

decât valoarea tabelară, deci diferența este ne semnificativă, ceea ce testează aceeași lipsă de semnificație între bărbații și femeile fumătoare sau bărbații și femeile nefumătoare.

**Exemplul 6.** Încercând să vedem dacă există diferențe semnificative între rezultatele tratamentului cu streptomycină și penicilină la două loturi de copii sub un an suferinzi de bronhopneumonie acută, am urmărit un număr de 20 copii pentru fiecare lot. Rezultatele obținute au fost introduse în tabelul 15.

Tabelul 15

Repartizarea bolnavilor cu bronhopneumonie în dependență de metoda tratamentului și rezultatul lui

Bolnavi cu bronhopneumonie	Vindecați	Complicații	Total
Tratament cu penicilină	a=14	b=6	a+b=n <sub>1</sub> =20
Tratament cu 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

$$\chi_c^2 = \frac{(14 \times 4 - 6 \times 16)^2 \times 39}{20 \times 20 \times 30 \times 10} = 0,52, \chi_t^2 = 3,8 \text{ pentru } \alpha = 5\%$$

Rezultă că valoarea  $\chi_c^2 < \chi_t^2$ , ceea ce înseamnă că rezultatele dintre cele două tratamente (penicilină și streptomycină) folosite în tratamentul bronhopneumoniei nu diferă semnificativ la cele două loturi de copii bolnavi.

### 1.8.7. STANDARDIZAREA PRIN METODA DIRECTĂ

Interpretarea comparativă a indicilor statistici nu este întotdeauna posibilă, chiar dacă dispunem de informații suplimentare despre caracterul fenomenului studiat. Valoarea autentică a fenomenului trebuie căutată în esența calitativă a lui.

Pentru a face o comparație a indicilor letalității în două spitale și a trage o concluzie privind cauzele divergenței acestor indici se cere mai întâi să analizăm omogenitatea nosologică a grupelor de bolnavi din aceste spitale.

Indiscutabil, indicele letalității va fi mai mare în spitalul, în care au fost internați bolnavii mai gravi. Prezența diverselor categorii de bolnavi după starea sănătății, vârstă, sex și alte caractere, face imposibilă concluzia corectă privind cauzele divergenței indicilor letalității în aceste spitale.

În multe investigații medico-sociale, precum și clinice este prea mică posibilitatea de a opera cu grupe omogene producătoare de fenomene, care necesită comparație. Mai frecvent aceasta se referă la indicii morbidității, natalității, mortalității generale pe orașe, județe și alte localități, care sunt diferite după componența de vârstă, sex etc.

**Exemplu:** Este necesar de a compara indicii mortalității generale dintre două localități „A” și „B”. Calculul simplu arată că mortalitatea în localitatea „A” este de 10%, iar în B – de 8,5%. Dar pentru a trage concluzii privind factorii, care au provocat diferitele niveluri ale mortalității se cere o analiză a totalităților, din care s-au obținut acești indici. Un factor influent asupra mortalității este componența de vârstă a totalităților investigate. Cu cât sunt mai multe persoane vârstnice, cu atât

mai mare poate fi nivelul mortalității generale. În cazul nostru, în localitatea „A” bătrânii constituie 30% din toată populația, pe când în localitatea „B” numai 18%. Prin urmare, totalitățile comparate fiind diferite prin componența de vârstă, produc diferite niveluri de mortalitate (10,0 și 8,5%), de aceea ele nu pot fi comparate așa cum sunt.

Pentru compararea indicilor intensivi generali, produși de totalități neomogene, vom utiliza o metodă specială – așa-numita **metodă de standardizare**.

**Standardizarea** este o metodă convențională, care permite calcularea indicilor standardizați ce înlocuiesc indicii generali intensivi sau medii când din cauza neomogenității totalităților statistice este imposibilă compararea lor.

Indicii standardizați sunt convenționali, deoarece, înlăturând acțiunea factorilor asupra indicilor reali, ei arată ce valoare ar fi avut aceștia dacă nu ar fi fost influențați de unul sau mai mulți factori. Așadar, indicii standardizați pot fi folosiți numai cu scop de comparare.

Se cunosc mai multe metode de standardizare. Cea mai frecvent utilizată în statistica medicală este metoda directă. Ea constă din 5 etape.

**Etapa I** – calculul indicilor intensivi (medii) generali (pentru fiecare grupă – pe sexe, vârstă etc.) pentru totalitățile care se compară.

**Etapa a II-a** – alegerea și calculul standardului.

**Etapa a III-a** – calculul „valorilor așteptate” pentru fiecare grupă de standard.

**Etapa a IV-a** – determinarea indicilor standardizați.

**Etapa a V-a** – compararea totalităților după indicii generali intensivi (sau medii) și standardizați. Concluzii.

Pentru o mai bună intuitivitate calculele etapelor se efectuează consecutiv, sub formă de tabel (*tabelul 16*).

*Tabelul 16*

Morbiditatea hepatică în secțiile A și B ale uzinei chimice (anul 2005)									
Sexul	Secția A		Secția B		Etapa I		Etapa II Standard (rubrica 2 + rubrica 4)	Etapa III	
	Nr.de muncitori	dintre ei bolnavi	Nr. de muncitori	dintre ei bolnavi	% bolnavilor de hepatită în secții	Nr. de bolnavi așteptați în grupa standard		A	B
	A	B							
B	50	1	170	4	2,0	2,3	220	4,4	5,06
F	200	10	30	3	5,0	10,0	230	11,5	23,0
Total	250	11	200	7	4,4	3,3	450	15,9	28,06
					<b>Etapa IV</b>		100	3,5%	6,2%

**Etapa I – Calculul indicilor intensivi.** Determinăm frecvența îmbolnăvirilor cu hepatită pe ambele sexe, la muncitorii secțiilor A și B ale uzinei chimice. Acești indici sunt rapoarte procentuale, calculați în felul următor: numărul bolnavilor raportat la numărul de muncitori înmulțit cu 100.

Astfel, în secția A bărbați bolnavi de hepatită sunt:  $(1 : 50) \times 100 = 2\%$ , iar femeii bolnave de hepatită:  $(10 : 200) \times 100 = 5,0\%$ . Pe ambele sexe în secția A vor fi:  $(11 : 250) \times 100 = 4,4\%$ . Astfel de calcule se efectuează respectiv și pentru secția B.

După aceasta analizăm indicii intensivi din ambele secții. Din tabelul de mai sus (etapa I) observăm că morbiditatea atât la bărbații cât și la femeile din secția A este mai mică decât la bărbații și femeile din secția B. Urmărind acest fenomen în cadrul colectivelor întregi de muncitori pe secții, observăm un paradox : frecvența îmbolnăvirilor muncitorilor în secția A este mai înaltă – 4,4% vizavi de 3,3% în secția B. Aceasta înseamnă, că asupra indicilor intensivi generali a influențat diferența structurii pe sexe a muncitorilor din ambele secții. În secția A lucrează mai multe femei (acestea s-au îmbolnăvit mai frecvent de hepatită decât bărbații), în timp ce în secția B majoritatea o constituie bărbații, la care această maladie se atestă mai rar. Așadar, în cazul dat, diferența indicilor intensivi este cauzată de neomogenitatea de sex a muncitorilor în ambele secții.

Pentru eliminarea acestui factor influent asupra nivelului morbidității hepatice în grupele neomogene de muncitori este necesar de a efectua standardizarea.

**Etapa a II-a – Alegerea și calcularea standardului.** Drept standard este considerată acea totalitate statistică, în care se vor reflecta toate particularitățile și caracterele grupelor comparate. În exemplul nostru, standardul este numărul total al muncitorilor din ambele secții, inclusiv pe sexe (rubrica 2 + rubrica 4). Astfel, standardul pentru grupul de bărbați în ambele secții va fi 220 (50 + 170), iar pentru femei 230 (200 + 30). În total standardul muncitorilor pe ambele secții va fi 450 (220 + 230). Cunoscând indicii intensivi (etapa I) și standardul (etapa a II-a) trecem la etapa a III-a.

**Etapa a III-a – Determinarea numărului așteptat de bolnavi în fiecare secție reieșind din grupul standard.** Numărul de bolnavi așteptați îl vom afla dând răspuns la întrebarea: câți bolnavi de hepatită așteptăm în grupul standard de 220 de muncitori, dacă frecvența îmbolnăvirilor va fi ca la bărbații din secția A – 2%? Facem proporția și efectuăm calculul: dacă la 100 vom avea 2 bolnavi, atunci la 220 vor fi „X”, de unde „X” va fi egal cu 4,4 bolnavi.

Așadar, numărul de bolnavi hepatici așteptați printre cei 220 de bărbați ai grupei standard va fi 4,4.

Respectând același principiu, efectuăm calcule analogice pentru ambele secții în toate trei grupuri de standard, luând în calcul frecvența îmbolnăvirilor pe sexe în ambele secții. Rezultatele calculelor sunt reprezentate în tabelul de mai sus.

După calcularea indicilor în grupele standard de femei și bărbați pentru ambele secții trecem la următoarea etapă.

**Etapa a IV-a – Determinarea indicilor standardizați.** Suma de bolnavi așteptați (bărbați + femei) pe secții o raportăm la numărul total de muncitori din ambele secții. Exprimând aceste valori în % obținem indicii standardizați. Deci, în secția A numărul total de bolnavi „așteptați” va fi:  $4,4 + 11,5 = 15,9$ . Indicele standardizat în această secție va constitui:  $(15,9 : 450) \times 100 = 3,5\%$ . În secția B respectiv:  $5,06 + 23,0 = 28,06$ , iar indicele standardizat va fi egal cu:  $(28,06 : 450) \times 100 = 6,2\%$ . În baza rezultatelor etapelor I și IV vom efectua următoarea etapă.

Tabelul 17

Compararea nivelurilor morbidității hepatice la muncitorii secțiilor A și B			
Indicii	Secția A	Secția B	Rezultatele comparării
Intensivi	4,4	3,3	A > B
Standardizați	3,5	6,2	A < B

**Etapa a V-a. Compararea nivelurilor morbidității hepatice în secțiile A și B** după indicii generali de frecvență și cei standardizați. Pentru o analiză mai elocventă construim un tabel analitic (*tabelul 17*), în care includem indicii intensivi generali și standardizați pe ambele secții. În baza acestor indici vom formula concluzia respectivă.

**Concluzie:** Comparând indicii standardizați în secțiile A și B, concluzionăm: dacă am fi avut totalități omogene (pe sexe) de muncitori în ambele secții, morbiditatea hepatică în secția A ar fi fost comparativ mai joasă decât în secția B ( $A < B$  sau  $3,5 < 6,2$ ).

Comparând, însă, indicii generali intensivi constatăm un fapt diametral opus ( $A > B$ ). Aceasta se explică prin influența raportului de sexe asupra indicilor intensivi în ambele secții. În legătură cu faptul că femeile sunt mai receptive la hepatită decât bărbații, iar în secția A s-au dovedit a fi mai multe femei (200) decât bărbați (50), am și constatat o morbiditate hepatică mai mare decât în secția B. În secția B predomină bărbații (170), care fiind mai rezistenți la hepatită, au cauzat un nivel mai jos de morbiditate hepatică.

Metoda de standardizare poate fi utilizată în cazul diferenței considerabile a nivelurilor indicilor de grup (diferite niveluri de morbiditate pe sexe, letalitate în spitale, diferite durate de tratament etc.).

Dacă indicii pe anumite grupe în totalități neomogene sunt aproximativ aceiași, standardizarea nu are sens. De asemenea, standardizarea nu se recomandă și în cazurile când indicii comparați pe anumite grupe sunt diferiți, iar totalitățile omogene.

Când operăm cu indici standardizați trebuie să reținem, că ei nu reprezintă nivelul real al fenomenului studiat, ci sunt doar convenționali și pot fi utilizați numai cu scop de comparare, eliminând influența diverșilor factori asupra indicatorilor statistici și că ei arată care ar fi fost nivelul fenomenului studiat în totalități cu structură omogenă.

## 1.9. ANALIZA CORELAȚIEI DINTRE FENOMENE ȘI SEMNE

În medicină și în 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, al 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 cauză și efect între fenomene. Asemenea corelații sunt studiate în cadrul științelor exacte, unde având de-a face cu fenomene simple, legătura de cauză și 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 existentă între fenomene. În cazul corelației statistice fiecărei valori numerice a variabilei X îi corespunde nu una ci mai multe valori ale variabilei Y, adică o totalitate statistică a acestei valori, care se grupează în jurul mediei Y.

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

În cazul corelației lineare dintre schimbările uniforme ale valorilor medii ale unei variabile au loc schimbări egale ale altei variabile. Pentru corelația nelineară schimbărilor uniforme ale unei valori le corespund valori medii ale altei variabile, care poartă caracter de creștere ori de micșorare. Aprecierea legăturilor de corelație lineară se realizează cu ajutorul coeficientului de corelație  $r_{xy}$ , iar a celor nelinare – 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. De 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. De 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. De 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 temporare și într-un singur sens. Așa, de 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.

### 1.9.1. 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.

Exemple:

Tabelul 18

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

Nr.	Timpul x	Nr. 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	

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

Exemple:

Tabelul 19

**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;
- $\sum$  – 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.

### 1.9.2. 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 ±1 denotă o corelație foarte puternică între fenomene;
- valoarea coeficientului de corelație cuprinsă între ±0,99 și ±0,70 denotă o corelație puternică;
- valoarea coeficientului de corelație cuprinsă între ±0,69 și ±0,30 denotă o corelație medie între fenomene;
- valoarea coeficientului de corelație cuprinsă între ±0,0 și ±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 dintre 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 frecvenței eșantioanelor să se facă la întâmplare.

### 1.9.3. 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 fenomenele obținute pe eșantioane, folosind eroarea coeficientului de corelație, care se notează cu  $m_r$ .

Formula de calcul pentru  $n < 30$ : 
$$m = \sqrt{\frac{1 - r_{xy}^2}{n - 2}}$$

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;
- $l$  – 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 dintre 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.

### 1.9.4. CORELOGRAMA

Existența sau inexistența unei corelații dintre 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 înscriem 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.



ză, 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 mărimea 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.

### 1.9.5. CORELAȚIA MULTIPLĂ

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

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

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

- $r_{xyz}$  – 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.

### 1.9.6. CORELAȚIA RANGURILOR (SPEARMAN)

În cazul în care dorim să stabilim legătura de dependență dintre fenomenele 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:

$$\rho = 1 - \frac{6 \sum d^2}{n \cdot (n^2 - 1)}, \text{ în care:}$$

- 1 – valoarea absolută a coeficientului de corelație;  
 $\rho$  – coeficientul de corelație Spearman;  
 6 – valoare constantă;  
 $\Sigma d^2$  – 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.

Exemplu:

Tabelul 20

Timpul trecut din momentul accesului de pancreatită acută (x) și numărul complicațiilor postoperatorii (y)						
Nr.	Timpul x	Nr. 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			$\rho = 1$	

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

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

Exemplu:

Tabelul 21

Legătura de corelație dintre copiii cu deficiență mentală (la 100 mii de copii)  
și invaliditatea copiilor (la 100 mii de copii)

Nr.	Țara	Deficiența mentală x	Invaliditatea 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	Kârgâzstan	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	Belarus	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						$\Sigma = -24,5$ $\Sigma = +24,5$	$\Sigma = 253,5$ $\rho = 0,55$

### 1.9.7. COEFICIENTUL DE ASOCIERE

În domeniul medicinei este necesar în multe cazuri de a stabili legătura nu numai între indicatorii 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 dintre 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:

Tabelul 22

**Vaccinarea contra unei maladii și morbiditatea populației în orașul A**

	S-au îmbolnăvit	Sănătoși	Total
Vaccinați	40 a	5 900 b	5 940 a + b
Nevaccinați	250 c	9 620 d	9 870 c + d
Total	290 a + c	15520 b + d	15 810 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 o legătură inversă, astfel, cu cât mai multe persoane sunt vaccinate, cu atât mai puține cazuri de îmbolnăvire se înregistrează.

### 1.10. REGRESIA

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

Coeficientul 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  $Rg_{xy} = r_{xy} \frac{\delta_x}{\delta_y}$  în care:

$Rg_{yx}$  – 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;

$Rg_{xy}$  – 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:

Tabelul 23

Înălțimea $x$	Frecvențe $f_x$	Greutatea $y$	Frecvențe $f_y$
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 ( $y$ ) au 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:

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

**Concluzie:** Creșterea înălțimii medii cu 1 cm la copiii studiați va avea drept urmare creșterea greutății 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:

$$y = \bar{Y} + Rg_{yx}(x - \bar{X}), \text{ în care:}$$

$y$  – 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 135 cm.

Înlocuind datele în formulă, obținem:

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

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, greutății 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 -  $\delta_{Rg}$ , care este calculată cu ajutorul formulei următoare:

$$\delta_{Rg} = \sqrt{\delta_y \cdot 1 - r_{xy}^2}$$

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:  $\delta_{Rg} = \sqrt{\delta_y \cdot 1 - r_{xy}^2} = 1,97 \cdot 1 - 0,82^2 = 1,97 \cdot 0,33 = \pm 0,65 \text{ kg/cm}$

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

### 1.11. ALTE METODE STATISTICO-MATEMATICE DE ANALIZĂ A INFORMAȚIEI

În afară de metodele matematice și statistice de prelucrare a informației descrise mai sus, există și alte metode, mai complicate, a căror utilizare permite analiza și sinteza mai aprofundată a fenomenului cercetat și a factorilor determinați. În cele ce urmează le vom caracteriza succint. Cei interesați să le studieze aprofundat vor consulta surse științifice suplimentare.

**1. Analiza discriminantă.** Esența metodei constă în acordarea de ponderi, ce derivă din nivelurile fenomenului studiat sau „note” – diferiților factori de risc. Notele se acordă pe baza datelor generale pentru întreaga totalitate sau pe baza riscului relativ al eșantionului cercetat pentru fiecare factor de risc în parte.

Reiese că factorii cu o pondere (notă) mai mare dau și un indice majorat al fenomenului. În metoda dată de analiză se evidențiază anumite trepte, unde factorii de risc sunt grupați după ponderile lor (mici, medii, mari etc.). Această analiză discriminantă numită „pas cu pas” poate fi efectuată descendent (prin eliminare) după note sau ascendent (aditiv). Aici se utilizează formula de funcție lineară:

$$D = b_1x_1 + b_2x_2 + \dots + b_kx_k, \text{ unde:}$$

D – funcția lineară discriminantă;

x – semnul fenomenului studiat (mediu, rang, greutate, vârstă);

b – coeficientul de ponderare.

**2. Analiza prin componente principale.** Este o metodă specială de corelație dintre evenimentul produs și multitudinea de factori calitativi și cantitativi.

Prin combinarea liniară a semnelor, ultimele sunt concentrate în mai puține variabile și astfel cele cu o pondere mai mare în influențarea evenimentului devin componente principale. Aceste componente principale corespund axelor principale după care se desfășoară punctele variabilelor, în cazul nostru ale factorilor de risc, eliminându-se astfel factorii cu o mai mică influență asupra fenomenului.

**3. Analiza factorială.** Este, în principiu, similară analizei prin componente principale, exprimând variabilele inițiale sub forma unei combinații lineare de factori comuni și specifici.

Analiza factorilor poate fi efectuată prin două metode tehnico-tematice:

- metoda factorială a lui Spearman, cu un singur factor comun;
- metoda de analiză multifactorială, care include un număr mai redus de factori comuni.

Analiza factorială, pornind de la corelație și regresie, caută să sintetizeze acțiunile sinergice sau antagoniste ale mai multor factori de risc și astfel evidențiază pe cei cu pondere mai însemnată în determinarea fenomenului.

Aici, însă cercetătorul trebuie să fie prudent în analiza rezultatelor căpătate, care nu întotdeauna pot fi considerate definitive. Deci, pe parcurs se pot găsi și alte soluții, prin rotarea spațială a vectorilor reprezentativi ai factorilor, care să determine logic legăturile de dependență multifactorială.

**4. Corelația multidimensională și regresia liniară multiplă.** A fost deja descris că corelația și regresia se utilizează mai ales în cazul dependenței a două variabile cantitative, precis măsurabile, dar există și posibilitatea evidențierii dependenței și în cazurile variabilelor cu caracter calitativ.

În cazul corelației multidimensionale proiectarea tridimensională permite calculul probabilist al fiecărui cuplu de variabile de două dimensiuni, calculându-se astfel coeficienții de corelație totală. Această corelație totală poate fi aplicată în special în cazul repartițiilor normale de frecvență la un număr mare de dimensiuni, similar ca în corelațiile tridimensionale, existând o mare gamă de corelații selective ale variabilelor.

Pentru calculul regresiei multiple, cea mai simplă este metoda de programare liniară, ce se transpune ușor în programul pentru calculator. Metoda aceasta permite să se introducă variabilele rând pe rând în regresia multiplă, până când orice majorare nu mai modifică esențial rezultatele.

**5. Analiza secvențială.** Se poate aplica în orice studiu, inclusiv epidemiologic și se bazează pe regula de oprire a investigației în funcție de rezultatele obținute până la data respectivă. Adică se utilizează mai des în domeniul observațiilor clinice și în experimente.

Această metodă se bazează pe faptul că eroarea medie scade pe măsură ce este mărit numărul de cazuri și rezultă din formula:

$$m_m = \frac{\sigma}{\sqrt{n}}$$

**6. Analiza dispersională.** În selecțiile aleatorii, când se formează grupe de eșantion, analiza dispersională ne demonstrează dacă aceste grupe fac parte sau nu din aceeași populație. Cu alte cuvinte, când avem o totalitate selectivă, formată din mai multe eșantioane mici, prin intermediul acestei metode de analiză, apreciem dacă acest eșantion este omogen.

Se analizează, deci, variația întregii populații cu ajutorul a două mărimi: una este variația în interiorul familiilor și a doua – variația de la familie la familie. În fond, este o analiză a deviațiilor.

De exemplu, formula lui Fischer de analiză dispersională este următoarea:

$$F = \frac{S_1^2}{S_2^2}$$

Când acest raport va fi supraunitar, deci coeficientul „ $F$ ” va fi mai mare sau mai mic ca unu, se respinge ipoteza nulă ( $H_0$ ) că cele două eșantioane comparate ar fi din aceeași populație și se acceptă ipoteza alternativă ( $H_1$ ) că ele sunt semnificativ diferențiate.

Variația acestor eșantioane depinde de metodologie și de unii factori întâmplători. Dispersia generală, la rândul său, depinde de suma dispersiilor factoriale „ $S_x$ ” (a metodologiei) și a dispersiei reziduale „ $S_z$ ” (a factorilor întâmplători).

Raportul lor (coeficientul Fisher) se compară cu rezultatele standard calculate special și dacă ele sunt egale sau primul e mai mare decât cel tabelar, atunci influența factorului cercetat se consideră acceptată. Astfel, noi putem analiza și determina prioritatea influenței factorilor metodologici sau întâmplători, sau a diferitelor građații și combinații ale lor referitor la fenomenul studiat.

În practică se folosesc mai multe metode de analiză dispersională în caz de:

- a) complex dispersional al unui factor;
- b) complex a doi factori interdependenți;
- c) complex multifactorial.

Ultima metodă permite de a analiza mai aprofundat influența celor mai diverși factori și condiții asupra fenomenului cercetat, de aceea ea este utilizată mai frecvent.

Aplicarea corectă a metodelor statistice și a tendinței generale a schimbării fenomenelor în dinamică permite extrapolarea acestora pe o anumită perioadă (mai scurtă sau mai îndelungată) în funcție de modificările factorilor determinanți și favorizanți cu acțiune pozitivă și negativă. Aceasta prezintă interes în planificarea serviciilor sanitare, în conformitate cu modele previzionale, care pot fi grupate în: inductive; deductive; normative; mixte.

**Metodele inductive** fac apel la opinia unor persoane experimentate în specialitatea problemei în care se efectuează investigația, sau se află părerea altor personalități din diferite domenii de preocupare, comparând-o cu cea a specialiștilor. Pentru aceasta se pregătesc, de obicei, anchete speciale pentru investigarea opiniei experților (specialiștilor), personalului operativ, populației. Aceste anchetări pot fi repetate în dinamică la același grup de persoane (metoda Delfi) sau folosite pe o cale multidisciplinară, având, însă, un scop comun, bine determinat și fiind efectuate de una și aceeași instituție.

**Metodele deductive** prevăd utilizarea metodelor matematice, teoretice, pe baza studiului curent, în dinamică, al fenomenului și extrapolării lui pe o perioadă bine determinată (extrapolarea ritmului mediu anual de creștere sau scădere, prin regresia liniară cu ajutorul parabolei, hiperbolei, prin metode exponențiale și logistice etc.).

**Metodele normative** se bazează pe metodele obișnuite de analiză a fenomenului studiat, a metodei inductive și celei deductive, fapt ce permite de a elabora normativele necesare de activitate a serviciilor pentru sănătate, a cadrelor etc.

Aceste metode previzionale în combinație cu cele obișnuite permit analiza aprofundată a fenomenului cercetat și formularea concluziilor și propunerilor practice respective adecvat situației sociale, economice, spațiale și temporare în totalitatea integrală pentru prezent și viitor.

## 1.12. PREZENTAREA MATERIALULUI STATISTIC

După ce materialul obținut din cercetare a fost prelucrat, se trece la prezentarea lui în tabele statistice și în formă grafică.

### 1.12.1. PREZENTAREA PRIN TABELE

Tabelele statistice constituie cea mai exactă formă de prezentare a informațiilor obținute prin observație sau experiment, pentru un bun cunoscător al statisticii nefiind necesară prezentarea descriptivă a cercetării.

*Tabelele statistice* sunt constituite din machete formate din coloane verticale și rânduri orizontale, din întretăierea cărora rezultă o serie de rubrici, fiecare corespunzând unei valori sau interval de valori al variantelor respective, exprimând o anumită valoare medie sau relativă.

Tabelarea constituie o metodă comodă, sintetică și sistematică:

- *comodă*, pentru că rezultatele cercetării sunt redade sub forma unor tabele cifrice, permițând observarea cu ușurință a aspectelor principale ale problemei cercetate;
- *sintetică*, pentru că datele cifrice redau aspectele esențiale ale fenomenului studiat;
- *sistemică*, pentru că între diferitele date cantitative sau calitative prezentate în tabel, există o înlănțuire logică, ușurând înțelegerea lor.

Cerințele de întocmire a unui tabel statistic sunt:

1. Fiecare tabel trebuie să aibă denumire, în care se include timpul.
2. Datele în tabel se aranjează corespunzător conținutului acestora.
3. Fiecare tabel trebuie să aibă totalizări pe orizontală și pe verticală.
4. În tabel se prezintă indicatori relativi sau medii aritmetici, dar nu date absolute.

**Fiecare tabel trebuie să aibă denumire**, care printr-o frază clară să redea conținutul tabelului, locul și perioada de timp la care se referă 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 poate răspunde printr-un asterix sub macheta tabelului (exemplu, *tabelul 24*).

*Tabelul 24*

**Mortalitatea generală a populației  
Republicii Moldova pentru anii 1996 – 2000 la 100 mii de locuitori**

Nr.	Genul	Anii analizați					Media pe 5 ani
		1996	1997	1998	1999	2000	
1.	Ambele sexe	1 149,6	1 175,7	1 093,1	1 133,0	1 132,8	1 136,8
2.	Feminin	1 082,0	1 079,8	1 018,6	1 050,6	1 057,6	1 057,7
3.	Masculin	1 237,9	1 239,3	1 174,1	1 224,5	1 214,8	1 218,1

**2. Datele în tabel se aranjează corespunzător conținutului acestora.** În rândurile orizontale, în stânga machetei sunt cuprinse subgrupurile din colectivita-



tea cercetată (populația repartizată după vârstă, sex, nou-născuți pe județe, bolnavi după diagnostic etc.), constituind subiectul tabelului, și care corespund obiectului de studiu. În coloanele machetei, în dreptul fiecărui rând al subiectului sunt prezentate valorile care corespund caracteristicilor de grupare, deci însușirile exprimate în valorile variantelor, constituind predicatul tabelului (valorile ce corespund înălțimii, greutateii nou-născuților, mortalității infantile, morbidității etc.).

**3. Fiecare tabel trebuie să aibă totalizări pe orizontală și pe verticală.** Totalul în tabelele statistice poate fi înscris în primul rând și în prima coloană sau în ultimul rând și în ultima coloană, neexistând standarde unificate privind acest criteriu. 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 când nu cunoaștem datele necesare pentru a fi înscrise într-o căsuță a tabelului, deși ele există, acest lucru se exprimă prin câteva puncte succesive (*tabelul 24*).

**4. În tabel sunt prezentați indicatorii relativi sau mediile aritmetice, dar nu datele absolute.** Deoarece compararea datelor absolute nu este reprezentativă din punct de vedere statistic, nu putem constitui tabelele, care în special ne permit compararea rapidă din astfel de date. Valorile medii și indicatorii relativi permit efectuarea comparațiilor veridice, deci anume pe acestea le includem în tabele. Excepții servesc cazurile când valoarea fenomenului constituie numere mici (de exemplu 2 cazuri de letalitate), care sunt percepute mai ușor decât valoarea relativă a acestor cazuri unice (0,0002%). Uneori în tabel pot fi incluse cifrele absolute, dar numai dacă paralel se arată și valoarea relativă a acestei cifre.

**Tipuri 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, care au 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 clasificarea dichotomică;
- tabele pentru distribuția de frecvențe;
- tabele de corelație.

**Tabelele pentru clasificarea dichotomică** le utilizăm în situația în care cercetăm fenomenele care se caracterizează prin însușiri sau caracteristici diametral opuse, excluzându-se unul pe altul (*tabelul 25*).

*Tabelul 25*

**Distribuția elevilor din colectivitatea X  
în anul Y pe sexe, locul de trai și prezența sau absența cariei dentare**

Total	Sex		Domiciliul		Caria dentară	
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

„A dichotomiza”, în limba greacă, înseamnă a divide, a împărți în două. De 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 mediu se poate împărți în populație rurală și urbană; după reacția la tuberculină (IDR) poate fi pozitivă sau negativă etc.

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.

**Tabelele pentru distribuția de frecvențe** le utilizăm atunci când dorim să prezentăm 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 (*tabelul 26*) 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 aceleași 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 în perechi.

*Tabelul 26*

Distribuția bolnavilor cu HTA în funcție de grupe de vârstă și valori ale TA				
Grupe de vârstă	Valorile TA			
	160-169,9	170-179,9	180-189,9	200-209,9
40-44	5			
45-47		10		
50-54			15	
55-59				20

**Exemplu:** 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 din care una din variabile (factorială) este prezentată de grupa de vârstă, iar cealaltă (determinată sau rezultantă) este prezentată de grupele de valori ale TA.

Se observă o legătură între creșterea valorilor TA și vârstă, legătură care va fi măsurată printr-un coeficient de corelație studiat la capitolul respectiv.

### 1.12.2. REPREZENTAREA GRAFICĂ A DATELOR STATISTICE

Reprezentarea grafică face posibilă înțelegerea fenomenelor studiate (natalitatea, morbiditatea, mortalitatea) prin diferite semne – simboluri (linii, figuri, puncte etc.).

**Graficele sunt folosite pentru a prezenta:**

- Dinamica fenomenului.
- Compararea indicatorilor care se referă la unul și același timp, dar în obiecte diferite (N., M., M.).
- Structura fenomenului studiat.
- Dependența unor indicatori de alții.

- Gradul de răspândire a unui anumit fenomen.
- Controlul asupra îndeplinirii obiectivelor planificate.

**Rolul graficelor:**

- să orienteze utilizatorul în selectarea informațiilor importante din raportul statistic;
- să dezvolte idei ulterioare și să le explice;
- să ajute utilizatorul să privească în profunzime problema prezentată în raportul statistic;
- să încurajeze compararea și analiza informațiilor.

Graficul este de fapt o hartă care „vorbește” direct ochiului și este foarte eficientă în crearea unei imagini în mintea receptorului.

Principalele componente ale unui grafic sunt (figura 4):

1. suprafața de reprezentare;
2. suprafața graficului mărginită de axe și cadran;
3. aria graficului;
4. legenda graficului;
5. rețeaua de axe – liniile orizontale și verticale;
6. etichetele corespunzătoare axelor.

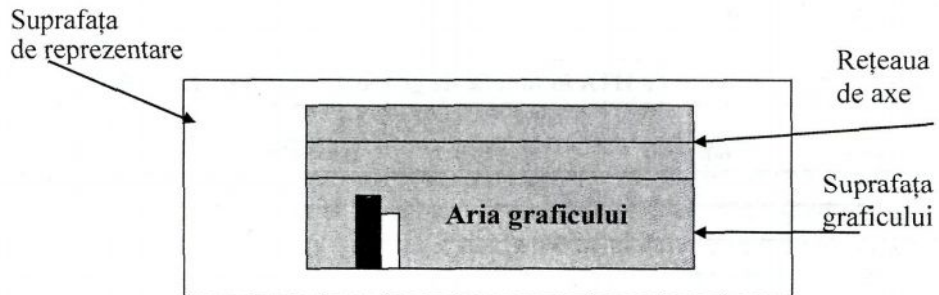


Figura 4. Componentele principalele ale unui grafic

**1. Suprafața de reprezentare** este o suprafață imaginată ce conține întregul grafic, incluzând titlul și textele explicative. În mod normal suprafața de reprezentare nu este marcată în nici un fel. Suprafața de reprezentare include:

- Titlul graficului;
- Axa verticală Y;
- Eticheta atribuită axei Y (nr. decese la 1 000 n/v);
- Valoarea maximă;
- Valoarea scalei (0, 2, 4 ...18);
- Marcajul care indică limita dintre valorile scalei;
- Valoarea minimă (0);
- Axa orizontală X;
- Suprafața graficului;
- Curbe, coloane etc.;
- Legenda;
- Sursa.

**2. Suprafața graficului** este suprafața ocupată de grafic fără titlu, valorile scalei etc. Raportul dintre X și Y este  $X:Y = 1,6 : 1$  sau  $3 : 5$ .

**3. Aria graficului** este o arie imaginară, care conține desenul propriu-zis (coloane, benzi, curbe etc.).

**4. Rețeaua de axe orizontale și verticale.** Rețeaua de axe permite citirea graficului cu mai multă ușurință. Axele orizontale se utilizează în cazul reprezentării prin bare verticale, iar axele verticale – prin bare orizontale.

**5. Textul.** Toate informațiile necesare înțelegerii graficului trebuie incluse în suprafața de reprezentare.

**6. Titlul și alte texte.** Titlul trebuie să descrie într-o formă concisă conținutul graficului:

- grupul descris (sub 5 ani);
- variabile implicate (mortalitate);
- anul / ani la care se referă datele (1993-2000);
- tipul de date (% , ‰ ...); (la 1000 n/v);

Titlul trebuie:

- să fie ușor de citit;
- să înceapă cu informația cea mai importantă.

**7. Legenda și etichetele** trebuie scrise cu caractere mai mici decât cele folosite pentru titlu.

**8. Sursa de proveniență a datelor:** axe, scale, marcaje.

Axele sunt reprezentate prin linii drepte. Pentru axa Y textul se plasează deasupra axei. Pentru axa X – sub axă. Valoarea min. a axei Y= 0, valoarea max. trebuie să fie mai mare decât cea mai mare valoare prezentată.

Pentru a facilita citirea graficului axele sunt marcate cu intervale egale, iar marcajele sunt însoțite de valorile corespunzătoare: 1,2, ...5 sau 10,20, ...50 ori zecimale: 0,1, 0,2, ...0,5.

**9. Legenda.**

Pentru graficele Bare, legenda poate fi inclusă:

- în suprafața graficului;
- sub grafic;
- lateral, în partea dreaptă a graficului.

Pentru Curbe, legenda poate fi inclusă:

- sub suprafața grafică sau în partea dreaptă (când sunt prezente 3-4 curbe);
- deasupra lor: când curbele sunt vizibil distincte și nu se intersectează;
- în afara suprafeței grafice: când curbele sunt apropiate sau se intersectează.

**10. Hașurarea și colorarea graficelor.**

- Se utilizează 4 sau 5 tipuri de hașurări.
- Accentuarea gradată a intensității hașurilor pornește de la partea superioară a graficului către partea inferioară.
- Nuanțele de culori trebuie aranjate de la cea mai deschisă către cea mai închisă.
- Culorile trebuie aplicate cu economie.
- Este recomandată folosirea unei singure culori cu diferite nuanțe ale acesteia.

**Tehnicile de construire a unui grafic:**

- Elementul de bază de la care se pornește în construirea unui grafic este sistemul de coordonate (*figura 5*).

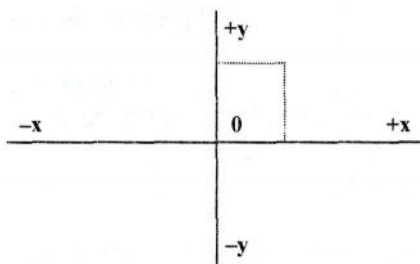


Figura 5. Sistemul cartezian

– În mod tradițional axa stângă verticală este numită ordonată, sau axa  $Y$  și axa dreapta orizontală – abscisa sau axa  $X$ .

– Acest sistem de coordonate împarte planul în patru cadrane.

**Grafice de frecvență:**

**Variabile cantitative.**

– Polare: histograma, diagrama lui Scatter.

– Diagramele.

**Histograma** este o reprezentare prin bare alăturate a distribuției frecvențelor. Scala categoriilor variabilei este reprezentată pe axa  $X$  prin perioade egale. Scala frecvențelor este reprezentată pe axa  $Y$  și se împarte în segmente.

Poligonul de frecvență.

**Graficul liniar** – seria cronologică, dinamică. Istoriograma /cronograma.

Diagramele cu bare – pentru comparații (statică).

**Variabile calitative.**

– Prin coloane, bare.

– Reprezentarea mai multor fenomene în același loc și în același timp.

– Același fenomen în mai multe locuri (urban, rural).

**Tipuri de grafice.**

1. Cu coordonate rectangulare:

– cu scara aritmetică;

– cu scara semilogaritmică.

2. Cu coordonate polare (histograma, diagrama lui Scatter).

3. Speciale.

**DIAGRAMELE**

Diagramele sunt metode de prezentare statistică a informației cu simboluri, utilizând o coordonată.

**Tipuri de diagrame.**

1. **Diagrame bazate pe lungimi:**

– cu bare (verticale, orizontale);

– pictograme.

2. Diagrame de proporții:

– cu bare;

– cu discuri.

3. Diagrame cu coordonate geografice.

4. Diagrame speciale.

**1.13. ORGANIZAREA CERCETĂRII STATISTICE**

Cunoașterea stării de sănătate a populației necesită o metodologie complexă de cercetare. Noțiunea de **cercetare înseamnă** investigația efectuată în vederea cunoașterii stării sau dinamicii fenomenelor, a distribuției și determinanților bolii și sănătății în populațiile umane.

Există 4 etape diferențiale de care trebuie să se țină cont în cadrul oricărei cercetări (figura 6).

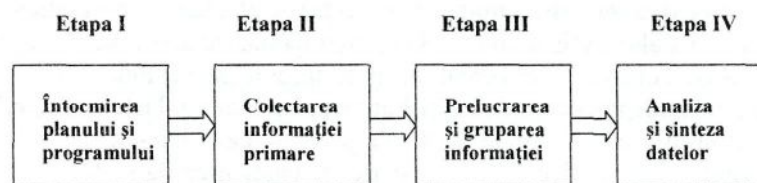


Figura 6. Etapele cercetării statistice

- I. Întocmirea planului și programului de studiu.
- II. Colectarea informației primare conform planului.
- III. Prelucrarea și gruparea informației.
- IV. Analiza și sinteza datelor obținute.

**Etapa I.** Întocmirea planului și programului de studiu. **Planul** constă în fixarea scopului și a sarcinilor de studiu. **Scopul** studiului trebuie să elucideze unele probleme de sănătate ce urmează să fie testate pe calea observațiilor sau pe cale experimentală.

**Sarcinile** (premisele, obiectivele) cercetării științifice trebuie să fixeze verificarea unor ipoteze sau a legăturilor de corelație dintre fenomene.

**Obiectul de studiu** este compus, mai frecvent, dintr-o colectivitate de bolnavi, de populație, de grupuri profesionale etc. și o colectivitate de instituții medicale. Obiectul de studiu este compus din unități de studiu cu caracteristicile sale (un pacient, un lucrător, o policlinică etc.).

**Delimitarea în volum a colectivității cercetate.** Oricare ar fi obiectul de cercetări și scopul urmărit, este necesară delimitarea ei ca mărime, volum, număr de fenomene, ființe sau lucruri studiate. Delimitarea se face prin observarea unui fenomen demografic sau de morbiditate, fie prin examinarea integrală a unei populații, fie doar parțială.

După ce am apreciat volumul cercetării, determinăm locul (delimitare în spațiu) cercetării și delimităm timpul în care va avea loc cercetarea. Din acest punct de vedere, cercetarea poate avea un caracter permanent în cazul observațiilor sau experimentărilor curente, fie un caracter periodic (ex. recensământul), la anumite intervale de timp, fie un caracter de cercetare de o singură dată.

**Programul de studiu** prevede:

- Elaborarea suportului de informație – un document-anchetă în care se fixează toate caracteristicile unității de studiu.
- Elaborarea machetelor tabelor statistice.

**Etapa II.** Colectarea informației primare conform planului. La această etapă există două cerințe: colectarea se face în bază unitară, după o instrucțiune bine determinată anticipat; trebuie să se facă cât mai puține erori de atenție.

Colectarea informației poate fi efectuată prin mai multe metode:

**1. Metoda observației nemijlocite** (directe) se utilizează de obicei în clinică, când medicul, după ce precizează anamneza 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 cer-

cetătorul statistic vrea să folosească această metodă, el trebuie să fie prezent alături de medic și să înregistreze datele de care are nevoie.

**2. Metoda extragerii informației** (indirectă) se efectuează prin culegerea datelor dintr-un formular statistic oficial. Când se efectuează acest lucru, se ține cont de programul de culegere a informației, unde trebuie să fie indicat concret care este unitatea de evidență, semnele ei atributive și cantitative. Datele privind aceste semne se extrag din documentele medicale și se notează într-un registru special sau în mai multe fișe, pe fiecare semn aparte. Această metodă se folosește des în studiile statistice și poate fi utilizată de însuși cercetătorul sau de persoane special instruite.

**3. Metoda interviului standard** 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, deoarece, în caz contrar, respondentul poate să refuze anchetarea.

*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 o să-l formuleze singur;
- semideschis, când 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.

Ancheta poate să conțină toate tipurile de întrebări sau numai 1–2; trebuie să fie nu prea voluminoasă ca să nu-l plictisească pe respondent. La anchetă, de obicei, se anexează o instrucțiune în ceea ce privește comportamentul respondentului la completarea ei.

Din punctul de vedere al scopului urmărit sunt mai multe tipuri de anchete:

*1. De tip transversal* – urmărește scopul de a determina structura, la un moment dat, a unor fenomene observate în grupul de populație studiat. Anchetarea de acest tip se repetă, de obicei, la anumite perioade (starea de sănătate, prevalența unor boli cronice etc.).

*2. De tip longitudinal* – permite de a studia în dinamică unele fenomene. De exemplu, incidența morbidității unor boli cu o mare variabilitate în timp (bolile infecțioase). Aceste anchete, la rândul lor, pot fi prospective și retrospective.

Cele prospective sunt mai avantajoase sub aspectul cunoașterii decât cele retrospective, deoarece permit stabilirea unui plan de cercetare dorit în condițiile fixate de cercetător.

Cele retrospective sunt mai ușor de efectuat, fiindcă se ia o perioadă dinamică și datele despre incidența sau prevalența morbidității ne permit să determinăm precis pe cine să interogăm.

*3. De tip epidemiologic* – permite de a studia în mod obligatoriu loturi de persoane în paralel cu loturi martor și astfel se pot determina corelațiile cauzale multifactoriale ale morbidității. Aceste anchete sunt detalizate la capitolul respectiv.

**Etapa III.** Prelucrarea și gruparea informațiilor. Informația colectată primar urmează a fi prelucrată prin intermediul unui computer și a programelor specializate. Prelucrarea constă din următoarele momente:

*Verificarea cantitativă și calitativă* – verificarea logică a datelor. Prin verificarea cantitativă se are în vedere inventarierea lucrărilor rezultate din cercetare, a înregistrării corecte în fișe, buletine etc.; prin cea calitativă – verificarea logică a datelor (confruntând de ex. vârsta, sexul cu diagnozele etc.) pentru a descoperi anumite incompatibilități. În cazul studierii morbidității se poate verifica dacă diagnosticul este corect înregistrat și nu este doar un diagnostic simptomatic. Tot ceea ce este ieșit din comun, mult „dispersat” față de o valoare „medie” a coeficienților respectivi trebuie verificat din nou cu atenție, pentru a găsi eventuale erori survenite fie în etapa de cercetare propriu-zisă, fie cu ocazia înregistrării.

*Codificarea generală și codificarea medicală* – notarea cu un „cifru”, sau mai corect „cod” a tuturor caracteristicilor. Codificarea generală și medicală diferă doar prin natura caracteristicilor, ultima utilizându-se în cazul codificării bolilor, deceselor, intervențiilor chirurgicale etc. Codificarea medicală se efectuează după Clasificarea Internațională a Maladiilor elaborată de OMS, în prezent revizia a X-a.

*Gruparea* – împărțirea materialului obținut în cercetare după caracteristici cantitative și calitative și poate fi:

**simplă** – când colectivitatea cercetată se analizează după o singură caracteristică: sex sau grupe de vârste, boli etc. În general ea servește nu atât la aprofundarea corelativă a fenomenelor studiate, cât mai mult la cunoașterea detaliată a unei caracteristici tipice pentru fenomenul respectiv;

**complexă** (combinată) – se efectuează în baza mai multor caracteristici de grupare și are ca scop studiul corelativ al obiectului cercetării. Astfel de exemple sunt: structura morbidității pe medii (urban, rural), județe, sexe și grupe de vârstă etc.;

**repetată** – scindarea sau reunirea grupelor obținute după alte caracteristici de grupare în vederea unei analize mai aprofundate, respectiv unei sinteze generale.

*Calcularea indicatorilor relativi și valorilor medii, a dispersiilor, a erorilor, compararea, testarea diferențelor.*

**Etapa IV.** Analiza și sinteza datelor obținute. Analiza trebuie să scoată în evidență structura și dinamica fenomenelor cercetate, tendința acestor fenomene, legăturile funcționale de cauzalitate, ritmul de apariție, dezvoltare și stingere.

Sinteza constă în generalizarea rezultatelor, analizate în mod complex, compararea lor, evidențierea legăturilor cauzale, descoperirea legăturilor existente în cadrul fenomenului studiat, efectuarea concluziilor și propunerilor pentru implementarea în practică.