificării ce revine pe un procent din ritmul dinamicii sau exprimă câte unități de măsură revin unui procent din ritmul dinamicii. Se calculează sub forma unui raport între modificarea absolută și ritmul modificării și se exprimă în unitatea de măsură a caracteristicii.

Variante de calcul:

- cu bază fixă se calculează ca raportul modificării absolute cu bază fixă la ritmul dinamicii cu bază fixă. **Are aceleași valoare pentru toată perioada analizată.**
- cu bază mobilă se calculează ca raportul modificării absolute cu bază mobilă la ritmul dinamicii cu bază mobilă.

#### Indicatori medii

1) Nivelul mediu, calculat ca medie a valorilor seriei. Calculul se justifică numai dacă nivelurile seriei sunt omogene.

- pentru o serie cronologică de intervale nivelul mediu se află calculând media aritmetică simplă a valorilor seriei.
- pentru o serie cronologică de momente nivelul mediu se află calculând media cronologică simplă pentru momente egal distanțate:

$$y = \frac{\frac{y_1}{2} + y_2 + \dots + y_{n-1} + \frac{y_n}{2}}{n-1}$$

ponderată pentru momente inegal distanțate:

$$y = \frac{y_1 \times \frac{t_1}{2} + y_2 \times \frac{t_1+t_2}{2} + \dots + y_n \times \frac{t_{n-1}}{2}}{t_1 + t_2 + \dots + t_{n-1}}$$

2) Modificarea medie absolută este media aritmetică simplă a modificărilor absolute cu bază mobilă –  $y_n - y_1/n-1$ . Indicatorul arată cu cât crește/scade fenomenul în medie (în valoare absolută) de la o perioadă de timp la alta.

**Calcularea acestui indicator are sens atunci când modificările absolute cu bază în lanț nu diferă prea mult ca mărime.** Modificarea medie absolută poartă numele de spor mediu, dacă este calculată pentru un fenomen cu tendință de creștere. În caz contrar vorbim despre scădere medie.

3) Indicele mediu de dinamică se calculează ca medie geometrică simplă a indicilor de dinamică cu bază mobilă. Arată de câte ori s-a modificat (a crescut sau a scăzut) în medie fenomenul analizat pe întreaga perioadă luată în calcul. Valoarea rezultată din calcul este semnificativă îndeosebi pentru fenomenele care evoluează în progresie geometrică (indicii cu bază mobilă au valori apropiate între ele). Valori mai mari de 100% ale acestui indicator arată tendința de creștere a fenomenului analizat. Valori mai mici de 100% corespund unei scăderi pe ansamblul perioadei considerate.

4) Ritmul mediu de dinamică arată cu câte procente se modifică în medie fenomenul pe întreaga perioadă analizată.

Ultimii doi indicatori nu sunt reprezentativi pentru seriile cronologice cu un număr mare de termeni și cu oscilații mari.

#### Ajustarea seriilor cronologice

Specifică seriilor cronologice este variabilitatea mare a termenilor. Aceste variații sunt produse de factori esențiali și întâmplători. Acțiunea factorilor esențiali determină tendința majoră de evoluție în timp amăririlor înregistrate de fenomenul analizat. Această tendință (*trend*) interferează cu cauzele neesențiale, întâmplătoare, efectul obținut fiind valorile reale ale fenomenului.

Scopul ajustării seriilor cronologice îl reprezintă evidențierea factorilor esențiali, cu acțiune sistematică, care urmăresc o legitate matematică de evoluție.

**Ajustarea este operația de înlocuire a termenilor reali ai seriei cronologice cu termeni teoretici care exprimă o anumită legitate matematică de evoluție a fenomenului considerat.**

Pentru seriile cronologice, această legitate de evoluție se realizează în funcție de timp. Întrucât abaterea termenilor reali de la cei teoretici calculați este efectul cauzelor neesențiale, întâmplătoare, prin ajustare se evidențiază mai bine tendința de evoluție în timp a fenomenului.

Există mai multe procedee de ajustare:

- ajustarea prin metoda grafică;
- ajustarea prin metoda mediilor mobile;
- ajustarea prin metoda modificării medii absolute;
- ajustarea prin metoda indicelui mediu de dinamică;
- ajustarea prin metode analitice.

Primele patru procedee formează grupul metodelor mecanice de ajustare.

1) Metoda grafică de ajustare constă în reprezentarea grafică a seriei (cronograma), pe axa Ox (abscisa) fiind trecute momentele sau intervalele succesive de timp, iar pe axa Oy (ordonata) înscriindu-se valorile numerice ale termenilor seriei. Se construiește pe același grafic o dreaptă sau curbă care să unească cele două puncte extreme ale seriei cronologice astfel încât să prezinte abateri minime față de poziția valorilor reale de pe grafic. Forma curbei astfel trasate indică legitatea matematică, forma de evoluție a fenomenului, după o dreaptă sau o funcție curbilinie.

**Metoda este simplă și rapidă, dar există pericolul interpretării subiective a graficului.**

**Metoda grafică precede obligatoriu aplicarea metodelor analitice de ajustare.**

2) Metoda mediilor mobile constă în înlocuirea termenilor reali ai SCR cu valori teoretice, numite **medii mobile** (medii glisante sau alunecătoare). Mediile mobile se calculează ca medii aritmetice parțiale dintr-un anumit număr de termeni succesivi ai seriei. Acest număr depinde de periodicitatea oscilațiilor și este ales astfel încât fiecare medie să cuprindă toți termenii la care se manifestă o oscilație completă. Aceasta poate fi evidențiată cu ajutorul reprezentării grafice, observând mărimea distanței medii dintre punctele de inflexiune ale graficului.

Mediile mobile asigură compensarea abaterilor, a oscilațiilor periodice. Noua serie obținută prin ajustare are o variație lină, continuă, evidențiind tendința de evoluție a fenomenului (trendul), independent de acțiunea factorilor sezonieri.

Mediile mobile pot fi calculate dintr-un număr impar sau par de termeni. Dacă media mobilă se calculează dintr-un număr impar (de exemplu  $k = 3$ ) de termeni, schema de calcul este următoarea:  $MM_1 = y_1 + y_2 + y_3 : 3$ ;  $MM_2 = y_2 + y_3 + y_4 : 3$ ;  $MM_3 = y_3 + y_4 + y_5 : 3$ ;  $MM_{n-2} = y_{n-2} + y_{n-1} + y_n : 3$ ;

Numărul de medii mobile obținut este mai mic decât numărul de termeni reali ai seriei. Primul și ultimul termen real nu vor avea corespondent o valoare ajustată, adică o medie mobilă.

Pentru cazul general, prin această metodă se pierd  $k-1$  termeni ai seriei (la începutul și la sfârșitul șirului), acesta fiind principalul dezavantaj al metodei.

**Exemplu pentru  $k=3$  (Tab.11).**

Tabelul 11

Avorturi la vârsta până la 15 ani, RM, aa.1995 – 2002

Anii	1995	1996	1997	1998	1999	2000	2001	2002
Termenul	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_{n-2}$	$y_{n-1}$	$y_n$
Nr.avorturi	115	75	47	49	57	33	30	33
Media mobilă	----	79	57	51	46	40	32	----

Dacă mediile mobile se calculează din număr par de termeni, calculul se realizează în două faze:

- se obțin medii mobile provizorii, care se plasează între termenii seriei reale;
- se determină medii mobile definitive (centrate) din câte două medii mobile provizorii succesive care se plasează în dreptul termenilor reali (pe care îi înlocuiesc).

Și în acest caz, numărul de medii mobile obținute este mai mic decât numărul de termeni reali ai seriei. Numărul termenilor reali care se pierd este în acest caz  $k$ . În prima fază se pierd  $k - 1$  termeni, iar în a doua fază un termen. Prima medie mobilă centrată se va plasa în dreptul celui de-al  $k + 2 : 2$ -lea termen al seriei.

Pierderea de informație produsă de termenii lipsă afectează concluziile analizei, în special în cazul unui număr redus de observații.

**Exemplu pentru k=4 (Tab.12).**

Tabelul 12

Avorturi la vârsta până la 15 ani, RM, aa.1995 – 2002

	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$	$y_7$	$y_8$
Anii	1995	1996	1997	1998	1999	2000	2001	2002
Nr. avorturi	115	75	47	49	57	33	30	33
Medii mobile parțiale			71,5	57	46,5	42,3	38,3	
Medii mobile centrate	---	---	64	52	44	40	---	---

## 3) Ajustarea prin metoda modificării medii absolute.

Această metodă este indicată atunci când seria cronologică prezintă tendința de creștere sub forma unei progresii aritmetice, situație evidențiată prin valorile relativ apropiate ale modificărilor absolute cu bază mobilă.

Termenii ajustați se determină cu relația:

$$Y_i = y_1 + t_i \times \Delta$$

unde:  $Y_i$  – valorile ajustate, care înlocuiesc valorile reale;

$y_1$  – primul termen al seriei cronologice reale (sau un alt termen luat ca bază de ajustare);

$\Delta$  – sporul mediu (modificarea absolută medie);

$t_i$  – variația timpului:  $t_1 = 0, t_2 = 1, t_3 = 2, \dots, t_n = n-1$  dacă se ia ca bază primul termen al seriei.

Primul termen ajustat este egal cu primul termen al seriei reale iar ultimul termen ajustat este egal cu ultima valoare a seriei reale. Metoda înseamnă netezirea evoluției fenomenului după a linie dreaptă, care unește primul și ultimul termen observat al seriei cronologice.

**Exemplu:**  $\Delta = 17$  (Tab.13).

Tabelul 13

Cazuri de narcomanie în localitatea N, aa. t (reale și ajustate)

	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$
Anii	$t_0$	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$
Prevalența narcomanie, real	122	140	168	176	192	206
Prevalența narcomanie, ajustat	122	139	156	173	190	206

## 4) Ajustarea pe baza indicelui mediu de dinamică

Această metodă este recomandabilă în situațiile în care seria cronologică are forma unei progresii geometrice. În acest caz indicii de dinamică cu bază mobilă au valori relativ apropiate.

Primul și ultimul termen ajustat sunt egali cu termenii corespunzători din seria reală.

Metoda modificării absolute medii și metoda indicelui mediu de dinamică sunt simple și rapide, dar nu iau în calcul toate valorile absolute ale termenilor seriei.

## 5) Ajustarea prin metode analitice

Metodele analitice de estimare a tendinței se bazează pe folosirea funcțiilor matematice. Alegerea funcției de ajustare se face pe baza analizei graficului și a indicatorilor seriei cronologice. Situațiile cele mai frecvent întâlnite sunt:

- fenomenul evoluează după o *funcție liniară* atunci când graficul arată o tendință de creștere absolută constantă și modificările cu bază mobilă au valori apropiate;
- fenomenul evoluează după o *funcție exponențială* atunci când graficul arată o tendință de creștere relativă constantă și se obțin valori apropiate ale indicilor cu bază în lanț;
- fenomenul evoluează după o *parabolă* atunci când graficul are punct de maxim sau de minim iar diferențele dintre modificările succesive cu bază mobilă (numite modificări cu bază în lanț de ordinul doi) au valori apropiate; frecvent, pe grafic, se evidențiază numai fragmente de parabolă.

După ce se alege forma cea mai potrivită pentru funcția de ajustare, se determină parametrii prin intermediul metodei celor mai mici pătrate. Această metodă pornește de la condiția minimizării pătratelor abaterilor valorilor ajustate ( $Y_i$ ) de la valorile reale ( $y_i$ ).

**EXTRAPOLAREA**

Un obiectiv important al analizei seriilor cronologice îl reprezintă estimarea evoluției probabile în viitor a fenomenului analizat. Extrapolarea reprezintă o prelungire a seriei cronologice în viitor, pe baza trendului observat din analiza perioadei anterioare. Mărimile obținute prin extrapolare sunt valori probabile, orientative. Nu se poate face o predicție exactă a viitorului din mai multe motive:

- pe lângă trendul pe baza căruia se face previziunea acționează și factori aleatori, întâmplători, care influențează nivelul real al fenomenului analizat;
- factorii de influență evidențiați prin analiza seriei cronologice își pot modifica acțiunea în viitor;

- există anumite limite (minime sau maxime) în evoluția fenomenelor. Aceste restricții impun o analiză critică, calitativă a rezultatelor obținute prin extrapolare.

**Metodele de extrapolare** sunt similare celor utilizate pentru estimarea trendului. Diferența constă în perioada de timp implicată în calcule.

Astfel, dacă analiza seriei cronologice a relevat o tendință de creștere constantă, extrapolarea se poate face prin **metoda modificării absolute medii**.

Atunci când evoluția seriei cronologice tinde către o exponențială, extrapolarea se poate face pe baza **indicelui mediu de dinamică**

Atunci când evoluția fenomenului a fost analizată cu ajutorul **metodelor analitice** și s-a constatat că reflectă o anumită funcție matematică, extrapolarea se poate face utilizând forma funcției de ajustare.

Atunci când în evoluția fenomenului considerat s-a constatat o variație sezonieră, valorile extrapolate vor cuprinde și această tendință, prin însumarea componentei sezoniere (în cazul modelului aditiv) sau prin înmulțire (pentru modelul multiplicativ).

**Exemplu:**

Tabelul 14

Evoluția numărului de avorturi la vârsta până la 15 ani, RM, aa. 1991-2003

Anul	Număr avorturi	Modificare absolută		Ritm dinamică		Indice dinamică		Val.absolută 1% ritm dinamică	
		fix	mobil	fix	mobil	fix	mobil	fix	mobil
1991	58	0	-	100	-	100	-	0	-
1992	72	14	14	24	24	124	124	0,58	0,58
1993	76	18	4	31	6	131	106	0,58	0,66
1994	81	23	5	40	7	140	107	0,58	0,71
1995	115	57	34	98	42	198	142	0,58	0,81
1996	75	17	-40	29	-35	129	65	0,59	1,14
1997	47	-11	-28	-19	-37	81	63	0,58	0,75
1998	48	-10	1	-17	2	83	102	0,59	0,5
1999	56	-2	8	-3	17	97	117	0,66	0,47
2000	32	-26	-24	-45	-43	55	57	0,58	0,55
2001	30	-28	-2	-48	-6	52	94	0,58	0,33
2002	33	-25	3	-43	10	57	110	0,58	0,3
2003	16	-42	-17	-72	-52	28	48	0,58	0,29

## 2.8. Standardizarea

Importanță deosebită pentru asigurarea comparabilității datelor statistice au metodele de standardizare a valorilor relative și medii.

Regula de bază a statisticii – „compară comparabilul” presupune compararea indicatorilor obținuți în totalități statistice omogene. Însă sunt frecvente cazurile când trebuie de comparat indicatori, asupra mărimii cărora a influențat neomogenitatea totalităților studiate.

În multe investigații socio-medice, precum și clinice, este exclusă posibilitatea de a opera cu colectivități omogene, producătoare de fenomene care se cer comparate. Mai frecvent aceasta se referă la compararea indicatorilor sănătății populației (natalitate, mortalitate generală, morbiditate) în țări, regiuni, raioane, orașe cu structură diferită a populației pe vârstă, sex; compararea activității instituțiilor medico-sanitare cu structură diferită a contingentelor de bolnavi.

### Exemplul 1.

- Pentru a compara letalitatea în două spitale și a face concluzie privitor la cauzele diferenței acestor indicatori este necesar mai întâi să analizăm dacă bolnavii în aceste spitale au fost omogeni nozologic. Evident, indicatorul letalității va fi mai mare în spitalul unde au fost internați mai mulți bolnavi cu forme grave ale maladiilor. Structura neomogenă a contingentelor de bolnavi după starea sănătății, vârstă, sex și alte caractere face imposibilă concluzia corectă privind cauzele diferenței mărimii indicatorilor letalității în aceste spitale.

### Exemplul 2.

- Se compară indicii mortalității generale în două raioane. Calculul simplu arată că în raionul A nivelul mortalității generale este 10‰, iar în raionul B – 8,1‰. Însă, până a formula concluzia finală referitor la factorii care au influențat diferența dintre indicatori, trebuie analizate totalitățile comparate. Se acordă atenție structurii pe vârste a populației din raioanele corespunzătoare, din motivul că structura populației pe vârste incontestabil influențează indicatorul mortalității generale. Cu cât sunt mai multe persoane vârstnice, cu atât mai mare poate fi nivelul mortalității. În cazul nostru, în raionul A, ponderea persoanelor vârsta cărora a depășit 60 de ani este de 30%, iar în raionul B – 16%. Prin urmare, în raioanele comparate structura populației pe vârste este diferită și, în acest



caz, indicatorii mortalității generale (10‰ și 8,1‰) nu pot fi comparați în formă brută.

Pentru a compara indicatorii menționați se va utiliza metoda de standardizare.

**Standardizare – metodă de calcul a indicatorilor standardizați (ipotetici), care substituie valorile relative sau medii care sunt incomparabile din cauza neomogenității structurale a totalităților comparate.**

Există trei metode de standardizare:

- metoda directă.
- metoda indirectă;
- metoda tangențială

Cea mai frecvent utilizată este metoda directă de standardizare. În cazul utilizării acestei metode drept standard este structura populației, care se consideră că este identică în ambele totalități comparate. Esența metodei este în eliminarea factorilor ce influențează mărimea indicatorilor obținuți.

Etapele metodei directe de standardizare:

1. Calculul indicatorilor intensivi (rata) sau medii speciali (pentru fiecare grupă – pe sexe, vârstă, durata de spitalizare, termenul de internare etc.) și generali pentru fiecare totalitate.
2. Selectarea și calculul standardului.
3. Calculul “valorilor așteptate” pentru fiecare grupă de standard
4. Calculul indicatorilor standardizați
5. Compararea totalităților după indicatorii intensivi sau medii generali și indicatorii standardizați. Concluzii.

Calculule efectuate, pentru o mai bună intuiuitate, se vor prezenta în formă de tabel (Tab.15)

Tabelul 15

Repartizarea bolnavilor și decedaților conform profilului secției în spitalele A și B (datele sunt convenționale)

Secția	Spitalul A		Spitalul B		Standardul	Spitalul A		Spitalul B	
	bolnavi	decedați	bolnavi	decedați		letalitatea	valorile așteptate	letalitatea	valorile așteptate
Terapie	600	30	200	12	800	5,0	40	6,0	48
Chirurgie	300	6	700	21	1000	2,0	20	3,0	30
Boli infecțioase	100	4	100	5	200	4,0	8	5,0	10
Total	1000	40	1000	38	2000	4,0	68	3,8	88
Letalitatea %	4,0		3,8		Indicatorii standardizați	3,4		4,4	

#### Exemplu:

**Etapa I – Calculul indicilor intensivi.** Determinăm rata decedaților, în dependență de profilul secției, în ambele spitale. Spitalul A, secția terapie –  $(30/600) \times 100$ ; secția chirurgie –  $(6/300) \times 100$ ; boli infecțioase –  $(4/100) \times 100$ . Analogic se fac calculele și pentru spitalul B. Ulterior determinăm indicatorul general al letalității pentru spitalul A –  $(40/1000) \times 100$  și pentru spitalul B –  $(38/1000) \times 100$ .

**Etapa II – Selectarea și calculul standardului.** Drept standard este considerată acea componentă a totalităților, care reflectă toate particularitățile totalităților comparate.

!!!Standard poate fi: componenta unei totalități comparate; suma totalităților comparate; semisuma totalităților comparate; o valoare empirică.

În exemplul dat drept standard pe secția a fost luat numărul de bolnavi din secțiile respective a fiecărui spital (coloana 1+coloana3), iar standard general – suma bolnavilor din ambele spitale (1000+1000).

**Etapa III – Calculul “valorilor așteptate” pentru fiecare grupă de standard.** Calculul se face în modul următor: în spitalul A, secția terapie, au decedat 5% bolnavi. Câți decedați vor fi la 800 de bolnavi nivelul letalității va fi același?

$$5\% - 100$$

$$x\% - 800$$

$$x = (5 \times 800) / 100$$

Astfel, 40 este numărul așteptat de decedați în secția terapie la 800 de bolnavi. Analogic se fac calculele pentru secțiile chirurgie și boli infecțioase a spitalului A și secțiile spitalului B.

**Etapa IV – Calculul indicatorilor standardizați.** Indicatorul standardizat este raportul procentual al sumei valorilor așteptate din fiecare totalitate la standardul general.

$$\text{Spitalul A} - (68/2000) \times 100 = 3,4\%$$

$$\text{Spitalul B} - (88/2000) \times 100 = 4,4\%$$

**Etapa V – Compararea indicatorilor** (Tab.16).

Tabelul 16

Compararea indicatorilor intensive și standardizați

Indicatori	Spitalul A	Spitalul B	Comparare
Intensivi	4,0	3,8	A>B
Standardizați	3,4	4,4	A<B

### Concluzii:

1. Compararea indicatorilor standardizați, calculați pentru spitalele A și B, ne permite să facem concluzia următoare: dacă profilul secției în spitalele A și B ar fi fost identic, letalitatea în spitalul A ar fi fost mai mică decât în spitalul B.
2. Compararea indicatorilor intensivi a oferit rezultate inverse, din motivul că mărimea acestor indicatori a fost influențată de profilul secției.

**Metoda tangențială de standardizare** se utilizează în cazurile când în totalitățile comparate lipsesc date privind bolnavii, decedații etc. Drept standard este luat un indicator cunoscut: exemplu – indicatorii mortalității generale, mortalității pe vârste, letalității etc. din literatura de specialitate, statistică oficială și cu acesta sunt comparați indicatorii primari.

**Metoda indirectă de standardizare** se utilizează în cazurile când indicatorii necesari pentru comparare și analiză lipsesc. Acestea sunt „reconstruiți” în mod invers, spre exemplu în baza datelor despre mortalitate, morbiditate, letalitate, care trebuie redați cât mai obiectiv în corespundere cu informația cunoscută despre numărul și structura populației.

**!!! Indicatorii standardizați sunt convenționali (ipotetici) și nu oferă informație despre mărimea reală a fenomenului studiat.**

**Fără indicatorii intensivi sau valorile medii nu sunt valabili.**

*Avantajul acestor indicatori este faptul că permit de a elimina influența asupra mărimii indicatorilor intensivi sau valorilor medii a neomogenității structurale a totalităților comparate și oferă răspuns la întrebarea „Care era să fie mărimea fenomenului studiat dacă totalitățile comparate ar fi fost identice”.*

## 2.9. Corelația și regresia

### 2.9.1. Corelația

În medicină și biologie, ca de altfel în toate domeniile de activitate, există o interdependență între fenomene. Apariția și evoluția unui fenomen este în strânsă legătură cu o serie de alte fenomene sau factori care intervin în determinarea sau favorizarea acestuia. Corelația este o metodă care ne permite să cunoaștem fenomenele din natură și societate sub raportul interferenței lor, a conexiunilor în care se găsesc.

În statistică, pentru studierea legăturilor multiple ce au loc între diferite fenomene, se folosește noțiunea de funcție  $f$ , care constă în faptul că fiecărei valori a variabilei independente (X), numită argument, îi corespunde valoarea altei variabile numită funcție (Y).

În general, deosebim două tipuri de corelații: **corelații funcționale** sau matematice și **corelații statistice** sau stohastice (întâmplătoare).

Corelațiile funcționale sunt perfecte, rigide, exprimând legătura de la cauză la efect între fenomene. Asemenea corelații sunt studiate în cadrul științelor exacte, unde având de-a face cu fenomene simple, legătura de la cauză la efect se evidențiază mai ușor și se exprimă sub formă de lege. În cazul corelației funcționale unei valori determinate a unei variabile independente X (argument) îi corespunde strict o valoare a variabilei dependente Y (funcție).

Corelațiile statistice sunt mai puțin perfecte, se evidențiază mai greu, exprimând legătura de dependență care există între fenomene. În cazul corelației statistice fiecărei valori numerice a variabilei X corespund nu una ci mai multe valori a variabilei Y, adică o totalitate statistică a acestei valori, care se grupează în jurul mediei  $\bar{Y}_x$ .

Legătura de corelație după formă poate fi lineară (rectilinie) sau nelinieră (curbilinie), iar după sens – directă (pozitivă) și inversă (negativă).

În cazul corelației lineare schimbărilor uniforme a valorilor medii a unei variabile, au loc schimbări egale a altei variabile. Pentru corelația nelinieră schimbărilor uniforme a unei valori îi corespund valori medii a altei variabile, care poartă caracter de creștere ori de micșorare. Aprecierile legăturilor de corelație lineare se realizează cu ajutorul coeficientului de corelație  $r_{xy}$ , iar celor neliniare – cu raportul de corelație  $\eta$  (eta).

Corelații directe se stabilesc între fenomene care evoluează în același sens, în aceeași direcție. Crește unul, crește și cel cu care are legătură de dependență; sau scade un fenomen, scade și cel cu care corelează. Spre exemplu, mărirea înălțimii copiilor determină mărirea greutateii lor. Corelațiile inverse se stabilesc între fenomene care evoluează în sens opus. Crește un fenomen și scade cel cu care are o legătură de dependență; sau scade un fenomen și crește cel cu care se corelează. Spre exemplu, cu cât e mai mare vârsta copiilor, cu atât e mai mică mortalitatea lor.

Corelațiile statistice directe presupun evoluția în același sens a fenomenelor ce se corelează, dar nu cu aceeași unitate de măsură. Spre exemplu, crește nivelul de trai al populației unei colectivități de două ori, crește și rezistența organismului la îmbolnăviri dar nu în aceeași măsură, pentru că în afară de nivelul de trai, receptivitatea organismului la îmbolnăviri este determinată și de alți factori (biologici, climatici, geografici etc.). Corelațiile statistice inverse presupun creșterea unui fenomen și scăderea celui cu care se corelează, dar nu în aceeași măsură. Spre exemplu, dacă într-o colectivitate am efectuat un număr dublu de vaccinări anti-pertussis, numărul copiilor ce vor contracta boala, în anul următor, va fi cu siguranță mai scăzut, dar nu va fi de două ori mai mic decât în anul premergător.

Corelațiile statistice, spre deosebire de cele funcționale, pot fi numai vremelnice și într-un singur sens. Așa spre exemplu, între înălțime și vârstă există o legătură de dependență directă numai până la vârsta de 18-20 de ani. După această vârstă, înălțimea rămâne nemodificată sau eventual către bătrânețe scade. Sau alt exemplu: nivelul de trai influențează nivelul mortalității infantile, dar aceasta din urmă nu poate influența nivelul de trai.

### CALCULAREA COEFICIENTULUI DE CORELAȚIE

Legătura de dependență dintre două sau mai multe fenomene, sensul și intensitatea acesteia, se stabilesc cu ajutorul coeficientului de corelație lineară (simplă sau multiplă) al lui Bravais-Pearson.

În seriile statistice simple, când  $n < 30$ , coeficientul de corelație se obține raportând suma produselor dintre abaterile de la media aritmetică a valorilor frecvențelor primului fenomen și abaterile de la media aritmetică a valorilor frecvențelor celui de al doilea fenomen, la rădăcina pătrată din produsul realizat între suma pătratelor abaterilor de la media aritmetică a valorilor frecvențelor primului fenomen și suma pătratelor abaterilor de la media aritmetică a valorilor frecvențelor celui de al doilea fenomen, cu care se corelează.

Formula de calcul:

$$r_{xy} = \pm \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}}$$

în care:

- $r_{xy}$  = coeficientul de corelație;
- $\sum d_x d_y$  = suma produselor dintre abaterile de la media aritmetică a valorilor frecvențelor celor două fenomene ( $x$  și  $y$ ) ce se corelează;
- $\sum d_x^2$  = suma pătratelor abaterilor de la media aritmetică a valorilor frecvențelor fenomenului  $x$ ;
- $\sum d_y^2$  = suma pătratelor abaterilor de la media aritmetică a valorilor frecvențelor fenomenului  $y$ . **Exemplu:** (Tab.17, Tab.18).

Tabelul 17

Timpul trecut din momentul accesului de pancreatită acută ( $x$ ) și numărul complicațiilor postoperatorii ( $y$ )

Nr.	Timpul $x$	Număr complicații, $y$	$d_x$	$d_y$	$d_x^2$	$d_y^2$	$d_x d_y$
1	3	6	-6	-8,8	36	77,4	52,8
2	5	8	-4	-6,8	16	46,2	27,2
3	7	12	-2	-2,8	4	7,8	5,6
4	10	19	1	4,2	1	17,6	4,2
5	13	20	4	5,2	16	27	20,8
6	16	24	7	9,2	49	84,6	64,4
Total	54	89			122	260,6	175
	9	14,8				178,3	

$$r_{xy} = 0,98$$

Tabelul 18

Nivelul de asigurare cu medici stomatologi (x) și ponderea copiilor sanați  
în cadrul a 5 raioane (y)

Nr.	Asigurare medici, x	Copii sanați, y	$d_x$	$d_y$	$d_x^2$	$d_y^2$	$d_x d_y$
1	2,3	65,7	-1	-19,7	1	388,1	19,7
2	3,2	91,7	-0,1	6,3	0,01	39,7	-0,63
3	3,4	88,0	0,2	2,6	0,04	6,8	0,52
4	3,6	91,4	0,3	6,0	0,09	36	1,8
5	3,9	90,3	0,6	4,9	0,36	24,0	2,94
Total	16,4	427,1			1,5	494,6	24,3
	3,3	85,4			27,2		
							$r_{xy} = 0,89$

În seriile statistice grupate coeficientul de corelație se obține raportând suma produselor dintre abaterile valorilor variantelor de la media ponderată a celor două fenomene ce se corelează și frecvențele perechi de valori ale variantelor la rădăcina pătrată din suma produselor dintre pătratele abaterilor valorilor variantelor de la media ponderată și frecvențele corespunzătoare fiecărei variante a primului fenomen, înmulțită cu suma produselor dintre pătratele abaterilor valorilor variantelor de la media ponderată și frecvențele corespunzătoare fiecărei variante a celui de al doilea fenomen, cu care se corelează.

Formula de calcul:

$$r_{xy} = \pm \frac{\sum d_x d_y f_{xy}}{\sqrt{\sum (d_x^2 f_x) \sum (d_y^2 f_y)}}$$

în care:

$r_{xy}$  = coeficientul de corelație;

$\Sigma$  = semnul însumării;

$d_x d_y f_{xy}$  = produsul dintre abaterile de la media ponderată a variantelor celor două fenomene ce se corelează și frecvențele perechi corespunzătoare variantelor fenomenelor x și y;

$d_x^2 f_x$  = produsul dintre pătratele abaterilor de la media ponderată a valorilor variantelor fenomenului x și numărul de frecvențe corespunzătoare fiecărei variante;

$d_y^2 f_y$  = produsul dintre pătratele abaterilor de la media ponderată a valorilor variantelor fenomenului y și numărul de frecvențe corespunzătoare fiecărei variante.

### INTERPRETAREA COEFICIENTULUI DE CORELAȚIE

Coeficientul de corelație poate fi cuprins între minus unu, zero și plus unu. Când valoarea coeficientului de corelație se apropie de +1, înseamnă că între cele două fenomene ce se corelează există o legătură foarte puternică. Semnul + al coeficientului de corelație denotă că legătura de dependență dintre fenomene este directă. Deci ambele fenomene evoluează în același sens, în aceeași direcție. Când valoarea coeficientului de corelație se apropie de -1, înseamnă că între cele două fenomene există o legătură foarte puternică, dar inversă, în sens opus: crește un fenomen, scade cel cu care se corelează.

În medicină, întâlnim de obicei valori ale coeficientului de corelație intermediare valorilor -1 și +1. Pentru interpretarea intensității legăturii de dependență dintre fenomene se utilizează următoarele **CRITERII**:

- valoarea coeficientului de corelație cuprinsă între  $\pm 1$  denotă o corelație foarte puternică între fenomene;
- valoarea coeficientului de corelație cuprinsă între  $\pm 0,99$  și  $\pm 0,70$  denotă o corelație puternică;
- valoarea coeficientului de corelație cuprinsă între  $\pm 0,69$  și  $\pm 0,30$  denotă o corelație medie între fenomene;
- valoarea coeficientului de corelație cuprinsă între  $\pm 0,0$  și  $\pm 0,29$  exprimă existența unei corelații slabe între fenomene;
- valoarea coeficientului de corelație 0 denotă că legătura dintre fenomene în mod practic o considerăm inexistentă. Cele două fenomene evoluează deci independent unul de altul.

Coeficientul de corelație între fenomene poate fi corect interpretat dacă se ține seama de următoarele **ASPECTE**:

- între fenomenele ce se corelează să existe, în mod logic, o legătură;
- cele două fenomene să fie cercetate pe eșantioane omogene;
- alegerea sau selecționarea frecvenței eșantioanelor să se facă la întâmplare.

### EROAREA COEFICIENTULUI DE CORELAȚIE LINEARĂ

Coeficientul de corelație, ce exprimă legătura de dependență dintre două fenomene, se obține de obicei pe eșantioane și nu pe univers. Valorile acestuia diferă mai mult sau mai puțin față de valoarea coeficientului de corelație pe care am fi obținut-o studiind fenomenele pe întreaga populație.

Pentru a ne convinge de fidelitatea coeficientului de corelație, în interpretarea legăturii dintre fenomene obținute pe eșantioane, ne servim de eroarea coeficientului de corelație, care se notează cu  $m_r$ .

Formula de calcul pentru  $n < 30$ :

$$m_r = \frac{1 - r_{xy}^2}{\sqrt{n - 2}}$$

În cazul când  $30 < n < 100$  se utilizează formula:

$$m_r = \frac{1 - r_{xy}^2}{\sqrt{n - 1}}$$

pentru  $n > 100$ :

$$m_r = \frac{1 - r_{xy}^2}{\sqrt{n}}$$

în care:

$m_r$  = eroarea coeficientului de corelație;

$r_{xy}^2$  = pătratul valorii coeficientului de corelație, obținut pe eșantioane;

1 = valoarea absolută a coeficientului de corelație, obținut pe univers, pe întreaga populație;

$n$  = numărul variantelor perechi ale fenomenelor ce se corelează.

Dacă  $n < 30$  sau valoarea coeficientului de corelație nu este mare, trebuie de decis cât de reală este legătura dintre fenomenele ce se corelează. În general, dacă valoarea coeficientului de corelație obținut pe eșantioane este mai mare decât triplul erorii sale, înseamnă că acesta a fost obținut pe eșantioane reprezentative, este deci real și ne putem bizui pe el în interpretarea legăturii de dependență între fenomene. Dacă  $(r_{xy}/m_r) < 3$  legătura dintre fenomene rămâne nedeterminată și se consideră că valoarea coeficientului de corelație, care diferă de zero, a fost obținută întâmplător.

Raportul dintre coeficientul de corelație și eroarea lui se numește criteriu de exactitate al coeficientului de corelație –  $t_r$ . Criteriul în cauză se stabilește cu ajutorul tabelului valorilor criteriului  $t$ . Dacă  $t_{real} > t_{tabel}$ , coeficientul de corelație se consideră semnificativ.

### CORELOGRAMA

Existența sau inexistența unei corelații între fenomene se poate evidenția aproximativ fără prea multe calcule, cu ajutorul reprezentărilor grafice. În acest caz, folosim un grafic cu două scări, ordonată și abscisă, pe care înscrinem valorile variantelor celor două fenomene  $x$  și  $y$ .

În rețeaua graficului este reprezentată prin câte un punct fiecare frecvență la nivelul valorii variantei corespunzătoare fenomenului  $x$  de pe abscisă și la înălțimea valorii variantei corespunzătoare fenomenului  $y$  de pe ordonată. Se realizează astfel „norul de puncte”. În funcție de poziția acestui nor de puncte putem aprecia, aproximativ, existența sau absența corelației dintre fenomene, precum și sensul și intensitatea aproximativă a acesteia. Dacă norul de puncte se va dispune fuziform, oblic de jos în sus și de la stânga la dreapta, între cele două fenomene există o corelație directă. Crește un fenomen, crește și cel de al doilea, cu care se corelează, sau ambele fenomene scad, evoluând în aceeași direcție. Intensitatea legăturii corelative dintre cele două fenomene se apreciază, aproximativ, după unghiul pe care dreapta ce trece prin mijlocul norului de puncte îl realizează cu abscisa. Cu cât acest unghi ascuțit este mai mare, având tendința să se apropie de  $45^\circ$ , cu atât corelația dintre fenomene este mai puternică. Cu cât unghiul format de această dreaptă și abscisă este mai mic, dreapta oblică având tendința să se apropie de orizontală, cu atât corelația dintre fenomene va fi mai slabă.

Dacă norul de puncte se dispune fuziform, oblic de sus în jos și de la stânga la dreapta, între cele două fenomene există o corelație inversă. Intensitatea acestei corelații inverse se apreciază după mărirea unghiului ascuțit pe care dreapta ce trece prin mijlocul norului de puncte îl formează cu abscisa, în partea opusă ordonatei.

Dacă punctele se dispun pe toată rețeaua grafică, neavând nici o tendință de a se grupa, înseamnă că între fenomene nu există nici o legătură de dependență, fenomenele evoluând independent unul față de celălalt. În cazul acesta, dreapta care trece prin mijlocul punctelor este paralelă fie cu ordonata, fie cu abscisa.

**CORELAȚIA MULTIPLĂ**

Se stabilește între mai mult de două fenomene, care au legătură de dependență între ele. În asemenea situații, existența legăturii de dependență între fenomene, sensul și intensitatea acestei legături, se stabilesc cu ajutorul coeficientului de corelație lineară multiplă.

Spre exemplu, dacă dorim să evidențiem legătura de dependență între înălțimea (fenomen x), greutatea (fenomen y) și vârsta (fenomen z) copiilor sau legătura de dependență între incidența dinților cariatiți (fenomen x), dinților cu parodontopatii marginale (fenomen y) și dinților absenți (fenomen z) ne folosim de următoarea formulă:

$$r_{xyz} = \sqrt{\frac{r_{xy}^2 + r_{xz}^2 - 2(r_{xy}) \times r_{yz} \times r_{xz}}{1 - r_{xz}^2}}$$

în care:

- $r_{xyz}$  = coeficientul de corelație lineară multiplă între cele trei fenomene;
- $r_{xy}$  = coeficientul de corelație lineară simplă între fenomenele x și y;
- $r_{yz}$  = coeficientul de corelație lineară simplă între fenomenele y și z;
- $r_{xz}$  = coeficientul de corelație lineară simplă între fenomenele x și z;

**Coeficientul de corelație lineară multiplă are valori mai mari decât valorile coeficienților de corelație lineară simplă luați în parte și totdeauna semn pozitiv. Interpretarea intensității corelației multiple ține seama de aceleași criterii enunțate la interpretarea coeficientului de corelație lineară simplă.**

**CORELAȚIA RANGURILOR (SPEARMAN)**

În cazul în care dorim să stabilim legătura de dependență între fenomene cercetate pe eșantioane mici ( $n < 30$ ), deci pe un număr redus de frecvențe, utilizăm coeficientul de corelație al rangurilor, propus de Spearman (1904). Acest coeficient se notează cu litera greacă  $\rho$  (ro) și se determină după formula propusă de Spearman:

$$\rho = 1 - \frac{6 \sum d^2}{n \times (n^2 - 1)}$$

în care:

- 1 = valoarea absolută a coeficientului de corelație;
- $\rho$  = coeficientul de corelație Spearman;

6 = valoare constantă;

$\sum d^2$  = suma pătratelor diferențelor dintre rangurile primului șir de variante și rangurile celui de al doilea șir de variante, cu care se corelează;

n = numărul variantelor perechi variantelor ce se corelează.

Coeficientul de corelație al rangurilor poate avea valori cuprinse între „-1” „0”- +”1”. El exprimă o legătură perfectă când are valoarea +1. În această situație, rangurile au valori egale, iar diferența între ranguri este egală cu 0.

În general, cu cât numărul rangurilor cu aceeași valoare, în cele două clasamente, este mai mare, cu atât suma pătratelor diferențelor între ranguri este mai mică, iar valoarea coeficientului de corelație Spearman va fi mai mare.

**Exemplu (Tab.19, Tab.20).**

Tabelul 19

Timpul trecut din momentul accesului de pancreatită acută (x) și numărul complicațiilor postoperatorii (y)

Nr.	Timpul, x	Număr complicații, y	Rangul x	Rangul y	d	d <sup>2</sup>
1	3	6	1	1	0	0
2	5	8	2	2	0	0
3	7	12	3	3	0	0
4	10	19	4	4	0	0
5	13	20	5	5	0	0
6	16	24	6	6	0	0
Total	54	89				0
					$\rho = 1$	

Tabelul 20

Legătura de corelație între copiii cu deficiență mintală (la 100 mii copii) și invaliditatea copiilor (la 100 mii copii)

Nr.	Țara	Deficiență mintală x	Invalidi-tatea copiilor y	Rangul x	Rangul y	d	d <sup>2</sup>
1	Armenia	120	33	1	6,5	-5,5	30,25
2	Tadjikistan	138	27	2	1	1	1
3	Azerbaidjan	167	38	3	9,5	-6,5	42,25
4	Turkmenistan	198	30	4	4,5	-0,5	0,25
5	Georgia	245	43	5	11	-6	36
6	Kîrgîzstan	318	28	6	2	4	16
7	Uzbekistan	342	30	7	4,5	2,5	6,25
8	Kazahstan	455	29	8	3	5	25
9	Bielarusi	538	36	9	8	1	1
10	Ucraina	625	48	10	14	-4	16
11	Rusia	795	38	11	9,5	1,5	2,25
12	Moldova	1018	44	12	12	0	0
13	Estonia	1110	47	14	13	1	1
14	Lituania	1021	57	13	15	-2	4
15	Letonia	1127	33	15	6,5	8,5	72,25
Total						$\Sigma=-24,5$ $\Sigma=+24,5$	$\Sigma=253,5$
							$\rho = 0,55$

### COEFICIENTUL DE ASOCIERE

În domeniul medicinei trebuie în multe cazuri de a stabili legătura nu numai între indicatori cantitativi, ci și între cei calitativi, care au caracter alternativ prezența fenomenului – lipsa fenomenului). În asemenea cazuri se utilizează tabelul de asociere format din patru câmpuri (2x2), care exprimă legătura de dependență între fenomenele cu caracter alternativ.

Legătura de corelație se determină prin intermediul **coeficientului de asociere**, utilizând formula următoare:

$$q = \frac{ad - bc}{ad + bc}$$

unde: a, b, c, d corespund valorilor situate în cele patru câmpuri.

**Exemplu:** (Tab. 21).

Tabelul 21

Vaccinarea contra unei maladii și morbiditatea populației în orașul A

	S-au îmbolnăvit	Sănătoși	TOTAL
Vaccinați	40 a	5900 b	5940 a + b
Nevaccinați	250 c	9620 d	9870 c + d
TOTAL	290 a + c	15520 b + d	15810 a + b + c + d

$$q = \frac{ad - bc}{ad + bc} = \frac{(40 \times 9620) - (5900 \times 250)}{(40 \times 9620) + (5900 \times 250)} = \frac{-1090200}{1859800} = -0,58$$

**Concluzie:** între vaccinare și morbiditate este legătură inversă, astfel cu cât mai multe persoane sunt vaccinate, cu atât mai puține cazuri de îmbolnăvire se înregistrează.

### Exemplu de legătură de corelație nelineară.

A fost efectuată analiza activității a 21 de medici pentru a determina dacă există legătură între calificarea medicului și costul analizelor necesare pentru diagnostic.\*

Comisia specială a evaluat calificarea fiecărui medic. Ulterior medicii au primit ranguri de la 1 (cel mai calificat) până la 21 (cel mai necalificat).

Costul mediu al analizelor necesare pentru fiecare medic în primele 3 zile de aflare a bolnavului în clinică a primit în mod analogic rangul de la 1 (cel mai mic) până la 21 (cel mai mare).

Perechile de ranguri pentru fiecare medic sunt prezentate în fig.4.

Coeficientul Spearman = -0,13. Valoarea absolută a coeficientului este mai mică decât cea critică chiar pentru pragul de semnificație 0,05, ceea ce denotă lipsa semnificației statistice pentru valoarea obținută.

Există sau nu legătură între calificarea medicului și costul analizelor necesare pentru stabilirea diagnosticului?

Există, însă legătură nelineară. Analizând atent repartiția din fig. 4, vom observa că cel mai mic cost mediu al analizelor se înregistrează

\* S.A. Schroeder, A. Schliftman, T.E. Piemine. Variation among physicians in use of laboratory tests: relation to quality of care. Med. Care, 12: 709-713, 1974

la cei mai calificați și la cei mai necalificați medici. Costurile cele mai înalte sunt la medicii de calificare medie.

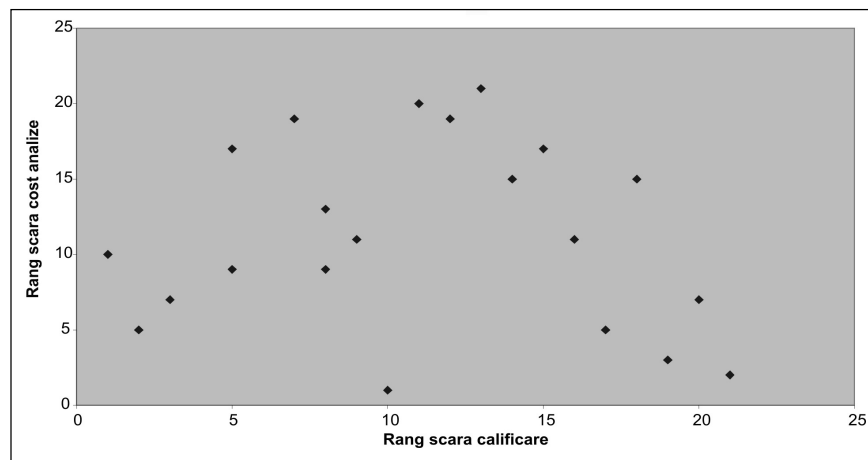


Fig.4. Calificarea medicilor și costul analizelor

### 2.9.2. Regresia

Termenul de regresie a fost introdus de F. Galton, care a observat că înălțimea descendenților regresează către înălțimea părinților.

Coefficientul de corelație ne dă indicații asupra sensului și intensității legăturii de dependență dintre fenomene, fără a putea preciza, sub aspect cantitativ, cu cât crește sau scade un fenomen când cel cu care se corelează crește sau scade cu o anumită cantitate.

Regresia, noțiune strâns legată de noțiunea de corelație, completează corelația și prin intermediul coeficientului de regresie, stabilește cu cât crește sau descrește sub aspect cantitativ, un fenomen, când cel cu care se corelează crește sau descrește cu o unitate de măsură.

Regresia poate fi simplă și multiplă; liniară și neliniară. Ca și corelația, regresia poate fi directă, când fenomenele evoluează în același sens (crește x, crește y sau scade x scade și y), sau indirectă, când fenomenul evoluează în sens opus (crește x scade y sau scade x crește y).

Formula coeficientului de regresie este:

$$Rg_{yx} = r_{xy} \frac{\delta_y}{\delta_x}$$

sau

$$Rg_{yx} = r_{xy} \frac{\delta_x}{\delta_y}$$

în care:

$Rg_{xy}$  = coeficientul de regresie a lui x în funcție de y. El exprimă, cantitativ, cu cât crește sau scade fenomenul x când y crește sau scade cu o unitate de măsură;

$Rg_{yx}$  = coeficientul de regresie a lui y în funcție de x. El exprimă, cantitativ, cu cât crește sau scade fenomenul y când x crește sau scade cu o unitate de măsură;

$r_{xy}$  = coeficientul de corelație liniară Bravais-Pearson;

$\delta_x$  = deviația standard a fenomenului x;

$\delta_y$  = deviația standard a fenomenului y.

**Exemplu:** (Tab. 22).

Tabelul 22

Înălțimea și greutatea copiilor de vârstă Z (n=22)

Înălțimea x	Frecvențe fx	Greutatea y	Frecvențe fy
130	3	29	3
132	5	30	4
135	7	31	3
136	4	32	4
137	3	33	4
		34	2
		35	1
		36	1

Repartizarea a 22 de copii conform înălțimii (x) și greutății (y) a oferit următoarele date: înălțimea medie – 134 cm; greutatea medie – 31,8 kg; abaterea standard pentru înălțime (x) – 2,37; abaterea standard pentru greutate (y) – 1,97; coeficientul de corelație – 0,82.

Cu cât se va modifica greutatea corpului la copiii în cauză, dacă înălțimea acestora va crește cu 1 cm?

Substituind datele în formula coeficientului de regresie, obținem:

$$Rg_{yx} = r_{xy} \frac{\delta_y}{\delta_x} = 0,82 \frac{1,97}{2,37} = 0,82 \times 0,83 = 0,68 \text{ kg / cm}$$



**Concluzie:** Creșterea înălțimii medii la copiii studiați cu 1cm va avea drept urmare creșterea greutateii lor cu 0,68 kg.

Utilizând coeficientul de regresie se poate afla mărimea fenomenului  $y$  (în cazul analizat anterior – greutatea), fără a recurge la măsurarea acestuia, utilizând în acest scop numai fenomenul  $x$  (greutatea). Se folosește următoarea ecuație a regresiei:

$$y = \bar{Y} + Rg_{yx} (x - \bar{X})$$

în care:

$y$  – greutatea cercetată;

$x$  – mărimea cunoscută a înălțimii;

$Rg_{yx}$  – coeficientul de regresie al înălțimii în raport cu greutatea;

$\bar{Y}$  – greutatea medie a colectivității cercetate;

$\bar{X}$  – înălțimea medie a colectivității cercetate.

În cazul dat înălțimea medie – 134 cm; greutatea medie – 31,8 kg;  $Rg_{yx}$  – 0,68. Se cere de aflat care va fi greutatea copiilor care au înălțimea de 135cm.

Înlocuim datele în formulă, obținem:

$$y = \bar{Y} + Rg_{yx} (x - \bar{X}) = 31,8 + 0,68(135 - 134) = 31,8 + 0,68 \times 1 = 32,5 \text{ kg}$$

Astfel, înălțimii de 135 cm îi corespunde greutatea de 32,5 kg.

**Scara regresiei:** În domeniul cercetărilor somatometrice a copiilor și adolescenților este foarte importantă metoda de estimare a indicatorilor înălțimii, greutateii corpului, perimetrului toracic etc. Valorile individuale ale acestor valori diferă uneori destul de evident. La persoanele care au aceeași înălțime greutatea corpului poate varia în limite destul de mari.

Mărimea diversităților individuale a acestor caracteristici este redată de abaterea standard a regresiei –  $\bar{\sigma}_{Rg}$ , care este calculată cu ajutorul formulei următoare:

$$\sigma_{Rg} = \pm \sigma_y \times (1 - r_{xy}^2)$$

Cu cât va fi mai mică valoarea abaterii standard a regresiei, cu atât mai mici vor fi limitele de variație a valorilor individuale față de media lor.

În exemplul anterior, conform datelor obținute, valoarea abaterii standard a regresiei va fi:

$$\sigma_{Rg} = \pm \sigma_y \times (1 - r_{xy}^2) = 1,97 \times (1 - 0,82^2) = 1,97 \times 0,33 = \pm 0,65 \text{ kg/cm}$$

Cunoscând valoarea coeficientului de regresie, folosind ecuația regresiei și abaterea standard a regresiei se poate de format scara regresiei.

## 2.10. Analiza statistică a variabilității

Caracterizarea unei colectivități statistice prin indicatorii tendinței centrale ne ajută să depistăm ceea ce este comun, esențial în manifestarea unui fenomen. Orice colectivitate are o anumită organizare internă, definită prin felul în care valorile individuale se împrăștie sau se concentrează în jurul valorii centrale. Astfel se poate întâmpla ca două colectivități analizate după aceeași variabilă să fie diferite prin tendința centrală (Fig. 5a), prin dispersie (Fig. 5b) sau prin amândouă (Fig. 5c). În felul acesta o valoare centrală poate fi credibilă, o alta nu. Din acest motiv se impune ca analiza prin indicatorii tendinței centrale să fie completată cu indicatorii ai variației și ai formei de distribuție.

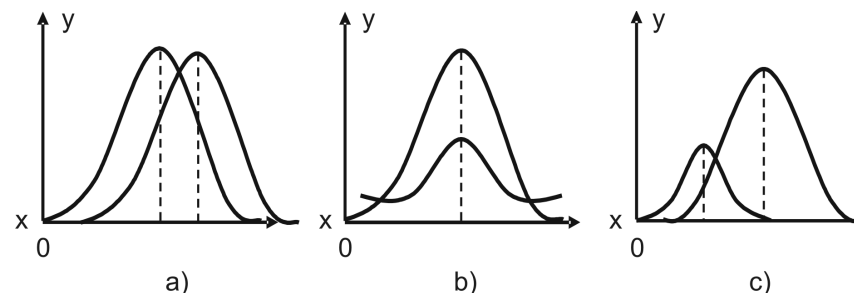


Fig.5: a) Distribuții cu tendință centrală diferită; b) Distribuții cu variabilitate diferită; c) Distribuții cu tendință centrală și variabilitate diferite.

**În statistică prin noțiunea generală de variabilitate se au în vedere abaterile măsurabile ale valorilor individuale față de o valoare centrală (tipică).**

De exemplu seriile:

{2, 2, 2, 10, 18, 18, 18}

{9, 9, 9, 9, 10, 11, 11, 11, 11}

au aceeași medie și mediană dar dispersie diferită. Acest fapt este evidențiat de abaterile înregistrate față de medie (10) sau mediană (10). Seria a doua este mai omogenă (înregistrează o variabilitate mai mică a valorilor individuale), iar valorile sale tipice sunt mai reprezentative.

Calculul și analiza indicatorilor variației valorilor individuale față de medie oferă posibilitatea de rezolvare a unor probleme de cunoaștere statistică. Dintre acestea se disting:

- 1) analiza gradului de omogenitate a datelor din care s-au calculat indicatorii tendinței centrale și verificarea reprezentativității acestora;
- 2) compararea în timp și (sau) spații a mai multor serii de distribuție după caracteristici independente și (sau) interdependente;
- 3) selectarea obiectivă a factorilor semnificativi de influență după care se structurează unitățile unei colectivități statistice;
- 4) separarea acțiunii factorilor esențiali de factorii întâmplători;
- 5) concentrarea valorilor individuale ale caracteristicilor și deplasarea acestora față de valorile tipice;
- 6) aplicarea diferitelor teste ale statisticii matematice.

Indicatorii variației utilizați în analizele statistice sunt clasificați după mai multe criterii:

- după numărul variantelor luate în calcul există *indicatori simpli și indicatori sintetici*;
- după modul de sistematizare a datelor primare există *indicatori ai variației calculați pentru serii de distribuție unidimensionale și indicatori calculați pentru serii multidimensionale*;
- după modul de calcul și exprimare există indicatori ai variației calculați ca *mărimi absolute și ca mărimi relative*.

#### Indicatori simpli:

- amplitudinea variației;
- abaterea valorilor individuale de la medie;
- abaterea intercuantilică.

#### Indicatori sintetici:

- dispersia;
- abaterea medie pătratică;
- coeficientul de variație;

#### Amplitudinea variației (A)

- Se mai numește câmp de variație sau amplitudine absolută.
- Se determină prin diferența dintre cea mai mare și cea mai mică valoare individuală înregistrată.

$$A = x_{\max} - x_{\min}$$

**Exemplu:** Un grup de bolnavi înregistrează următoarele durate de tratament în zile:

8, 7, 9, 8, 9, 9, 10, 11, 12, 11, 10, 10, 9, 9, 7, 8, 10, 10, 11, 10

Amplitudinea variației pentru durata tratamentului:  $12 - 7 = 5$  zile.

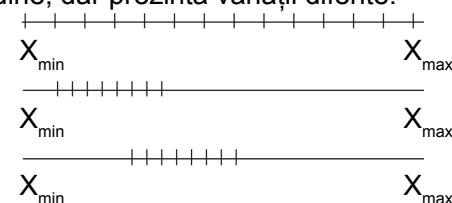
Amplitudinea relativă ( $A_{\%}$ ) este raportul procentual dintre amplitudinea absolută și media aritmetică a valorilor analizate.

$$A_{\%} = \frac{A}{x} \times 100 = \frac{x_{\max} - x_{\min}}{x} \times 100$$

În exemplul de mai sus  $A_{\%} = (5/9) \times 100 = 55,5\%$

Observații referitor la amplitudine:

- 1) Se exprimă în unitatea de măsură a caracteristicii studiate sau în procente în cazul amplitudinii relative;
- 2) Amplitudinea nu ține seama de toate valorile, ci numai de cele extreme, care adeseori sunt aberante;
- 3) Este sensibilă la valorile aberante, se consideră un indicator mai puțin relevant nepermițând cunoașterea structurii interne de variație. De exemplu, următoarele trei serii vizualizate au aceeași amplitudine, dar prezintă variații diferite:



Aceasta înseamnă că amplitudinea variației este un indicator destul de rezonabil numai dacă valorile individuale ale seriei sunt repartizate în mod uniform;

- 4) În cazul seriei de distribuție pe intervale valorice calculul amplitudinii este lipsit de sens;
- 5) Amplitudinea se utilizează frecvent la alegerea numărului de intervale de grupare a datelor și la stabilirea mărimii intervalelor.

#### Abaterea valorilor individuale de la medie ( $d_i$ ).

Exprimă cu câte unități de măsură sau de câte ori (sau cât la sută) valoarea caracteristicii individuale a colectivității se abate de la mărimea unui indicator al tendinței centrale.

- în forma absolută:  $d_i = x_i - \bar{x}$

- în forma relativă:

$$d_{\%} = \frac{d_i}{\bar{x}} \times 100 = \frac{x_i - \bar{x}}{\bar{x}} \times 100$$

**Exemplu:** considerând aceleași date din exemplul anterior, bolnavul care s-a tratat 12 zile se abate de la media grupei cu 3 zile ceea ce reprezintă 33,3% din medie iar pacientul care s-a tratat 7 zile se abate de la media semigrupului cu 2 zile ceea ce reprezintă 22,2% din medie.

Cu cât obținem mai multe abateri relative mari în analiza unui volum de date (câtre 100% sau peste acest nivel) cu atât colectivitatea este mai eterogenă.

Observații:

- 1) valorile individuale se compară cu valoarea lor medie, dar la fel de bine poate fi utilizat pentru comparare oricare alt indicator al tendinței centrale (de ex. mediana);
- 2) În seriile de distribuție pe intervale valorice pentru calculul abaterilor individuale se iau în considerare centrele de interval;
- 3) Abaterea valorilor individuale nu poate da informații decât la nivelul fiecărei variante pierzând imaginea împrăștierii pe ansamblul distribuției.
- 4) În analizele statistice se urmăresc în mod deosebit abaterile maxime pozitive ( $d_{\max}^+$ ) și abaterile maxime negative ( $d_{\max}^-$ ) calculate în cifre absolute și relative astfel:

$$d_{\max}^+ = x_{\max} - \bar{x} \text{ sau } d_{\max}^+ \% = (d_{\max}^+ / \bar{x}) \times 100$$

$$d_{\max}^- = x_{\min} - \bar{x} \text{ sau } d_{\max}^- \% = (d_{\max}^- / \bar{x}) \times 100$$

### Abaterea intercuantilică

Se calculează ca diferență între cuantila superioară și cuantila inferioară de același ordin. Astfel pentru  $r = 4$   $Q_3 - Q_1$  conține 50% din numărul observațiilor; pentru  $r = 10$   $D_9 - D_1$  conține 80% din numărul observațiilor (Fig.6).

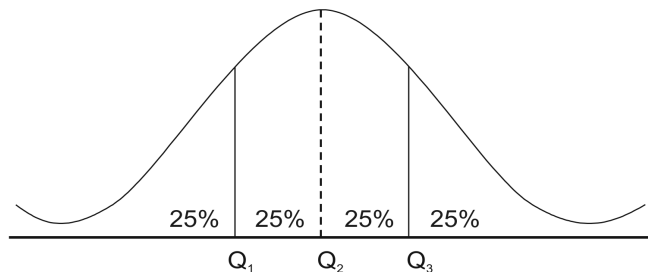


Fig.6. Abaterea  $Q_3 - Q_1$  în distribuția normală

Observații:

- 1) Acest indicator se exprimă în unitatea de măsură a caracteristicii;
- 2) Calculul abaterii intercuantilice, spre deosebire de cel al amplitudinii, prezintă avantajul că evită valorile individuale extreme sau aberante. Prin calculul acestui indicator se pierde informații dar are câștig de cauză omogenitatea.
- 3) Ca și amplitudinea, abaterea intercuantilică nu poate fi utilizată în calcule algebrice.

### Dispersia ( $\sigma^2$ )

Se calculează ca media aritmetică a patratelor abaterilor valorilor individuale de la tendința centrală (uzual de la medie).

Pentru o serie simplă formula dispersiei este:

$$\sigma^2 = \frac{\sum d^2}{n}$$

Pentru o serie de distribuție pe frecvențe formula dispersiei este:

$$\sigma^2 = \frac{\sum d^2 f}{n}$$

Observații:

- 1) Se mai numește varianță sau moment centrat de ordin doi;
- 2) Este indicatorul care măsoară variația totală a unei caracteristici studiate datorată atât cauzelor esențiale cât și celor întâmplătoare;
- 3) Este un indicator cu valoare teoretică, util în verificări de ipoteze statistice, o mărime abstractă folosită ca bază de calcul pentru abaterea medie pătratică;
- 4) Nu are formă concretă de exprimare;
- 5) Cu cât valorile individuale ale caracteristicii sunt mai apropiate între ele, cu atât mărimea dispersiei este mai mică. La limită, dacă toate valorile individuale sunt egale între ele, dispersia este nulă;
- 6) Dispersia, ca și media valorilor individuale, este sensibilă la prezența valorilor extreme, aberante.

### Abaterea medie pătratică ( $\sigma$ )

Este calculată ca o medie pătratică din abaterile valorilor individuale de la media lor.

În seriile statistice simple:

$$\sigma = \sqrt{\frac{\sum d^2}{n}}$$

În seriile de distribuție pe frecvențe:

$$\sigma = \sqrt{\frac{\sum d^2 f}{n}}$$

Menționăm, că în cazul în care numărul de frecvențe este mai mic de 120, numitorul formulelor deviației standard devine  $n-1$ .

Deviația standard este cel mai util și mai important indicator de dispersie. Ea pune în evidență intervalul valoric, în jurul mediei, în care s-au distribuit valorile individuale ale fenomenului studiat. O deviație standard cu valoare mică, pune în evidență strânsă o distribuție strânsă a valorilor frecvențelor fenomenului cercetat în jurul mediei, deci evidențiază un eșantion omogen. O valoare mare a deviației standard pune în evidență o dispersie mare a valorilor individuale în jurul mediei, dovedind că eșantionul este neomogen și ca atare rezultatele obținute pe un așa eșantion nu pot fi generalizate.

În general, cu cât deviația standard are valori mai mari, cu atât media caracterizează mai puțin fenomenul cercetat. Deviația standard are o valoare semnificativă atunci când eșantionul pe care am lucrat are o distribuție normală, adică distribuția se face conform curbei Gauss – Laplace (Fig. 7).

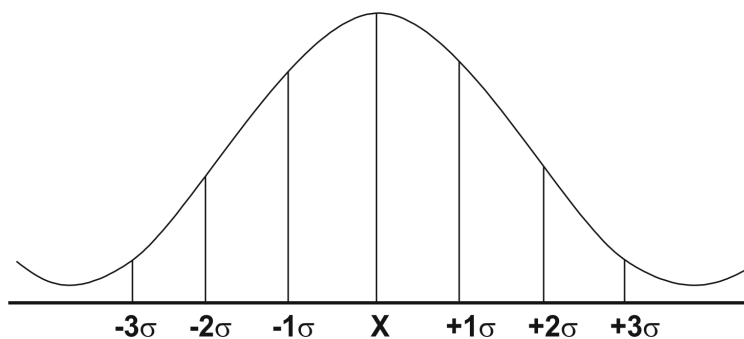


Fig.7. Curba Gauss – Laplace.

În cadrul unei asemenea distribuții, în mod normal, între valoarea mediei și cea a deviației standard se distribuie 68,26% din totalul valorilor unei colectivități statistice; între medie și două valori ale sigmei se distribuie 95,45% din valori, iar între medie și trei valori ale sigmei sunt cuprinse 99,73% din valori.

De obicei, în practica, datorită variabilității mari a fenomenelor biologice, se consideră ca normale valorile cuprinse în intervalul determinat de  $X \pm 2\sigma$ . În cercetările experimentale, care necesită o exactitate mai mare, intervalul în care considerăm ca normale valorile frecvențelor fenomenului cercetat este mai mare, determinat de  $X \pm 3\sigma$ .

Observații:

- 1) Se mai numește **abatere standard, deviație standard, abatere tip, standard deviation**;
- 2) Este un indicator important în analiza variației, se folosește la estimarea erorilor de selecție; stabilirea eșantionului; în calcule de corelație;
- 3) Se utilizează pentru elaborarea standardelor de dezvoltare fizică, pentru delimitarea stărilor normale de cele patologice;
- 4) Exprimă cu cât se abate în medie fiecare valoare individuală de la media valorilor evidențiind influența abaterilor mari;
- 5) Se calculează ca o medie pătratică simplă sau ponderată a pătratelor abaterilor valorilor individuale de la media lor;
- 6) Se exprimă în unitatea de măsură în care se exprimă și caracteristica.

### Coefficientul de variație (Cv)

Pe lângă multiplele utilizări practice, deviația standard are și un mare neajuns. Datorită faptului că deviația standard este exprimată în unități de măsură diferite, specifice fenomenelor pe care le cercetăm, ea nu poate fi utilizată în aprecierea comparativă a două sau mai multe eșantioane, sub aspectul omogenității, care se exprimă în unități de măsură diferite.

Spre exemplu, dacă am măsurat înălțimea și greutatea unui lot de 30 de elevi și am găsit pentru înălțime o medie de 145 cm cu o deviație standard de  $\pm 5$  cm, iar pentru greutate am găsit o medie de 35 de kg cu o deviație de  $\pm 3$  kg, pe baza valorilor deviației standard, exprimate în centimetri și kilograme, nu putem aprecia sub ce aspect, al înălțimii

sau al greutatei, lotul de elevi este mai omogen, mai uniform dezvoltat.

În această situație se recurge la coeficientul de variație, care nu este altceva decât deviația standard exprimată procentual.

Ca definiție, coeficientul de variație este raportul procentual dintre valoarea deviației standard și media aritmetică.

Formula de calcul:

$$C_v = \frac{\sigma}{\bar{X}} \times 100\%$$

Coeficientul de variație arată cât la sută din medie reprezintă deviația standard. Astfel, exprimând procentual pe  $\sigma$ , scăpăm de influența unității de măsură, putând compara între ele, sub aspectul omogenității, câte eșantioane dorim.

În exemplul anterior, coeficientul de variație pentru înălțime este:

$$C_v = \pm (5\text{cm}/145\text{cm}) \times 100 = \pm 3,4\%$$

Pentru greutate coeficientul este:

$$C_v = \pm (3\text{kg}/35\text{kg}) \times 100 = \pm 8,6\%$$

Din datele obținute rezultă, că lotul de copii este mai omogen dezvoltat sub aspectul înălțimii, și nu sub aspectul greutății, cum eram tentați să apreciem pe baza valorilor deviației standard.

Cu cât valoarea procentuală a coeficientului de variație este mai mică, cu atât eșantionul cercetat este mai omogen.

În general, se consideră că un coeficient de variație cu valori sub  $\pm 10\%$  indică o variație mică, deci o omogenitate care ne permite să generalizăm rezultatele, dacă și din punct de vedere numeric eșantionul este corespunzător. Un coeficient de variație cu valori cuprinse între  $\pm 10\%$  și  $\pm 20\%$  indică o variație medie. Eșantionul este mai puțin omogen ca în primul caz, dar totuși permite generalizarea rezultatelor obținute, cu o anumită probabilitate de a greși.

Valorile peste  $\pm 20\%$  ale coeficientului de variație indică o variație mare, o lipsă aproape totală de omogenitate, fapt ce contraindică generalizarea rezultatelor obținute pe un asemenea eșantion.

De menționat că aceste limite ale coeficientului de variație, ce ne permit să apreciem omogenitatea sau lipsa de omogenitate a eșantionului, trebuie să reprezinte valorile duble ale acestuia, în cazul cercetărilor clinice, și valorilor triple, în cazul cercetărilor experimentale, așa cum am văzut și la deviația standard.

În exemplul anterior, valoarea dublă a coeficientului de variație pentru înălțime este de  $\pm 6,8$  și pentru greutate  $\pm 17,2$  din valorile mediilor respective.

### Variabile standardizate

- 1) Variabila  $[z = d/\sigma]$  se numește variabilă standardizată sau abatere normată.
- 2) Unitatea de măsură a abaterii normate se numește unitate standard sau valoare standard.
- 3) Exprimă poziția unității statistice într-o distribuție dată atât față de medie cât și față de dispersie.
- 4) Cu ajutorul acestei variabile se pot compara mărimi ce provin din distribuții diferite.

### 2.11. Aprecierea veridicității valorilor medii și relative

Constantele (medii sau de dispersie) obținute de noi pe eșantioane se numesc „statistici” și valorile lor sunt mai mult sau mai puțin apropiate de valorile constantelor colectivităților generale (parametrii) – în funcție de gradul de reprezentativitate (calitativă și cantitativă) al eșantioanelor, însă niciodată nu sunt identice.

Cu ajutorul acestor constante de eșantion, a „statisticilor”, se estimează constantele colectivităților generale, „parametrii”. Estimatul rezultat pe baza constantelor de eșantionare este aleator. Ca atare valorile exacte ale parametrilor colectivității generale rămân necunoscute, în schimb se poate preciza un interval valoric în care se va situa media colectivității generale (MA = media absolută), în jurul căreia, în acel interval valoric se vor distribui mediile de eșantion, cu o anumită probabilitate.

Dacă dintr-o colectivitate generală extragem mai multe eșantioane, valorile medii ale acestor eșantioane vor fi foarte apropiate de valoarea mediei absolute, distribuția acestora, în jurul mediei absolute, făcându-se conform aceleiași curbe Gauss-Laplace.

### Eroarea standard sau eroarea medie a mediilor

**Constanta care ne permite să stabilim intervalul valoric în care se găsește media absolută și în jurul căreia se distribuie valorile medii de eșantion, cu o anumită probabilitate, se numește eroare**

standard (ES) sau eroare medie a valorilor medii de eșantion față de media universului.

Formula de calcul:

$$ES_M = \pm \frac{\sigma_x}{\sqrt{n}}$$

în care:

$ES_M$  = eroarea standard;

$\sigma$  = abaterea standard;

$n$  = numărul frecvențelor.

În situația în care numărul de frecvențe al eșantionului este mai mic sau egal cu 120, atunci  $\sigma$  se raportează la numărul de frecvențe -1 și formula devine:

$$ES_M = \pm \frac{\sigma_x}{\sqrt{n-1}}$$

în care:

$ES_M$  = eroarea standard;

$\sigma$  = abaterea standard;

$n$  = numărul frecvențelor

### Exemplul 1.

La 120 persoane a fost determinată tensiunea arterială sistolică. În baza valorilor individuale s-a stabilit că  $M = 140,5$  mm Hg și  $\delta = 6,25$ .

### Exemplul 2.

În baza controlului medical la 508 persoane au fost stabilite afecțiuni dentare. În medie au fost afectați câte 5,84 dinți,  $\delta = 3,16$ .

În primul nostru exemplu, eroarea standard va fi:

$$ES_M = \pm \frac{6,25}{\sqrt{120}} = \frac{6,25}{10,90} = \pm 0,57$$

În cel de al doilea exemplu al nostru, eroarea standard va fi:

$$ES_M = \pm \frac{3,16}{\sqrt{508}} = \frac{3,16}{22,53} = \pm 0,14$$

Valoarea erorii standard, adăugată sau scăzută din valoarea mediei eșantionului, ne permite să obținem limitele intervalului în care se găsește valoarea mediei absolute și în jurul ei distribuite mediile eșantioanelor, cu o anumită probabilitate.

### Intervalul de siguranță

**Intervalul valoric, determinat cu ajutorul erorii standard, în care se estimează a se afla media absolută**, se numește interval de siguranță sau de încredere statistică, în acest interval de siguranță, determinat de media eșantionului plus/minus eroarea standard, media absolută se va găsi într-o proporție de 68,26%, deci probabilitatea ca media absolută să se găsească în interiorul acestui interval este de 68,26%.

$$M-ES < I < (M+ES) = 68,26\%$$

$$P-ES < I < (P+ES)$$

În primul nostru exemplu, media ponderată fiind 140,5 mm Hg, intervalul de siguranță va fi  $140,5 \pm 0,57$ , deci va fi cuprins între 139,93 și 141,07 mm Hg. În cel de al doilea exemplu al nostru, media ponderată fiind de 5,84 dinți iar eroarea standard 0,14, intervalul de siguranță va fi  $5,84 \pm 0,14$ , deci cuprins între 5,70 și 5,98.

$$2 \times ES$$

$$3 \times ES$$

### Pragul de semnificație

**Contraprobabilitatea sau probabilitatea ca mediile de eșantion să depășească limitele – maximă și minimă – intervalului de siguranță, situându-se în afara lor**, se numește prag de semnificație, în cazul în care intervalul de siguranță este determinat de  $M \pm ES$ , atunci contraprobabilitatea (pragul de semnificație) se obține scăzând din 100 valoarea probabilității, deci:

$$q = 100 - 68,26\% = 31,74\%$$

În exemplul nostru media ponderată fiind de 5,84 dinți afectați, intervalul de siguranță va fi:  $5,84 \pm 0,14$ , deci va fi cuprins între 5,70 -5,98.

Dacă vrem ca valoarea contraprobabilității, a posibilității de a greși, să fie mai mică, atunci trebuie să mărim intervalul de siguranță. Acest interval se mărește adăugând și scăzând din valoarea mediei de două ori valoarea erorii standard.

$$M \pm 2 \cdot ES = 5,84 + 2 \cdot 0,14 = 5,84 \pm 0,28$$

Intervalul de siguranță va fi cuprins deci între limitele de 5,56 și 6,12. În acest interval, ceva mai mare, valoarea mediei absolute se va găsi cu o probabilitate mai mare decât în primul caz, de 95,45%, iar posibilitatea de a se situa în afara acestui interval de siguranță se reduce la 4,55% ( $100\% - 95,45\% = 4,55\%$ ). Dacă vrem să reducem și mai mult probabilitatea de a greși, atunci mărim intervalul de siguranță, adăugând și scăzând din valoarea mediei de trei ori valoarea erorii standard. Intervalul de siguranță va fi:

$$M \pm 3.E.S = 5.84 \pm 3.0.14 = 5.84 \pm 0,42$$

Intervalul de siguranță în acest caz va fi cuprins între 5,42 și 6,26. În acest interval de siguranță, mult mai mare, media absolută se va găsi cu o probabilitate de 99,73%, iar pragul de semnificație va fi de 0,27% ( $100\% - 99,73\% = 0,27\%$ ).

Rezultă că intervalul de siguranță nu are limite fixe ci ele se modifică în funcție de dorința cu care vrem să asigurăm rezultatele noastre, în cazul în care acceptăm un prag de semnificație mai mare, deci o probabilitate mai mare de a greși, atunci intervalul de siguranță este mai mic. Cu cât vrem să lucrăm mai precis, deci să greșim mai puțin, cu atât intervalul de siguranță crește.

### Testul de semnificație

Mărimea intervalului de siguranță depinde de faptul dacă în jurul mediei luăm o singură dată, de două sau de trei ori valoarea erorii standard. **Multiplul erorii standard (1, 2 sau 3), care determină mărimea intervalului de siguranță**, se numește test de semnificație și se notează cu litera „t”. Ca atare, la o probabilitate de 68,26% și un prag de semnificație de 31,74% valoarea lui  $t=1$ ; la 95,45% probabilitate și 4,55% prag de semnificație, valoarea lui  $t=2$ ; la o probabilitate de 99,73% și un prag de semnificație de 0,27% valoarea lui  $t=3$ . În mod obișnuit, noi nu garantăm rezultatele sau concluziile, obținute pe eșantion, cu probabilitatea cu care acestea se găsesc în interiorul intervalului de siguranță și cu contraprobabilitatea, cu probabilitatea de a greși, deci cu pragul de semnificație.

În medicină și biologie pragurile de semnificație de 31,74%, de 4,55% și de 0,27%, corespunzând valorilor lui  $t$  de 1,2 sau 3, nu se prea folosesc, în schimb, rezultatele se garantează cu pragurile de semnificație de 0,05 (5%), 0,01 (1%) și 0,001 (0,1%). Pe bază de cal-

cule s-a stabilit că pentru aceste praguri de semnificație valorile corespunzătoare ale lui  $t$  sunt de 1,96, 2,58 și 3,29. Ca atare, la pragul de semnificație de 5%, valoarea lui  $t$  va fi 1,96, iar intervalul de siguranță va fi:  $M \pm 1.96.E.S$ . La 1% prag de semnificație valoarea lui  $t$  va fi de 2,58, iar mărimea intervalului de siguranță va fi dată de  $M \pm 2,58.E.S$ . La 0,1% prag de semnificație valoarea lui  $t$  va fi de 3,29, iar intervalul de siguranță va fi dat de  $M: 3,29.E.S$ .

Aceste valori ale lui  $t$  rămân nemodificate în situația în care lucrăm pe eșantioane al căror număr de frecvențe este mai mare de 120. În situația în care lucrăm pe eșantioane cu un număr de frecvențe mai mic de 120, atunci valoarea testului de semnificație se modifică și se ia din **tabelul testului t**, pe care o găsim în cărțile de statistică.

Acest tabel (Tab.23) are 4 coloane: prima coloană pentru gradul de libertate iar celelalte trei corespund celor trei praguri de semnificație: 5%, 1%, 0,1%. Prin grad de libertate înțelegem numărul de frecvențe ale eșantionului (sub 120)-1; deci **GL = n-1**.

Tabelul 23

Tabelul testului „t”

Grad de libertate	Ce t trebuie utilizat la un prag de semnificație		
	(0,05)=5%	(0,01)=1%	(0,001)=0,1%
1	12,706	63,657	
2	4,303	9,925	31,598
3	3,182	5,841	12,941
4	2,776	4,604	8,610
5	2,571	4,032	6,859
6	2,447	3,707	5,959
7	2,365	3,499	5,405
8	2,306	3,355	5,041
9	2,262	3,250	4,781
10	2,228	3,169	4,587
11	2,201	3,106	4,437
12	2,179	3,055	4,218
13	2,160	2,977	4,140
14	2,145	2,977	4,140
15	2,131	2,947	4,073
16	2,120	2,921	4,015
17	2,110	2,898	3,965

Continuare Tab. 23

Grad de libertate	Ce t trebuie utilizat la un prag de semnificație		
	(0,05)=5%	(0,01)=1%	(0,001)=0,1%
18	2,101	2,878	3,922
19	2,093	2,861	3,883
20	2,086	2,845	3,850
21	2,080	2,831	3,819
22	2,074	2,819	3,792
23	2,069	2,807	3,767
24	2,064	2,797	3,745
25	2,060	2,787	3,725
26	2,056	2,779	3,707
27	2,052	2,771	3,690
28	2,048	2,763	3,674
29	2,045	2,756	3,659
30	2,042	2,750	3,646
35	2,030	2,724	3,591
40	2,021	2,704	3,551
50	2,008	2,678	3,496
60	2,000	2,660	3,460
70	1,994	2,648	3,435
80	1,989	2,638	3,416
90	1,986	2,631	3,402
100	1,982	2,625	3,390
120	1,980	2,617	3,373
peste	1,960	2,580	3,290

În cazul în care rezultatele cercetărilor noastre sunt exprimate procentual, deci avem o probabilitate și o contraprobabilitate de eșantion, pentru a estima intervalul în care se găsește probabilitatea colectivității generale sau probabilitatea populației ne folosim tot de valoarea erorii standard. Intervalul valoric în care se găsește probabilitatea fundamentală sau a colectivității generale se obține adunând sau scăzând din probabilitatea eșantionului sau eșantioanelor valoarea erorii standard.

În această situație eroarea standard ES este egală cu rădăcina pătrată din produsul dintre probabilitatea și contraprobabilitatea eșantionului, raportat la numărul frecvențelor acestuia, în cazul când numărul

frecvențelor eșantionului este mai mic de 120, acest produs se raportează la numărul frecvențelor minus unu.

Formula erorii standard va fi deci:

$$ES = \pm \sqrt{\frac{P_e q_e}{n}} \text{ sau } ES = \pm \sqrt{\frac{P_e q_e}{n-1}} \quad n \leq 120$$

în care:

ES = eroarea standard;

$P_e$  = probabilitatea eșantionului;

$q_e$  = contraprobabilitatea eșantionului;

n = numărul de frecvențe ale eșantionului.

### Eroarea standard procentuală

Eroarea standard în cazul valorilor medii este exprimată în unități de măsură diferite, ca și deviația standard. Ea nu poate servi deci la compararea a două sau mai multe eșantioane, sub aspectul omogenității, în asemenea situații se calculează eroarea standard procentuală, pentru a scăpa de influența unității diferite de măsură.

$$ES\% = \frac{ES}{M} \times 100$$

în care:

ES% = eroarea standard procentuală;

M = media aritmetică sau ponderată.

În general, se consideră că o eroare standard procentuală cu valori sub 1% este acceptabilă, în jurul lui 2% dubioasă și peste 3% inacceptabilă, evidențiind un eșantion neomogen, deci necorespunzător pentru a lucra pe el și a generaliza rezultatele.

În primul nostru exemplu, ES% va fi:

$$ES\% = \pm \frac{0,57}{140,5} \times 100 = \pm \frac{57}{140,5} = \pm 0,40\%$$

deci corespunzătoare.

În cel de al doilea exemplu al nostru, ES% va fi:

$$ES\% = \pm \frac{0,14}{5,84} \times 100 \pm \frac{14}{5,84} = \pm 2,4\%$$

deci necorespunzătoare.



**Eroarea maximă admisă sau eroarea limită este eroarea pe care o acceptăm anticipat în alegerea eșantionului, în funcție de natura cercetării, de precizia cu care vrem să garantăm rezultatele și de volumul sau mărimea eșantionului.** Ea reprezintă deci intervalul de siguranță în care se va încadra media universului, cu o probabilitate dată.

Eroarea maximă admisă se notează cu litera grecească delta ( $\Delta$ ) și se obține înmulțind valoarea testului de semnificație cu valoarea erorii standard.

Formula de calcul:

$$\Delta_x = t_\alpha \times ES$$

în care:

$\Delta_x$  = eroarea maximă admisă;

$t_\alpha$  = testul de semnificație;

ES = eroarea standard.

În primul nostru exemplu, admitând un prag de semnificație de 5%,  
 $\Delta_x = \pm 1,96 \times 0,57 = \pm 1,11$ .

În cel de al doilea exemplu al nostru, admitând un prag de semnificație de 5%,  
 $\Delta_x = \pm 1,96 \times 0,14 = \pm 0,27$ .

Pornind însă de la formula erorii standard:

$$ES = \pm \frac{\sigma_x}{\sqrt{n}}$$

observăm că valoarea acesteia este direct proporțională cu valoarea lui sigma și invers proporțională cu rădăcina pătrată din numărul frecvențelor. Aceasta înseamnă că urmărind o scădere a erorii standard de "n" ori, numărul de frecvențe va trebui să crească de "n<sup>2</sup>" ori.

De exemplu, dacă am efectuat o cercetare pe 100 de cazuri clinice și am stabilit o  $\sigma = \pm 0,40$ , eroarea standard va fi:

$$ES = \pm \frac{0,40}{\sqrt{100}} = \pm \frac{0,40}{10} = \pm \frac{400}{1000} = \pm 0,4$$

Pentru a reduce eroarea standard de două ori, de la 0,40 la 0,20, numărul de frecvențe trebuie să crească de 2<sup>2</sup> ori, deci de la 100 la 400 cazuri. Dacă vrem să reducem de 4 ori valoarea erorii standard, numărul frecvențelor eșantionului trebuie să crească de 4<sup>2</sup> ori, deci de la 100 la 1600 cazuri,

În general valoarea erorii maxime admise nu trebuie să fie mai mare de 1, atunci când lucrăm cu valori absolute, sau de 1%, atunci când lucrăm cu valori procentuale.

Pornind de la formula  $\Delta_x = t_\alpha \times ES$ , putem stabili volumul sau mărimea eșantionului pe care, efectuând lucrarea sau cercetarea noastră, să putem asigura rezultatele cu un prag de semnificație de 0,05 (5%), 0,01 (1%) sau 0,001 (0,1%). Cum procedăm? Înlocuim eroarea standard cu formula din care o obținem, în cazul valorilor medii și în cazul valorilor procentuale; formula devine:

$$\Delta_x = \pm t_\alpha \frac{\sigma_x}{\sqrt{n}} \quad \text{sau} \quad \Delta_x = \pm t_\alpha \sqrt{\frac{Pq}{n}}$$

Trecând pe n de partea cealaltă a egalității, vom avea:

$$\sqrt{n} = \frac{t_\alpha \sigma_x}{\Delta_x} \quad \text{sau} \quad \sqrt{n} = \frac{t_\alpha \sqrt{Pq}}{\Delta_x}$$

de unde:

$$n = \frac{t_\alpha^2 \sigma_x^2}{\Delta_x^2} \quad \text{sau} \quad n = \frac{t_\alpha^2 Pq}{\Delta_x^2}$$

Din aceste formule de obținere a mărimii eșantionului cunoaștem: valoarea lui t, întrucât aceasta va fi 1,96, 2,58 sau 3,29, în funcție de pragul de semnificație cu care dorim să garantăm rezultatele (5%, 1% sau 0,1%); valoarea erorii limită, căreia îi atribuim anticipat valorile dorite – în funcție de precizia cu care vrem să lucrăm (0-1 sau 0-1%).

Ce nu cunoaștem, atunci când ne propunem să stabilim volumul sau mărimea eșantionului, este valoarea lui sigma. Valoarea lui sigma o luăm fie din literatura de specialitate, dacă există cercetări făcute anterior de alți cercetători în domeniul în care vrem să lucrăm noi, fie o stabilim efectuând cercetarea pe un eșantion pilot alcătuit dintr-un număr restrâns de cazuri (nu mai puțin de 5) și valoarea lui sigma, astfel obținută, se introduce în formula de calcul a mărimii eșantionului.

În primul nostru exemplu, admitând o eroare limită de 0,5, dorind să garantăm rezultatele cu un prag de semnificație de 5% (deci  $t = 1,96$ ) și având o valoare a lui sigma egală cu  $\pm 6,25$ , mărimea eșantionului ar trebui să fie:

$$\sqrt{n} = \frac{1,96 \times 6,25}{0,5} = \frac{12,25}{0,5} = 24,5$$

$$n = 24,5^2 = 600,25 \text{ cazuri.}$$

Deci, pentru a putea generaliza media valorilor TA (140,5 mm Hg) cu un prag de semnificație de 5% și un delta de 0,5, era necesar să examinăm 600 persoane, nu 120 câte au fost examinate.

În cel de al doilea exemplu al nostru, admitând o eroare limită de 0,3, dorind să garantăm rezultatele cu un prag de semnificație de 5% (deci  $t = 1,96$ ) și având o valoare a lui sigma egală cu  $\pm 3,16$ , mărimea eșantionului ar trebui să fie:

$$\sqrt{n} = \frac{1,96 \times 3,16}{0,3} = \frac{6,19}{0,3} = 20,6$$

$$n = 20,6^2 = 424,36, \text{ respectiv } 425 \text{ persoane.}$$

Deci, pentru a putea generaliza media dinților afectați (5,84) cu un prag de semnificație de 5% și un delta de 0,3, era suficient să examinăm, din punct de vedere stomatologic, numai 425 persoane, nu 508 câte au fost examinate.

Dacă la aceleași valori ale lui delta și ale lui sigma am fi dorit să garantăm rezultatele noastre cu un prag de semnificație de 1%, în loc de 5%, atunci mărimea eșantionului trebuia să fie:

#### în primul exemplu

$$\sqrt{n} = \frac{2,58 \times 6,25}{0,5} = \frac{16,125}{0,5} = 32,25$$

$$n = 32,25^2 = 1040 \text{ cazuri.}$$

#### în al doilea exemplu:

$$\sqrt{n} = \frac{2,58 \times 3,16}{0,3} = \frac{8,15}{0,3} = 27,2$$

$$n = 27,2^2 = 739,81, \text{ deci } 740 \text{ de persoane.}$$

Dacă am fi dorit ca pragul de semnificație să fie 0,1% ( $t=3,29$ ), păstrând celelalte valori constante, atunci numărul de persoane examinate ar fi trebuit să fie:

#### în primul exemplu:

$$\sqrt{n} = \frac{3,29 \times 6,25}{0,5} = \frac{20,56}{0,5} = 41,12$$

de unde:

$$n = 41,12^2 = 1691 \text{ persoane.}$$

#### în al doilea exemplu

$$\sqrt{n} = \frac{3,29 \times 3,16}{0,3} = \frac{10,39}{0,3} = 34,6$$

de unde:

$$n = 34,6^2 = 1197,16 \text{ deci } 1198 \text{ de persoane.}$$

Din aceste exemple putem conchide că numărul de frecvențe ce alcătuiesc eșantionul depinde pe de o parte de valoarea pragului de semnificație și de valoarea erorii limită, iar pe de altă parte de valoarea lui sigma. Cu cât pragul de semnificație și eroarea limită vor avea valori mai mari, cu atât numărul de frecvențe va putea fi mai restrâns și făcând cercetarea pe un astfel de eșantion putem generaliza rezultatele cu o posibilitate mai mare de a greși ( $p=5\%$  și  $\text{delta} = 1$ ). Cu cât pragul de semnificație și eroarea limită vor avea valori mai mici ( $p=0,1\%$  și  $\text{delta} = 0,01$ ) cu atât numărul de frecvențe va trebui să fie mai mare pentru ca lucrând pe un astfel de eșantion să putem garanta rezultatele cu o precizie mai mare (deci cu o posibilitate de a greși foarte mică).

La aceleași valori ale pragului de semnificație și ale erorii limită, eșantionul va fi cu atât mai mare, mai numeros, cu cât sigma va avea valori mai mari și cu cât sigma va fi mai mic, ca valoare, cu atât numărul de frecvențe va putea fi mai restrâns.

În cazul în care efectuăm un studiu longitudinal al morbidității stomatologice, deci eșantionul ales trebuie urmărit o perioadă mai îndelungată de timp (cinci, zece ani), atunci pentru a ne asigura că numărul de frecvențe nu va deveni insuficient (prin plecarea din localitate sau prin deces a unor persoane), la numărul de frecvențe stabilit prin calcul adăugăm încă 10%. Spre exemplu dacă am stabilit că pentru eșantionul nostru sunt suficiente 425 persoane și acest eșantion vrem să-l reexaminăm la intervale de 2 ani, timp de zece ani, la numărul de 425 mai adăugăm 42-43 persoane (10%) și eșantionul nostru va trebui să fie de 468 persoane, în cazul când rezultatele cercetării sunt exprimate procentual, atunci pentru stabilirea mărimii eșantionului se folosește formula:

$$n = \frac{t_{\alpha}^2 Pq}{\Delta_x^2}$$

**Exemplul 1.** Într-un studiu longitudinal (prospectiv) al persoanelor, pe o perioadă îndelungată de timp (5-10 ani), pentru a ne asigura că numărul de frecvențe nu va diminua prin plecarea sau decesul unor persoane, la numărul de frecvențe stabilit inițial adăugăm 10%; dacă numărul de cazuri stabilit inițial a fost de 600, adăugând 10% vom avea nevoie de 660 cazuri.

**Exemplul 2.** Dacă ne propunem să generalizăm incidența de 98% a afecțiunilor stomatologice, cu un prag de semnificație de 5% și o eroare maximă de 1%, stabilită pe un lot de populație, atunci volumul eșantionului va fi:

$$n = \frac{1,96^2 \times 98 \cdot 2}{1} = \frac{3,84 \times 196}{1} = \frac{752,74}{1} = 753 \text{ persoane}$$

Formulele de mai sus, folosite pentru determinarea volumului eșantionului, sunt valabile pentru situația în care extragem mai multe eșantioane din colectivitatea generală sau nu cunoaștem mărimea acesteia. Atunci când extragem un singur eșantion, cum se întâmplă de obicei, și cunoaștem mărimea colectivității generale, formulele de obținere a lui  $n$  devin:

$$n = \frac{t_{\alpha}^2 \cdot \sigma_x^2}{\Delta_x^2 + \frac{t_{\alpha}^2 \cdot \sigma_x^2}{N}} \text{ sau } n = \frac{t_{\alpha}^2 Pq}{\Delta_x^2 + \frac{t_{\alpha}^2 Pq}{N}}$$

în care:  $N$  = mărimea colectivității generale.

Pentru a stabili dacă două sau mai multe rezultate obținute pe eșantioane sau dacă rezultatele obținute pe eșantioane și cele obținute pe colectivitatea generală sunt asemănătoare sau, din contră, diferă semnificativ între ele, ne folosim de comparația statistică.

### Compararea statistică

Am văzut că în medicină și biologie efectuăm cercetări pe eșantion și rezultatele obținute le extrapolăm la nivelul colectivității generale, acestea fiind valabile ca și când am fi efectuat cercetarea pe întreaga colectivitate. Rezultatele obținute pe eșantion diferă însă de la un eșantion la altul, fiind cu atât mai reprezentative cu cât eșantionul este mai omogen și mai bine reprezentat numeric.

Pentru a stabili dacă două sau mai multe rezultate obținute pe eșantioane sau dacă rezultatele obținute pe eșantioane și cele obținute

te pe colectivitatea generală sunt asemănătoare sau, din contră, diferă semnificativ între ele, ne folosim de comparația statistică.

### Compararea mediei unei colectivități generale cu media unui eșantion prin criteriul diferenței și erorii diferenței

Când dispunem de o medie a colectivității generale cu care comparăm o medie de eșantion, se procedează la stabilirea diferenței între cele două medii scăzând valoarea cea mai mică din cea mai mare.

Se calculează după aceea valoarea erorii standard pentru eșantion. Cum se face interpretarea? Dacă diferența între cele două medii – absolută și de eșantion – este mai mare decât valoarea dublă sau triplă a erorii standard, atunci diferența este **semnificativă statistic**.

$D > 2(3) \times ES$  diferență **semnificativă din punct de vedere statistic** sau

$D > U_{\alpha} \times ES$   $U_{\alpha}$  coeficient tabelar în funcție de riscul  $\alpha$ ; a cărui valori se vor citi în Tab. 24:

Tabelul 24

Limitele de semnificație a distribuției  $U_{\alpha}$

$\alpha$	$U_{\alpha}$	$\alpha$	$U_{\alpha}$
0,00		0,09	1,695
0,01	2 576	0,001	3,29053
0,02	2.326	0,0001	3.89059
0,03	2,170	0,00001	4,41717
0,04	2.054	0,000001	4,89164
0,05	1,950	0,0000001	5.32672
0,06	1,881	0,00000001	5,73073
0,07	1,812	0,000000001	6,10941
0,08	1,751		

Acest lucru înseamnă că din colectivitatea din care am extras eșantionul există anumiți factori (de mediu fizic, de alimentație, de muncă, etc.) care au determinat această diferență de valori.

Dacă din contră, diferența între media universului și media eșantionului este mai mică decât valoarea dublă sau triplă a erorii standard, atunci diferența între cele două medii este nesemnificativă din punct de vedere statistic, este întâmplătoare și se datorește faptului că am efectuat cercetarea pe un număr limitat de frecvențe.

În această situație nu trebuie să căutăm, cu orice preț, o justificare pentru diferența existentă căci măbind numărul de frecvențe al eșantionului această diferență dispore.

$D < 2(3) \times ES$  diferența **nesemnificativă** din punct de vedere statistic.

**Exemplul 3.** Dacă am examinat din punct de vedere al valorilor TA sistolice populația adultă a unei colectivități generale și am obținut o medie a TA ( $\mu=147,5$  mm Hg), iar pe un eșantion de 120 de persoane am obținut o medie  $M = 140,5$  mm Hg și o eroare standard (ES) de  $\pm 0,57$  mm Hg, stabilind diferența dintre cele două medii obținem:

$$D = \mu - M = 147,5 - 140,5 = 7 \text{ mm Hg}$$

$$D > 2(3) \times ES \text{ sau } D > U_{\alpha} \times ES$$

$D = 7 \text{ mm Hg} > U_{\alpha} \times 0,57$ , iar pentru  $\alpha = 5\%$  vom citi în tabelul distribuției  $U_{\alpha}$  valoarea de 1,98. Deci  $7 > 1,98 \times 0,57$  deci diferența dintre cele două medii, a colectivității generale și a eșantionului este semnificativă statistic pentru un risc  $\alpha = 5\%$ . Dacă diferența dintre cele două medii ar fi fost mai mică decât  $U_{\alpha} \times ES$  atunci diferența ar fi fost nesemnificativă statistic și media eșantionului s-ar fi încadrat în limitele admisibile de variabilitate față de media colectivității generale, deci s-ar fi situat în interiorul intervalului de siguranță, în acest caz diferența dintre cele două medii s-ar fi datorat unui număr prea mic de cazuri examinate în cadrul eșantionului și nu unor condiții de mediu diferite, existente în cele două colectivități.

**Exemplul 4.** Dacă examinându-se din punct de vedere stomatologic populația unui județ (colectivitatea generală), s-a obținut o medie ( $M_a$ ) egală cu 7,50 dinți afectați, iar noi pe eșantionul nostru am găsit o medie de 5,84 dinți afectați și o eroare standard de  $\pm 0,14$  dinți afectați, stabilind diferența între cele două medii obținem:

$$D = \mu - M \text{ sau } D = M_a - M_e; D = 7,50 - 5,84 = 1,66$$

$D(1,66) > 2 \times 0,14$  ( $3 \times 0,14$ ), deci diferența între cele două medii este **semnificativă** din punct de vedere statistic.

Dar dacă media pe întreaga colectivitate ar fi fost de 6,2 dinți afectați și media eșantionului nostru de 5,84 dinți afectați, eroarea standard având aceeași valoare de 0,14, atunci:

$$D = 6,10 - 5,84 = 0,26; \quad D = 0,26 \text{ dinți afectați}$$

$D(0,26) < (2 \times 0,14)$  și decât ( $3 \times 0,14$ ), deși diferența ar fi fost nesemnificativă din punct de vedere statistic.

### Compararea mediilor a două eșantioane diferite (testul Student)

Când cele două medii care se compară sunt medii de eșantion, stabilim diferența dintre cele două medii:

$D = Me_1 - Me_2$ . Calculăm apoi eroarea diferenței ( $\sigma_D$ ) dintre cele două valori medii:

$$\sigma_D = \pm \sqrt{ES_1^2 + ES_2^2}$$

în care:  $\sigma_D$  = eroarea diferenței

$ES_1^2$  = eroarea standard a eșantionului unu la pătrat

$ES_2^2$  = eroarea standard doi la pătrat.

Interpretarea se face astfel: dacă diferența între cele două medii de eșantion este mai mare decât valoarea dublă sau triplă a erorii diferenței atunci diferența între medii este semnificativă din punct de vedere statistic, deci cele două eșantioane luate în studiu, nu aparțin aceleiași colectivități generale. Unul din eșantioane a avut deci condiții diferite de celălalt. Ca atare în cadrul discutării rezultatelor trebuie să stabilim care au fost acele condiții care au făcut ca rezultatele să difere semnificativ.

Dacă, din contră, diferența dintre cele două medii de eșantion este mai mică decât valoarea dublă sau triplă a erorii diferenței, atunci diferența este nesemnificativă din punct de vedere statistic, este întâmplătoare și se datorește diferenței de eșantionaj. Ca atare nu trebuie să căutăm neapărat o explicație a acestei diferențe.

**Exemplul 1.** Pe două eșantioane extrase din două colectivități diferite s-au măsurat valorile TA sistolice și am găsit pentru primul eșantion alcătuit din 120 persoane o medie  $M_1 = 140,5$  mm Hg și o eroare standard egală cu  $ES_1 = \pm 0,57$  mm Hg iar pentru al doilea eșantion alcătuit dintr-un număr de 160 de persoane am găsit o medie a TA sistolice  $M_2 = 154,5$  mm Hg și o eroare standard  $ES_2 = 1,2$ . Stabilind diferența dintre medii obținem:

$$|M_1 - M_2| = |140,5 - 154,5| = 14 \text{ mm Hg, iar}$$

$$\sigma_D = \pm \sqrt{ES_1^2 + ES_2^2} = \sqrt{(\pm 0,57)^2 + (\pm 1,2)^2} = 1,32$$

Diferența dintre cele două medii de eșantion fiind egală cu 14, iar eroarea diferenței înmulțită cu 1,96+2 care pentru un risc  $\alpha$  (coeficient tabelar) de 5% și

$D = 14 \text{ mm Hg} > 1,96 \sigma_D$  înseamnă că diferența este semnificativă din punct de vedere statistic. Dacă diferența dintre cele două medii a

celor două eșantioane ar fi fost mai mică decât valoarea  $1,96 \cdot \sigma_D$  atunci diferența ar fi fost ne semnificativă statistic, ambele eșantioane provenind din colectivități cu condiții asemănătoare.

**Exemplul 2.** Pe două eșantioane, extrase din două colectivități limitrofe, deci cu condiții asemănătoare (aceiași factori de mediu fizic, același mediu de alimentație, aceleași condiții de muncă, același nivel cultural-sanitar etc.), se stabilește media dinților absenți și se găsește că la primul eșantion, de 250 persoane ( $e_1$ ) media ( $Me_1$ ) este de 5,5 dinți absenți, cu o eroare standard de  $\pm 0,7$  iar la al doilea eșantion ( $e_2$ ), de 200 de persoane, media  $Me_2$  a dinților absenți este de 6,2 iar eroarea standard de  $\pm 0,2$ . Diferența între medii este:

$$D = Me_2 - Me_1 \quad D = 6,2 - 5,5 = 0,7 \text{ dinți absenți.}$$

$$\sigma_D = \pm \sqrt{0,7^2 + 0,2^2} = \pm \sqrt{0,49 + 0,04} = \pm \sqrt{0,53} = \pm 0,72$$

dinți absenți.

Diferența dintre medii (0,7) fiind mai mică decât valoarea diferenței ( $\sigma_D = 0,72$ ) înseamnă că între cele două medii nu există o diferență semnificativă din punct de vedere statistic.

**Compararea a două valori calitative prin criteriul diferenței și erorii diferenței** (procentuale)

Când avem de comparat o probabilitate a colectivității generale cu o probabilitate de eșantion, pentru a stabili semnificația sau lipsa de semnificație a diferenței dintre aceste probabilități, procedăm în felul următor:

Stabilim diferența între cele două probabilități (probabilitatea în colectivitatea generală și probabilitatea eșantionului). Calculăm apoi valoarea erorii standard pentru eșantion, conform formulei:

$$ES = \pm \sqrt{\frac{P_e q_e}{n-1}}$$

Interpretarea se face astfel: dacă diferența dintre cele două probabilități este mai mare decât valoarea dublă sau triplă ( $U_\alpha$ ), a erorii standard, atunci diferența este semnificativă și trebuie să stabilim căror factori se datorează.

$$D > U_\alpha \text{ sau}$$

$$D > 2(3)ES, \text{ atunci diferența este semnificativă statistic.}$$

Dacă din contră, diferența dintre cele două probabilități este mai mică decât valoarea dublă sau triplă a erorii standard, atunci diferența este ne semnificativă, este întâmplătoare și se datorește diferenței de eșantionaj.

**Exemplul 1.** La 278362 născuți vii în anul Z în țara S proporția băieților a fost 51,4% iar a fetelor de 43,5% eroarea standard a colectivității generale:

$$ES = \pm \sqrt{\frac{Pq}{n}} = \pm \sqrt{\frac{51,4 \cdot 48,6}{278362}} = \pm 0,009$$

Pe un eșantion de 16901 născuți vii în anul Z în țara F proporția băieților a fost de 51,2% iar a fetelor – 48,8%.

$$ES_1 = \pm \sqrt{\frac{Pq}{n}} = \pm \sqrt{\frac{51,2 \cdot 48,8}{16901}} = \pm 0,14$$

Diferența între cele două probabilități  $D = P - p = 51,4\% - 51,2\% = 0,2\%$  iar  $ES$  în eșantionul nostru este  $\pm 0,14\%$  diferența  $D = 0,2\%$  este mai mică decât  $U_\alpha \cdot ES = 1,96 \times 0,14$  deci probabilitatea băieților este aceeași în eșantion ca și în colectivitatea generală, diferența fiind ne semnificativă statistic. De fapt pe glob diferența între cele două sexe este aproximativ egală.

**Exemplul 2.** Dacă pe populația unui județ s-a stabilit că probabilitatea parodontopatiei parțiale este de 60% iar pe un eșantion de 350 de locuitori s-a stabilit că parodontopatia parțială este prezentă la 80% din populație, atunci diferența este:

$$D = 80\% - 60\%; \quad D = 20\%$$

Eroarea standard pentru eșantion va fi:

$$ES = \pm \sqrt{\frac{80 \times 20}{360}} = \pm \sqrt{\frac{1600}{360}} = \pm \sqrt{4,57} = \pm 2,13\%$$

$D (20\%) > 2 \times 2,13\% (3 \times 2,13\%) =$  diferența este semnificativă din punct de vedere statistic,

Între cele două probabilități există o diferență semnificativă din punct de vedere statistic, asigurând aceasta afirmație cu un prag de semnificație de 5%, respectiv 1%.

Când avem de comparat două probabilități obținute pe 2 eșantioane, procedăm în felul următor:

Stabilim diferența dintre cele două probabilități de eșantion:

$$D = Pe_1 - Pe_2$$

Calculăm după aceea eroarea diferenței dintre cele două probabilități, după formula:

$$\sigma_D = \pm \sqrt{\frac{P_{e1}q_{e1}}{n_1 - 1} + \frac{P_{e2}q_{e2}}{n_2 - 1}}$$

Dacă diferența dintre cele două probabilități de eșantion este mai mare decât valoarea dublă sau triplă a erorii diferenței, atunci diferența este semnificativă din punct de vedere statistic.

$D > 2(3)\sigma_D$  = diferență semnificativă statistic.

Dacă din contră, diferența dintre probabilitățile de eșantion este mai mică decât valoarea dublă sau triplă a erorii diferenței, atunci diferența este nesemnificativă din punct de vedere statistic, este întâmplătoare și se datorește diferenței de eșantionaj.

$D < 2(3)\sigma_D$  = diferență nesemnificativă din punct de vedere statistic.

### Compararea a două valori medii sau alternative prin criteriul testului student

Semnificația sau lipsa de semnificație a două valori medii sau procentuale, obținute pe eșantioane, se poate stabili și pe o altă cale și anume pe calea testului Student.

Atunci când, în cadrul eșantionării, am discutat și am definit testul  $t$  am spus că testul  $t$  este multiplul erorii standard și că cu ajutorul lui stabilim mărimea intervalului de siguranță. Acest test  $t$  a cărui valoare – în cazul în care lucrăm pe eșantioane cu un număr mai mic de 120 frecvențe – o luăm din tabela testului  $t$ , se numește “ $t$  tabelar”. Dar valoarea lui  $t$  o mai putem obține și prin calcul după formula:

$$t = \frac{D}{\sigma_D}$$

Acest  $t$  obținut raportând valoarea diferenței la eroarea sa, se numește “ $t$  calculat”.

În cazul în care dorim să apreciem semnificația sau lipsa de semnificație a diferenței dintre două valori medii sau două probabilități, obținute pe eșantioane, procedăm în felul următor: calculăm valoarea lui  $t$ , raportând diferența dintre cele două medii la eroarea sa.

$$t_{calc.} = \frac{D}{\sigma_D} = \frac{|M_1 - M_2|}{\sqrt{ES_1^2 + ES_2^2}}$$

în care:

$t$  = testul de semnificație

$D$  = diferența dintre valorile medii sau procentuale

$\sigma_D$  = eroarea diferenței.

Stabilim apoi valoarea lui “ $t$  tabelar” în felul următor: dacă numărul frecvențelor celor două eșantioane depășește suma de 120 atunci valoarea lui “ $t$  tabelar” o cunoaștem ca fiind 1,96 pentru un  $p = 0,05$  (5%); 2,58 pentru un  $p = 0,01$  (1%) sau 3,29 pentru un  $p = 0,001$  (0,1%).

Dacă numărul însumat de frecvențe al celor două eșantioane ce se compară este mai mic de 120 de frecvențe, atunci valoarea lui “ $t$  tabelar” o citim în tabela testului  $t$  în gradul de libertate dat de numărul însumat de frecvențe minus 2.

Interpretarea se face în felul următor: dacă valoarea lui “ $t$  calculat” este mai mare decât valoarea lui “ $t$  tabelar” atunci diferența între cele două valori medii sau între cele două probabilități este semnificativă din punct de vedere statistic.

“ $t$  calculat”  $>$  “ $t$  tabelar” = diferența **semnificativă** statistic. Dacă din contra, valoarea lui “ $t$  calculat” este mai mică decât valoarea lui “ $t$  tabelar”, atunci diferența dintre cele două medii sau dintre cele două probabilități este nesemnificativă din punct de vedere statistic.

“ $t$  calculat”  $<$  “ $t$  tabelar” = diferență **nesemnificativă**. Pentru exemplificare și verificare, în același timp, vom lua aceleași exemple pe care le-am apreciat, sub aspectul semnificației diferenței și cu ajutorul erorii diferenței.

În cazul comparării a două medii obținute pe eșantioane diferența între medii a fost de 0,7 dinți absenți iar eroarea diferenței a fost de  $\pm 0,72$  dinți absenți atestând o diferență nesemnificativă între medii, încercând să stabilim semnificația sau lipsa de semnificație pe calea testului  $t$ , obținem următoarele rezultate:

$$t_{calculat} = \pm \frac{D}{\sigma_D} = \pm \frac{0,70}{0,72} = \pm 0,09$$

t tabelar, în cazul exemplului nostru – datorită faptului că numărul de frecvențe este mult mai mare decât 120, are valori constante (1,96; 2,58; 3,29).

t calculat având o valoare mult mai mică (0,09) decât valoarea lui t tabelar (1,96; 2,58; 3,29) testează, așa după cum am arătat, o diferență ne semnificativă din punct de vedere statistic între cele două medii de eșantion.

În cazul comparării a două valori **procentuale**, de eșantion cu ajutorul **testului “t”** stabilim valoarea lui “t calculat” și valoarea lui “t tabelar”.

Spre exemplu, s-a examinat un eșantion de 84 de elevi și s-a stabilit că frecvența cariei dentare este de 90%, iar examinarea unui alt eșantion de 60 de elevi, extrase dintr-o altă colectivitate școlară, a evidențiat prezența cariei dentare doar la 60% dintre aceștia. Spre a stabili dacă diferența dintre cele două probabilități este semnificativă din punct de vedere statistic sau nu, stabilim mai întâi diferența dintre cele două probabilități:

$D = P_{e1} - P_{e2}$  în care;

D = diferența

$P_{e1}$  = probabilitatea primului eșantion

$P_{e2}$  = probabilitatea celui de al doilea eșantion

D = 90% – 60% = 30%

Diferența este egală cu 30%, calculăm apoi eroarea diferenței conform formulei:

$$\sigma_D = \pm \sqrt{\frac{P_{e1}q_{e1}}{n_1 - 1} + \frac{P_{e2}q_{e2}}{n_2 - 1}}$$

$$\sigma_D = \pm \sqrt{\frac{90 \times 10}{83} + \frac{60 \times 40}{59}} = \pm \sqrt{10,84 + 30,67} = \pm \sqrt{41,51} = \pm 6,40\%$$

Diferența dintre cele două probabilități (30%) fiind mai mare decât triplul erorii diferenței ( $3 \times 6,40\% = 19,20\%$ ) este semnificativă din punct de vedere statistic un prag de semnificație sub 1%.

D (30%) >  $\sigma_D$  (19,20%) = diferență semnificativă statistic pentru un  $p < 0,01$ .

Folosind metoda testului t evidențiem aceeași semnificație a diferenței:

$$t_{\text{calculat}} = \pm \frac{D}{\sigma_D} = \pm \frac{30}{6,40} = \pm 4,68$$

$t_{\text{tabelar}} = 1,96; 2,58; 3,29$  în funcție de pragul de semnificație.

Interpretarea “t calculat” (4,68) fiind mai mare decât oricare din valorile lui “t tabelar” evidențiază aceeași diferență semnificativă între cele două probabilități de eșantion.

## 2.12. Compararea valorilor absolute sau a distribuțiilor de frecvență: testul $X^2$

Când avem de comparat între ele valori absolute sau distribuții de frecvențe, folosim **testul  $X^2$  (chi pătrat) al lui Pearson**. Testul  $X^2$  se obține însumând rapoartele dintre pătratul diferențelor stabilite între frecvențele colectivității generale (frecvențele teoretice) și frecvențele observate și frecvențele colectivității generale.

Formula de calcul:

$$X^2 = \sum \frac{(FT - FO)^2}{FT}$$

în care:

$X^2$  = testul de comparație sau de concordanță chi pătrat;

FT = frecvențele universului sau frecvențele teoretice obținute prin calcul;

FO = frecvențele observate sau frecvențele eșantionului nostru.

În situația în care avem o distribuție a colectivității **generale** pe care o comparăm cu **distribuția unui eșantion**, atunci introducem datele respective în formulă și calculăm valoarea lui  $X^2$ , în situația în care nu avem datele colectivității generale, distribuția frecvențelor teoretice o obținem prin calcul – de aici denumirea de frecvențe teoretice – așa cum vom vedea ulterior.

Valoarea lui  $X^2$ , obținută pe baza formulei de mai sus, se numește valoarea lui “ $X^2$  calculat”. Pentru a putea stabili dacă între distribuția colectivității generale și cea a eșantionului observat de noi este sau nu o diferență semnificativă din punct de vedere statistic, avem nevoie – ca și în cazul comparației cu ajutorul testului t – de valoarea lui “ $X^2$  tabelar”. Această valoare o luăm din **tabelul testului  $X^2$**  (Tab. 25), care este asemănătoare cu tabelul „Testului t”.

Tabelul 25

Tabelul testului  $X^2$ 

Grad de libertate	Prag de semnificație		
	(0,05) = 5%	(0,01) = 1%	(0,001) = 0,1%
1	3,8	6,6	10,8
2	5,9	9,2	13,8
3	7,8	11,3	16,3
4	9,4	13,3	18,5
5	11,1	15,1	20,5
6	12,6	18,5	22,5
7	14,1	18,5	24,3
8	15,5	20,1	26,1
9	16,9	21,7	27,9
10	18,3	23,2	29,6
11	19,7	24,7	31,3
12	21,0	26,2	32,9
13	22,4	27,7	34,5
14	23,7	29,1	36,1
15	25,0	30,6	37,7
16	26,3	32,0	39,3
17	27,6	33,4	40,8
18	28,9	34,8	42,3
19	30,1	36,2	43,8
20	31,4	37,6	45,3
21	32,7	38,9	46,8
22	33,9	40,3	48,3
23	35,2	41,6	49,7
24	36,4	43,0	51,2
25	37,7	44,3	52,6
26	38,9	45,6	54,1
27	40,1	47,0	55,5
28	41,3	48,3	56,9
29	42,6	49,6	58,3
30	43,8	50,9	59,7
40	55,8	63,7	73,4
50	67,5	76,2	86,7
60	79,1	88,4	99,6
70	90,5	100,4	112,3
80	101,9	112,3	124,8
90	113,1	124,1	137,2
100	124,3	135,8	149,4

Valoarea lui  $X^2$  tabelar se citește în tabelul testului  $X^2$  ținând seama de gradul de libertate și de pragul de semnificație cu care vrem să garantăm semnificația sau lipsa de semnificație dintre cele două distribuții de frecvențe. Prin grad de libertate în tabelul testului  $X^2$  înțelegem produsul dintre numărul rândurilor tabelului (în care am introdus distribuțiile de frecvență) minus 1 și numărul coloanelor acestui tabel minus 1.

$$GL = (nrR - 1) - (nrC - 1)$$

**Exemplul 1.** Examinând întreaga populație adultă a unei colectivități de 1337 locuitori sub aspectul valorilor tensiunii arteriale sistolice și a unui eșantion de 120 de persoane am obținut următoarele distribuții de frecvențe pe grupe de valori ale tensiunii arteriale sistolice (Tab. 26).

Tabelul 26

Distribuții de frecvențe pe grupe de valori ale tensiunii arteriale sistolice

$x_i$ =variabila TA sistolică	$f_o$ = frecvențe observate	$f_t$ =frecvențe teoretice
127,5	6	180
132,5	17	210
137,5	31	290
142,5	42	320
147,5	17	230
152,5	5	65
157,5	2	42
	$\Sigma=120$	$\Sigma=1337$

Lăsând la o parte prima coloană a valorilor variabilei tensiunii arteriale sistolice și primul rând cu denumirea coloanelor, tabelul nostru conține 7 rânduri și două coloane, deci valoarea lui  $X^2$  tabelară o vom citi din tabelul  $X^2$  la gradul de libertate:  $GL = (nr. \text{ de rânduri} - 1) \times (nr. \text{ de coloane} - 1) = (7 - 1) \times (2 - 1) = 6$ . Valoarea tabelară a lui  $X^2$  pentru  $GL = 6$  și un risc  $\alpha = 5\%$  este de 12,6. Pentru un risc  $\alpha = 1\%$  este de 18,5 și pentru un risc  $\alpha = 1\%$  este de 22,5. 91 %  $X^2$  calculat se obține:

$$X^2 = \frac{(180-6)^2}{180} + \frac{(210-17)^2}{210} + \frac{(290-31)^2}{290} + \frac{(320-42)^2}{320} + \frac{(230-17)^2}{230} + \frac{(65-5)^2}{65} + \frac{(42-2)^2}{42} = \frac{30276}{180} + \frac{37249}{210} + \frac{67081}{290} + \frac{77284}{320} + \frac{45369}{230} + \frac{3600}{65} + \frac{1600}{42} = 168,2 + 177,3 + 231,3 + 241,5 + 197,2 + 55,3 + 38 = 1108,8$$



Deci valoarea lui  $X^2$  calculat (1108,8) este mult mai mare decât valorile găsite în tabelul  $X^2$  pentru oricare din cele trei riscuri asumate (5%, 1%, 0,1%), deci diferența între cele două distribuții de frecvențe teoretice și observate este semnificativă din punct de vedere statistic, denotând că în cadrul eșantionului există condiții diferite (factori de mediu, alimentari, de muncă etc.) care au determinat această diferență, față de cele din colectivitatea generală. Dacă valoarea lui  $X^2$  calculat ar fi fost mai mică decât valoarea lui  $X^2$  tabelar atunci diferența între cele două distribuții de frecvențe ar fi fost nesemnificativă din punct de vedere statistic, eșantionul provenind din aceeași colectivitate generală, reflectând condițiile acesteia.

**Exemplul 2.** Dacă am stabilit că pe teritoriul unui spital populația adultă (39.620 locuitori) se distribuie în raport cu dinții extrași ca în tabelul alăturat (coloana FT), iar examinarea unui eșantion de 5.500 persoane adulte evidențiază distribuția din coloana FO, atunci tabelul nostru – lăsând la o parte prima coloană, coloana variantelor, și primul rând, cu denumirea coloanelor, are 6 rânduri și 2 coloane, deci valoarea lui  $X^2$  tabelar o vom citi din tabelul testului  $X^2$  la gradul de libertate 5 (Tab. 27).

Tabelul 27

Determinarea gradelor de libertate.

Nr. dinți absenți	Nr. pers. cot. gen.	Nr. pers. eșantion
X	FT	FO
1	2850	860
2	5250	920
3	10320	1130
4	15810	1620
5	3860	650
6	1530	320

$$GL = (6-1) \times (2-1) = 5$$

La acest grad de libertate, valoarea lui  $X^2$  tabelar va fi de: 11,1 pentru pragul de semnificație 0,05 (5%); 15,1 pentru pragul de semnificație 0,01 (1%); și 20,5 pentru pragul de semnificație 0,001 (0,1%).

Valoarea lui  $X^2$  calculat o obținem prin calcul. Introducând datele în formula lui  $X^2$  calculat, avem:

$$\begin{aligned} X^2 \text{ calculat} &= (2850 - 860)^2 : 2850 + (5250 - 920)^2 : 5250 + (10320 - 1130)^2 : 10320 + \\ &+ (15810 - 1620)^2 : 15810 + (3860 - 650)^2 : 3860 + (1530 - 320)^2 : 1530 = 1990^2 : 1530 = \\ &= 3960100 : 2850 + 18748900 : 5250 + 84456100 : 10320 + 201356100 : 15810 + 10304100 : \\ &: 3860 + 1464100 : 1530 = 1389 + 3571 + 8183 + 12735 + 2669 + 956 = 29503 \end{aligned}$$

Deci  $X^2$  calculat are o valoare de 29503.

Interpretarea diferenței dintre cele două distribuții de frecvențe (a colectivității generale și a eșantionului nostru) se face comparând cele două valori ale lui  $X^2$  (obținute prin calcul și din tabele), ca și în cazul testului t. Dacă valoarea lui  $X^2$  calculat este mai mare decât a lui  $X^2$  tabelar, atunci între cele două distribuții de frecvențe este o diferență semnificativă din punct de vedere statistic. Dacă, din contră, valoarea lui  $X^2$  calculat este mai mică decât valoarea lui  $X^2$  tabelar, diferența între distribuții este nesemnificativă, în exemplul nostru  $X^2$  calculat, având valoarea de 38116, adică mai mare decât a lui  $X^2$  tabelar, testează o diferență semnificativă din punct de vedere statistic între cele două distribuții de frecvențe.

În situația în care nu cunoaștem distribuția colectivității generale, stabilim prin calcul valorile frecvențelor teoretice.

**Exemplul 3.** Într-o secție de oncologie au fost internați cu diagnostic de cancer pulmonar un număr de 78 de bolnavi. Dintre aceștia 69 erau bărbați și 9 femei. Din cei 69 bărbați bolnavi de cancer pulmonar 39 erau fumători iar 30 nefumători. Dintre cele 9 femei cu cancer pulmonar 6 erau fumătoare și 3 nefumătoare. Pentru a stabili dacă între bolnavii fumători și nefumători cu cancer bronhopulmonar există diferențe semnificative din punct de vedere statistic sau nu, vom introduce datele în Tab. 28.

Tabelul 28

Numărul de bolnavii fumători și nefumători cu cancer bronhopulmonar

Cancer pulmonar	Fumători		Nefumători		Total
	FO	FT	FO	FT	
Bărbați	39=a	39,8=A	30=c	29,2=C	69=a+c
Femei	6=b	5,2=B	3=d	3,8=D	9=b+d
Total	45=a+b	45=A+B	33=c+d	33=C+D	78=a+b+c+d=n

Admițând de la început că între fumători și nefumători nu există diferențe statistice, deci admițând ipoteza nulă ( $H_0$ ), pe baza unei re-

guli de trei simple, pornind de la valorile frecvențelor observate se pot stabili valorile frecvențelor teoretice astfel:

Dacă la 78 bolnavi avem 45 fumători  
la 69 bolnavi... avem x fumători

$$x = \frac{69 \times 45}{78} = 39,8(A)$$

În mod asemănător vom obține prin calcul și celelalte valori (B, C, D), pe care le vom introduce în tabel corespunzător frecvențelor teoretice. Disponând în acest fel atât de valorile frecvențelor observate cât și de frecvențele teoretice putem prin formula menționată anterior:

$$X_c^2 = \sum \frac{(FO - FT)^2}{FT}$$

se obține valoarea lui  $X^2$  calculat astfel:

$$X_c^2 = \sum \frac{(39 - 39,8)^2}{39,8} + \frac{(30 - 29,2)^2}{29,2} + \frac{(6 - 5,2)^2}{5,2} + \frac{(3 - 3,8)^2}{3,8} =$$

$$\frac{0,64}{39,8} + \frac{0,64}{29,2} + \frac{0,64}{3,8} = 0,016 + 0,0219 + 0,123 + 0,168 = 0,3289 = 0,33$$

deci  $X^2$  calculat = 0,33.

Pentru a vedea dacă diferența dintre bărbații și femeile fumătoare sau între bărbații și femeile nefumătoare este sau nu o diferență semnificativă din punct de vedere statistic comparăm valorile  $X_c^2 = 0,33$  cu valoarea  $X_{\alpha}^2$  din tabele care pentru un risc  $\alpha = 5\%$  este de 3,8, pentru  $\alpha = 1\%$  este 6,6, iar pentru  $\alpha = 0,1\%$  este 10,8. În exemplul nostru  $X^2$  calculat este mai mic decât  $X^2$  tabelar pentru oricare din riscurile  $\alpha$  de mai sus asumate, deci diferența este nesemnificativă statistic. Dacă  $X^2$  calculat ar fi fost mai mare decât  $X^2$  tabelar atunci diferența ar fi fost semnificativă statistic, ceea ce ar fi demonstrat că sexul joacă un rol important în cancerul bronhopulmonar.

**Exemplul 4.** Examinând din punct de vedere stomatologic un eșanțon de 810 elevi (470 băieți și 340 fete) dintr-o colectivitate școlară, am constatat următoarele:

=> dintre cei 470 băieți examinați 250 prezentau carii dentare, iar 220 erau fără carii;

=> la fete 180 prezentau carii dentare iar 160 erau fără carii.

Vrem să știm dacă între băieții cu carii și fetele cu carii există o diferență semnificativă din punct de vedere statistic sau nu. Cu alte cuvinte, vrem să vedem dacă această afecțiune, caria dentară, are o predispoziție legată de sex sau din contră sexul nu are nici o influență în determinarea cariei dentare.

Pentru a putea calcula valoarea lui  $X^2$  avem nevoie de frecvențele teoretice. Cum le obținem? Introducem datele cunoscute, deci frecvențele observate, într-un tabel quadruplu (un tabel cu două intrări și cu ieșiri: cu 2 rânduri și 2 coloane) (Tab. 29).

Tabelul 29

Numărul de elevi cu carie și fără carie în funcție de sex

LOTUL DE ELEVII	ELEVII CU CARIILE		ELEVII FĂRĂ CARIILE		TOTAL ELEVII
	FO	FT	FO	FT	
Băieți	250	249,50	220	220,50	470
Fete	180	480,50	160	159,50	340
TOTAL	430	430	380	380	810

Frecvențele teoretice le obținem prin calcul. Considerăm că între incidența cariei dentare la băieți și la fete nu există nici un fel de deosebire. Acceptăm, deci, de la început, principiul ipotezei nule și universul nostru este reprezentat de numărul total al frecvențelor. Cu ajutorul regulii de trei simple obținem frecvențele teoretice, făcând raționamentul: dacă la cei 810 elevi examinați corespund 430 elevi cu carii, dintre cei 470 băieți examinați câți vor avea carii?

$$X = 470 \times 430 : 810 = 249,50$$

Deci, frecvențele teoretice pentru băieții cu carii, 249,50, le introducem în căsuța corespunzătoare a tabelului.

Dacă la cei 810 elevi au fost găsiți 380 fără carii, la cei 470 băieți câți vor fi fără carii?

$$X = 470 \times 380 : 810 = 220,50$$

Același raționament îl facem și pentru obținerea frecvențelor teoretice ale fetelor cu și fără carii:

$$X = 340 \times 430 : 810 = 180,50$$

deci frecvențele teoretice ale fetelor cu carii, 180,50, le introducem în căsuța corespunzătoare a tabelului.

$$X = 340 \times 380 : 810 = 159,50$$

deci frecvențele teoretice ale fetelor fără carii le introducem în căsuța corespunzătoare a tabelului nostru.

Având acum frecvențele teoretice și pe cele observate, le introducem în formula lui  $X^2$  și obținem valoarea lui  $X^2$  calculat:

$$X^2 \text{ calculat} = (249,50 - 250) : 249,50 + (220,50 - 220) : 220,50 + (180,50 - 180) : 180,50 + (159,50 - 160) : 159,50 = (-0,50) : 249,50 + 0,50 : 220,50 + 0,50 : 180,50 + (-0,50) : 159,50 = 0,25 : 249,50 + 0,25 : 220,50 + 0,25 : 180,50 + 0,25 : 159,50 = 0,0010 + 0,0011 + 0,0015 = 0,0049,$$

deci  $X^2$  calculat = 0,0049.  $X^2$  tabelar, având în vedere că datele noastre le-am înscris într-un tabel cu două rânduri și cu două coloane, îl vom citi din tabelul testului  $X^2$  la gradul de libertate 1 și va avea valoarea: 3,84 pentru pragul de semnificație de 0,05 (5%); 6,63 pentru pragul de semnificație de 0,01 (1%); și 10,80 pentru pragul de semnificație de 0,001 (0,1%).  $X^2$  calculat având o valoare mult mai mică (0,0049) decât valoarea lui  $X^2$  tabelar (3,84; 6,63; 10,80) testează lipsa de semnificație a diferenței dintre frecvența cariei dentare la băieți și la fete.

Cu alte cuvinte, caria dentară este la fel de frecventă atât la băieți cât și la fete, deci nu are o predispoziție de sex.

### Calea de obținere a lui $x^2$ calculat

Situația în care frecvențele teoretice trebuie să le obținem prin calcul este anevoioasă, presupunând calcule laborioase, în această situație putem folosi o altă formulă, mai simplă, în activitatea practică și anume:

$$X^2 \text{ calculat} = \frac{(ad - bc)^2}{n_1 n_2 n_3 n_4} N(-1)$$

Această formulă rezultă din desfacerea binomului lui Newton  $(a+b)^n$ .

**Exemplul 5.** Bolnavii de cancer pulmonar fumători și nefumători (Tab.30).

Tabelul 30.

Numărul de bolnavii fumători și nefumători cu cancer bronhopulmonar

Cancer pulmonar	Fumători	Nefumători	Total
Bărbați	39=a	30=b	69=a+b=n <sub>1</sub>
Femei	6=c	3=d	9=c+d=n <sub>2</sub>
Total	45=a+c=n <sub>3</sub>	33=b+d=n <sub>4</sub>	78=a+b+c+d=N

Legendă:

a=bărbați fumători

b=bărbați nefumători

c=femei fumătoare

d=femei nefumătoare

a+b+c+d=N numărul total de cazuri

Aplicând formula simplificată obținem:

$$X_c^2 = \frac{(a \cdot d - b \cdot c)^2 \cdot N(-1)}{n_1 \cdot n_2 \cdot n_3 \cdot n_4} = \frac{(39 \cdot 3 - 30 \cdot 6)^2 \cdot (78 - 1)}{69 \cdot 9 \cdot 45 \cdot 33} = \frac{(117 - 180)^2 \cdot 77}{922185} = \frac{2045813}{922185} = 2,21$$

$X^2$  pentru un risc  $\alpha=5\%$  și pentru  $GL=(\text{numărul de rânduri}-1) \times (\text{număr de coloane} - 1)=(2-1) \times (2-1)=1$  are valoarea de 3,8. Valoarea calculată este mai mică decât valoarea tabelară deci diferența este nesemnificativă ceea ce testează aceeași lipsă de semnificație între bărbații și femeile fumătoare sau bărbații și femeile nefumătoare.

**Exemplul 6.** Să vedem dacă există diferențe semnificative între rezultatele tratamentului cu streptomycină și penicilină la două loturi de copii sub un an suferind de bronhopneumonie acută, am urmărit un număr de 20 de copii pentru fiecare lot. Rezultatele obținute au fost introduse în Tab. 31.

Tabelul 31

Rezultatele tratamentului bolnavilor cu bronhopneumonie

Bolnavi bronhopneumonie	Vindecați	Complicații	Total
Tratament penicilină	a=14	b=6	a+b=n <sub>1</sub> =20
Tratament streptomycină	c=16	d=4	c+d=n <sub>2</sub> =20
Total	a+c=n <sub>3</sub> =30	b+d=n <sub>4</sub> =10	a+b+c+d=N=40

$$X_c^2 = \frac{(14 \times 4 - 6 \times 16)^2 \times 39}{20 \times 20 \times 30 \times 10} = 0,52$$

$X_t^2=3,8$  pentru  $\alpha=5\%$

Rezultă că valoarea  $X_c^2 < X_t^2$  ceea ce înseamnă că rezultatele între cele două tratamente (penicilină și streptomycină) folosite în tratamentul bronhopneumoniei nu diferă semnificativ la cele două loturi de copii bolnavi.

Dezvoltarea istorică a statisticii, complexitatea obiectului său de cercetare, multitudinea proceselor și fenomenelor social-economice și culturale pe care le studiază au determinat apariția procesului de diferențiere pe ramuri a statisticii: statistica economică, statistica demografică, statistica medicală, statistica industrială etc.

Indiferent de domeniul de activitate, totalitatea operațiunilor folosite în cercetarea statistică poartă denumirea de – metodologia cercetării statistice. În cadrul metodologiei statistice, diferitele operațiuni se succed într-o ordine bine stabilită, reprezentând în desfășurarea lor cronologică – etapele cercetării statistice – ce trebuie respectate în cercetările întreprinse în toate domeniile de activitate.

Uneori, pentru elucidarea unor mecanisme sau precizarea acțiunii sau rolului anumitor substanțe în organism, întreprindem cercetări experimentale pe animale de laborator, a căror rezultate încercăm să le transpunem la om. Nu trebuie să uităm însă nicicând că rezultatele obținute pe animale nu sunt superpozabile cu cele înregistrate la om.

Astăzi dezvoltarea impetuoasă a științelor, în mod deosebit a celor tehnice, este indisolubil legată de utilizarea unor metode noi, complexe și precise, de evaluare și cuantificare a diferitelor aspecte calitative, obținute până de curând prin vechile metode descriptive, mai puțin precise și exacte.

Aceste câteva aspecte generale justifică și impun cunoașterea și aplicarea metodei statistice în general, a statisticii matematice îndeosebi, alături de celelalte metode de investigație, în cercetările clinice, de laborator, ca și în cele menite să contribuie la cunoașterea și aprecierea stării de sănătate a populației.

Etapele principale ale cercetării statistice sunt: adunarea sau observarea, prelucrarea, prezentarea și analiza materialului statistic.

### 3.1. Adunarea sau observarea materialului statistic

Este operația care ne permite să obținem datele necesare efectuării unei cercetări statistice. Termenul de “ADUNAREA” se utilizează atunci când datele necesare efectuării lucrării statistice le luăm din evidențe primare aflate la nivelul unităților sanitare, ca rezultat al activității personalului acestor unități.

**Exemplul 1.** Dacă ne propunem să stabilim incidența hipertensiunii arteriale într-o secție de spital, extragem fișele medicale ale bolnavilor din staționar, care prezintă valori ale TA peste limitele normale (140/90 mm Hg) din secția de boli interne. Aceste date le supunem ulterior operației de prelucrare a materialului statistic.

**Exemplul 2.** Dacă ne propunem să facem un studiu privind boala parodontală, atunci datele necesare efectuării studiului le extragem din registrele sau fișele de consultație ale bolnavilor examinați la nivelul unităților stomatologice dintr-un anumit teritoriu.

Termenul de “**observare**” se utilizează atunci când materialul necesar efectuării lucrării îl obținem ca rezultat nemijlocit al muncii proprii.

**Exemplul 3.** În cadrul stabilirii incidenței TA la populația dintr-un teritoriu, examinăm o parte (lot) din colectivitatea respectivă, prin examen clinic și de laborator, consemnând în mijloacele de înregistrare valorile tensionale obținute de noi pentru fiecare persoană examinată.

**Exemplul 4.** Dacă ne propunem să efectuăm o cercetare din care să reiasă rolul fluorului din apa potabilă asupra cariei dentare, în cazul “observării” materialului statistic, luăm un lot de animale cărora le administrăm apă cu fluor în diferite concentrații, o anumită perioadă de timp și consemnăm în protocoalele de lucru, întocmite în acest scop, toate modificările ce survin la animalele din lotul de cercetare, comparativ cu un lot martor.

Adunarea sau observarea materialului statistic constituie operațiunea de bază într-o cercetare statistică, deoarece corectitudinea și valabilitatea rezultatelor depind în cea mai mare măsură de modul în care se efectuează aceasta. Ca atare adunarea materialului se va face cu mare atenție, în mod obiectiv, fără a încerca să denaturăm datele, pe baza unor idei preconceptuate. Să nu se uite niciodată că în cercetarea științifică un “nu” categoric este mai valoros decât un “da” îndoielnic.

Buna reușită a etapei de adunare a materialului statistic depinde în egală măsură și de uniformitatea metodelor și mijloacelor de lucru. Întrucât cercetările statistice incumbă un volum mare de muncă, în efectuarea acestora sunt angrenate mai multe persoane. Pentru realizarea uniformității în munca de adunare a materialului se folosesc metode și mijloace de lucru judiciose întocmite, iar personalul care participă la operația de adunare a materialului este bine și amănunțit instruit asupra conținutului mijloacelor de înregistrare.

Operațiunea de adunare sau observare a materialului statistic cuprinde următoarele subetape:

**Subetapa datelor preliminare** sau a datelor pregătitoare în care, înainte de efectuarea propriu-zisă a operației de adunare a materialului statistic, se stabilesc și se precizează unele noțiuni absolut necesare bunei desfășurări a operației propriu-zise.

#### **Stabilirea scopului și obiectului lucrării**

În această subetapă se precizează scopul în care se efectuează lucrarea statistică și obiectul acesteia.

**Exemplul 5.** În cazul cercetării incidenței hipertensiunii arteriale scopul incidenței este acela de a cunoaște care sunt factorii biologici, de mediu, sociali-economici și comportamentali care au determinat o incidență crescută a bolii hipertensive, în vederea eliminării sau diminuării acestora, iar obiectul lucrării statistice este persoana căreia i s-au determinat valorile TA.

**Exemplul 6.** În cazul cercetării experimentale: “rolul fluorului în protejarea organismului împotriva cariei dentare”, scopul urmărit este acela de a cunoaște în ce măsură fluorul, administrat în anumite concentrații, în apa de băut, exercită asupra organismului un rol protector împotriva cariei dentare, iar obiectul lucrării îl constituie caria dentară.

Este necesar să stabilim – în cadrul datelor pregătitoare – scopul și obiectul lucrării, întrucât conținutul mijloacelor de observare a materialului este impus de aceste două noțiuni.

#### **Delimitarea cercetării**

După ce s-a stabilit scopul și obiectul cercetării statistice se trece la delimitarea cercetării. Această delimitare se face în:

- **timp**, stabilindu-se, în funcție de natura problemei cercetate, perioada de timp în care ne propunem să adunăm sau să observăm materialul necesar efectuării lucrării. Din acest punct de vedere trebuie precizat că sunt cercetări care se fac pe baza materialului obținut la un moment dat, pe baza unei singure examinări iar altele necesită perioade mai lungi sau mai scurte de timp. Spre exemplu starea de morbiditate prin HTA și respectiv prin afecțiuni stomatologice se poate cunoaște pe baza unei singure examinări a populației dintr-o colectivitate, fără a necesita o urmărire în timp a acesteia.

Ea e similară cu un instantaneu fotografic care exprimă ce se petrece la un moment dat în sânul populației din punct de vedere stomatologic, sau al valorilor tensiunii arteriale.

În cazul urmăririi în dinamică a TA la populația unei localități, examinarea populației din această colectivitate se face la anumite intervale de timp, pentru a stabili dacă măsurile preconizate pentru combaterea HTA sunt eficiente sau nu.

În cazul exemplului cu administrarea fluorului, în diferite concentrații, în apa potabilă, lotul de populație examinat inițial trebuie urmărit și examinat, comparativ cu un lot martor, la diferite intervale de timp, pentru a stabili dacă fluorul exercită sau nu un rol protector împotriva cariei dentare;

- **spațiu**; cu care ocazie se stabilește teritoriul pe care ne propunem să efectuăm cercetarea. Acest teritoriu poate să fie restrâns (o clasă, o școală, o întreprindere) sau mai extins (o circumscripție, o comună, teritoriul deservit de un anumit spital unificat etc.), în funcție de natura problemei pe care o cercetăm și de scopul și obiectul cercetării. Când cercetarea se extinde pe un teritoriu mai mare, atunci se respectă de obicei împărțirea administrativă a teritoriului: sat, comună, raion etc. Starea de morbiditate prin HTA se poate stabili fie la populația unei colectivități restrânse (întreprindere, sat, comună) fie la populația întregului raion. Starea de morbiditate prin afecțiuni stomatologice se poate stabili spre exemplu, fie la populația școlară – de o anumită vârstă – sau la toată populația colectivității, pe ani de vârstă sau pe grupe de vârstă;
- **volum**, urmărind să stabilim mărimea colectivității pe care o vom lua în studiu, sub raport cantitativ, deci numărul de unități statistice ce urmează a fi cercetate.

Cercetările statistice se fac fie pe totalul unităților statistice cuprinse într-o populație și atunci se spune că s-a lucrat pe “univers”, colectivitate statistică generală, fie pe un număr limitat de unități statistice din cadrul populației și atunci se spune că s-a lucrat pe “eșantion”. Cele mai veridice concluzii se obțin atunci când cercetarea a cuprins totalitatea unităților statistice ce compun o mulțime, o colectivitate. Asemenea cercetări însă se efectuează foarte rar întrucât sunt foarte costisitoare – necesitând sume mari de bani, sunt greu de organizat, datorită numărului mare de personal, de aparatură, reactivi și necesită timp îndelungat. Uneori nici nu este necesar să efectuăm cercetări pe “univers”

Întrucât statistica matematică ne oferă criterii exacte pentru stabilirea numerică a colectivității ce urmează a fi cercetată, pentru ca concluziile obținute în urma cercetării să poată fi generalizate ca și când am fi lucrat pe “univers” sau pe colectivitatea statistică generală.

### **Stabilirea unității statistice**

Stabilirea unității statistice – adică a părții componente a mulțimii sau colectivității expusă cercetării. Este necesar să stabilim unitatea statistică cu care lucrăm pentru că nu totdeauna aceasta este distinctă, iar uneori ea e impusă de natura cercetării și scopul urmărit.

**Exemplul 1.** Dacă dorim să stabilim incidența HTA într-o colectivitate atunci fiecare persoană examinată care prezintă valori crescute ale TA constituie o unitate statistică. Ea reprezintă un caz nou de îmbolnăvire. Prin caz nou de îmbolnăvire înțelegem orice îmbolnăvire din momentul depistării sau luării în evidență până în momentul vindecării sau decesului. Cazul nou de îmbolnăvire nu se confundă nici cu persoana bolnavă nici cu consultația, pentru că una și aceeași persoană poate prezenta mai multe îmbolnăviri noi într-o anumită perioadă de timp și pentru aceeași îmbolnăvire persoana respectivă se poate prezenta de mai multe ori la consultație.

În cazul recensământului de populație unitatea statistică este persoana recenziată cu caracteristicile: vârstă, sex, naționalitate, profesie etc. Alte exemple de unități statistice sunt: familia, căsătoria, nașcutul viu, gravida, instituția medicală etc.

**Exemplul 2.** Dacă inițiem o cercetare în scopul cunoașterii morbidității stomatologice pe afecțiuni, atunci fiecare afecțiune nouă constituie o unitate statistică și o folosim în calculul morbidității. Prin afecțiune nouă sau caz nou de îmbolnăvire înțelegem în cazul nostru, orice afecțiune stomatologică din momentul depistării până în momentul vindecării. Deci pentru a stabili care este morbiditatea stomatologică prin caria dentară, într-o anumită colectivitate, înregistrăm toate cariile dentare noi, depistate la populația examinată, pe care le etichetăm ca unități statistice și nu persoana care prezintă caria dentară constituie unitatea statistică, în cazul stabilirii morbidității pe contingente, prin afecțiuni stomatologice, unitatea statistică o constituie persoana care cu ocazia examinării prezintă afecțiuni stomatologice, în calcularea morbidității generale nu persoana bolnavă constituie unitatea statistică, ci cazul nou de îmbolnăvire.

Aceeași persoană poate prezenta, în decurs de un an, mai multe îmbolnăviri noi. Fiecare dintre aceste îmbolnăviri constituie unități statistice distincte. De asemenea pentru fiecare îmbolnăvire nouă, care poate dura uneori o perioadă mai lungă de timp, persoana bolnavă se poate prezenta de mai multe ori la medic, fără ca aceste consultații să fie etichetate ca îmbolnăviri noi și deci ca unități statistice distincte.

### **Stabilirea programului adunării sau observării materialului statistic**

După ce am făcut delimitarea lucrării statistice în timp, spațiu și volum, este necesar să stabilim caracteristicile fiecărei unități statistice cercetate.

Prin caracteristică (variabilă) se înțelege trăsătura sau însușirea proprie fiecărei unități statistice.

**Exemplul 1.** În cazul cercetării morbidității prin HTA într-o colectivitate se pot lua în considerare următoarele caracteristici (variabile): vârsta, sexul, localitatea de domiciliu, profesia, factorii de risc etc.

**Exemplul 2.** În cazul cercetării morbidității stomatologice într-o anume colectivitate, se vor lua în considerare următoarele caracteristici: cazul nou de boală, vârsta, sexul, localitatea de domiciliu, natura îmbolnăvirii noi, profesia etc. Caracteristicile sau variabilele pot fi de tip cantitativ exprimate în unități de măsură ca: înălțime, greutate, talie, perimetru toracic, în cazul dezvoltării fizice, glicemie, TA, temperatură, puls; și de tip calitativ când exprimă o calitate ce nu poate fi măsurată prin unități de măsură, de exemplu: sexul, culoarea tegumentelor, starea la externare a unor bolnavi (care poate fi vindecat, ameliorat, agravat). Aceste caracteristici calitative sau cantitative trebuie exprimate clar și precis, încât persoanele care participă la operația de adunare a materialului statistic să nu fie puse în situația de a le da o interpretare subiectivă sau greșită.

A stabili caracteristicile unităților statistice, ce vor fi adunate sau observate în cadrul efectuării lucrării, înseamnă a stabili programul operației de adunare a materialului statistic.

### **Planul observării sau adunării materialului statistic**

După ce programul de adunare a materialului statistic a fost stabilit se trece la elaborarea planului de adunare a materialului statistic. Planul adunării materialului statistic cuprinde date privitoare la meto-

dologia adunării materialului și date privitoare la organizarea adunării materialului statistic.

Datele metodologice ale adunării materialului statistic se referă la datele pe care le-am discutat mai înainte și anume: stabilirea scopului și obiectul lucrării sau cercetării, delimitarea în timp, spațiu și volum a cercetării, stabilirea unității statistice, elaborarea programului adunării materialului.

Datele organizatorice ale planului se referă la problemele privitoare la documentarea în tema luată în studiu, la întocmirea sau procurarea formularelor de înregistrare, la alegerea și instruirea personalului ce va participa la adunarea materialului statistic etc.

Planul cercetării este deci mai cuprinzător decât programul adunării materialului, el incluzând și planul prelucrării, al prezentării și analizei materialului, în concordanță cu ipoteza de lucru de la care am pornit.

### **Subetapa stabilirii metodelor de adunare sau înregistrare a materialului statistic**

Metodele folosite pentru înregistrarea materialului statistic depind de faptul dacă ne-am propus – în cadrul datelor pregătitoare – să lucrăm pe „univers” sau pe „eșantion” reprezentativ. Din acest punct de vedere metodele de înregistrare se împart în două grupe mari:

#### **Metode de înregistrare totală**

Dacă ne-am propus să lucrăm pe „univers” deci pe întreaga colectivitate, atunci metodele de înregistrare vor fi totale – înregistrând toate unitățile statistice ce compun o colectivitate – iar dacă ne-am propus să lucrăm pe „eșantion”, deci numai pe o parte din colectivitate, atunci metodele folosite pentru înregistrarea materialului vor fi metode parțiale. Metodele de înregistrare totală pot fi la rândul lor:

**Total continue**, când unitățile statistice se înregistrează în totalitate dar în măsura în care fenomenul are loc. Spre exemplu dacă ne propunem să facem un studiu asupra fenomenului de natalitate și mortalitate, într-un anumit teritoriu sau într-o anumită colectivitate, atunci se înregistrează toți nașcuții vii sau toate decesele din teritoriul sau colectivitatea respectivă însă pe măsură ce ele se produc. Aceste înregistrări sunt limitate în spațiu dar nu în timp.

**Total periodice**, când înregistrarea unităților statistice dintr-o colectivitate sau dintr-un teritoriu se face în totalitate dar numai la anumite

intervale de timp. Aceste înregistrări sunt limitate atât în timp cât și în spațiu. Ele redau aspecte statistice, la un moment dat, ale fenomenului studiat, spre deosebire de înregistrările totale continue care redau dinamica fenomenului. Un exemplu de înregistrare totală periodică o constituie recensămintele de populație, care de obicei se efectuează la intervale de 10 ani și cuprind populația întregii țări. Deși metoda de înregistrare totală continuă este superioară celei totale periodice, datorită volumului mare de muncă, costului ridicat, necesitând un mare număr de personal specializat, ea se aplică numai în situații aparte.

#### **Metode de înregistrare parțială**

Le utilizăm atunci când ne propunem să efectuăm cercetarea pe o parte din colectivitate, pe eșantion. Lucrăm pe o parte din colectivitate folosind metode de înregistrare parțială, atunci când fie că nu cunoaștem colectivitatea generală, deci nu o putem aborda, fie că volumul de muncă este prea mare și necesită fonduri mari, timp îndelungat, personal numeros, bază materială bogată și putem obține aceleași rezultate, aceleași concluzii, lucrând numai pe o parte din unitățile statistice ce compun colectivitatea.

Printre metodele de înregistrare parțială amintim:

**Estimația** – este o metodă de înregistrare parțială care, pe baza unor sondaje făcute în sânul colectivității generale, luată în studiu, încearcă să caracterizeze, să tragă concluzii de ordin general valabile pentru întreaga colectivitate.

**Exemplul 1.** Pentru a cunoaște morbiditatea prin HTA într-o colectivitate se examinează prin extragere aleatorie un număr limitat de cazuri (din toate grupele de vârstă, de ambele sexe, de diferite profesii) reprezentând un eșantion din populația generală, iar rezultatele obținute se vor putea generaliza, estima, în colectivitatea generală.

**Exemplul 2.** Pentru a cunoaște starea de morbiditate prin afecțiuni stomatologice la elevii dintr-o colectivitate școlară, se examinează din punct de vedere stomatologic, prin sondaj, un număr de elevi din fiecare clasă a școlii respective, iar rezultatele obținute în urma prelucrării datelor, investigațiilor de specialitate, sunt generalizate, ca și când studiul sau examinarea s-ar fi extins asupra tuturor elevilor din respectiva colectivitate școlară.

De menționat este faptul că eșantionul pe care s-a lucrat a trebuit să fie „reprezentativ” atât sub aspect cantitativ, cât și sub aspect ca-

litativ, adică al celorlalte atribute ca: sex, grupă de vârstă, condiții de alimentație, de locuință etc. Estimația, pe baza unor sondaje mai mult sau mai puțin numeroase făcute în colectivitate, aproximează frecvența unor fenomene în sânul colectivității fără a avea pretenția de exactitate (ea introduce întotdeauna un risc și o eroare maximă admisă). Cu cât numărul de investigații din eșantion este mai mic cu atât rezultatele sunt mai imprecise și cu cât numărul de cazuri din eșantion este mai mare, rezultatele sunt mai exacte.

**Metoda eșantionului sau selecția** – este metoda de înregistrare parțială care, pe baza examinării unei părți reprezentative din punct de vedere cantitativ (număr cazuri) și calitativ (atribute) din unitățile statistice ce compun colectivitatea generală, urmărește să caracterizeze întreaga colectivitate din care a fost extras eșantionul. Sub aspect cantitativ eșantionul va fi determinat pe baza unor formule matematice ce vor fi prezentate la capitolul „eșantionaj”, iar sub aspect calitativ variabilele vor fi stabilite în funcție de natura fenomenelor pe care le cercetăm.

**Exemplul 1.** Pentru a stabili incidența HTA într-o colectivitate vom examina un număr limitat de cazuri, în ceea ce privește TA și diferitele atribute sau variabile: sex, vârstă, domiciliu, profesie, urmărindu-se o stratificare a acestor cazuri în eșantion asemănătoare cu stratificarea din populația generală.

**Exemplul 2.** Dacă ne propunem să cunoaștem starea de morbiditate prin afecțiuni stomatologice la populația infantilă dintr-un raion, nu vom proceda la examinarea activă de specialitate a întregii populații infantile, ci vom selecționa colectivități reprezentative din raion (din mediul urban, din mediul industrial, din mediul rural, cu diferite forme de relief, cu nivel social-economic diferit, cu obiceiuri culinare specifice, cu nivel sanitar diferit, de vârste diferite etc.) pe care, examinându-le din punct de vedere stomatologic, vom putea stabili aspectele de morbiditate ce caracterizează întreaga populație infantilă din raion.

În anumite situații, dictate de specificul bolii cercetate, cu afinitate sau predilecție pentru anumite grupe de vârstă, vom lua în studiu numai acele vârste la care maladia respectivă este mai frecventă. Spre exemplu, dacă ne propunem să cunoaștem morbiditatea prin afecțiuni parodontale într-o colectivitate, eșantionul pe care vom lucra nu va cuprinde populația de toate vârstele ci cu precădere vârstele de 35-40 ani, întrucât incidența îmbolnăvirilor prin afecțiuni parodontale e mai

crescută la aceste vârste, în asemenea situație spunem că am examinat și înregistrat partea de bază a colectivității. La fel procedăm în cazul cercetării morbidității prin afecțiuni cardio-vasculare etc.

**Monografia** – este o metodă de înregistrare parțială utilizată pentru studiul unei singure colectivități (localitate, circumscripție, întreprindere, școală) sau a unui singur fenomen (demografic, de morbiditate). Ea urmărește să cunoască aspectele caracteristice ale colectivității sau fenomenului respectiv, fără a le putea generaliza în populația generală. Fiind o metodă descriptivă, monografia se mulțumește doar să sesizeze aspectele caracteristice ale colectivității sau fenomenului fără să le poată explica.

De exemplu, monografia unei localități descrie: problemele demografice, nivelul de trai, ocupația populației, nivelul cultural, obiceiurile culinare ale populației, starea de sănătate cu evidențierea problemelor de morbiditate și mortalitate, factorii de risc din colectivitate etc. Ea nu poate să explice relația dintre aceste caracteristici ale populației din localitatea respectivă.

**Ancheta** – este o metodă de investigație parțială a materialului statistic pe formulare adecvate problemei cercetate, în formularele de anchetă datele se înscriu fie de anchetator, și atunci vorbim de anchetă directă, fie de către persoana anchetată, și atunci vorbim de anchetă indirectă, în cazul anchetei indirecte exactitatea răspunsurilor depinde de nivelul de pregătire a celor anchetați. Ancheta de obicei succede monografiei, încercând să aprofundeze unele aspecte sesizate de monografie. Fiind o metodă de profunzime, ancheta se limitează la colectivități mai restrânse (secția unei întreprinderi, clasa unei școli etc.). În cazul în care ancheta stabilește rolul negativ al unor factori negativi din colectivitatea respectivă, ea se încheie cu propuneri de remediere a acestora.

**Exemplul 1.** Dacă în colectivitatea în care s-a întocmit monografia a rezultat că HTA este frecventă în sânul colectivității, atunci personalul medical, făcând un studiu pe eșantion, stabilește cauzele sau factorii de risc (alimentație, solicitări la locul de muncă, consum de alcool etc.) și propune măsuri concrete care să ducă la eliminarea acestora și la scăderea incidenței HTA în colectivitatea respectivă.

**Exemplul 2.** Dacă monografia a scos în relief faptul că populația prezintă o dantură deficitară, atunci ancheta respectivă caută să stabilească pe un eșantion cauzele acesteia și se încheie cu recomandarea



unor măsuri concrete privind modul de alimentație, condițiile de igienă buco-dentară etc.

### Mijloace de înregistrare

Adunarea și înregistrarea materialului statistic se face pe anumite formulare statistice numite mijloace de înregistrare. Conținutul mijloacelor de înregistrare trebuie să fie clar, concis și să cuprindă toate caracteristicile unităților statistice de care avem nevoie în lucrarea pe care o efectuăm. Deci conținutul mijloacelor de înregistrare diferă de la o lucrare la alta. El este dictat de scopul și obiectul lucrării, unitatea statistică aleasă și se stabilește odată cu elaborarea programului lucrării.

Aceste mijloace de înregistrare pot fi:

**CertIFICATELE** sunt formulare ce servesc pentru înscrierea unei singure unități statistice, o singură dată. De exemplu: certificatul de naștere, buletinul de identitate, certificatul de deces etc.

**Fișele** servesc la înregistrarea unei singure unități statistice dar nu o singură dată ci de mai multe ori, pe perioade de timp mai îndelungate. De exemplu: fișa copilului de 0-18 ani, fișă de consultație a adultului, fișă de consultație stomatologică, fișa gravidei sau foaia de observație clinică în spitale.

**Listele** sunt mijloace de înregistrare ce servesc pentru mai multe unități statistice. Sunt sinonime cu tabelele. Ele înscriu unitățile statistice pe câte un rând al listei, iar diferitele caracteristici sau variabile ale unităților statistice se înscriu o singură dată, în coloane. De exemplu dacă ne propunem să examinăm activ din punct de vedere al stării de sănătate elevii dintr-o clasă, folosim pentru înscrierea rezultatelor o listă în care se înscriu pe rânduri, nominal, toți elevii examinați, iar în partea dreaptă a fiecărui nume, în coloane, diferite variabile: sex, vârstă, afecțiuni etc.

**Registrele** sunt mijloace de înregistrare ce cuprind mai multe liste de același fel. De exemplu registrul de consultații medicale, de vaccinare, de evidență a gravidelor etc.

### 3.2. Prelucrarea materialului statistic

Datele adunate sau observate se referă de cele mai multe ori la fiecare unitate statistică în parte. Pentru a putea caracteriza însă colectivitatea luată în cercetare este necesar să supunem aceste date unei

operații de prelucrare care constă în aranjarea, verificarea, codificarea, sortarea și gruparea acestora în așa fel încât să permită o prezentare clară a materialului statistic.

După locul în care se efectuează prelucrarea deosebim:

- prelucrarea descentralizată care are loc la nivelul unităților sanitare care au efectuat adunarea materialului statistic, iar după aceea datele prelucrate sunt înaintate forurilor superioare interesate în centralizarea datelor provenite de la unitățile teritoriale;
- prelucrarea centralizată are loc la nivelul forurilor superioare care inițiază cercetarea, pe baza formularelor de înregistrare a materialului trimise de unitățile teritoriale. Prelucrarea centralizată se face la nivelul Centrului Național de Management în Sănătate. Acest mod de prelucrare are avantajul că se execută de către specialiști reușind o mai bună calitate a rezultatelor lucrării.

Etapa de prelucrare a materialului statistic cunoaște mai multe subetape:

**Verificarea materialului statistic** poate fi:

**Verificare cantitativă**, care stabilește dacă toate formularele de înregistrare sunt completate și dacă s-a răspuns la toate rubricile corespunzătoare caracteristicilor unităților statistice prevăzute în planul cercetării. De exemplu dacă ne propunem să cercetăm morbiditatea într-o colectivitate școlară, datele obținute în urma examinării de specialitate a elevilor le vom înscrie în fișele de consultație întocmite în acest sens. Fișele astfel completate pentru fiecare elev în parte sunt supuse unei verificări cantitative care stabilește dacă s-au înscris datele la toate caracteristicile elevilor examinați: vârstă, sex, domiciliu, afecțiuni medicale (HTA, infecții acute ale căilor respiratorii, boli diareice acute, otite), afecțiuni stomatologice (carii dentare, pulpite, gangrene, extracții). Formularele care nu cuprind datele necesare la toate rubricile prevăzute sunt înlăturate, neputând fi supuse operației de prelucrare propriu-zisă.

**Verificarea calitativă** urmărește să stabilească dacă datele înscrise în diferitele rubrici sunt exacte. Ea poate fi:

- **logică** – când urmărește să stabilească dacă nu există nepotriviri între vârstă și diagnostic, între sex și diagnostic etc. Se întâmplă uneori ca diagnosticul unui bolnav de sex masculin (ex. Cancer de prostată) să fie trecut din greșeală pe formularul unei femei. Există programe speciale de verificare logică a acestor neconcordanțe pe calculator.

- **cifrică** – atunci când urmărește să stabilească dacă datele înregistrate corespund realității. Spre exemplu, pe baza anului nașterii se poate calcula caracteristica de vârstă a individului în anul în curs și verifica astfel concordanța acestor date.

#### Codificarea materialului statistic

Codificarea nu este altceva decât notarea prin cifre (numerică) sau litere (alfabetică), a diferitelor caracteristici ale unităților statistice, în scopul ușurării prelucrării și interpretării rezultatelor obținute. Codificarea se folosește de obicei în cadrul lucrărilor de amploare, care cuprind un număr mare de unități statistice. Spre exemplu în cadrul cercetării morbidității din teritoriul unui spital se pot codifica cu litere localitățile, iar diagnosticele de boală se codifică numeric conform clasificării și codificării cauzelor de boală și deces – revizia a-X-a OMS pe 999 cauze de îmbolnăvire. De exemplu, infarctul miocardic acut are codul 459, pneumonia virală are codul 506, caria dentară are codul 544 etc.

**Sortarea materialului statistic** constă în repartizarea unităților statistice după caracteristici individuale. Sortarea poate fi făcută manual sau cu ajutorul calculatoarelor electronice, mai ales în lucrările mari când sunt înscrise mai multe caracteristici ale unităților statistice sub forma unor chestionare, sortarea acestor caracteristici făcându-se automat cu ajutorul calculatorului.

**Centralizarea datelor statistice** urmează sortării sau repartizării unităților statistice pe diferite caracteristici. Ea constă în însumarea caracteristicilor statistice și a unităților statistice. Aceasta se poate face sub forma unor tabele care să cuprindă rubrica totală. Centralizarea poate fi simplă când la baza însumării stă un singur criteriu (ex. centralizarea cazurilor noi de boală dintr-un teritoriu) sau grupată când la baza însumării unităților statistice stau mai multe criterii (ex. centralizarea îmbolnăvirilor pe sexe, grupe de vârstă, cauze de boală etc.).

**Gruparea materialului statistic** constă în aranjarea unităților statistice, în funcție de diferitele caracteristici sau variabile, în grupe cât mai omogene, pentru a le putea scoate cât mai ușor în evidență. Gruparea este operația statistică ce permite trecerea de la aspectele particulare, individuale ale unităților statistice la aspecte generale, comune, esențiale ale materialului cercetat.

Așa, spre exemplu, dacă ne propunem să cercetăm morbiditatea prin HTA sau afecțiuni stomatologice într-o colectivitate organizată

(școală, instituție), datele obținute în urma examinării diferă de la o unitate statistică la alta sub aspectul vârstei, sexului, profesiei, stării de sănătate sau boală etc., neputându-ne permite observarea a ceea ce este general, comun, caracteristic întregii colectivități.

Repartizarea datelor pe diferite variabile: grupă de vârstă, sex, profesie, stare de sănătate sau boală, mod de alimentare permite, în cadrul grupării, observarea a ceea ce este general, esențial, oferindu-ne posibilitatea să întvedem concluziile ce se degajă în urma studiului efectuat.

Criteriile sau caracteristicile de grupare a materialului statistic depind de scopul urmărit prin cercetarea întreprinsă. Este ceea ce afirma Claude Bernard: “Când nu știm ceea ce căutăm, nu vedem ceea ce am găsit.”

Spre exemplu dacă nu cunoaștem rolul suprasolicitării nervoase în boala hipertensivă și nu stabilim acest criteriu în gruparea bolnavilor, sau dacă nu cunoaștem rolul sursei de apă potabilă în influențarea morbidității în afecțiuni stomatologice și nu grupăm materialul statistic și după acest criteriu, putem trece pe lângă o concluzie esențială fără a o putea sesiza.

**Gruparea** poate fi **simplă** când la baza repartizării materialului statistic stă un singur criteriu.

**Exemplul 1.** Gruparea pe sexe a bolnavilor cu HTA (Tab. 32).

Tabelul 32

Numărul de bolnavi cu HTA în funcție de sex

	Sex	Nr. bolnavi cu HTA
	M	30
	F	20
TOTAL		50

**Exemplul 2.** Repartizarea bolnavilor cu afecțiuni stomatologice pe cauze de boală (Tab.33).

Tabelul 33

Numărul de bolnavi cu afecțiuni stomatologice pe cauze de boală

Nr. crt.	Afecțiunea	Nr. îmbolnăviri
1.	Caria dentară	60
2.	Pulpită	15
3.	Gangrenă	8
4.	Parodontopatii	26
	TOTAL	109

**Gruparea** poate fi **complexă** când la baza repartizării materialului statistic stau două sau mai multe criterii.

**Exemplul 3.** Distribuția pe sexe și grupe de vârstă a bolnavilor cu HA, în colectivitatea X anul Y (Tab.34).

Tabelul 34

Repartizarea bolnavilor cu HTA în funcție de sex și grupe de vârstă

		Grupa de vârstă			Total
		0-15	16-49	50-90	
Sex	M	2	8	20	30
	F	3	7	10	20
					50

**Exemplul 4.** Distribuția pe cauze și sexe a afecțiunilor stomatologice în colectivitatea X anul Y (Tab.35).

Tabelul 35

Repartizarea bolnavilor cu afecțiuni stomatologice în funcție de sex în localitatea X, anul Y

Nr. crt.	Afecțiunea	Număr îmbolnăviri pe sexe		
		Total	M	F
1.	Total	100	45	55
2.	Carie dentară	60	25	35
3.	Pulpită	25	17	8
4.	Gangrenă	8	6	2
5.	Parodontopatii	16	6	10

După natura caracteristicii gruparea poate fi:

- **în timp**, când repartizarea materialului statistic are la bază anumite intervale de timp, de exemplu pe semestre (semestrul I, semestrul II) sau pe trimestre etc.
- **în spațiu**, când repartizarea materialului statistic se face în raport cu locul sau teritoriul unde s-a examinat populația, de exemplu mediu rural și mediu urban.

**Limitele grupării** nu se stabilesc mecanic ci depind de natura fenomenului cercetat și scopul lucrării, în principiu o grupare prea amănunțită a materialului statistic duce la o fărâmițare ce ne împiedică să observăm ce este caracteristic, esențial. De asemenea dacă o caracteristică are prea multe valori vom repartiza materialul statistic pe grupe de valori care se vor numi intervale de clasă. De exemplu, ar fi inutilă

cercetarea mortalității pe ani de vârstă, de la 1 la 80 de ani, întrucât mortalitatea nu diferă de la un an de vârstă la altul. Dar este importantă gruparea deceselor pe grupe de vârstă de câte 5 ani sau de câte 10 ani, întrucât fenomenul cercetat „mortalitatea generală” permite unele interpretări și concluzii în raport cu aceste grupe de vârstă.

Fiecare grupă de valori sau „clasă” are la rândul ei o limită inferioară, o limită superioară și un centru al clasei egal cu semisuma celor două limite.

**Exemplul 5.** Distribuția pe grupe de valori ale TA a bolnavilor cu HTA în colectivitatea X anul Y (Tab.36).

Tabelul 36

Repartizarea bolnavilor cu HTA pe grupe de valori ale TA în colectivitatea X anul Y

Nr. crt.	Grupe de valori	Centrul clasei	Frecvența
	TA		
1.	160-169,9	165	18
2.	170-179,9	175	22
3.	180-189,9	185	10
TOTAL			50

**Exemplul 6.** Distribuția pe grupe de vârstă a bolnavilor cu afecțiuni stomatologice în colectivitatea X anul Y (Tab.37).

Tabelul 37

Repartizarea bolnavilor cu afecțiuni stomatologice pe grupe de vârstă în colectivitatea X anul Y

Nr. crt.	Grupe de vârstă	Centrul clasei	Frecvența
1.	0-19	10	79
2.	20-39	30	38
3.	40-59	50	42
4.	60-80	70	64
TOTAL			223

De menționat că în cadrul grupării trebuie să cunoaștem exact valorile extreme – maximă și minimă – ale caracteristicii și trebuie să stabilim intervale de clasă egale iar limitele intervalelor sau claselor să fie distincte pentru a nu crea confuzii cu ocazia repartizării materialului statistic. De exemplu, dacă dorim să distribuim elevii unei clase pe grupe de înălțimi vom stabili înălțimea maximă și minimă a elevilor din clasa respectivă iar în interiorul intervalului respectiv vom delimita, la

intervale egale, grupe de valori ale înălțimii. Cel mai scund elev, având înălțimea de 150 cm și cel mai înalt de 170 cm, în interiorul acestor valori extreme vom stabili grupe de valori ale înălțimii din 5 în 5 cm, limitele grupelor de înălțime fiind și ele distincte.

**Exemplul 7.** Distribuția pe grupe de înălțimi a elevilor din clasa X în anul Y (Tab.38).

Tabelul 38

Repartizarea elevilor pe grupe de înălțimi din clasa X în anul Y

Nr.crt.	Grupe de înălțime cm	Număr elevi
1.	150-154,9	4
2.	155-159,9	12
3.	160-164,9	18
4.	165-169,9	9
	TOTAL	43

### Elaborarea indicatorilor

Gruparea de obicei cuprinde cifre brute, valori absolute, care rareori pot fi interpretate ca atare. De aceea materialul statistic este supus în continuare unei prelucrări cu ajutorul metodelor de statistică matematică în vederea obținerii unor indicatori statistici ca: mărimi medii, mărimi relative, indici etc., ce vor permite aprecieri comparative, corecte și concluzii semnificative.

### 3.3. Eșantionarea

Prin **eșantionare** (eșantionaj) înțelegem totalitatea tehnicilor și procedeele statistice cu ajutorul cărora se poate caracteriza o colectivitate (N) studiind o parte a acesteia. Se va studia un număr limitat de cazuri (n), care vor fi extrase aleator, întâmplător, din întreaga colectivitate.

Elementele efectiv studiate constituie **eșantionul sau mostra**; populația din care provine eșantionul poartă numele de **colectivitate generală sau univers**, bază de sondaj. Elementele individuale extrase din colectivitatea generală, la care se studiază caracteristicile sau însușirile, și care compun eșantionul, se numesc **unități de selecție**.

Scopul unui astfel de studiu pe eșantion este de a nu studia întreaga colectivitate generală, ci o copie de dimensiuni mai mici a acesteia, iar caracteristicile obținute pe eșantion să poată fi extrapolate cu o anumită probabilitate asupra populației generale. Această probabilitate (P) poate fi de 90%, 95%, 99%, dar niciodată de 100%, deci presupune

asumarea unui risc ( $\alpha = 100 - P$ ) în efectuarea studiului, în biologie se lucrează de obicei cu o probabilitate de 95%, deci cu un risc „ $\alpha$ ” de 5%.

Două condiții trebuie asigurate în vederea unui studiu pe un eșantion: reprezentativitatea eșantionului și precizia studiului.

### Reprezentativitatea

Este asigurată de alegerea aleatoare, întâmplătoare, a unităților de selecție din baza de sondaj (univers), astfel încât fiecare din elementele colectivității generale (univers) să aibă aceeași șansă de a fi cuprinse în eșantion. Prin această alegere aleatoare se va reproduce populația generală la nivel de eșantion cu anumite erori. Aceste erori se numesc „**erori aleatoare**” și ele pot fi măsurate și controlate de cercetător. Nerespectarea acestei condiții de extragere întâmplătoare a unităților de selecție din baza de sondaj, din colectivitatea generală, duce la apariția unui alt tip de erori, și anume „**erori sistematice**” (bi-ais), care sunt imposibil de determinat de către cercetător. De exemplu: un studiu experimental pentru verificarea unui nou medicament, efectuat pe persoane care se prezintă voluntar, va introduce astfel de erori sistematice, pentru că acestea sunt persoane cu structuri psihice modificate sau suferă de boala pentru care se testează medicamentul, deci nu vor putea reprezenta trăsăturile colectivității generale. De asemenea, un studiu efectuat asupra morbidității printr-o afecțiune într-o secție clinică a unui spital nu poate reproduce structura morbidității la nivelul unui raion sau al țării. Deci ele nu pot constitui eșantioane reprezentative pentru teritoriul respectiv.

### Precizia

Se referă la volumul eșantionului. Aceasta presupune determinarea unui număr (n) minim de cazuri necesare pentru constituirea eșantionului. Pentru determinarea acestuia, să presupunem că pornim de la o colectivitate generală, bază de sondaj, univers, care are un volum de N cazuri și este caracterizată de indicatorul central (media  $\mu$ ) și de dispersie (deviație standard  $\sigma_x$ ). Din această populație generală extragem aleator un eșantion de volum  $n_1$  și care va putea fi caracterizat de media  $\bar{X}_1$  și abaterea standard  $\sigma_1$ . Din populația generală pot fi extrase și un al doilea eșantion, de volum  $n_2$  și caracterizat de media  $\bar{X}_2$  și abaterea standard  $\sigma_2$ ; al treilea eșantion, de volum  $n_3$  și caracterizat prin media

$\bar{X}_3$  și abaterea standard  $\sigma_3$ ; ș.a.m.d. Mediile eșantioanelor ( $\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots$ ) extrase din populația generală se vor comporta ca niște variabile aleatoare și se vor distribui în stânga și în dreapta mediei populației generale. Comparând media primului eșantion  $\bar{X}_1$  cu media populației generale  $\mu$ , se observă că între acestea există o anumită diferență sau eroare, care poartă numele de eroare maximă admisă și se notează cu  $\Delta x$ . Se poate scrie deci relația:

$$\mu = \bar{X}_1 \pm \Delta x$$

unde:

$\mu$  = media populației generale;

$\bar{X}_1$  = media eșantionului I;

$\Delta x$  = eroarea maximă admisă.

Această eroare maximă admisă ( $\Delta x$ ) depinde de:

- un coeficient de probabilitate  $t_\alpha$ , în funcție de riscul  $\alpha$  existent în studiu (exemplu: 5%) sau probabilitatea ( $P = 1 - \alpha$ ) cu care vom garanta rezultatele ( $P = 95\%$  pentru  $\alpha = 5\%$ );
- abaterea standard în populația generală ( $\sigma_x$ );
- numărul de cazuri conținut în eșantioane ( $n$ ).

Deci:

$$\Delta x = \pm t_\alpha \frac{\sigma_x}{\sqrt{n}}$$

Introducând în formula anterioară, se poate scrie:

$$\mu = \bar{X}_1 \pm t_\alpha \frac{\sigma_x}{\sqrt{n}}$$

Această presupune, de altfel, că pentru un risc „ $\alpha$ ” cu care proiectăm sondajul, media populației generale se va situa într-un anumit interval, numit interval de încredere.

Numărul minim de cazuri necesar unui studiu pentru a înregistra o eroare inferioară sau cel mult egală celei standard se va calcula conform uneia dintre formulele următoare:

**pentru caracteristici cantitative:**

$$n = \frac{t_\alpha^2 \sigma_x^2}{\Delta_x^2}$$

**pentru caracteristici calitative:**

$$n = \frac{t_\alpha^2 Pq}{\Delta_x^2}$$

De exemplu, pentru studiul unei afecțiuni, boli, în populație,  $P$  poate fi reprezentată de prevalență (%) bolii în populația generală.

Dacă în colectivitatea noastră, pe care am efectuat studiul, probabilitatea hipertensiunii arteriale a fost de 17%, dorind ca această probabilitate să o generalizăm la un risc de  $\alpha=5\%$  și  $\Delta x=0,8\%$ , atunci numărul de persoane ce vor trebui examinate va fi:

$$n = \frac{t_\alpha^2 Pq}{\Delta_x^2} = \frac{1,96^2 \times 17 \times 83}{0,8^2} = \frac{3,84 \times 17 \times 83}{0,64} = \frac{5418,24}{0,64} = 8466$$

Sau, dacă ne propunem să generalizăm incidența de 98% a afecțiunilor stomatologice, cu un prag de semnificație de 5% și o eroare maximă de 1%, stabilită pe un lot de populație, atunci volumul eșantionului va fi:

$$\frac{1,96^2 \times 98 \times 2}{1} = \frac{3,84 \times 196}{1} = \frac{752,74}{1} = 753 \text{ persoane}$$

Formulele de calcul al volumului eșantionului, de mai sus, se referă la selecția simplă cu repetiție, deci în care, după extragere, fiecare unitate este reintrodusă în colectivitatea generală.

**Pentru selecția simplă fără repetiție se folosește formula:**

$$n = \frac{t_\alpha^2 \sigma_x^2}{\Delta_x^2 + \frac{t_\alpha^2 \sigma_x^2}{N}}$$

**Pentru selecția cu stratificare se folosește formula:**

$$n = \frac{t_\alpha^2 \sum_N \sigma_j^2}{\Delta_x^2 + \frac{t_\alpha^2 \sum_N \sigma_j^2}{N}}$$

în care  $N$  = mărimea colectivității generale.

Pentru a caracteriza fenomenele biologice sau medicale, sub aspectul legităților care stau la baza producerii sau influențării lor, din motivele amintite, nu facem cercetări sau observații pe „univers”, adică

pe întreaga „populație”, pe colectivitatea generală, ci pe părți reprezentative din colectivitate, din populație, pe eșantioane.

Aprecierea reprezentativității eșantionului sub aspect calitativ, sub aspectul omogenității deci, am văzut că se face cu ajutorul deviației standard, respectiv al coeficientului de variație. Aprecierea reprezentativității eșantionului sub aspect cantitativ, sub aspectul numărului de frecvențe sau de unități statistice, pe care făcând cercetarea să putem generaliza rezultatele, se face cu ajutorul unor formule matematice pe care le vom prezenta în continuare.

Alegerea părții din „populație” sau din colectivitatea generală – reprezentativă din punct de vedere calitativ și cantitativ – pe care făcând observația sau efectuând cercetarea să putem ajunge la rezultate sau concluzii similare cu cele obținute pe întreaga colectivitate, se numește eșantionare.

### METODE DE EȘANTIONARE

Alegerea frecvențelor sau unităților statistice care alcătuiesc eșantionul trebuie făcută de așa manieră, încât aceasta să ofere șanse egale de selecționare tuturor unităților, părțile componente ale colectivității generale să fie proporțional reprezentate, iar eșantionul – analog cu întregul – să reflecte toate caracteristicile, atributele și aspectele colectivității generale (populației).

Alegerea poate fi:

- **Întâmplătoare** când alegerea unităților statistice se efectuează prin tragere la sorți. În asemenea selecții trebuie asigurate șanse egale fiecărei unități a colectivității generale de a fi selecționate.
- **Mecanică** atunci când selecția unităților statistice se face respectând un anumit criteriu. Spre exemplu, de pe o anumită listă în care unitățile statistice ale colectivității au fost înscrise într-o anumită ordine (alfabetică, vârstă, localitate etc.), se aleg unitățile statistice ale eșantionului din 5 în 5 sau din 10 în 10 unități:
- **Stratificată** atunci când extragerea unităților statistice se face pe grupe omogene (straturi) ce reflectă întocmai structura pe aceleași grupe ale colectivității generale. Spre exemplu, dacă colectivitatea generală este reprezentată de 65% adulți, 25% copii și 10% bătrâni, atunci și eșantionul ales din colectivitatea generală să aibă proporțional aceeași structură pe grupe de vârstă a populației.

### Tipuri de eșantionare

În funcție de natura schemelor de eșantionaj abordate putem să clasificăm eșantionajele în eșantionaje aleatoare și nealeatoare sau empirice.

Din categoria eșantionajelor aleatoare fac parte:

- eșantionaje simple (elementare)
- eșantionaje stratificate
- eșantionaje în grupuri
- eșantionaje în trepte (faze)
- eșantionaje multifazice.

În ceea ce privește **eșantionajul simplu** (elementar), procedeul constă în extragerea la sorți a unităților care vor compune eșantionul (în general identice cu unitățile de selecție din populație) din baza de sondaj; procedeul presupune deci existența bazei de sondaj, adică o listă sau un fișier din care se extrag, într-o manieră aleatorie, unitățile de selecție. Baza de sondaj poate fi o listă a tuturor născuților vii, a femeilor însărcinate din țară sau dintr-un anumit teritoriu, o listă a tuturor persoanelor spitalizate, o listă a tuturor raioanelor, a comunelor, a familiilor dintr-un teritoriu etc.

**Eșantionajul stratificat** se utilizează când în interiorul populației de referință unitățile statistice se găsesc plasate pe categorii. Fiecare categorie este omogenă din punct de vedere al conținutului intern, dar categoriile diferă între ele. Așa este cazul populațiilor clasate după caracteristicile vârstă, mediu de proveniență, diagnostic al bolnavilor externi, clasați pe secții sau grupuri de boli etc. Se adoptă în acest caz o selecție proporțională cu volumul stratului sau cu volumul optim când din unele straturi mai puțin reprezentate se ia un număr sporit de unități.

Dacă unitățile statistice se găsesc plasate în grupuri asemănătoare, fiecare grup în parte având însă o compoziție eterogenă (cazul familiei, gospodăriei, dispensarului medical etc.) se consideră ca unitate de selecție grupul (familia, respectiv gospodăria) și nu individul, înregistrarea ulterioară a eșantionului beneficiind de un plus de operativitate și precizie, comparativ cu selecția simplă a unităților individuale.

**Eșantionajul în grupuri** este un procedeu folosit când lipsește baza de sondaj iar constituirea ei este dificilă și foarte costisitoare și în sfârșit când dispersia teritorială a populației de investigat este exagerată. Se pornește de la faptul că populația de studiat poate fi imaginată ca fiind constituită din unități de selecție ierarhizate.

De exemplu născuții vii sau gravidele aparțin la familii, acestea aparțin la o colectivitate care poate fi definită (cartier, comună) care la rândul lor aparțin la orașe sau raioane în cadrul unei țări.

**Eșantionajul în trepte**, este un procedeu în care unitățile de selecție sunt ierarhizate ca în cadrul eșantionajului în grupuri, diferența constând în faptul că în timp ce în cadrul cuiburilor se realizează o singură extragere aleatorie, în eșantionajul în trepte se efectuează mai multe trageri la sorți.

Astfel, în exemplu, țara-raioane-comune-familie-noi născuți, următoarele extrageri sunt posibile: sondajul de gradul I – raioane, rezultând raioane selecționate, sondajul de gradul II – comune, rezultând comune selecționate la prima extragere, sondajul de gradul III – familii, rezultând familii selectate din comunele alese în selecția în treapta a II-a. Familiile selectate constituie o bază de sondaj, din care se va face o nouă selecție, faza a IV-a, aceasta fiind eșantionul de investigat efectiv. Este recomandabil a se evita realizarea unui număr prea mare de extrageri pentru ca în eșantionajul cu mai multe faze, fiecare operație de extragere antrenează o eroare de eșantionaj. În exemplul descris mai sus, se observă că mai multe baze de sondaj sunt necesare pentru fiecare fază dar se evită necesitatea existenței unei liste a tuturor unităților de observare pentru țara întreagă, de exemplu.

**Eșantionajul multifazic** este un procedeu care combină metodele anterior prezentate. El constă în formarea unui eșantion primar pentru a obține informații cu caracter general urmat de constituirea unuia sau mai multor eșantioane din eșantionul primar pentru a obține informații suplimentare, mai specifice. Există și alte metode empirice de eșantionaj, mai mult sau mai puțin utilizate, care au unele facilități demne de avut în vedere, dar având o valoare științifică mai redusă, vor fi numai menționate. Este vorba de:

- metoda cotelor prin care se cere fiecărui investigator numărul și caracteristicile persoanelor care trebuie să fie intervievate
- metoda eșantioanelor fixe prin care se colectează periodic o serie de date din același eșantion
- metoda eșantioanelor tipice unde se consideră, de exemplu, că o localitate este reprezentativă pentru situația zonei.

#### Determinarea efectivelor unui eșantion

Tehnicile de determinare a efectivelor eșantioanelor pot fi complicate în cadrul unor tipuri de eșantionaj. În aceste situații se poate recurge

la 2 soluții. Prima constă în utilizarea din plin a echipei de informaticieni împreună cu tehnica de calcul din dotare, iar cea de-a doua se adresează la o serie de tabele prefabricate care dau pentru diferite tipuri de eșantionaj, mărimea eșantionului de studiu deja calculat, din literatura de specialitate.

Dacă baza de sondaj există sau poate fi constituită și dispunerea unităților de observare în bază este întâmplătoare există 3 modalități de alegere aleatorie:

- procedeul tragerii la sorți cu/fără restituire
- procedeul tabelor de nr. întâmplătoare
- procedeul pasului mecanic.

În condițiile prezentate mai sus și în funcție de tipul caracteristicii cel mai puțin dispersată (cât mai omogenă), volumul minim al eșantionului poate fi calculat prin formulele prezentate în Tab.39.

Tabelul 39.

Calculul volumului eșantionului

Nr. crt.	Tip de eșantionaj	Volumul eșantionului	
		Caract. cantitativă*	Caract. calitativă
1.	Simplu cu repetiție	$n = t_{\alpha,\gamma}^2 \cdot x \frac{S_x^2}{\Delta_x^2}$	$n = t_{\alpha,\gamma}^2 \cdot x \frac{p(1-p)}{\Delta_p^2}$
2.	Simplu fără repetiție	$n = \frac{t_{\alpha,\gamma}^2 \cdot S_x^2}{\Delta_x^2 + \frac{t_{\alpha,\gamma}^2 \cdot S_x^2}{N}}$	$n = \frac{t_{\alpha,\gamma}^2 \cdot p(1-p)}{\Delta_x^2 + \frac{t_{\alpha,\gamma}^2 \cdot p(1-p)}{N}}$
3.	Stratif. proporțională	$n = \frac{t_{\alpha,\gamma}^2 \sum N_j \sigma_j^2}{\Delta_x^2 + \frac{ac}{N}}$	$n = \frac{t_{\alpha,\gamma}^2 \sum N_j p_j (1-p_j)}{\Delta_x^2 + \frac{ac}{N}}$
4.	Stratif. cu volum optim	$n = \frac{t_{\alpha,\gamma}^2 (\sum N_j \sigma_j)^2}{t_{\alpha,\gamma}^2 (\sum N_j \sigma_j)^2 + N^2 \Delta_x^2}$	$n = \frac{t_{\alpha,\gamma}^2 (\sum N_j p_j (1-p_j))^2}{t_{\alpha,\gamma}^2 (\sum N_j p_j (1-p_j))^2 + N^2 \Delta_p^2}$

\* Dacă sunt necesare eșantioanele de volume mai mari de 30 unități de observare, în locul coeficientului  $t_{\alpha,\gamma}$ , se folosește  $U_{\alpha}$  pentru care:

$U_{0,05} = 1,96$ ;  $U_{0,01} = 2,57$ ;  $U_{0,001} = 3,29$

Notațiile din tabelul 39 au semnificațiile următoare:

$n$  = volumul eșantionului

$N$  = volumul populației de referință

$N_j$  = volumul stratului  $j$ ;  $1 \leq j \leq k$

$S_x(\sigma_j)$  = abaterea standard a caracteristicii în populație, respectiv în stratul  $j$ ;  $1 \leq j \leq k$

$p(p_j)$  = proporția caracteristicii

$\Delta_x(\Delta_p)$  = eroarea maximă limită acceptată

$t_{\alpha\gamma}$  = coeficient ce caracterizează distribuția lui Student, pentru riscul  $\alpha$  dat și  $\gamma$  = grade de libertate.

Aplicarea formulelor de determinare a volumului eșantionului din tabelul nr. 39, necesită unele judecăți și anume: se pornește de la dorința oricărui investigator ca volumul eșantionului să fie "optim" înțelegând prin aceasta că acest volum să nu fie nici prea mare, nici prea mic. Se știe că:

- un eșantion prea voluminos are, în general, o precizie bună, dar poate fi foarte costisitor sub aspect financiar, al timpului, al personalului etc.
- un eșantion de volum prea mic este mai eficient sub aspect economic dar poate conduce la rezultate cu o precizie mică, pe care nu le putem accepta; de aici în orice problemă de calcul al volumului eșantionului este nevoie de un compromis.

Analizând prima formulă din tabelul 39, corespunzătoare unui eșantionaj simplu cu repetiție, se vede că dimensionarea volumului eșantionului depinde de variabilitatea caracteristicii studiate (exprimată prin varianta  $S_x^2$ ), de precizia dorită (exprimată prin mărimea intervalului de încredere  $\Delta_x$ ) cât și prin riscul acceptat în estimare (exprimat prin valoarea coeficientului  $t_{\alpha\gamma}$ ).

În ceea ce privește variabilitatea, aceasta nu poate fi influențată pentru că ține de natura însăși a fenomenului biologic studiat; privind riscul acceptat în estimare ( $\alpha$ ) în orice studiu din domeniul biologicului sau al cercetării medicale se ia invariabil  $\alpha=0,05$  și atunci pentru dimensionarea optimă a volumului eșantionului ("realizarea compromisului") s-au adus unele modificări, acceptate în practică, privind eroarea maximă ca măsură a preciziei estimării parametrilor din populație prin parametrii corespunzători din eșantion.

Prezentăm în continuare în Tab.40 diferite variante obținute pentru valori ale erorii maxime, menținând constante atât variabilitatea fenomenului cât și probabilitatea (riscul  $\alpha$ ) de garantare a afirmațiilor.

Tabelul 40

Dimensionarea volumului eșantionului în funcție de eroarea acceptată

Nr. crt.	Eroarea maximă ( $\Delta_p$ )	$\sigma_p^2=p(1-p)$	Risc ( $\alpha$ )	Volumul eșantionului
1.	0,05	0,4x0,6	0,05	369
2.	0,06	0,4x0,6	0,05	256
3.	0,07	0,4x0,6	0,05	188
4.	0,08	0,4x0,6	0,05	144
5.	0,09	0,4x0,6	0,05	114
6.	0,10	0,4x0,6	0,05	92

Datele din Tab. 40 se referă la o caracteristică de tip calitativ, de genul bolnav-sănătos, alimentat, corespunzător-nealimentat, masculin-feminin etc.

După cum a fost menționat, când baza de sondaj nu există și populația este organizată în unități ierarhizate, pentru evaluarea volumului eșantionului pot fi folosite o serie de tabele din literatura de specialitate care prezintă numărul minim de grupuri ce trebuie selectate în funcție de diverse proporții așteptate și de volumul populației de referință.

Exemplificăm în Tab. 41 evaluarea numărului de grupuri necesare, pentru o eroare maximă așteptată de 0,05; calculele fiind datorate lui W. Lutz, 1982. (Calculele sunt efectuate pe baza formulei corespunzătoare unui eșantionaj simplu cu repetiție).

**Exemplul I.** Eșantion pentru populație.

Se dorește realizarea unui eșantion de populație reprezentativ la nivel național, necesar pentru efectuarea aceluiași studiu privind opiniile în legătură cu serviciile sanitare din Moldova în rândul beneficiarilor de servicii de sănătate.

Pentru realizarea acestui obiectiv pot fi parcurse următoarele etape:

- așa cum s-a menționat și în cadrul eșantionajului pentru medici, teritoriul țării a fost împărțit în 3 zone considerate tipice din punct de vedere geografic și al tradițiilor de existență ale populației;
- din aceste zone pot fi alese, prin tragere la sorți, 9 raioane plus municipiul Chișinău;



- se va adopta ca fracțiune de eșantionaj, valoarea de  $5 \text{‰}_{0000}$  ceea ce a condus la un efectiv total minim al eșantionului de 1200 persoane; acest efectiv poate fi confirmat printr-un calcul de determinare a volumului eșantionului pentru o precizie de 3% și o probabilitate de lucru de 95% ( $\alpha=0,05$ );
- eșantionul total poate fi distribuit la nivelul celor 9 raioane revenind în medie pentru fiecare raion selectat în jur de 120 persoane; pentru municipiul Chișinău se va lua un număr mai mare de persoane;
- la nivelul fiecărui raion se va avea în vedere distribuția aproximativ egală a populației în mediul urban și rural; aceasta înseamnă că din totalitatea străzilor și/sau comunelor trebuie selectate străzile și comunele care vor intra în eșantion;
- pentru obținerea celor aproximativ 60 de persoane din mediul urban și 60 din mediul rural, la nivelul fiecărui raion selectat, sunt suficiente selectarea a 3 străzi și 3 comune, iar din fiecare stradă/comună selectată vor fi alese la întâmplare prin metoda pasului mecanic sau prin folosirea unui tabel cu numere întâmplătoare, câte 10 familii;
- din familiile selectate, atât din mediul urban cât și rural, operatorul de interviu va consemna datele din fișa de anchetă pentru toate persoanele sau numai pentru persoanele adulte;
- prin însumarea datelor se obține eșantionul pentru copii, adulți, vârstnici, pe medii, la nivelul fiecărui raion iar în final eșantionul la nivel național.

Dacă aceste inconveniente nu întotdeauna sunt prezentate, se pot imagina și alte scheme de selecție adecvate în care principiul aleatorului se poate îmbina cu principiul selecției dirijate sau al metodei cotelor cu unele avantaje practice, demne de luat în seamă.

Tabelul 41

Volumul eșantionului pentru sondaj în grupuri\*

Nr. total de grupe formate	0,05	0,1	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
400	26	24	23	21	20	18	17	15	14	13
400	25	24	22	21	19	18	17	15	14	13
300	25	23	22	21	19	18	16	15	14	12
250	24	23	22	20	19	17	16	15	14	12
200	24	23	21	20	19	17	16	15	13	12
150	23	22	21	19	18	17	16	14	13	12
100	22	20	19	18	17	16	15	14	13	12
90	21	20	19	18	17	16	15	14	12	11
80	20	19	18	17	16	15	14	13	12	11
70	20	19	18	17	16	15	14	13	12	11
60	19	18	17	16	15	14	14	13	12	11
50	18	17	16	15	15	14	13	12	11	10
40	16	16	15	14	14	13	12	12	11	10
35	15	15	14	14	13	12	12	11	10	10
30	14	14	13	13	12	12	11	11	10	9
25	13	13	12	12	11	11	10	10	9	9
20	12	12	11	11	10	10	10	9	9	8
15	10	10	9	9	9	9	9	8	8	7
10	8	8	7	7	7	7	7	7	6	6

Deci, colectivitățile dinamice sunt formate din fluxuri de evenimente.

Pentru o înțelegere clară a relației dintre colectivitățile statice și cele dinamice trebuie subliniat faptul că în ambele cazuri există o mulțime de elemente variabile; în cazul colectivităților statice timpul și forma organizatorică sunt constante, iar în cazul celor dinamice, spațiul și forma organizatorică sunt constante.

*Unitățile statistice* reprezintă elementele constitutive ale colectivităților. Ca și colectivitățile, unitățile pot fi separate în *statice și dinamice*, în cazul celor statice unitățile compun efectivul (de persoane, de bolnavi etc.). În cazul unităților din colectivitățile dinamice, unitățile aparțin aceleiași structuri organizatorice, dar în condiții diferite de timp.

\* După W. Lutz "Sampling: how to select people households places to study community health", I.E.A., 1982.

Unitățile statistice pot fi *simple și complexe*. Cele simple sunt elementele constitutive ale colectivității (persoana, bolnavul etc.), iar cele complexe sunt rezultatul organizării sociale și economice a colectivității (familia, echipa, secția, anul de studiu etc.). Unitățile statistice sunt unități *independente* care pot fi studiate separat, pe subcolectivități sau pe întreaga colectivitate.

Unitățile colectivității pentru care se culeg date în procesul cunoașterii se deosebesc de unitățile de raportare, ce pot fi, de exemplu, instituții și altele, care potrivit legislației în vigoare informează sistematic asupra activității unităților aflate în structura lor organizatorică. Unitățile sunt purtătoare ale unor trăsături variabile în timp și spațiu.

*Caracteristicile statistice*, denumite și *variabile statistice*, reprezintă criteriile pe baza cărora se caracterizează unitățile colectivității. Aceste criterii pot fi însușiri sau trăsături ale unităților care definesc și delimitează între ele unitățile colectivității și care urmează să fie înregistrate. Formele concrete de manifestare ale caracteristicilor la nivelul fiecărei unități a colectivității se numesc *variante sau valori*.

Caracteristicile statistice se diferențiază după mai multe *criterii*.

După conținutul lor pot fi: de timp, de spațiu și atributive.

*Caracteristicile de timp* arată apartenența unităților la un moment sau o perioadă de timp.

*Caracteristicile de spațiu* arată situarea în teritoriu a unității. Se exprimă prin cuvinte pe baza unui nomenclator al unităților teritoriale.

Toate celelalte sunt *caracteristici atributive* și servesc pentru definierea fenomenelor studiate.

După modul de exprimare, ele se separă în: *caracteristici calitative* (exprimate prin cuvinte) și *caracteristici cantitative* (exprimate numeric), de exemplu, profesia și vârsta.

După natura variației, cele numerice se împart în caracteristici cu *variație continuă* și cu *variație discontinuă* sau *discretă*. Cele cu variație continuă, de exemplu, vârsta, care poate lua orice valoare într-un interval dat și numărul de copii, ca variabilă discretă ce dimensionează familiile și care nu poate lua decât valori strict determinate în intervalul de valori înregistrate la un moment dat. De reținut că, în statisticile practice toate variabilele numerice se înregistrează ca variabile discrete, de exemplu, vârsta în ani împliniți.

După modul de manifestare la nivelul unităților simple pot fi *alternative* – manifestarea directă sau opus ei, de exemplu, urban-rural – și

*nealternative* – cu variante distincte numerice sau calitative, de exemplu, salariile și profesia.

După modul de obținere și folosire a datelor pot fi *primare* – obținute în procesul de culegere a datelor și cu care se caracterizează nivelul de dezvoltare atins de unitățile simple și complexe și caracteristici *derivate* – obținute prin aplicarea unui model de calcul.

*Datele statistice* sunt caracterizări numerice ale unităților, grupelor și colectivității, obținute din observare și prelucrare. În statistică, datele sunt întotdeauna mărimi concrete, caracterizate printr-o parte noțională care definește conținutul calitativ, valoarea numerică și elementele de cuantificare (de timp, de spațiu etc.).

Mesajul datelor îl reprezintă *informația*. Datele ce caracterizează un fenomen sau proces economic sau social ce se repetă cu regularitate se numesc *indicatori statistici*.

*Indicatorii statistici* reprezintă expresia numerică a unei determinări calitative obiective, obținută în urma efectuării unei cercetări statistice raportată la condiții specifice de timp, spațiu și organizatorice.

Indicatorul statistic exprimă de regulă, numeric, o categorie economică, medicală etc.

Exprimarea numerică a unei categorii presupune folosirea mai multor indicatori, fiecare punând în evidență anumite aspecte esențiale ale acesteia.

### 3.4. Prezentarea materialului statistic

Prezentarea materialului statistic se poate face cu ajutorul tabelelor și cu ajutorul reprezentărilor grafice.

#### 3.4.1. Prezentarea prin tabele

**Prezentarea prin tabele** constituie o metodă comodă, sintetică și sistematică:

- **comodă**, pentru că rezultatele cercetării sunt redacte sub forma unor date cifrice, permițând observarea cu ușurință a aspectelor principale ale problemei cercetate;
- **sintetică**, pentru că datele cifrice redau aspectele esențiale ale fenomenului studiat;
- **sistematică**, pentru că între diferitele date cantitative sau calitative prezentate în tabel, există o înlănțuire logică ușurând înțelegerea lor.

Pentru ca un tabel să corespundă cerințelor, el trebuie să îndeplinească următoarele cerințe:

**Să aibă un titlu** care, printr-o frază clară, să redea conținutul tabelului, locul și perioada de timp la care se referă, ca și modul în care au fost obținute datele. Datele titlului tabelului trebuie să răspundă la 4 întrebări: ce, unde, când, cum? Ce cuprinde tabelul, unde s-a efectuat cercetarea, când s-a efectuat cercetarea și cum au fost obținute datele la întrebarea "cum" se răspunde de obicei printr-un asterix sub macheta tabelului.

**Exemplul 1.** Distribuția pe sexe și grupe de valori ale HTA a bolnavilor cu HTA din localitatea X în anul Y (Tab. 42).

Tabelul 42

Repartizarea pe sexe și grupe de valori ale TA a bolnavilor cu HTA din localitatea X în anul Y

Nr. crt.	Sex	Total	Grupe de valori TA			
			160-169,9	170-179,9	180-189,9	200-209,9
1.	Total	50	28	15	7	-
2.	F.	32	20	8	4	...
3.	M.	18	8	7	3	-

**Exemplul 2.** Distribuția pe cauze, grupe de vârstă și sex a îmbolnăvirilor stomatologice a elevilor liceului X anul Y (Tab.43).

Tabelul 43

Repartizarea pe cauze, grupe de vârstă și sex a îmbolnăvirilor stomatologice a elevilor liceului X anul Y

Nr. crt.	Afecțiunea	Număr elevi								
		Total	Grupe de vârstă		15-16		17-18		19-20	
			Sexul		Sexul		Sexul		Sexul	
			M	F	M	F	M	F	M	F
1.	Caria simplă	118	54	64	20	24	18	21	16	10
2.	Pulpită	15	7	8	3	4	3	2	1	2
3.	Gangrena	4	3	1	-	1	2	-	1	
4.	Extracții	18	10	8	5	2	1	3	4	3
	Total	155	74	81	28	31	24	26	22	24

**Rândurile și coloanele tabelului** trebuie să fie logic așezate și să fie notate corespunzător conținutului acestora. Spre exemplu, în tabelul 43 grupele de vârstă ca și grupele de afecțiuni sunt așezate într-o

ordine logică: crescând la grupele de vârstă și în ordinea gravității (de la caria simplă la extracții), în cazul grupelor de afecțiuni.

Totalurile în tabelele statistice se înscriu de obicei în primul rând și în prima coloană, în tabelele matematice totalurile se înscriu în ultimul rând și în ultima coloană, în statistică însă avem uneori situații în care tabelele sunt extinse depășind dimensiunile atât în lungime cât și în lățime. Dacă am înscrie totalurile în ultimul rând și în ultima coloană a tabelului am putea observa mai greu căror caracteristici le corespund aceste totaluri.

Atunci când nu dispunem de datele necesare pentru a le înscrie în căsuțele tabelului se trage o linie orizontală, semn al inexistenței datelor respective, iar atunci când nu cunoaștem datele necesare înscrierii într-o căsuță a tabelului, deși ele există, acest lucru se exprimă prin câteva puncte succesive.

### TIPURILE DE TABELE STATISTICE

Există în general o mare diversitate de tabele statistice, de la cele mai simple, care au la bază un singur criteriu de clasificare, până la cele mai complexe, având la bază mai multe caracteristici, depinzând atât de natura materialului cercetat cât și de ingeniozitatea celui care le întocmește. Este recomandabil însă să nu uităm lucrul cel mai important și anume faptul că tabelul trebuie să redea cu ușurință esențialul. În scop didactic am împărțit tabelele în 3 tipuri:

- tabele pentru clasificare dichotomică;
- tabele pentru distribuția de frecvențe;
- tabele de corelație.

**Tabelele pentru clasificarea dichotomică** le utilizăm în situația în care cercetăm fenomene care se caracterizează prin însușiri sau caracteristici diametral opuse, excluzându-se unul pe altul. A dichotomiza, în limba greacă, înseamnă a divide, a împărți în două. Spre exemplu, populația unei colectivități, distribuită sau repartizată pe sexe, se împarte în două părți: de sex masculin și de sex feminin; după starea de sănătate populația poate fi sănătoasă sau bolnavă; după mediul se poate împărți în populație rurală și urbană; după reacția la tuberculină (IDR) poate fi pozitivă sau negativă la tuberculină etc.

**Exemplul 1.** Distribuția elevilor din colectivitatea X în anul Y pe sexe, locul de domiciliu și prezența sau absența cariei dentare (Tab. 44).

Tabelul 44

Distribuția elevilor pe sexe, locul de domiciliu și prezența sau absența cariei dentare din colectivitatea X în anul Y

TOTAL	Sex	Domiciliu	Carii dentare	
500	M 300	R 200	P	150
			A	50
	U 100	P	60	
		A	40	
F 200	R 150	P	100	
		A	50	
	U 50	P	30	
		A	20	

În cazul utilizării unei astfel de prezentări nu trebuie să admitem mai mult de două, maxim trei dichotomizări, întrucât prin supraaglomerare ele devin confuze, greu de urmărit și interpretat.

**Tabele pentru distribuția de frecvențe** le utilizăm atunci când dorim să prezentăm în tabele rezultatele grupării după anumite criterii.

**Tabele de corelație** le folosim pentru evidențierea corelației dintre două fenomene între care în mod logic există o legătură de dependență. Aceste tabele se caracterizează prin aceea că au două variabile: una determinantă (factorială) și cealaltă determinată (rezultantă), în cazul tabelului de corelație valorile înscrise în căsuțele tabelului corespund în același timp unei anumite grupe de valori a primului fenomen determinant și unei anumite grupe de valori a celui de-al doilea fenomen determinat de primul. Ca atare cele două variabile ale tabelului de corelație trebuie să fie împărțite într-un număr egal de grupe de valori pentru ca ele să se coreleze perechi.

**Exemplul 2.** Încercând să stabilim dacă între vârsta bolnavilor hipertensivi și valorile TA există sau nu o legătură de dependență, utilizăm un astfel de tabel de corelație în care una din variabile (factorială) este reprezentată de grupa de vârstă, iar cealaltă (determinată sau rezultantă) este reprezentată de grupele de valori ale TA (Tab. 45).

Tabelul 45

Distribuția bolnavilor hipertensivi în raport cu grupa de vârstă și valorile TA

Grupe de vârstă	Valori TA			
	160-169,9	170-179,9	180-189,9	190-199,9
40-44	5			
45-47		10		
50-54			15	
55-59				20

Se observă o legătură strânsă între creșterea valorilor TA și înaintarea în vârstă, legătură care va fi măsurată printr-un coeficient de corelație, care va fi studiat la capitolul respectiv.

**Exemplul 3.** Dacă dorim să stabilim într-o colectivitate dacă între vârstă și parodontopatie există o legătură de dependență sau nu, utilizăm un astfel de tabel de corelație în care una din variabile (factorială) este reprezentată de grupele de vârstă, iar cealaltă variabilă (determinată sau rezultantă) este reprezentată de valori sau frecvențe ale dinților afectați (Tab. 46).

Tabelul 46

Corelația între grupele de vârstă și numărul de dinți afectați

Grupa de vârstă	Nr. dinți cu parodontită					
	1-3	4-6	7-9	10-12	13-15	16-19
40-44	2					
45-49		3				
50-54			8			
55-59				12		
60-64					25	
65-69						50

Distribuția frecvențelor în raport cu grupele de vârstă și numărul de dinți afectați de parodontopatie reliefează faptul că între cele două fenomene există o legătură strânsă și directă de dependență întrucât pe măsura înaintării în vârstă numărul dinților afectați este mai mare.

### 3.4.2. Reprezentarea grafică a materialului statistic

Tabelele, deși redau foarte exact aspecte cantitative ale unui fenomen cercetat, fiind în același timp sintetice și sistematice, necesită totuși o anumită pregătire de specialitate pentru a putea fi urmărite și

înțelese. Ca atare ele pot fi utilizate în lucrări de cercetare științifică prezentate sub formă de articole în reviste de specialitate care se adresează unui public cu pregătire corespunzătoare. Ele redau mai mult aspecte statice ale fenomenului cercetat, încărcarea tabelor cu date referitoare la perioade lungi de timp îngreunează urmărirea și înțelegera acestora.

Graficele, deși mai puțin exacte decât tabelele, au o arie de utilizare mai largă întrucât sunt mai intuitive, putând fi urmărite și înțelese mai ușor de către un public mai larg.

### Componentele principale ale unui grafic

Graficul este de fapt o hartă care “vorbește” direct ochiului și este foarte eficientă în crearea unei imagini în mintea receptorului. Construirea unui grafic înseamnă atât știință cât și artă; înseamnă cunoașterea graficelor de bază, modalitatea de realizare din punct de vedere tehnic și mai ales cum se combină elementele de bază pentru a crea o hartă corespunzătoare; alegerea tipului de reprezentare grafică pentru situații dificile, complexe ale evoluției fenomenelor înseamnă de fapt artă. Persoana care realizează graficul își asumă o responsabilitate mare pentru că un grafic corect conduce la informare corectă iar un grafic gândit și construit incorect conduce la dezinformare (Anders Wallgren, Britt Wallgren, Rolf Persson, 1990). Atunci când un grafic este privit, ochiul trebuie să înregistreze imediat caracteristicile principale precum și unele detalii; pentru aceasta persoana care construiește graficul trebuie să realizeze un echilibru între detaliu și ansamblu. Un grafic stimulează receptorul să facă conexiuni și să observe imediat modelele care pot apărea în evoluția unui fenomen.

#### Rolul graficului:

- să orienteze utilizatorul în selectarea informațiilor importante din raportul statistic (“gateway”)
- să dezvolte idei ulterioare și să le explice
- să încurajeze utilizatorul să privească în profunzime problema prezentată în raportul statistic
- să încurajeze compararea și analiza informațiilor

Graficele sunt construite cu ajutorul unor elemente simple cum ar fi linii, arii, text etc. Toate aceste elemente simple trebuie combinate astfel încât graficul rezultat să aibă sens și să poată fi citit cu ușurință.

Persoana care are responsabilitatea reprezentării grafice trebuie să-și pună următoarea întrebare:

### Care este tipul de grafic potrivit pentru reprezentarea grafică a datelor statistice?

- Alegerea depinde în primul rând de problema pe care dorim să o punem în evidență prin reprezentarea grafică.
- Nu există o regulă care să stabilească o corespondență între un anumit tip de relație între variabile și un anumit tip de grafic.
- Trebuie luați în considerare diferiți factori pentru fiecare situație în parte.
- Avem de ales uneori între:
  - grafice corespunzătoare și necorespunzătoare unei anumite situații sau,
  - între mai multe tipuri de grafice potrivite aceleiași situații
- Trebuie să decidem în primul rând ce dorim să punem în evidență cu ajutorul aceluia grafic: evoluția în timp, variațiile, etc.
- Trebuie subliniate principalele caracteristici ale datelor și să ținem cont de limitele impuse de acestea. Caracteristicile importante în acest context sunt:
  - structura datelor
  - tipul variabilelor
  - caracteristicile măsurătorilor.

### Principalele componente ale unui grafic sunt( fig. 8, fig. 9):

1. suprafața de reprezentare (“chart area”)
2. suprafața graficului mărginită de axe și cadran (“plot area”)
3. aria graficului
4. legenda graficului
5. rețeaua de axe – liniile orizontale și verticale (“gridlines”)
6. etichetele corespunzătoare axelor

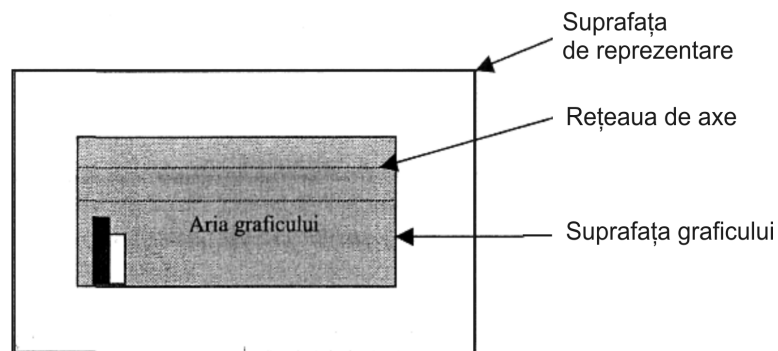


Fig. 8. Componentele principale ale unui grafic

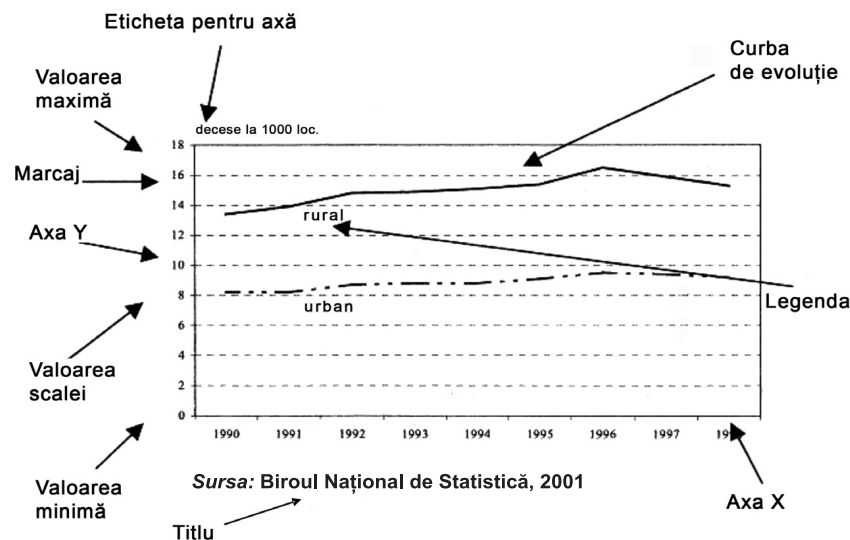


Fig. 9. Rata mortalității populației pe medii în localitatea X pe anii 1990-1999.

**Suprafața de reprezentare** este o suprafață imaginară ce conține întregul grafic incluzând titlul și textele explicative. În mod normal suprafața de reprezentare nu este marcată în nici un fel, cu excepția cazului în care fundalului i se atribuie o altă culoare.

**Suprafața de reprezentare este mai mare decât cea a graficului. Include graficul și textul corespunzător acestuia astfel încât să constituie o unitate coerentă.**

Suprafața de reprezentare include:

- titlul graficului (hărții)
- axa verticală Y
- eticheta atribuită axei Y
- valoarea maximă
- valorile scalei
- marcajul care indică limita dintre valorile scalei ("tick")
- valoare minimă
- axa orizontală X
- eticheta atribuită axei X
- suprafața încadrată
- curbe, coloane etc.
- legenda
- sursa

**Suprafața graficului** este suprafața ocupată de grafic fără titlu, valorile scalei etc. Suprafața trebuie să fie suficient de mare pentru a obține un grafic vizibil. Graficele care conțin multe detalii necesită o suprafață mai mare. Proporțiile suprafeței sunt determinate până la un anumit punct de date – o serie cronologică mare se reprezintă de obicei într-un dreptunghi. Ca regulă, formatul orizontal este mult mai potrivit și o alegere bună este dată de proporția 1:1,6.

1,6



Uneori șirul 1:1,2 – 1:2,2 este dat ca reper pentru a stabili raportul dintre cele două laturi ale dreptunghiului format de axa verticală și cea orizontală.

**Aria graficului**

Aria graficului este o arie imaginară care conține desenul propriu-zis (coloane, benzi, curbe etc.). Poate fi colorată într-o nuanță deschisă pentru a pune mai bine în evidență conținutul graficului.

**Rețeaua de axe orizontale și verticale.**

În unele cărți este denumită “rețea ajutătoare de reprezentare grafică” pentru a nu fi confundată cu rețeaua rectangulară formată din sistemul coordonatelor carteziane.

Rețeaua de axe permite citirea graficului cu mai multă ușurință și se pot aprecia cu aproximație valorile. Axele orizontale se utilizează în cazul reprezentării prin dreptunghiuri verticale, în timp ce axele verticale se utilizează în cazul reprezentării prin dreptunghiuri orizontale. În ambele cazuri este mult mai ușor de apreciat lungimea dreptunghiurilor. În diagramele în care reprezentarea este prin curbe se recomandă utilizarea ambelor rețele de axe.

Rețeaua de axe nu trebuie să fie obstructivă și nu trebuie să distragă atenția de la datele reprezentate. Axele trebuie să fie cât mai subțiri. Dacă fundalul este colorat atunci culoarea albă pentru axe este o alegere potrivită.

**Textul**

Toate informațiile necesare înțelegerii graficului trebuie incluse în suprafața de reprezentare astfel încât aceasta să fie dominată de grafic și nu de text. Textul trebuie să cuprindă strictul necesar.

**Titlul și alte texte**

Titlul trebuie să descrie într-o formă concisă conținutul graficului. Din titlu trebuie să reiasă foarte clar:

- grupul descris (femei între 25-29 ani)
- variabilele implicate (cauza decesului)
- anul la care se referă datele 2005
- tipul de date reprezentate (valori absolute, procente).

Titlul trebuie:

- să fie ușor de citit
- să înceapă cu informația cea mai importantă

Este recomandat ca titlul să apară deasupra graficului pentru a respecta o ordine naturală în citirea lui. Alinierea titlului cu textul corespunzător axei Y conferă unitate graficului și paginii de text în care este inserat acesta.

Centrarea titlului poate fi utilizată atunci când graficul este prezentat pe foile destinate retroproiectorului. Titlul poate fi scris cu alte caractere decât cele folosite pentru text.

Legenda și etichetele trebuie scrise cu caractere mai mici decât cele folosite pentru titlu.

**Sursa**

Pentru fiecare grafic se menționează sursa de proveniență a datelor, sursa trebuie precizată sub grafic iar cuvântul “sursă” este recomandat să fie scris cu litere italice.

*Orice text trebuie scris pe orizontală indiferent dacă se referă la axa verticală sau la curba reprezentată grafic.*

**Axe, scale, marcaje**

Axele sunt reprezentate prin linii drepte. Nu există motive pentru utilizarea săgeților. Pentru “axa Y” textul se plasează deasupra axei. Pentru axa X” textul se plasează în partea dreaptă sub axă. În cazul seriilor cronologice nu este necesar ca anii să fie însoțiți de cuvântul “anul”. Valoarea minimă a axei Y este întotdeauna “0”. Valoarea maximă este aleasă astfel încât să existe un mic spațiu între grafic și valoarea maximă a scalei. În cazul în care avem mai multe grafice unul lângă altul trebuie păstrată aceeași scală. În acest fel comparațiile pot fi realizate în mod direct.

Pentru a ajuta citirea graficului axele sunt marcate la intervale egale, iar marcajele sunt însoțite de valorile corespunzătoare. Pentru a evita încărcarea graficului, o parte din valorile scalei pot fi omise păstrându-se marcajele.

În general o singură axă verticală și o singură axă orizontală sunt suficiente. În cazul seriilor cronologice foarte lungi, scala din stânga poate fi repetată în partea dreaptă.

Valorile scalei sunt alese astfel încât să se refere la un sistem numeric natural. Practica standard prevede ca dimensiunea intervalului pe scală să fie 1, 2,..., 5, sau alternativ 10, 20,..., 50 etc.; pentru numere zecimale se folosește 0.1, 0.2,..., 0.5. Scalele construite de tipul 30, 60, 90 se citesc mai greu. Excepție fac scalele care reprezintă procente. În acest caz valorile scalei 25, 50, 75% sunt naturale. O altă excepție este atunci când valorile scalei sunt importante prin ele însele, de exemplu grupele de vârstă. Este recomandat să nu se reprezinte pe scală valorizări; se recomandă folosirea categoriilor mii astfel, încât pe axă să se poată scrie 100, 200 etc., în loc de 100000, 200000.

**Legenda**

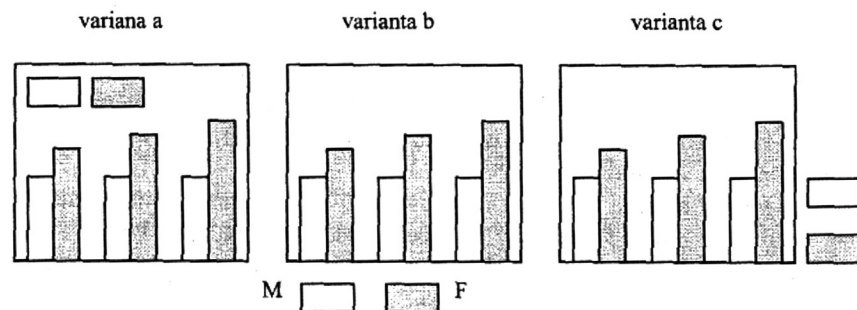
Aceleași principii se aplică atât pentru arii cât și pentru curbe. Trebuie specificat ce reprezintă fiecare arie sau curbă.

În cazul *ariilor*, legenda poate fi scrisă direct pe aria respectivă sau în afara ei în partea dreaptă. Reprezentarea ariilor în legendă trebuie făcută în aceeași direcție cu reprezentarea din grafic, de la stânga la dreapta sau de sus în jos.

Pentru *dreptunghiurile* verticale sau orizontale grupate, trebuie să explicăm ce reprezintă fiecare grup în parte. Legenda trebuie poziționată în suprafața grafică, sub grafic, sau lateral în partea dreaptă a graficului (variante a, varianta b, varianta c).

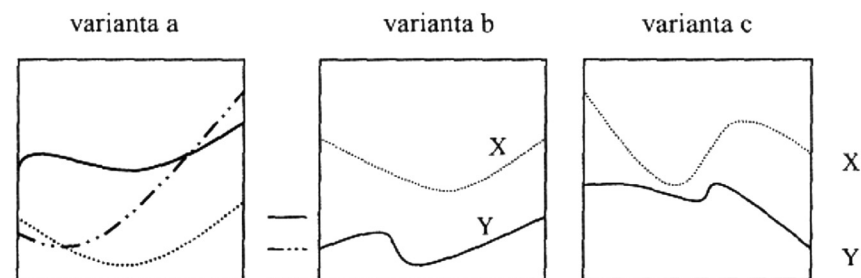
Pentru barele verticale grupate, legenda poate fi așezată orizontal în suprafața grafică, dedesubtul ei sau în partea dreaptă.

În cazul dreptunghiurilor orizontale, legenda este poziționată vertical fie în suprafața grafică fie în afara ei, fie în partea dreaptă.



În cazul *curbelor*, trebuie specificat în mod clar ce reprezintă fiecare curbă. Legenda poate fi poziționată astfel:

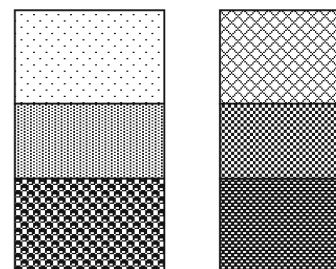
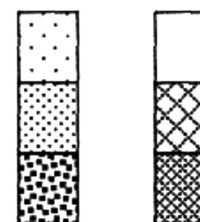
- sub suprafața grafică sau în partea dreaptă a ei – în cazul în care avem reprezentat un număr mare de curbe, pentru a nu încărca graficul (variante a);
- deasupra lor atunci când sunt vizibil distincte și nu se intersectează (variante b);
- în afara suprafeței grafice, în partea dreaptă când sunt apropiate sau se intersectează (variante c).

**Hașurări/ culori**

Pentru a putea face comparații cu ușurință se recomandă respectarea unor principii elementare în utilizarea culorilor și hașurărilor.



Se recomandă folosirea a cel mult 4 sau 5 tipuri de hașurări.



Se recomandă accentuarea gradată a intensității hașurilor pornind de la partea superioară a graficului către partea inferioară.





Nuanțele trebuie să se distingă cu ușurință. Trebuie aranjate de la nuanța cea mai deschisă către cea mai închisă. Pot fi completate cu alb, iar în cazuri excepționale cu negru.

#### Hașurarea ariilor

Liniile trebuie înclinare în aceeași direcție.

Trebuie evitate

- liniile verticale, deoarece creează un efect de lungire a coloanelor/benzilor
- liniile orizontale, deoarece creează un efect de micșorare a acestora.

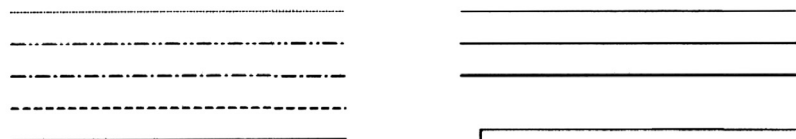
Când avem mai multe categorii, modelele trebuie să se diferențieze în mod vizibil. Dacă există o ordine între categorii, hașurările trebuie să arate acest lucru.

#### Culorile:

- trebuie aplicate cu economie – culorile puternice îndepărtează mesajul graficului, iar costurile de editare sunt ridicate;
- este recomandată folosirea unei singure culori discrete și eventual a diferitelor nuanțe ale acesteia.

#### Tipuri de linii

Reprezentarea grafică a unei singure serii cronologice se realizează printr-o curbă continuă de culoare neagră. Curba trebuie să fie suficient de groasă pentru a nu se confunda cu rețeaua de axe dar nu



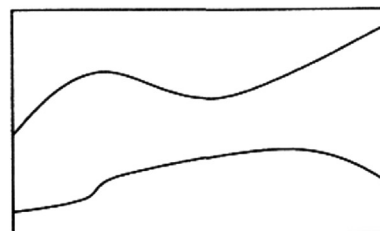
Grosimea liniei trebuie să fie adaptată la dimensiunea graficului.

foarte groasă. Ca dimensiuni potrivite se recomandă 0,5 mm pentru curbă și 0,1 mm pentru rețeaua de axe.

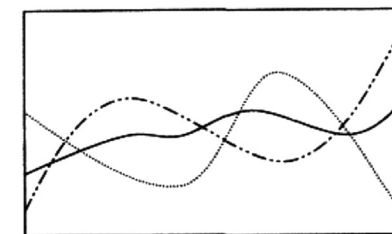
Pentru reprezentarea în același grafic a mai multor curbe trebuie să folosim tipuri diferite de linii pentru a le putea distinge. Teoretic există un număr mare de combinații de linii și puncte; Practic este dificil de folosit mai mult de 4-5 (varianta a, varianta b). De asemenea, putem folosi curbe de dimensiuni diferite. Cea mai subțire trebuie să fie diferită de rețeaua de axe, iar cea mai groasă nu trebuie să fie prea groasă.

În cazul în care curbele nu se intersectează și sunt vizibil distincte se poate folosi un singur model. În cazul în care curbele se intersectează este necesară utilizarea mai multor modele astfel încât utilizatorul să nu fie obligat să stabilească ce porțiuni aparțin aceleiași curbe pe baza intuiției.

varianta a



varianta b



#### Simbolurile

Seriile cronologice sunt alcătuite din observații unite prin linii. Uneori observațiile corespunzătoare fiecărei perioade (lună, an) pot fi marcate prin simboluri.

x - - - - x - - - - x  
o - - - - o - - - - o

Simbolurile au un rol important în cartograme. În aceste cazuri trebuie să fie simple – cercuri sau pătrate, în general figuri geometrice simple de dimensiuni relativ mici.

#### Tehnicile de bază în construirea unui grafic

Elementul de bază de la care se pornește în construirea unui grafic este sistemul de coordonate (sistem de referință). Sistemul de coor-

donate uzual constă în linii așezate în unghi drept și numite sistem de coordonate cartezian. Sistemul de coordonate din figura 10 este bidimensional dar, ideea poate fi generalizată pentru trei sau mai multe axe (dimensiuni). În mod tradițional dreapta verticală este numită ordonată (sau “axa Y”) și dreapta orizontală este numită abscisă (“axa X”). Acest sistem de coordonate împarte planul în patru cadrane. Originea sau punctul “0” este punctul unde se intersectează axele și începe scala numerică. Valorile din partea superioară a axei Y sunt pozitive, în timp ce valorile din partea inferioară sunt negative. Valorile din partea stângă a axei X sunt negative iar valorile din partea dreaptă sunt pozitive.

În general, statistica folosește pentru reprezentarea grafică a informațiilor cadrantul unu și foarte rar cadrantul unu împreună cu cadrantul patru. De exemplu majoritatea indicatorilor sociali sunt pe o scală pozitivă, ceea ce determină utilizarea mai frecventă a cadrantului unu, celelalte cadrane fiind prin convenție omise. Fiecărui punct “A” din cadran i se asociază perechea de valori (x, y).

Intersecția celor două axe indică nivelul fenomenului într-o anumită perioadă a evoluției sale.

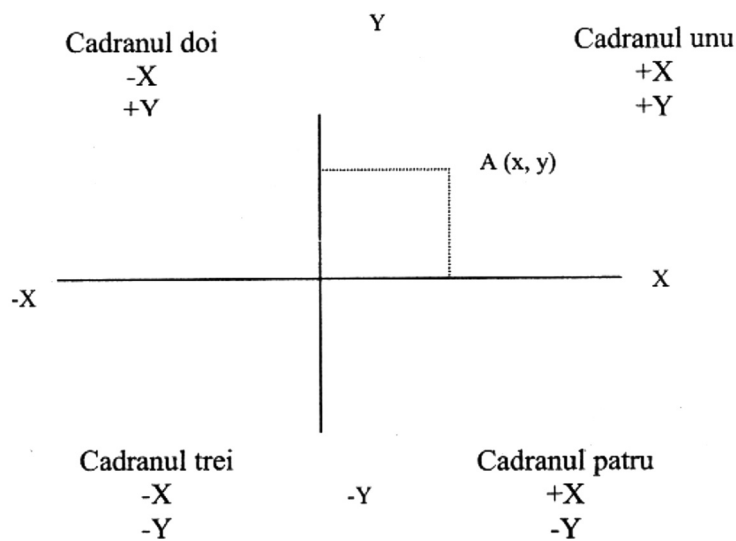


Fig.10. Sistemul de coordonate cartezian.

Din punct de vedere al tehnicii de construire graficele pot fi grupate în mod neconvențional în două tipuri de grafice:

1. **Grafice fundamentale**, de bază (histograma, poligonul, ogiva, diagrama lineară)
2. **Grafice speciale** (diagrama prin coloane, piramida populației/vârstelor, graficele de structură care utilizează cercul („pie chart”), hărți statistice, cartograme cu diferite simboluri, grafice semilogaritmice etc.)

### 1. Grafice fundamentale

Seriile de variație pot fi reprezentate prin următoarele tipuri de diagrame de distribuție: histograma, poligonul, ogiva și diagrama liniară, printre aceste tipuri se vor descrie histograma și diagrama lineară care sunt frecvent utilizate în cercetarea din domeniul sănătății.

#### **Histograma**

Histograma este o reprezentare grafică, prin dreptunghiuri alăturate, a distribuției frecvențelor sau procentelor. Ea se utilizează când autorul graficului dorește să arate frecvența pentru o variabilă continuă (de exemplu vârsta). Histograma poate descrie și distribuția procentuală. Frecvența cazurilor (sau procentajul) pentru fiecare categorie este reprezentată de un dreptunghi. Aria dreptunghiului este proporțională cu frecvența cazurilor din categoria respectivă.

#### **Aria dreptunghiului = Baza × înălțimea**

Scala categoriilor variabilei este reprezentată pe “axa X” prin distanțe egale. Scala frecvențelor este reprezentată pe “axa Y” prin distanțe egale corespunzătoare frecvențelor egale. Axa Y începe întotdeauna din punctul “0” pentru a evita problemele ce ar putea să apară în compararea bazelor. Axa X începe însă din orice punct convenabil (inferior) de pe axă. Obiectivul este acela de a obține o figură cu aria totală egală cu frecvența totală N (sau procentul total 100%).

**Exemplul 1.** În Tab. 46 sunt prezentate datele necesare pentru construirea unei histogramme.

Tabelul 46

Frecvența în cadrul fiecărei categorii a persoanelor investigate (%)

Ani de școlarizare	Distribuția frecvențelor	Frecvența în cadrul fiecărei categorii a persoanelor investigate (%)
4 ani	21	9,4
5 ani	53	23,8
6 ani	69	30,9
7 ani	47	21,1
8 ani	33	14,8
	N = 223	100,0%

Histograma utilizează valori absolute sau relative.

Tehnica este aceeași, dar pe axa Y sunt reprezentate valori absolute sau relative în funcție de situație. Pe axa X este reprezentată variabila pe care o măsurăm (intervalul de clasă), iar pe axa Y este reprezentată frecvența pentru fiecare interval de clasă; înălțimea dreptunghiului este proporțională cu numărul de persoane din categoria respectivă.

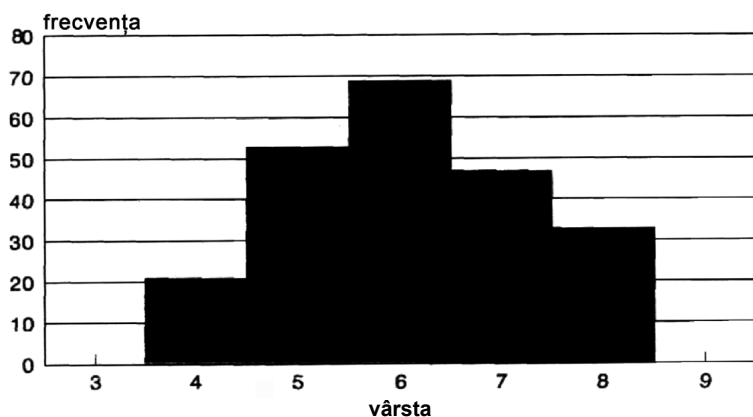


Fig.11. Histograma.

Unele variații ale histogramei de bază sunt folosite pentru a reflecta anumite caracteristici ale datelor. Spre exemplu, dacă variabila este "nominală" dreptunghiurile din histogramă pot fi separate astfel încât să vizualizeze separat și distinct categoriile. De obicei acest grafic mai este numit și diagrama prin coloane iar dreptunghiurile au lățimile egale.

Pentru variabilele "ordinale", unde nu sunt definite distanțe egale dreptunghiurile pot avea spații între ele pentru a evidenția acest lucru (în particular dacă variabila este discretă și mai puțin în cazul în care este continuă). În acest caz, histograma poate fi de tip scară și da impresia de ordonare a categoriilor (marcajul este în interiorul fiecărei coloane). Pentru variabilele ordinale există tradiția de a lua o lățime standard, deși distanțele nu sunt definite. Acest lucru duce la evitarea distorsiunilor aparente ce pot să apară datorită variației lățimilor.

Histograma este potrivită în cazul variabilelor de tip "interval de variație", discrete. Dacă variabilele sunt de tip "continuu" există alte tipuri de grafice mai potrivite (exemplu, poligonul frecvențelor).

#### Poligonul frecvențelor

Poligonul frecvențelor sau al procentelor este o figură închisă ce unește punctele dintre centrele de interval și frecvențele lor. Este o alternativă la histogramă.

O altă modalitate de reprezentare grafică este distribuția cumulativă a frecvențelor sau procentelor. Acest tip de grafic arată frecvența, respectiv procentele cazurilor situate sub marginea superioară a fiecărei clase succesive.

#### Diagrama liniară (cronogramă sau historiogramă)

Reprezintă forma specifică pentru reprezentarea seriilor dinamice, cronologice. Este utilizată pentru descrierea evoluției în timp a unui fenomen (natalitate, fertilitate, mortalitate etc.).

Diagrama liniară prezintă valoarea anumitor variabile dependente (reprezentate pe axa Y) pentru fiecare categorie a variabilei independente (reprezentate pe axa X). Punctele de pe grafic sunt unite printr-o linie dreaptă, iar figura nu este închisă cu axa X. Aria de sub curbă nu are un înțeles particular așa cum se întâmplă în cazul histogramei sau a poligonului.

În acest tip de grafic ochiul urmărește ușor curba de evoluție și sesizează imediat modificările care s-au produs în timp în evoluția fenomenului observat. Este frecvent utilizată în monitorizarea stării de sănătate a unei comunități (natalitate, mortalitate, morbiditate).

Fig. 12 este un exemplu de diagramă liniară. Variabila independentă este timpul reprezentat pe axa X iar variabila dependentă este rata (reprezentată pe axa Y).

Acest tip de grafic este util atunci când una dintre variabile (de pe axa X) este continuă, este deci, variabilă de tip interval (exemplu vârsta și timpul).

Se poate folosi și o diagramă prin coloane în cazul în care variabilele de pe axa X sunt *nominale* sau *ordinale*. În acest caz barele verticale sunt mai indicate pentru comparații. Pentru seriile cronologice urmărirea vizuală a coloanelor în scopul observării evoluției este mai dificilă.

În diagrama lineară curbele pentru seriile cronologice sunt mult mai ușor de urmărit și în același timp oferă o descriere (un desen) a evoluției în timp a fenomenului observat.

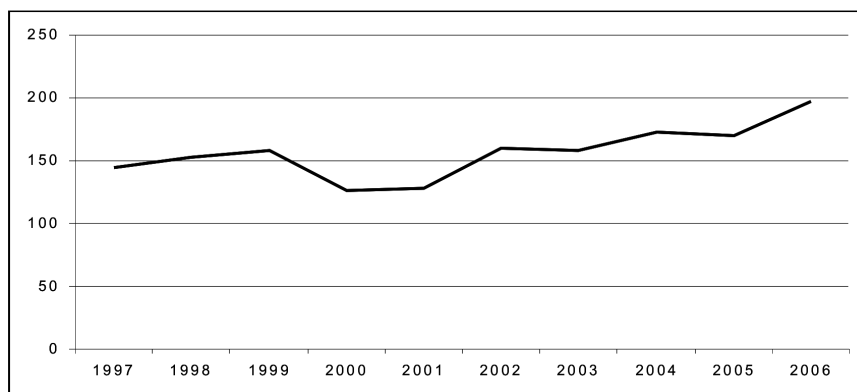


Fig 12. Incidența prin maladiile tractului gastrointestinal în Republica Moldova, 1997-2006 (la 10000 locuitori).

### Atenție

- Axa Y trebuie să fie  $\frac{3}{4}$  din lungimea axei X (sau de aceeași lungime).
- Întotdeauna axele frecvențelor sau ale procentelor încep de la punctul 0 sau de la origine. Axele scorurilor pot începe de la orice scor convenabil pentru a obține o diagramă clară.
- Fiți siguri că diferențele numerice egale sunt reprezentate prin distanțe fizice egale pe toate scalele.
- Etichetarea corectă a graficului – includerea scalelor, sursa datelor, titlul explicativ, note explicative, etc. Evitați confuziile: nu reprezentați mai multe grafice diferite utilizând aceleași sistem de axe.

## 2. Grafice speciale

### Diagrama prin coloane

Este cea mai simplă formă de grafic. Acest tip de diagramă este utilizat atunci când se urmărește:

- reprezentarea mai multor fenomene în același loc și în același timp, *sau*
- același fenomen în mai multe locuri dar în același timp (ex. rata de mortalitate generală pe raioane în anul 2006).

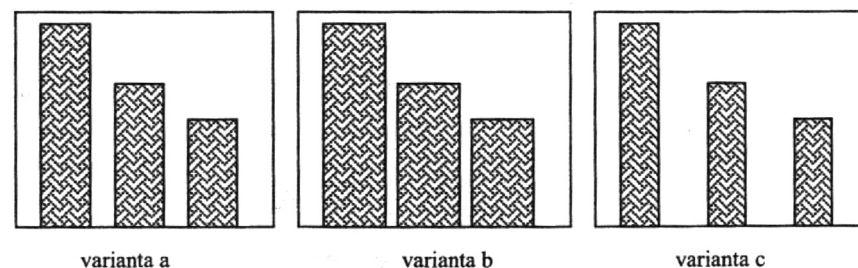
Diagrama prin coloane este ușor de reprezentat grafic și ușor de citit. Se utilizează atunci când dorim să reprezentăm valori distincte ale variabilelor – variabile *calitative* sau *discrete*. Pentru ilustrarea acestui lucru coloanele sunt separate de spații.

Sunt utilizate pentru reprezentarea grafică a frecvențelor absolute sau a frecvențelor relative, a sumelor sau a mediilor.

Pe axa X sunt reprezentate variabilele, în timp ce pe axa Y sunt reprezentate frecvențele.

*Utilizarea procentelor permite mult mai bine compararea mulțimilor de date de dimensiuni diferite.*

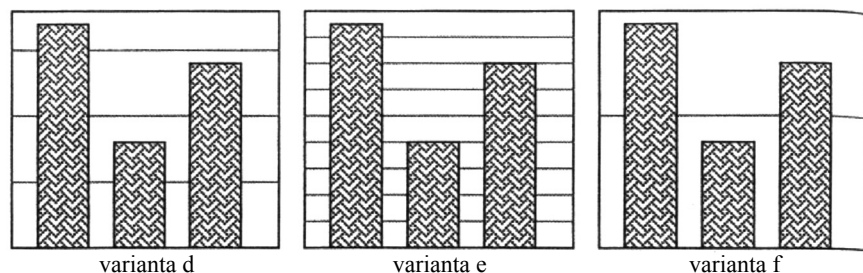
Coloanele trebuie să fie mai late decât spațiile dintre ele, iar spațiile trebuie să fie bine definite, astfel încât graficul de tip coloană să nu poată fi confundat cu o histogramă (varianta a este corectă)



Rețeaua de axe este utilă pentru comparații și pentru citirea aproximativă a valorilor. Dacă axele rețelei sunt în număr prea mare graficul este greu de citit, iar dacă sunt în număr prea mic prezența lor nu se justifică (varianta e și varianta f).

Din punct de vedere al orientării barele pot fi:

- verticale
- orizontale

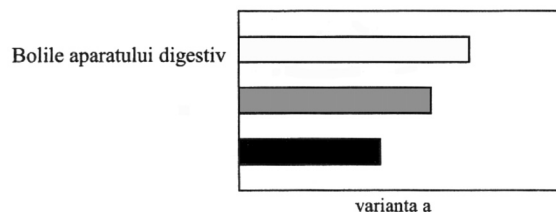


### Diagrama prin bare orizontale

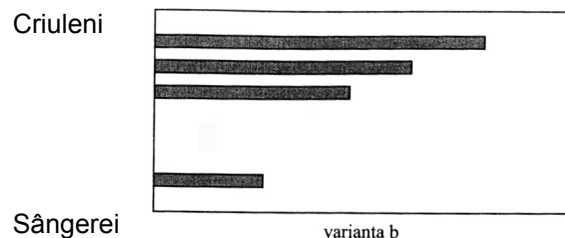
În acest tip de grafic, pe axa X sunt reprezentate frecvențele, în timp ce pe axa Y sunt reprezentate variabilele. Barele orizontale sunt preferate în locul barelor verticale în două situații:

- variabilele au șiruri lungi de caractere (de exemplu – bolile aparatului digestiv) – varianta a
- număr mare de valori a variabilei (de exemplu sunt trecute 30 de raioane) – varianta b

Pentru un număr mare de valori ale variabilei nu este suficient spațiu pe axa X pentru a reprezenta dreptunghiurile cu valorile corespunzătoare.



Problemele de reprezentare a dreptunghiurilor încep de la 6-8 valori ale variabilei.



Pentru a pune în evidență conținutul graficului se vor respecta câteva principii.

- *Ordonarea barelor*
  - ordonarea adecvată a valorilor variabilei oferă un grafic mai bun
  - pentru o variabilă calitativă există o libertate în ordonarea valorilor sale
  - aranjarea valorilor în ordinea descrescătoare a frecvențelor creează un interes mai mare
  - neordonarea valorilor conferă o imagine haotică
- *Gruparea barelor* este recomandată în două situații:
  - pentru a descrie simultan două sau mai multe categorii
  - diferite categorii sunt reprezentate în același sistem de axe. Pentru diferențierea categoriilor se folosesc hașurări sau culori diferite și se atașează graficului o legendă.

Se recomandă să nu se reprezinte mai mult de două-trei categorii, un număr mai mare va face dificilă înțelegerea graficului (de exemplu: masculin, feminin; urban, rural; salariat, patron, lucrător pe cont propriu).

Dacă există un număr mai mare de categorii se vor folosi mai multe grafice cu coloane obișnuite.

Coloanele grupate sunt reprezentate fie pe verticală fie pe orizontală. Cele două grafice produc efecte diferite, chiar dacă aria are aceeași dimensiune în ambele cazuri.

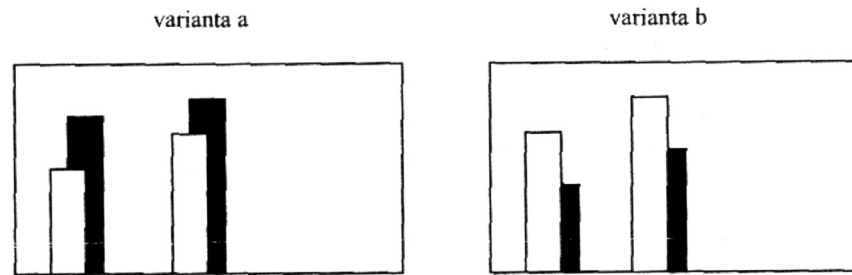
În cazul coloanelor orizontale, axa procentelor este lungă, astfel încât diferențele dintre valori pot fi observate mai clar.

O alternativă la coloanele grupate o reprezintă *coloanele suprapuse parțial* ("overlapping bars").

Avantajul în acest caz este dat de economisirea spațiului, iar reprezentarea grafică devine mai interesantă.

Există totuși riscul ca suprapunerea să facă graficul greu de înțeles sau să conducă la neînțelegeri (varianta b) Se recomandă în cazul în care:

- toate valorile unei categorii sunt mai mici decât valorile celeilalte categorii (varianta a)
- există posibilitatea de a reprezenta categoria cu valori mai mici în față cu o culoare mai deschisă (varianta a)



O alternativă de grupare a coloanelor o reprezintă coloanele aranjate sub formă de piramidă, “stivuite” (“stacked charts”). Este cunoscută sub denumirea de diagramă de structură prin dreptunghiuri (Fig. 13). Dreptunghiurile sunt reprezentate unul deasupra celuilalt. În acest caz, suprafața dreptunghiului care reprezintă întreaga colectivitate (100%) este divizată în părți proporționale cu ponderea specifică a fiecărei componente care alcătuiește întregul. Se recomandă un număr relativ mic de componente (variabile).

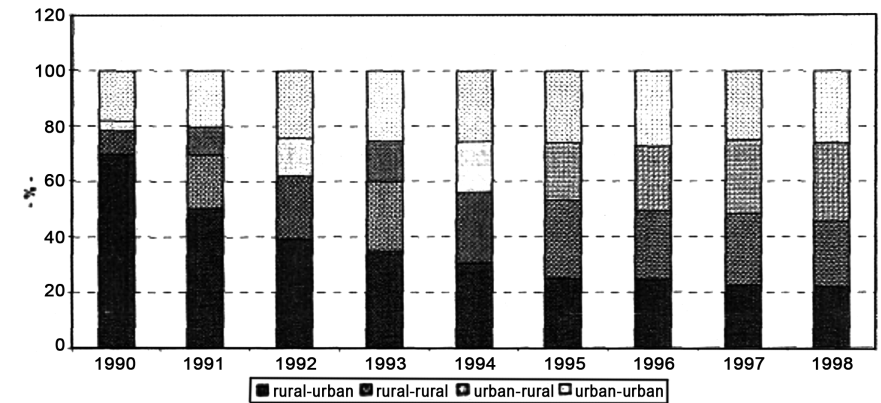
Pentru valoarea fiecărei variabile, înălțimea coloanei corespunde cu frecvența totală a categoriei respective. Cu precizie, poate fi citită numai dimensiunea categoriei de la bază; celelalte categorii pot fi apreciate cu aproximație.

Hașurările sau culorile diferite indică divizarea totalului în categoriile componente.

*Coloane suprapuse vs. coloane “stivuite”*

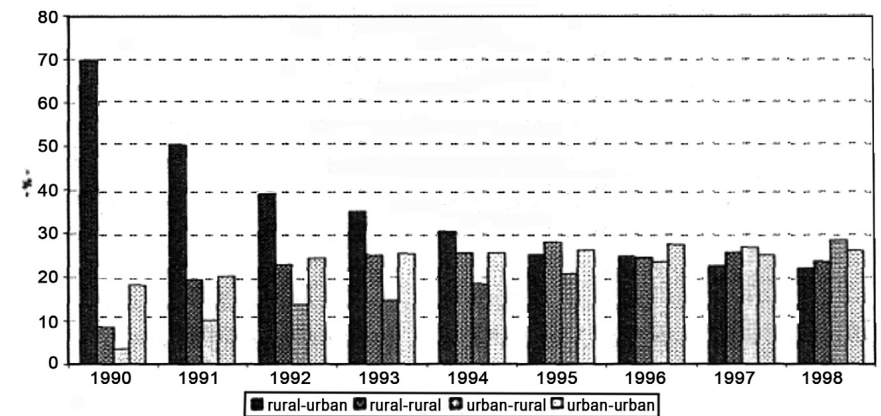
- ambele reprezintă situații similare
- alegerea se limitează la ceea ce dorim să subliniem cel mai mult
- în cazul coloanelor suprapuse este ușor să compari între ele categorii diferite dar mai dificil să înțelegi ce se întâmplă la nivelul întregii categorii
- în cazul coloanelor “stivuite” ansamblul este vizibil, în timp ce dimensiunea fiecărei categorii este secundară.

Aceași informație poate fi reprezentată din două puncte de vedere utilizând două tipuri de grafice.



Sursa: Tendințe Sociale, INS, UNICEF, 2001

Fig.13. Ponderea fluxurilor de migrație în migrația totală.



Sursa: Tendințe Sociale, INS, UNICEF, 2001

Fig. 14. Frecvența fluxurilor de migrație în migrația totală.

**Exemplul 2:** (Fig. 13 și Fig. 14).

În reprezentarea din Fig. 13 se observă că ponderea fluxului de migrație rural-urban a fost de trei ori mai mare în 1990 decât în 1998. Ponderea fiecărei categorii este comparată cu întregul.

În Fig. 14 se observă distribuția fluxului de migrație pe diverse categorii.

*Diagrama de structură care utilizează ca reprezentare cercul* ("pie chart")

Aceasta este tipică pentru reprezentarea distribuției procentelor în cazul variabilelor calitative și constituie o alternativă la diagrama prin coloane (Fig. 15). Acest tip de grafic permite să se stabilească în ce raport se găsesc grupele din cadrul colectivității față de colectivitatea în ansamblul ei (proporția bolilor aparatului digestiv din totalul bolilor).

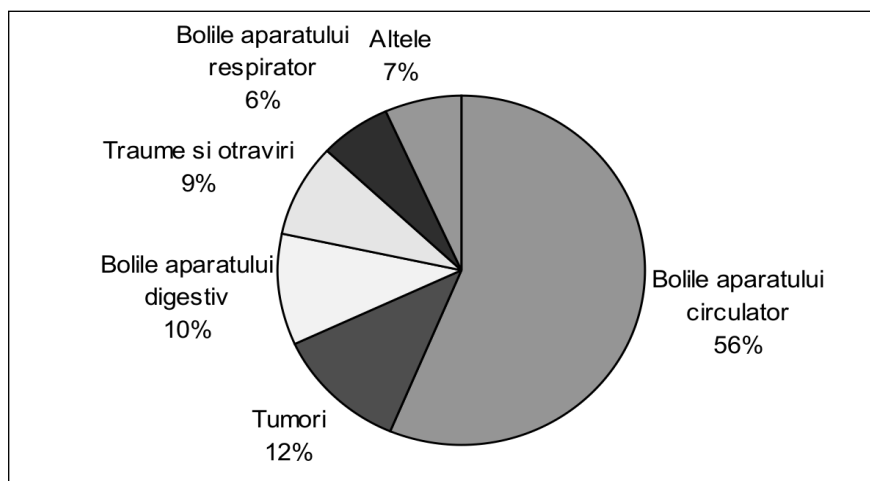


Fig. 15. Structura deceselor pe cauze medicale în Republica Moldova, 2006 (%).

Caracteristicile acestei diagrame sunt:

- totalul ariei reprezintă 100% iar 1% corespunde la 3,6 grade;
- în aria cercului se scriu procentele și nu gradele;
- diagrama se citește în sensul acelor de ceasornic începând cu punctul care poate fi asociat orei 12;
- cercul nu trebuie să aibă mai mult de 5-6 sectoare;
- culorile trebuie să fie atribuite începând cu cea mai închisă până la cea mai deschisă;
- cum această diagramă oferă o vedere de ansamblu, ultima categorie trebuie să fie "altele" și culoarea sau hașurările nu trebuie să fie dominante.

Dacă dorim să comparăm mai multe grupuri de date folosind această diagramă, atunci vom construi o diagramă pentru fiecare grup în parte. Aria fiecărui cerc va fi proporțională cu dimensiunea grupului. În acest fel putem compara atât dimensiunile grupurilor cât și distribuțiile procentuale din cadrul fiecărui grup simultan.

*Diagrama prin benzi*

Diagrama prin benzi reprezintă o alternativă a histogramei. În acest tip de grafic dreptunghiurile sunt foarte înguste și capătă aspectul de benzi. Benzile sunt orizontale. Se poate alege între numere și procente. Sunt două modalități de reprezentare a procentelor:

- procentele pot fi procente din total bărbați și separat procente din total femei
- procente din totalul populației

Un exemplu de diagramă prin benzi este "*piramida vârstelor*" ("population pyramid") – Piramida vârstelor descrie populația unei țări sau regiuni pe sexe și grupe de vârstă. Constă din două histograme orizontale, una pentru bărbați și una pentru femei. Histogramele sunt așezate în oglindă pentru a se face compararea între sexe, în populația unei țări/regiuni. Pe axa orizontală putem reprezenta atât valori absolute cât și procente.

Lungimea benzilor este proporțională cu valorile reprezentate iar lățimea lor este aceeași pentru toate benzile.

Tehnica este utilizată și în alte situații, de exemplu:

- pentru a reprezenta proporția de fumători în diferite grupuri de vârstă la bărbați și respectiv la femei
- status marital etc.

**Hărți statistice**

Hărțile statistice ocupă un loc special în reprezentarea variațiilor spațiale (geografice) a diverselor fenomene sau probleme sociale de sănătate. Dintre acestea, în sistemul de sănătate sunt frecvent utilizate *cartogramele*, care reprezintă o combinație dintre grafic și hartă (Fig. 16).

Cartogramele exprimă distribuția nivelelor unui fenomen într-o arie geografică (țară, raioane, regiuni). Diferențele dintre proporții, rate, medii etc., sunt puse ușor în evidență prin colorarea/hașurarea diferită a ariilor geografice. Permite compararea valorilor dintre diferitele zone în același timp; compararea pune în evidență atât zonele cu valorile cele

mai ridicate, cât și zonele cu valorile cele mai scăzute. Se obține o vizualizare spațială a informațiilor despre fenomenul observat (fertilitate, mortalitate etc.)

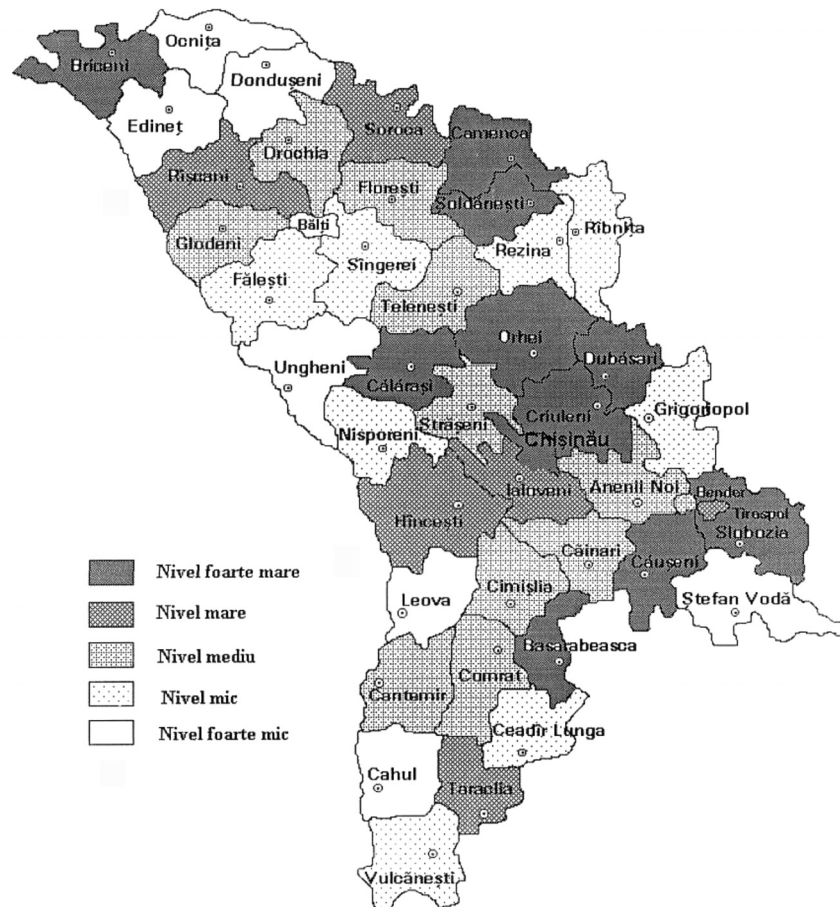


Fig. 16. Clasificarea raioanelor după nivelul infirmității motorii cerebrale, anul X(%).

În tehnica de construire a unei cartograme pașii importanți sunt:

- se grupează pe intervale de mărime valorile fenomenului observat;

- se stabilește un cod de culori sau hașurări prin care vor fi exprimate valorile corespunzătoare fiecărui interval de mărime;
- se colorează sau se hașurează fiecare zonă conform grupului de valori din care face parte;
- în legendă se va explica codul utilizat.

În domeniul sănătății publice s-a dezvoltat, în ultimii ani, conceptul de HEGIS (Health Environment Geographic Information System) care are la bază GIS (Geographic Information Systems).

### Aranjarea graficelor în raportul statistic

În procesul de elaborare a raportului statistic, *obiectivele autorului* au un rol major. Aceleași date pot fi reprezentate în moduri diferite conform mesajului pe care autorul dorește să-l transmită.

**Exemplul 3.** Mai jos este prezentat același indicator – rata de incidență prin ciroză hepatică – din două perspective diferite conform intențiilor autorului.

În **varianta a**, scopul este de a prezenta evoluția ratei de incidență pe o anumită perioadă relevantă de timp, 5 ani. În **varianta b**, mesajul este complet diferit și corespunde unui alt scop – compararea ratei de incidență pe 5 ani. Deoarece este vorba de o serie temporală mesajul transmis de aceste date se referă la evoluție. Diferența se poate observa în modul în care sunt organizate graficele și din explicațiile prezentate în text.

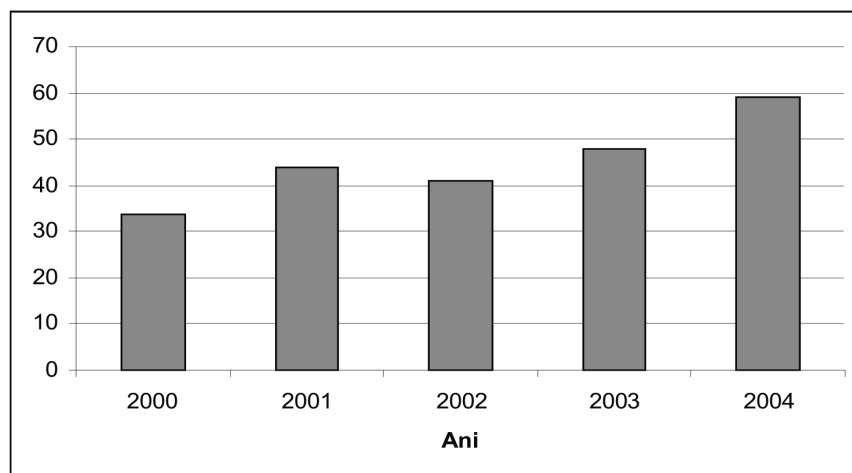
În **varianta a**, evoluția este subliniată de reprezentarea anilor pe axa orizontală. Acesta este modul obișnuit de reprezentare a datelor temporale. De asemenea pentru a evidenția evoluția, anii sunt sortați în ordine crescătoare. Textul asociat descrie evoluția în ansamblu și subliniază neregularitățile remarcabile.

În **varianta b**, intențiile autorului sunt complet diferite, (obiectivul este **compararea**).

Pentru comparații sunt necesare schimbări structurale. În **varianta b** anii sunt reprezentați pe axa verticală eliminând impresia de evoluție. Această schimbare este importantă pentru perceperea graficului, deoarece eliminând diferențele false, mesajul devine mai clar.



## Varianta a: un singur obiectiv – evoluția



## Varianta b: un singur obiectiv – compararea

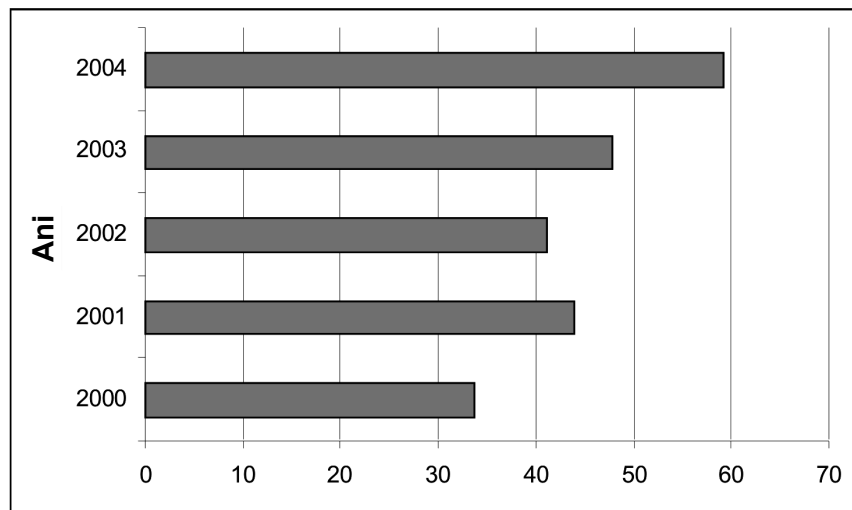


Fig. 17. Rata de incidență prin ciroză hepatică în Republica Moldova, 2000-2004 (la 100 mii locuitori).

## Interpretare:

**Varianta a** – Tendința incidenței este de creștere din 2000 până în 2004, observându-se un nivel mai scăzut în 2000 și 2002.

**Varianta b** – în perioada 2000-2004 nivelul cel mai scăzut al ratei de incidență prin ciroză hepatică s-a înregistrat în 2000 iar cel mai ridicat în 2004.

În *varianta b* anii sunt tratați ca variabile nominale în loc de variabile ordinale și astfel, sortate în ordine crescătoare după valorile variabilelor. Această reordonare are două efecte pozitive: pe de o parte, distruge imaginea evoluției, deoarece anii nu sunt prezentați în succesiune și pe de altă parte, permite o mai bună comparare a valorilor.

La rândul lui textul asociat este diferit de cel din *varianta a* – în locul descrierii modului în care are loc evoluția ratei de incidență, se menționează numai anii cei mai buni și cei mai slabi. Această diferență este importantă pentru autor, mai ales atunci când dorește să transmită mai multe mesaje.

Există situații în care două obiective pot fi combinate pentru a fi reprezentate grafic. În acest caz reprezentarea grafică ține cont în primul rând de obiectivul principal.

**O prezentare grafică greșită are consecințe mai grave decât inexistența ei.**

*Anders Wallgren*

**Lista de verificare (“Check-list”)**

Înainte de a începe construirea graficului

- Care este grupul țintă?
- Care este rolul graficului?
- Ce tip de grafic trebuie ales?
- Cum trebuie prezentat graficul?
- Cât de mare trebuie să fie graficul?
- Este reprezentarea grafică cea mai bună alegere de prezentare a informațiilor?

După ce construirea graficului a fost finalizată

- Este graficul ușor de citit?
- Poate fi graficul interpretabil (interpretat greșit)?
- Sunt potrivite forma și mărimea graficului?
- Este integrat graficul corect în text ?

Se recomandă testarea graficului: se solicită unei persoane care are caracteristicile grupului țintă citirea și interpretarea graficului. For-

*mulați întrebări despre grafic astfel încât să înțelegeți cum este el perceput de persoana aleasă.*

### 3.5. Analiza și interpretarea materialului statistic

Datele statistice, obținute în cadrul operației de adunare și prelucrare a materialului statistic, necesare efectuării unei lucrări științifice, sunt de obicei date brute sau mărimi absolute. Operația de prelucrare continuă cu unele operații de finețe în care datele brute sunt transformate în mărimi relative sau mărimi medii.

Sub această formă de indicatori statistici, deci raportate la aceeași bază, datele obținute pot fi supuse unor operații de comparare, corelare, abstractizare și generalizare, ceea ce ne permite o interpretare cât mai realistă a materialului cercetat. După ce materialul statistic a fost prelucrat și datele obținute sub formă de valori absolute, relative sau medii, au fost prezentate în tabele sau grafice, urmează etapa de analiză și interpretare a materialului statistic. Această operație constă dintr-o serie de aprecieri, comparații și corelații a rezultatelor obținute.

#### Aprecierea tendințelor în timp

Valorile înregistrate de noi în raport cu timpul (an, semestru, trimestru, luni) pot fi mai mari sau mai mici, pozitive sau negative, în funcție de modificarea unor factori care pot contribui la determinarea sau influențarea evoluției fenomenului cercetat, într-un sens sau altul.

Folosind aceste aprecieri comparative putem sesiza rolul pozitiv sau negativ pe care factorii respectivi (factori de risc sau de protecție) l-au exercitat asupra fenomenului cercetat.

**Exemplul 1.** Cercetând morbiditatea prin HTA într-o colectivitate și determinând anumite valori ale acesteia, pentru a vedea în ce măsură recomandările făcute pentru eliminarea factorilor de risc au fost eficiente, comparăm valorile găsite de noi în momentul examinării cu valorile morbidității prin HTA în aceeași colectivitate cu 5 sau 10 ani în urmă. Dacă valorile găsite de noi ale morbidității (sub raportul indicelui de incidență sau prevalență) sunt mai reduse decât cele înregistrate cu 5 sau 10 ani în urmă, deci s-au redus de la o prevalență de 18% spre exemplu la o valoarea a indicelui de prevalență de 12%, înseamnă că măsurile recomandate, în timp, pentru reducerea factorilor de risc (alimentari, efort fizic, stres) au fost eficiente.

**Exemplul 2.** Cercetând morbiditatea prin afecțiuni stomatologice într-o colectivitate industrială care lucrează în mediu cu vapori de mercur, trebuie să comparăm valorile morbidității stomatologice găsite de noi în momentul examinării cu valorile pe care această morbiditate le-a înregistrat cu 5 sau 10 ani în urmă și cu condițiile de muncă existente în perioada respectivă de timp.

#### Aprecierea tendințelor fenomenelor în spațiu

Rezultatele obținute în cercetarea noastră trebuie apreciate apoi comparativ cu rezultatele pe care alți cercetători le-au obținut în colectivități spațiale diferite cu condiții de muncă și de trai similare sau diferite. Aprecierea rezultatelor obținute de noi în raport cu rezultatele înregistrate în colectivități diferite în care există factori de risc mai numeroși, ne permite să reliefăm rolul factorilor de risc în determinarea sau influențarea acestei patologii.

Spre exemplu dacă în colectivitatea A examinată de noi, valorile prevalenței prin HTA sunt de 12% iar în colectivitatea B, cu condiții similare cu cele ale colectivității A, înregistrăm valori ale prevalenței de 18% înseamnă că munca de prevenire și combatere a unor factori de risc existenți în colectivitatea B lasă de dorit. Dacă există condiții geografice, social-economice, comportamentale diferite, iar asistența medicală este de aceeași calitate, diferența valorică poate fi pusă pe seama unor factori de risc insuficient cunoscuți și stăpâniți, în felul acesta putem sesiza și evalua rolul pe care unii factori climatici, social-economici, culturali, igienico-sanitari, de asistență medicală etc. îl exercită în determinarea sau influențarea stării de sănătate a populației.

**Corelarea rezultatelor obținute cu factorii de risc** care sunt în măsură să influențeze fenomenul cercetat. Asemenea factori pot fi factori naturali ai mediului fizic extern (clima, altitudinea, condiții hidrologice, microclimat, apă, aer etc.) factori biologici, factori social-economici (locuință, alimentație, obiceiuri, mentalități, condiții de muncă etc.). În urma operației de corelare se poate stabili măsura în care unul dintre factorii de risc influențează, într-o mai mare sau mai mică măsură, starea de morbiditate printr-o anumită afecțiune și totodată se poate face o ierarhizare a acestora stabilindu-se și ordinea de influențare a acestora.

După ce am efectuat, în cadrul operației de analiză și interpretare, aprecieri, comparații și corelații ale rezultatelor obținute în raport cu complexul de factori de mediu extern, socio-economici, biologici, ce ar

fi putut contribui la determinarea sau influențarea, într-un sens sau altul a fenomenului cercetat, se trece la formularea de concluzii.

**Exemplul 3.** Dacă în colectivitatea A factorul de risc pentru HTA nu influențează semnificativ valorile TA, iar în colectivitatea B valorile crescute ale TA se corelează cu o solicitare nervoasă de lungă durată și cu o alimentație nerațională se impune o schimbare a ritmului de lucru în întreprinderea respectivă și corectarea modului de alimentație ale angajaților

**Exemplul 4.** Dacă morbiditatea prin stomatită mercurială a crescut față de valorile anilor trecuți, fiind mai mare decât morbiditatea unor colectivități cu condiții asemănătoare, rezultă că au existat deficiențe în supravegherea stării de sănătate a colectivității respective.

Lucrarea **se încheie** cu capitolul **propuneri**.

Întrucât orice lucrare cu caracter aplicativ, pe care o efectuăm, urmărește un scop practic, este firesc ca ea să se încheie cu acest capitol de propuneri, în cadrul acestui capitol propunerile elaborate trebuie să fie cât mai concrete, să fie ușor aplicabile, vizând totodată și căile de rezolvare a problemelor.

**Exemplul 5.** În cazul frecvenței crescute a bolii hipertensive în colectivitatea B se vor face propuneri de modificare a proceselor tehnologice cu automatizarea unor operații solicitante, se vor indica micropauze la anumite intervale de timp, se va institui un program de educație pentru sănătate în vederea instituirii unui regim alimentar rațional etc.

**Exemplul 6.** Pentru reducerea cazurilor de stomatită mercurială se vor lua măsuri cu caracter tehnic, menite să limiteze concentrația vaporilor de mercur în aerul atmosferic, la locul de muncă, se vor indica măsuri de depistare activă prin consultații periodice active, de tratament prompt și eficient al cazurilor recent depistate, de educație sanitară a muncitorilor expuși, cu accent pe măsurile de igienă individuală, buco-dentară etc.

## BIBLIOGRAFIE

1. Albert, J. M., Yun, H. (2001). Statistical advances in AIDS therapy trials. *Statistical Methods in Medical Research* 10:85–100.
2. Andersen, P. K., Klein, J. P., Knudsen, K. M., Tabanera y Palacios, R. (1997). Estimation of variance in Cox's regression model with shared gamma frailties. *Biometrics* 53:1475–1484.
3. Babiker, A. G., Cuzick, J. (1994). A simple frailty model for family studies with covariates. *Statistics in Medicine* 13:1679–1692.
4. Bailar JC III. The promise and problems of meta-analysis. *New England Journal of Medicine* 337:559-560,1997.
5. Bjorling, L. E., Hodges, J. S. (1997). Rule-based ranking schemes for antiretroviral trials. *Statistics in Medicine* 16:1175–1191.
6. Bloom BS, et al. A reappraisal of hepatitis B virus vaccination strategies using cost-effectiveness analysis. *Annals of Internal Medicine* 118:298-306,1993.
7. Borzak S, Kidker PM. Discordance between metaanalyses and large-scale randomized controlled trials: examples from the management of acute myocardial infarction. *Annals of Internal Medicine* 123:873-877,1995.
8. Buyse, M., Molenberghs, G. (1998). Criteria for the validation of surrogate end-points in randomized experiments. *Biometrics* 54:1014–1029.
9. Carr, A., Samaras, K., Thorisdottir, A., Kaufmann, G. R., Chisholm, D. J., Carsenten, B. (1996). Regression models for interval censored survival data: application to HIV infection in Danish homosexual men. *Statistics in Medicine* 15:2177–2189.
11. Clayton, D., Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society A* 148:82–117.
12. Clemens JD, et al. The BCG controversy: a methodological and statistical reappraisal. *Journal of the American Medical Association* 249:2362-2369, 1983.
13. Daniels, M. J., Hughes, M. D. (1997). Meta analysis for the evaluation of potential surrogate markers. *Statistics in Medicine* 16:1965–1982.
14. DeGruttola, V., Fleming, T. R., Lin, D. Y., Coombs, R. (1997). Validating surrogate markers – are we being naive? *Journal of Infectious Diseases* 175:237–246.
15. Delta Coordinating Committee (1996). Delta: a randomised double-blind controlled trial comparing combinations of zidovudine plus didanosine or zalcitabine with zidovudine alone in HIV-infected individuals. *Lancet* 348:283–291.
16. Delta Coordinating Committee and Virology Group (1999). An evaluation of HIV RNA and CD4 cell count as surrogates for clinical outcome. *AIDS* 13: 565–573.
17. DeMasi, R. A., Babiker, A. G. (1998). Nonparametric estimation of the proportion of treatment effect explained by a surrogate marker. Presented at the 1998 ENAR conference, April 1, 1998.

18. Dempster, A. P., Laird, N.M., Rubin, D. R. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39:1–38.
19. Efron, B. (1988). Logistic regression, survival analysis and the Kaplan Meier curve. *Journal of the American Statistical Association* 83:414–425.
20. Ellenberg, S. S., Finkelstein, D. M., Schoenfeld, D. A. (1992). Statistical issues arising in AIDS clinical trials (with discussion). *Journal of the American Statistical Association* 87:562–583.
21. Fieller, E. C. (1940). The biological standardisation of insulin. *Journal of the Royal Statistical Society* 7:S1–S15.
22. Finkelstein, D. M., Schoenfeld, D. A., Stamenovic, E. (1997). Analysis of multivariate failure time data from an AIDS clinical trial. *Statistics in Medicine* 16:951–961.
23. Finkelstein, D. M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics* 42:845–854.
24. Foulkes, M. A. (1998). Advances in HIV/AIDS statistical methodology over the past decade. *Statistics in Medicine* 17:1–25.
25. Freedman, L. S., Graubard, B. I., Schatzkin, A. (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine* 11:167–178.
26. Gehan, E. A. (1965). A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika* 52:620–653.
27. Hughes, M. D. (2000). Analysis and design issues for studies using censored biomarker measurements with an example of viral load measurements in HIV clinical trials. *Statistics in Medicine* 19:3171–3191.
28. Jekel IF, et al. Influence of the prevalence of infection on tuberculin skin testing programs. *Public Health Reports* 84:883-886, 1969.
29. Journot, V., Chene, G., Joly, P., Saves, M., Jacqmin-Gadda, H., Molina, J.-M., Salamon, R., the ALBI Study Group (2001). Viral load as a primary outcome in Human Immunodeficiency Virus trials: a review of statistical analysis methods. *Controlled Clinical Trials* 22:639–658.
30. Kay, R. (1984). Multistate survival analysis: An application in breast cancer. *Methods of Information in Medicine* 23:157–162.
31. Klein, J. P., Moeschberger, M. L. (1997). *Survival analysis: techniques for censored and truncated data*. New York: Springer-Verlag.
32. Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* 48:795–806.
33. Kulinskaya E.V., Satarov G.A. Testing the Hypothesis about the Quality of Classification in the Data-Analysis System Clams // *Statistical Data Analysis. Abstracts of International Conference*. Sofia, 1990.
34. Munink M, et al. *Decision Making in Health and Medicine: Integrating Evidence and Values*. Cambridge: Cambridge University Press, 2001.
35. Mureşan P. *Manual de metode matematice în analiza stării de sănătate*. Bucureşti, 1989, – 574 pag.

36. Newcombe, R. G. (1988). Explanatory and pragmatic estimates of the treatment effect when deviations from allocated treatment occur. *Statistics in Medicine* 7:1179–1186.
37. O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics* 40:1079–1087.
38. Pickles, A., Crouchley, R. (1995). A comparison of frailty models for multivariate survival data. *Statistics in Medicine* 14:1447–1461.
39. Prentice, R. L., Williams, B. J., Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* 68:373–379.
40. Raab, G. M., Parpia, T. (2001). Random effects models for HIV marker data: Practical approaches with currently available software. *Statistical Methods in Medical Research* 10:101–116.
41. Robins, J. M., Greenland, S. (1994). Adjusting for differential rates of prophylaxis therapy for PCP in high-dose versus low-dose AZT treatment arms in an AIDS randomized trial. *Journal of the American Statistical Association* 89:737–749.
42. Smith, P. J., Thompson, T. J., Jereb, J. A. (1997). A model for interval-censored tuberculosis outbreak data. *Statistics in Medicine* 16:485–496.
43. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Medical Decision Making* 13:322-338, 1993.
44. Sun, J. (1997). Regression analysis of interval-censored failure time data. *Statistics in Medicine* 16:497–504.
45. *Survival Analysis: State of the Art*. Boston: Kluwer Academic Publishers, pp. 99–120.
46. Thall, P. F., Cheng, S. C. (1999). Treatment comparisons based on two-dimensional safety and efficacy alternatives in oncology. *Biometrics* 55:746–753.
47. Touloumi, G., Pocock, S. J., Babiker, A. G., Darbyshire, J. H. (1999). Estimation and comparison of rates of change in longitudinal studies with informative drop-outs. *Statistics in Medicine* 18:1215–1233.
48. Touloumi, G., Pocock, S. J., Babiker, A. G., Darbyshire, J. H. (2002). Impact of missing data due to selective drop-outs in cohort studies and clinical trials. *Epidemiology* 13:347–355.
49. Tubert-Bitter, P., Bloch, D. A., Raynauld, J. P. (1995). Comparing the bivariate effects of toxicity and efficacy of treatments. *Statistics in Medicine* 14:1129–1141.
50. Volberding, P. A., Lagakos, S. W., Koch, M. A., Pettinelli, C. B., et al. (1990).
51. Walker, A. S., Babiker, A. G., Darbyshire, J. H. (2000). Analysis of multivariate failure-time data from HIV clinical trials. *Controlled Clinical Trials* 21:75–93.
52. Walker, A. S. (1999). The analysis of multivariate failure time data with application to multiple endpoints in trials in HIV infection. Unpublished PhD thesis, University College, London.

53. Wang, S. T., Klein, J. P., Moeschberger, M. L. (1995). Semi-parametric estimation of covariate effects using the positive stable frailty model. *Applied Stochastic Models and Data Analysis* 11:121–133.
54. Wei, L. J., Glidden, D. V. (1997). An overview of statistical methods for multiple failure time data in clinical trials. *Statistics in Medicine* 16:833–839.
55. Wei, L. J., Lin, D. Y., Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modelling marginal distributions. *Journal of the American Statistical Association* 84:1065–1073.
56. Weinstein MC, Fineberg HV. *Clinical Decision Analysis*. Philadelphia: Saunders, 1980.
57. White, I. R., Babiker, A. G., Walker, A. S., Darbyshire, J. H. (1999). Randomization-based methods for correcting for treatment changes: examples from the Concorde trial. *Statistics in Medicine* 18:2617–2634.
58. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. М.: Статистика, 1974.
59. Анализ нечисловой информации в социологических исследованиях. М.: Наука, 1985.
60. Бала Ю.М., Фуки В.Б., Рог А.И., Савченко Т.Л., Савченко А.В. О возможности автоматизации процесса дифференциальной диагностики атеросклеротического кардиосклероза и ревматических пороков сердца, осложненные мерцательной аритмией // *Кардиология*, 1977. Т. 17. №7.
61. Большев Л.Л.; Смирнов Н.В. *Таблицы математической статистики*. М.: Наука, 1983.
62. Воронин Ю.А. Теория классифицирования и ее приложения. Новосибирск: Наука, 1985.
63. Гельфанд И.М., Алексеевская М.А., Губерман Ш.А., Сыркин А.Л., Головня Л.Д., Извекова М.А. Прогнозирование исхода инфаркта миокарда с помощью программы «Кора-3» // *Кардиология*. 1977. Т.17. № 6, 7.
64. Горелик А.Л., Скрипкин. В.А. Методы распознавания. Учебное пособие для вузов. М.: Высшая школа, 1984.
65. Кендалл М.Дж., Стьюарт А. Многомерный статистический анализ и временные ряды. М.: Наука, 1976.
66. Кулинская Е.В. Об эмпирических индексах качества классификации и их реализации в пакете программ CLAMS для IBM PC // Всесоюз. симпозиум с международным участием «Теория и практика классификации и систематики в народном хозяйстве». Тез. докл. М.: ВНИИТИ, 1980.
67. Кулинская Е.В, Сатаров Г.А. Проверка гипотез о качестве классификации в пакете программ CLAMS для IBM PC // Всесоюз. симпозиум с международным участием «Теория и практика классификации и систематики в народном хозяйстве». Тез. докл. М: ВНИИТИ, 1990.
68. Куперштох В.Л., Миркин Б.Г., Трофимов В.А. Сумма внутренних связей как показатель качества классификации // *А и Т*. 1976. № 3.

69. Любищев А.А. Проблемы формы, систематики и эволюции организмов. М.: Наука, 1982.
70. Маамяги А.В. Некоторые задачи статистического анализа классификаций. Таллинн: Изд-во АН ЭССР, 1982.
71. Орлов А.И. Общий взгляд на статистику объектов нечисловой природы // [2].
72. Орлов А.И. Статистика объектов нечисловой природы // *Статистика. Вероятность. Экономика*. М.: Наука, 1985.
73. Орлов А.И. Статистика объектов нечисловой природы // *Тр. 1 Всемирного Конгресса Общества им. Бернулли «Математическая статистика, теория вероятностей, комбинаторика и ее применения»*. Вып. 1. М.: МИАН СССР, Советский Комитет Общества им. Бернулли, 1988.
74. Орлов А.И. Махаланобиса расстояние // *Математическая энциклопедия*. Т.8. М.: Советская Энциклопедия, 1982.
75. Орлов А.И., Рыданова Г.В. О некоторых результатах статистики объектов нечисловой природы // *Материалы I Всесоюз. школы-семинара «Программно-алгоритмическое обеспечение анализа данных в медико-биологических исследованиях (8-в июня 1985 г., Пущино)»*. Пущино: НИВЦ АН СССР, 1986.
76. Орлов А.И. Математические методы классификации, статистика объектов нечисловой природы и медико-биологические исследования // *Доклады МОИП 1984 г. Общая биология. Цитогенетический и математический подход к изучению биосистем*. М.: Наука, 1986.
77. Орлов А.И. Некоторые неклассические постановки в регрессионном анализе и теории классификации // *Программно-алгоритмическое обеспечение анализа данных в медико-биологических исследованиях*. М.: Наука, 1987.
78. Орлов А.И. Некоторые вероятностные вопросы теории классификации // *Прикладная статистика*. М.: Наука, 1983.
79. Орлов А.И. О сравнении алгоритмов классификации по результатам обработки реальных данных // *Общая биология. Новые данные исследований структуры и функций биологических систем. Доклады МОИП, 1985*. М.: Наука, 1987.
80. Орлов А.И. Оценка размерности модели в регрессии // *Алгоритмическое и программное обеспечение прикладного статистического анализа*. М.: Наука, 1980.
81. Орлов А.И. Асимптотика некоторых оценок размерности модели в регрессии // *Прикладная статистика*. М.: Наука, 1983.
82. Орлов А.И. Некоторые вероятностные вопросы кластер-анализа // *Общая биология. Новые данные исследований структуры и функций биологических систем. Доклады МОИП, 1985*. М.: Наука, 1987.
83. Орлов А.И. Устойчивость в социально-экономических моделях. М.: Наука, 1979.

84. Орлов А.И., Гусейнов Г.А. Математические Методы в изучении способных к математике школьников //Исследования по вероятностно-статистическому моделированию реальных систем. М.: ЦЭМИ АН СССР, 1977.
85. Орлов А.И. Математика нечеткости //Наука и жизнь. 1982. № 7.
86. Орлов А.И. Парные сравнения в асимптотике Колмогорова //Экспертные оценки в задачах управления. М.: ИПУ, 1982.
87. Орлов А.И. Классификация объектов нечисловой природы на основе непараметрических оценок плотности //Проблемы компьютерного анализа данных и моделирования. Минск: Белорусск. Гос. ун-т, 1991.
88. Плоткин А.А. Устойчивость разбиения как критерий оптимальности Построенной классификации //Статистические методы анализа экспертных оценок. М.: Наука, 1977.
89. Райская Н.Н., Гостилин Н.Н., Френкель А.А. Об одном способе проверки обоснованности разбиения в кластерном анализе //Всесоюз. конф. «Применение многомерной статистического анализа в экономике и оценке качеств) продукции». Тез. докл. Тарту, 1977.
90. Раушенбах Г.В. Меры близости и сходства //[2].
91. Рекомендации. Прикладная статистика. Методы обработки данных. Основные требования и характеристики. М.: ВНИИСИ, 1987.
92. Розова С.С. Классификационная проблема в современной науке. Новосибирск: Наука, 1986.
93. Фоменко А.Т. Новая эмпирико-статистическая методика обнаружения параллелизмов в датировании дубликатов //Проблемы устойчивости стохастических моделей. М.: ВНИИСИ, 1984.
94. Шорников Б.С. Классификация и диагностика в биологическом эксперименте. Проблема оценки и классификации интерьерных признаков человека. М.: Наука, 1979.
95. Шрейдер Ю.А., Шаров А.А. Системы и модели. М.: Радио и связь, 1982.
96. Шурыгин А.М. Статистический кластер-критерий //Алгоритмическое и программное обеспечение прикладного статистического анализа. М.: Наука, 1980.