

PARTEA I

BIOSTATISTICA ȘI METODELE DE STUDIU ALE SĂNĂȚĂȚII PUBLICE

CAPITOLUL 1

BAZELE TEORETICE ȘI METODOLOGICE ALE BIOSTATISTICII

1.1. SCURT ISTORIC

Statistica a apărut din nevoia reală de a cunoaște în expresie numerică o serie de activități, fenomene și procese social-economice. Începuturile statisticii se întind până în antichitate sub forma unor evidențe necesare statului. Termenul de statistică derivă din cuvântul latin „status”, ceea ce înseamnă în traducere poziție, stare, situație și din cuvântul italian „statista” cu semnificație de om de stat.

Apariția proprietății private asupra mijloacelor de producție și a formațiunilor statale sclavagiste a impus ținerea unor evidente, a numărului populației și a activităților productive. Evidențierea lor, la început ca înregistrări izolate (*de unde și denumirea de statistică practică*), datează din cele mai vechi timpuri ale dezvoltării societății omenești. Astfel, primul recensământ a fost efectuat în anul 2300 î. Chr. la chinezi. Recensămintele efectuate de romani (Servius Tullius) (din 5 în 5 ani, apoi din 10 în 10 ani) erau cunoscute sub denumirea de „cens” (numărătoare). În Dacia evidențele populației se numeau „tabularium”.

În societatea feudală statistica nu a putut lua o dezvoltare mai importantă, ca în general toate celelalte științe, din cauza stării de fărâmițare teritorială și se reducea mai mult la inventarierea bunurilor și mai ales a pământurilor feudale, bisericesti. De asemenea, apar date statistice despre numărul decedaților în timpul epidemiilor, îndeosebi a epidemiilor de ciumă. Cu toate acestea în secolele XVI-XVII au fost elaborate un mare număr de lucrări, în care se descria detaliat situația social-economică folosind datele statistice. Apare „**statistica descriptivă**”. Dintre numeroasele lucrări, putem evidenția strălucita monografie de talie europeană a lui Dimitrie Cantemir „*Descriptio Moldaviae*” (1716). Tot atunci, la romani, apar pentru prima dată registrele pentru născuți, decedați, căsătoriți, care se păstrau în temple.

Dezvoltarea modului de producție capitalist a condus la apariția unor noi concepte și metode statistice de investigație. Englezul William Petty, prin lucrarea sa „**Aritmetica politică**”, contribuie substanțial la afirmarea statisticii ca știință.

Începând cu a II-a jumătate a sec. XVIII rolul metodelor statistico-matematice în investigarea și interpretarea rezultatelor fenomenelor și proceselor se amplifică odată cu apariția „**calculului probabilităților**”. Merită să amintim numele lui P.S. Laplace, care elaborează teoria probabilității, K.F. Gauss, care definește „Legea normală de repartiție”, P.L. Cebâșev, care a formulat „Legea numerelor mari”, dezvoltată ulterior de A.N. Colmogorov, S.D. Poisson, care definește „Legea numerelor mici”, „Legea de repartiție a evenimentelor rare” etc.

Paralel cu statistica generală se dezvoltă și statistica medicală. Începuturile au fost determinate de nevoia obținerii celor mai simple informații cantitative, formulate de regulă sub forma „câți bolnavi?”, „câți decedați?”, „câte celule sau microbi?” etc. Ulterior apar o serie de lucrări privind procesele demografice, precum și morbiditatea, mortalitatea și cauzele acestora. Se organizează sistemul de evidență medicală, precum și declararea și raportarea obligatorie a bolilor transmișibile. Spre sfârșitul sec. XIX apare statistica sanitară a zemstvelor.

Biostatistica, în școala medicală a Republicii Moldova, a apărut și s-a dezvoltat în cadrul Medicinii sociale, deoarece fără ea sunt de neconceput aprecierea stării de sănătate a populației, măsurarea corelației dintre factorii de risc și boală, evaluarea impactului unor factori de protecție asupra sănătății colectivității, fundamentarea cât mai obiectivă a unor programe de intervenție etc. Integrarea Biostatisticii în Medicina socială se explică și prin faptul că legile și procedeele ei sunt aplicabile deopotrivă în demografie și epidemiologie, domenii inseparabile legate de teoria și practica Sănătății publice. Totodată, biostatistica, alături de alte discipline de bază, este necesară înainte de toate pentru aportul său la formarea generală a medicului, deoarece îl învață să judece în termeni probabili, să înțeleagă și să măsoare riscurile deciziilor sale în condiții de incertitudine, să verifice ipoteze cauzale sau prognostice, să compare rezultatele unor observări sistematice și să afirme pentru un anumit grad de siguranță dacă diferențele rezultate din observări și măsurători se pot datora numai șanseii. În felul acesta biostatistica devine un instrument al cercetării biomedicale, clinice și medico-sociale, dar și al aplicării gândirii științifice în practica medicală.

Deci, dezvoltarea continuă a vieții social-economice a făcut ca statistica și biostatistica să capete, atât în plan național, cât și internațional, un cadru instituțional.

Începând cu sfârșitul sec. XVIII apar birouri oficiale de statistică și în particular de statistică sanitară.

În finalul acestui scurt istoric am putea afirma, că rădăcinile statisticii moderne sunt: statistica practică; statistica descriptivă; aritmetica politică; calculul probabilităților ș. a.

1.2. NOȚIUNI GENERALE

Biostatistica este o ramură a statisticii generale specializată în studiul fenomenelor biologice și medicale. Ea se ocupă de culegerea, centralizarea și gruparea datelor, de prelucrarea lor și de determinarea unor indicatori pentru descrierea fe-

nomenelor biomedicale studiate, pe baza evidențierii unor legități sau variabilități statistice.

Biostatistica este știința, care are ca obiect cunoașterea în structură și dinamică a sănătății unei populații, în corelație cu factorii social-economici, culturali, sanitaro-igienici și medico-biologici determinanți, având ca scop detectarea tendințelor acestei stări, în condițiile activității rețelei sanitare, a cărei eficiență și eficacitate este chemată s-o evalueze.

Nici o știință nu are ca obiect principal și ca posibilitate metodologică de studiu diversele aspecte ale sănătății, noțiune definită vag, dar atât de cuprinzător de către OMS ca: "Bunăstare fizică, psihică și socială".

Este necesar de a evidenția și constituirea biostatisticii ca știință, deoarece ea își are obiectul său propriu de studiu și metoda sa proprie, parțial comună cu metodele altor științe, dar aplicată necesităților de cunoaștere logică a fenomenelor din domeniul sănătății.

Un specific al biostatisticii este, că ea se ocupă de studiul fenomenelor în dinamică, identificând tendințele fenomenelor, acestea fiind unul din scopurile principale ale statisticii sanitare.

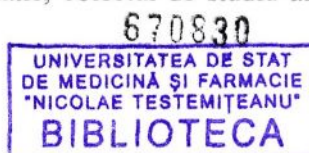
Obiectul de studiu al biostatisticii îl constituie fenomenele și procesele, care reprezintă următoarele particularități:

Studiul stării de sănătate a populației.	Cunoașterea condițiilor de mediu.	Cunoașterea resurselor sanitare.	Cunoașterea activităților medico-sanitare.
Indicii de morbiditate.	Aer.	Instituții medicale.	Asistența Medicală Primară.
Indicii demografici.	Apă.	Personalul medical.	Asistența Medicală Spitalizată.
Dezvoltarea somatometrică și studiul constantelor biometrice.	Alimentație.	Bugetul pentru sănătate.	Asistența Medicamentoasă.
Aspectele de invaliditate și handicap.	Habitat.		Asistența medico-socială.
Sănătatea mintală.	Proces de muncă și învățământ.		
	Radiații ionizante și alte noxe.		

Figura 2. Obiectul de studiu al biostatisticii. Compartimentele de studiu

- se produc într-un număr mare de cazuri;
- variază de la un caz la altul;
- sunt forme individuale de manifestare concretă în timp, în spațiu și sub raport organizatoric.

Astfel, rezultă că obiectul de studiu al statisticii îl constituie fenomenele de masă, care au caracter variabil ca formă de manifestare în timp, în spațiu și sub raport organizatoric. Schematic, obiectul de studiu al statisticii sanitare poate fi



divizat în 4 compartimente mari (*figura 2*). Toate aceste aspecte trebuie studiate în structură și dinamică, stabilindu-se previzional atât tendința fenomenelor, cât și necesarul de asistență medicală și resurse.

1.3. METODE DE CERCETARE

Metodele de cercetare aplicate în statistica sanitară nu diferă cu mult de cele folosite în alte domenii, unele dintre ele au fost chiar împrumutate de la acestea (cele matematice, economice). Utilizarea acestor metode se face ținând cont de specificul lor. De exemplu, statistica sanitară în cadrul studiului Sănătății publice are ca scop nu numai depistarea fenomenelor, dar și a dinamicii lor, evidențierea tendințelor acestor fenomene, corelației lor cu factorii ce le provoacă etc. Studiarea acestor particularități ale fenomenelor poate evolua prin intermediul următoarelor metode:

1. Metoda observației – constă în monitorizarea 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 prin această metodă în studiul Sănătății publice sau al unor cazuri clinice se realizează pe diferite căi:

a) statistică – când se acumulează informația despre schimbările fenomenului studiat sau activitatea organelor, instituțiilor sanitare și altor servicii pentru sănătate sub formă de valori numerice.

b) aprecierea prin expertiză – prezintă un supliment la calea statistică, 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 va fi studiat după externare în baza documentației medicale. Experții se pronunță cu privire la termenul spitalizării bolnavului, dată fiind starea lui, apreciază calitatea investigațiilor, corectitudinea diagnosticului și tratamentul prescris în fiecare caz concret de patologie. Același lucru se poate efectua și în cadrul serviciului de ambulatoriu, de urgență etc. Datele generalizate servesc la calcularea unor coeficienți de corijare obținuți pe cale statistică. Expertiza este o metodă mai mult calitativă de investigație, permițând elaborarea măsurilor concrete de ajustare a fenomenului studiat.

În funcție de timpul când se efectuează observația, deosebim:

a) cercetarea curentă (continuă) – se utilizează atunci când variabilitatea fenomenului studiat se poate schimba destul de des, când aceste schimbări 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, în secțiile de boli infecțioase, unde observarea trebuie să se facă de câteva ori pe zi sau peste un anumit interval de timp;

b) 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 unele boli se acutizează în anumite perioade ale anului (boala ulceroasă, 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.);

c) 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 observațiilor cercetările pot fi:

- de o singură dată;
- repetate sau cu revenire la aceeași totalitate.

2. Metoda epidemiologică sintetizează cele evidențiate prin metoda observației și prezintă un studiu corelativ al fenomenelor din cadrul Sănătății publice cu factorii (cunoscuți sau presupuși) de risc. Această metodă, ca și cea precedentă, utilizează 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 este strâns legată cu primele două metode, fundamentând studiul sanitar al Sănătății publice sau al activității serviciilor pentru sănătate în cadrul dezvoltării istorice a societății. Aici mai frecvent se utilizează comparările 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 este metoda economico-matematică, când se utilizează în combinație cu o serie de metode matematice ce ne permit să optimizăm acțiunile sanitare ale rețelei sanitare legate de problemele de sănătate a populației cu analiza de sistem, de prognozare etc.

5. Metoda experimentală are ca scop elaborarea și aprobarea noilor forme și metode de organizare a muncii, a asistenței medicale, aplicarea în practică a experienței înaintate, verificarea reciprocă a diferitor 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 constituie o problemă de ordin social. Soluționarea lui va cere din partea acestuia îmbinarea mai multor metode de cercetare.

6. Metoda demografică permite obținerea, analiza și sinteza informațiilor privind evenimentele, fenomenele din sânul populației, în structura și dinamica lor.

7. Metoda sociologică utilizează anchetarea și interviul. Ea permite de a studia opinia populației și a medicilor față de calitatea, oportunitatea, reformarea etc. serviciilor de sănătate.

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 la ce etapă.

1.4. TOTALITATEA STATISTICĂ

Studiul oricărui fenomen privind starea sănătății populației sau al altor noțiuni cu folosirea metodelor statistice cere de la medic alegerea în cunoștință de cauză a obiectului de studiu (adică a totalității statistice), unității de observare și a semnelor ei caracteristice.

Totalitatea statistică reprezintă un număr mare de elemente (unități de observație) relativ omogene, luate împreună în anumite hotare de spațiu și timp. Această totalitate poate servi drept eșantion.

Totalitatea statistică constă din unități de observație, grupate printr-o modalitate specifică, dar nu în urma unei sume mecanice a acestor unități. Numărul de unități de observație determină *volumul totalității* destinate studiului și simbolic se notează prin „*n*”.

În funcție de scopul și sarcinile studiului urmează să se delimiteze elementul primar al totalității – *unitatea de observație*. De exemplu, în studiul morbidității în populație, unitatea de observație va fi omul, în cel al mortalității – cazul de deces. Fiecare unitate de observație are mai multe caracteristici (*caractere*) sau particularități. Pentru început se va stabili, care particularitate va fi luată în evidență și care nu. Această problemă se rezolvă în cadrul oricărui studiu în parte și se aleg numai caracteristicile tipice necesare pentru atingerea scopului și elucidarea sarcinilor concrete ale cercetării. Astfel de caracteristici, cum sunt sexul, vârsta, locul de trai, durata bolii, spitalizării, rezultatele investigațiilor, tratamentul bolnavului și altele ne permit să dăm o caracterizare multilaterală nu numai a unității de evidență, dar și a totalității în ansamblu. Fiecare caracteristică poate avea la rândul ei diverse gradații (sexul – bărbat, femeie; vârsta – până la 15 ani, 16-20, 21-25, 26-30 ani etc.; locul de trai – urban, rural; rezultatele investigațiilor pot fi gradate după valorile lor numerice sau după nivelurile normei etc.). Între caractere pot fi găsite relații corelative mai mult sau mai puțin complexe și care pot fi evidențiate sau verificate pe parcursul cercetării întregului fenomen (sănătatea publică, morbiditatea, mortalitatea, natalitatea, dezvoltarea somatometrică, psihomotorie a copilului etc.).

După particularitățile sale de exprimare, fiecare caracteristică (*figura 3*) poate fi: **calitativă** (atributivă, descriptivă), exprimată prin cuvinte, sau **cantitativă** (numerică), exprimată prin cifre. Caracteristici calitative sunt: sexul, etnia, profesia,

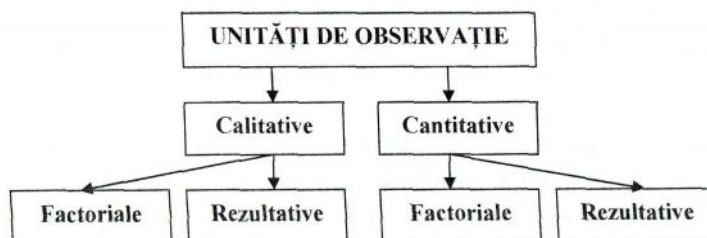


Figura 3. Clasificarea caracteristicilor unității de observație

locul de trai, entitatea nosologică, rezultatul final al tratamentului, calitatea lui etc. Caracteristicile cantitative includ: vârsta, indicii antropometrici, temperatura, durata de tratament, spitalizării, rezultatele investigațiilor etc. În acest caz fiecare caracteristică numerică, la fiecare unitate de observație, va avea valoarea sa specifică, pe care o vom numi variabilă sau variantă, notată simbolic prin „v”.

Pe parcursul studiului poate fi evidențiată corelația unor caracteristici ale fenomenului studiat, de care depinde într-o măsură mai mare sau mai mică valoarea altor caracteristici. De exemplu, masa corpului va depinde de înălțime, iar ultima este corelată de vârstă.

Caracteristicile, sub influența cărora se schimbă alte caracteristici, se numesc **factoriale**, iar celelalte – **rezultative**. Caracteristici factoriale sunt: metodele de profilaxie, diagnosticare, tratament, dozele medicamentelor, sexul, vârsta, profesia, studiile, venitul etc., cele rezultative includ: diagnosticul, rezultatul tratamentului (însănătoșire, ameliorare, fără schimbări, înrăutățire, invaliditate, deces), rezultatele analizelor (nivelul hemoglobinei, colesterolului etc.).

După ce și-a clarificat scopul și sarcinile studiului, cercetătorul trebuie să determine unitatea de evidență cu toate caracteristicile sale (descriptive sau numerice; factoriale sau rezultative) ce prezintă interes pentru cercetarea dată; toate aceste noțiuni urmează să fie oglindite în programul de culegere a materialului informativ, adică în registrul (chestionarul) respectiv.

Oricare ar fi obiectul de cercetări și scopul urmărit, este necesară delimitarea unității de observație 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. Deci, conform genului său, cunoaștem 2 tipuri de cercetare:

Cercetarea integrală – se subînțelege studiul întregului volum sau număr din colectivitatea cercetată, adică întreaga „populație” sau „universul statistic” (se iau toate unitățile de studiu existente la momentul dat). Cercetarea integrală este utilizată în studiul unor fenomene cu apariție rară, cu un număr mic de cazuri sau când cunoașterea este necesară pentru luarea măsurilor imediate (unele boli infecțioase și parazitare), sau pentru determinarea structurii populației prin recensământ.

Cercetarea parțială se limitează la studiul unei părți din „universul statistic”, ales pe bază de selecție, spre a putea fi reprezentativ pentru întreaga populație. Totalitatea parțială trebuie: să dețină particularitățile caracteristice de bază de care dispune cea integrală; să dispună de un volum satisfăcător, după numărul de observații, ca să exprime cât mai precis calitățile totalității integrale.

1.5. ELEMENTE DE TEORIE A PROBABILITĂȚII

Teoria probabilităților este o teorie matematică ce se ocupă cu studiul fenomenelor întâmplătoare ce pot apărea în gruparea unui semn studiat 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 legitate în fiecare zi, pe ore. Studiind această legitate, 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 măsura 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 prin două procedee:

– clasic, sau abordare a priori – dacă fenomenul se produce în „h” posibilități din „n” experimente, atunci probabilitatea acestui fenomen va fi „h/n”;

– frecvența empirică, sau abordare a posteriori – când experimentul se repetă de mai multe ori – „n”, iar fenomenul se produce cu o frecvență de acum fixată „h” (precum în exemplul cu adresările la serviciul de urgență). Aici probabilitatea apariției fenomenului va fi „h/n”.

Conform definiției clasice a probabilității, adoptate de P.S. Laplace, probabilitatea apariției în totalitatea parțială a unui fenomen „p” se determină prin raportul fenomenelor deja apărute „m” la numărul tuturor cazurilor posibile „n”:

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

Drept contraprobabilitate a evenimentului determinăm alternativa – probabilitatea lipsei fenomenului, însemnată prin litera „q”:

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

Din formulă reiese că suma probabilităților prezenței fenomenului cu contraprobabilitatea lui este egală cu unitatea, iar în procente cu 100%.

Rezultatele de mai sus pot fi exprimate și altfel: 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, și viceversa. Această teorie argumentează legea cifrelor mari.

Legea cifrelor mari are două aplicații importante pentru determinarea totalităților selective:

1. Pe măsura majorării numărului de observații rezultatele cercetării obținute pe baza totalității selective tind să reproducă datele totalității integrale.

2. La atingerea unui anumit număr de observări 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 observări se manifestă legitatea ce stă la baza totalității integrale și care nu poate fi observată în cazul unui număr mic de observații.

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%. La întâmplare se extrage câte o bilă, se înseamnă culoarea ei și apoi ea este pusă înapoi în urnă. Operația se repetă de 10, 20, 50 ori și mai mult. La început proporția bilelor 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 „P1” diferă de cea din totalitatea integrală „P” cu o mărime ce exprimă eroarea reprezentativă „m” și se explică prin faptul că nu toate unitățile de observație au fost studiate. Comparând aceste ponderi, dintre care una e obținută 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 încât poate fi neglijată.

Statistic este demonstrat faptul că în caz de o totalitate selectivă mare ($n > 30$) cu probabilitatea de 95% ponderea fenomenului „P1” va fi diferită față de cea din totalitatea integrală „P” cu „2m”; cu probabilitatea de 99,7% diferența „P1 – 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 înscamnă 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 „ Δ ” reprezintă eroarea limită admisă pentru studiul dat. Deci,

$$P = P_1 \pm \Delta$$

În cazul experimentului cu bilele s-a constatat legitatea că la un anumit număr de extrageri „n” într-o totalitate selectivă „P1” se înregistrează raportul aproape de cel real al bilelor albe și negre din totalitatea generală „P”. Odată cu majorarea volumului totalității selective, se poate influența asupra mărimii maxime a erorii, aducând-o la valori minime. Eroarea dată începe a fi aproape aceeași de la un anumit număr, suficient de mare, care este în dependență de numărul totalității generale.

Pentru a determina acest număr „n”, de la care indiferent cât vom majora numărul de observări, rezultatele nu vor fi semnificativ diferite, trebuie să cunoaștem cu ce exactitate este necesar de a primi rezultatele, adică eroarea limită admisă „ Δ ” pentru studiul dat. Astfel, putem determina volumul eșantionului.

1.6. EȘANTIONAJUL

Prin eșantion sau „totalitate selectată” se subînțelege un număr „n” de elemente extrase pentru o cercetare parțială dintr-o populație care, dacă ar fi studiată, ar reprezenta o cercetare integrală.

Eșantionajul prezintă două serii de avantaje. Astfel, datorită faptului că din totalitatea unităților de observare nu sunt studiate decât o parte, efortul depus în cercetare este mai redus. Există, adică, un plus de operativitate și un cost mai mic.

A doua categorie de avantaje este generată de calitatea de reprezentativitate a eșantionului. Deși nu se înregistrează toată populația, datele care se obțin sunt reprezentative pentru aceasta. În alți termeni, rezultatele cercetării efectuate pe un eșantion pot fi extrapolate la populația-țintă, care este reprodusă – în mic – în structura colectivității de selecție.

Însă datele eșantionului nu sunt întru totul certe: ele sunt valabile pentru populația de origine, numai cu o probabilitate, care este totdeauna inferioară unității.

De asemenea, eşantionul reproduce populația cu unele erori. Ele pot fi divizate în două categorii:

1. datorate subiectivității în măsurarea valorilor semnelor caracteristice a informațiilor celor investigați, a non-răspunsurilor și a utilajului de investigație, așa-ziselor erori de implementare și măsurare;

2. apărute în procesul de extrapolare a rezultatelor eşantionului la întreaga populație, numite erori de selecție și care pot fi mai grave decât cele de măsurare.

Deci, pentru a realiza avantajele eşantionajului, este necesară efectuarea unei selecții stricte, care ar îndeplini condițiile cercetării și o determinare corectă a volumului cercetării.

1.6.1. TEORIA SELECȚIEI

Teoria selecției, bazată pe aceeași teorie a probabilităților, constă din studiul relațiilor existente între o populație și eşantionul sau eşantioanele posibil de a fi extrase din aceasta, scopul selecției fiind cunoașterea întregii populații, întregi cu minimum de efort.

Tipurile de selecție sunt:

1. **Selecția aleatorie sau Randomizată simplă** (la întâmplare) se efectuează prin extrageri întâmplătoare din liste, în care sunt înregistrate toate cauzele individuale fără nici o grupare sistemică prealabilă. Selecția aleatorie poate fi **simplă**, după exemplul de mai sus, dar poate fi și **stratificată**, după o grupare prealabilă și extrăgând apoi la întâmplare cazurile, care trebuie cercetate din fiecare grupă, proporțional cu volumul fiecărui grup.

2. **Selecția mecanică – Sistemică** – o selecție probabilistică în care elementele dintr-o listă sunt selectate aplicând un pas fix (interval constant) după determinarea unui punct de plecare random. Dacă proporția subcolectivității studiate se stabilește la 5% sau 10% pe baza calculelor de probabilitate, se stabilește pasul de numărare, luându-se fiecare al 20-lea sau al 10-lea individ sau bolnav respectiv. Tot pe baza selecției mecanice sunt alese, de exemplu, localitățile dispuse după modelul tablei de șah.

Pasul de numărare se stabilește împărțind numărul cazurilor care formează populația de referință din care se extrage eşantionul la mărimea eşantionului.

Avantajul acestei selecții este simplitatea cu care se poate alege eşantionul, dar acest avantaj este uneori în detrimentul exactității rezultatelor cercetării, îndeosebi când densitatea în localitățile supuse cercetării este diferită.

3. **Selecția tipică (stratificată)** – selecție proporțională cu mărimea populației – se utilizează frecvent în cercetări din domeniul sănătății, unde variabilitatea fenomenelor este mare, fiind necesară zonarea, gruparea pentru o mai bună organizare. Teritoriul se împarte în zone, în care fenomenul respectiv are intensitate maximă, minimă și medie. Ulterior din fiecare zonă se extrage un număr proporțional cu intensitatea fenomenului.

Dacă în cadrul selecției tipice este nevoie de a se face o nouă grupare, stratificare, pentru a obține zone și mai omogene, metoda ia denumirea de selecție în mai multe faze, sau cu mai multe trepte.

Avantajul acestei metode pentru cercetări este net superior, fiindcă oferă mai mari șanse ca fenomenele studiate să fie uniform reprezentate, eşantionul obținut fiind reprezentativ pentru întreaga colectivitate.

Selecția tipică poate fi combinată, la rândul ei, cu selecția aleatorie sau mecanică în faza a doua sau a treia, obținându-se astfel o selecție mixtă sau combinată.

Este adevărat că, așa cum spunea statisticianul rus Paievski, un eșantion mare (50 000 unități de selecție), extras „oricum” dintr-o populație este reprezentativ pentru orice populație, totuși, în special în fenomenele cu o mare variabilitate a caracteristicilor studiate, numai o selecție combinată: tipică stratificată și pe ultimele sau ultima treaptă, aleatorie sau mecanică, poate garanta fiabilitatea rezultatelor și reprezentativitatea față de populația din care s-au extras eșantioanele.

4. **Selecția în cuiburi (în serii) – cluster** – o selecție probabilistică, în care fiecare unitate este de fapt o colecție, o microzonă (un cuib de „ n ” elemente), ce urmează să fie cercetate integral. Organizația Mondială a Sănătății recomandă această metodă de selecție drept una dintre cele mai reprezentative.

Tot pentru a satisface cerințele teoriei probabilităților, trebuie să se calculeze foarte corect dacă structura pe diferite caracteristici ale eșantionului (sex, vârstă, ocupație, pe total, zone, medii etc.) este similară cu caracteristicile populației, în caz contrar trebuie să fie eliminate o parte din unitățile de selecție (gospodării, persoane etc.) pentru o standardizare a eșantionului după modelul populației (deci structurile procentuale pe sex, vârstă etc. ale eșantionului să fie similare cu cele ale populației de referință).

1.6.2. MĂRIMEA EȘANTIONULUI REPREZENTATIV

În subcapitolul precedent s-a vorbit despre căile de selecție a eșantionului, avantajul fiecărei metode în atingerea scopului final al studiului. O importanță și mai mare are posibilitatea de a determina numărul de cazuri de observație al eșantionului, ceea ce permite să se precizeze nu numai limitele, dar și să se asigure reprezentativitatea totalității selective. Determinarea unui număr optim de cazuri prezintă importanță și sub aspectul eliminării sau reducerii erorilor, permite de a repartiza just necesitățile în cadre și mijloace și de a reduce termenele studiului.

Este necesar ca fenomenul studiat să fie cunoscut sub toate aspectele sale, în profunzime, în acest sens fiind bine ca fenomenul să fie supus în prealabil unei analize statistico-matematice. Dar, de cele mai multe ori, nu există o posibilitate de a cunoaște în profunzime fenomenul studiat. În aceste cazuri, stabilind o eroare limită cu care să se lucreze „ Δx ” și o probabilitate admisă „ P ”, putem, conform recomandărilor lui P. Mureșan, să stabilim eșantionul reprezentativ după formula:

$$n = \frac{N t^2 p q}{N \Delta x^2 + t^2 p q}, \text{ unde:}$$

n – volumul eșantionului reprezentativ;

t – factorul de probabilitate, care poate fi egal cu 1,96 sau 3 pentru o probabilitate de 95% ori 99% respectiv;

p și q – probabilitatea și contraprobabilitatea de apariție (sau neapariție) a fenomenului cercetat.

În cazurile când nu avem date despre fenomenul cercetat, se constată că „ n ” este maxim când produsul „ pq ” este maxim, or, ținând seama de faptul că $0 \leq p \leq 1$ și $q = 1 - p$ produsul este maxim, atunci când $p = q = 0,5$;

Δ – eroarea limită admisă, valoarea maximă acceptată a căreia este de 5%, pentru care „ Δx ” va fi egală cu 0,05;
N – volumul colectivității generale.

În literatura de specialitate putem găsi mai multe formule, în baza cărora se apreciază mărimea eșantionului, toate luând în considerare factorul de probabilitate și eroarea admisă. De asemenea, pot fi utilizate tabele speciale, care au fost elaborate și permit de a stabili numărul cazurilor necesare de a fi investigate, adică volumul eșantionului.

1.7. CALCULAREA INDICATORILOR STATISTICI

Datele statistice obținute în cursul cercetării nu exprimă decât însușiri selective, individuale sau de grup și sunt, de obicei, prezentate prin valori absolute. Aceste valori sunt greu de interpretat, lipsind elementul comun de comparare, corelare și sinteză. Cu toate că uneori în anuarele statistice întâlnim în loc de indici speciali valori absolute, așa-numiții indicatori primari, aceștia pot fi utilizați atunci când studiem evenimente sau fenomene cu o frecvență rară, cifra absolută exprimând mai bine esența structurii și dinamicii fenomenului studiat. În celelalte cazuri sunt calculați anumiți indicatori cu ajutorul cărora se generalizează, abstractizează pentru o interpretare comparativă, corelativă a diferitor subgrupe de semne ale totalității cercetate, în felul acesta analizându-se mai profund caracteristicile individuale și de grup ale semnelor, iar pe baza lor și ale totalității întregi. Acești indicatori se numesc derivați; ei pot fi prezentați sub formă de valori relative sau medii.

1.7.1. VALORILE RELATIVE

Pentru a face o caracteristică a totalității cercetate, în prealabil, după semnele ei calitative (mai ales în caz de repartitie alternativă) se utilizează valorile relative.

Mai frecvent folosim 4 grupe de indicatori:

1. Indicatorii intensivi (de nivel, de frecvență) arată care este frecvența unui fenomen dintr-o anumită perioadă și teritoriu față de o colectivitate care nemijlocit a produs acest fenomen. Acești indici pot fi numiți rate, deoarece fenomenul provine din populație.

$$I_{int} = \frac{\text{Nr. abs. al fenomenului}}{\text{Nr. abs. al populației}} \times 100 \quad (1\ 000; 10\ 000; 100\ 000)$$

În funcție de înmulțitor, indicatorul va fi prezentat:

100 – ‰ procent

1 000 – ‰ promile

10 000 – ‰ prodecimile

100 000 – ‰ prosantimile

Indicatorii intensivi se utilizează pentru stabilirea fenomenului cercetat (morbiditatea, natalitatea, mortalitatea) – unicul ce poate fi utilizat pentru compararea frecvențelor cercetate din diferite teritorii. Se utilizează pentru analiza succesivă, în dinamică a fenomenelor cercetate.

Exemplu. În orașul D au fost înregistrate 350 de cazuri de deces la o populație de 50 000 de locuitori. Indicatorii intensivi în majoritatea cazurilor de studiu al

morbidității, proceselor demografice, invalidității etc. se calculează în medie la o mie de locuitori. Deci, în cazul dat indicele va fi:

$$x = \frac{350 \times 1000}{50\ 000} = 7\ ‰$$

La analiza cazurilor de morbiditate cu pierderea temporară a capacității de muncă și a letalității spitalicești cota mediului față de care se determină nivelul fenomenului va fi o sută. Acest fapt e legat în primul caz de necesitatea comparării nivelului morbidității date în diferite colectivități după volum mare, iar în al doilea – mediul (numărul de bolnavi în colectivitate, în spitale etc.) e mic.

În genere, când analizăm un fenomen cu o frecvență de răspândire mică într-un mediu mare indicele intensiv poate fi calculat față de 10 000, 100 000. Aici urmărim două scopuri:

1. obținem rezultatul, fie chiar și fracționar, dar sub o valoare întreagă înaintea virgulei – 1, 2 și mai mult;
2. acest indice, fiind un număr întreg și nu fracționar mic (de tipul 0,00001), poate fi ușor citit sau reprezentat grafic.

Deci, în așa cazuri unitatea de măsură a valorii indicelui va fi exprimată în promile, prodecimile, prosantimile etc. Orice indice căpătat sub formă de promile, prodecimile etc. poate fi ușor schimbat, fără să-l recalculăm, mutând virgula cu o cifră în dreapta sau în stânga.

Când calculăm indicii anuali ai mortalității, natalității, morbidității, la numărător luăm populația medie anuală (semisuma numărului de populație în localitate la început și la sfârșit de an).

La calcularea coeficientului de intensitate a fenomenului pe o perioadă a anului (o lună, un trimestru) pentru a-l raporta la indicii similari din anii precedenți, compararea se va face cu perioada analogică a anilor trecuți sau dacă încercăm să-l comparăm cu indicii anuali din trecut, atunci și pe cel dat trebuie să-l transferăm la nivelul indicelui anual așteptat.

Pentru aceasta, în formula obișnuită de calcul al indicatorului intensiv numărătorul se înmulțește la 12 (numărul de luni ale anului), iar numitorul – la numărul de luni pe baza căruia a fost determinat indicele selectiv.

De exemplu: pe parcursul primului trimestru al anului de dare de seamă în localitatea E, cu o populație de 400 000 de oameni, s-au născut 1 500 de copii. Indicele de intensitate va fi:

$$x = \frac{1500 \times 1000}{400\ 000} = 3,75\ ‰$$

Pentru al readuce la nivelul indicelui anual facem următoarea recalculare:

$$x = \frac{1500 \times 1000 \times 12}{400\ 000 \times 3} = 15\ ‰$$

Deci, nivelul anual al natalității în localitatea dată va fi estimat la 15‰.

Utilizarea indicatorilor intensivi în biostatistică este mai frecventă comparativ cu alți indicatori și, de obicei, ei sunt necesari la:

- determinarea nivelului sau frecvenței de răspândire a unui fenomen într-o colectivitate concretă cercetată;
- compararea diferitor totalități (colectivități) după nivelul de frecvență al unui fenomen omogen;

- evidențierea schimbărilor dinamice în frecvența răspândirii fenomenului cercetat pe baza unei totalități concrete.

Pentru aplicarea în practică a acestor aspecte de utilizare, indicatorii intensivi trebuie să fie calculați de cercetător după o metodă identică (referitor la înmulțitor), în caz contrar, indicii devin incomparabili.

2. Indicatorii extensivi (de repartiție, de structură, de pondere) arată raportul dintr-o parte integrantă a fenomenului și fenomenul integru (structura, proporția, cota). Acești indicatori sunt reprezentați prin %, calculați după regula a trei simple și sunt dispuși în tabele sau în text, totalul subgrupurilor fiind egal cu 100,0%. Uneori acești indicatori pot fi exprimați în promile (‰).

Acest tip de indicatori niciodată nu face asociere între mediu și fenomen și nu permite să fie evaluate schimbările dinamice ale lui, făcând numai bilanțul static la un anumit moment sau pentru o perioadă anumită de timp.

De exemplu, în a. 2007 în localitatea „H” au fost înregistrate 2 500 de cazuri de îmbolnăviri, inclusiv 800 cazuri de maladii cardiovasculare, 500 – pulmonare, 600 – infecțioase și 450 – traumatisme.

Metoda de calcul. Totalul de boli înregistrate – 2 500 = 100%. Ponderea maladiilor cardiovasculare nu este cunoscută și se notează prin „x”.

Deci, rezultă proporția: 2 500 boli constituie100%
800 boli din tot întregul, vor constitui x %,

$$\text{unde: } x = \frac{800 \times 100}{2500} = 32,0\%$$

În aceeași consecutivitate calculăm ponderea procentuală și la celelalte componente ale fenomenului, care va fi egală cu:

– maladii pulmonare	– 20,0%
– maladii gastrice	– 24,0%
– boli infecțioase	– 6,0%
– traumatisme	– 18,0%

Analizând rezultatele obținute, observăm că ponderea cea mai mare revine maladiilor cardiovasculare – locul I în structură, pe locul II se plasează bolile gastrice, apoi cele pulmonare, traumatismele și bolile infecțioase. Analiza aceluiași fenomen pe baza datelor aceleiași localități numai că într-o altă perioadă de timp arată că structura lui s-a schimbat și locurile nosologiilor, conform ponderii, sunt altele. Este un lucru firesc, deoarece pe baza acestui indicator noi am studiat structura fenomenului, fără a ține însă cont de schimbările ce au avut loc în timp și în mediul (populația) care l-a produs. Deci, comparând un fenomen omogen pe diferite perioade de timp în aceeași totalitate cercetată sau în totalități diferite pe baza indicatorilor extensivi trebuie să apreciem numai locul ocupat de fiecare parte componentă în structura fenomenului după ponderea respectivă și nicidecum să operăm cu cota lui procentuală, deoarece în acest răstimp au suferit schimbări esențiale atât fenomenul, cât și mediul. La micșorarea ponderii unei părți a totalului se va mări numai de câțiva cota procentuală a alteia sau altor părți, dar în toate cazurile suma părților va fi egală cu 100%.

La compararea ponderilor se va ține cont de faptul că valoarea mare a indicatorului extensiv nu ne denotă valoarea lui absolută de repartiție (distribuție) reală în

totalitate. În același timp, comparând numai indicatorii extensivi, nu putem stabili cauza schimbărilor produse în fenomenul dat sau a factorilor ce l-au cauzat. Pentru aceasta este necesar de a cunoaște alți indicatori.

3. Indicatorii de raport simplu arată raportul dintre două valori independente una față de alta și care nu se produc reciproc. Mai frecvent se utilizează pentru a demonstra nivelul de asigurare a populației cu resurse (medici, medici de diferite profiluri, paturi, aparataj etc.).

$$I_{\text{rap}} = \frac{\text{Nr. abs. de resurse}}{\text{Nr. abs. al popul.}} \times 1\,000$$

Dacă după tehnica de calcul acest indice este similar cu cel intensiv, apoi după esență el este opus.

De exemplu: În localitatea F cu o populație de 30 000 de locuitori au fost amplasate 300 de paturi. Care este nivelul de asigurare a populației cu ele?

$$\text{Răspuns: } N_{\text{asigurare}} = \frac{300 \times 1000}{30\,000} = 10,0\%$$

Comparând nivelul de asigurare a populației cu paturi spitalicești, calculat prin metoda de mai sus, cu normativul respectiv în vigoare, putem trage concluzii referitoare la nivelul de asigurare a populației din localitatea dată cu asistență medicală spitalicească. Posibilitățile de utilizare a acestui indicator sunt asemănătoare cu cele ale indicelui intensiv, numai că scopul se referă în primul rând la aprecierea utilizării mijloacelor serviciului pentru sănătate. Acest indice nu este o rată, ci un raport simplu, deoarece fenomenul (paturile) nu este o parte a populației.

4. Indicatorii demonstrativi (ilustrativi) arată raportul unei serii de valori față de o mărime luată ca bază și considerată egală cu 100%. Acest indice ne demonstrează de câte ori sau cu câte % s-a micșorat sau s-a mărit fenomenul cercetat și se utilizează atunci când apare necesitatea de a analiza schimbările dinamice ale unui semn sau fenomen omogen, sau a totalității în studii repetate.

De exemplu, dinamica numărului de medici în ultimii 20 de ani.

1985 – 15 568 – 100 %

1990 – 13 915 – x

1995 – 13 302

2000 – 11 718

2005 – 10 833

Datele anului 1985 sunt luate drept bază de calcul și sunt considerate egale cu 100%. Pentru anul 1990, calculăm după regula a trei simple:

$$x = \frac{13915 \times 100}{15568} = 89,4\%$$

Similar calculăm pentru fiecare an, luând tot anul 1985 drept bază fixă. Deci, pentru 2005 valoarea va fi:

$$x = \frac{10833 \times 100}{15568} = 69,6\%$$

Analizând rezultatele obținute constatăm, că numărul medicilor în a. 1990 față de 1985 s-a micșorat cu 10,6% (100 - 89,4), iar în 2005 față de același an – cu 30,4%. Deci, se observă o tendință vădită de scădere a numărului absolut al medicilor din Republica Moldova pe parcursul anilor cercetați.

Diferența dintre nivelurile absolute ale perioadei curente și celei precedente se numește spor absolut sau creștere absolută.

Raportul procentual al nivelului următor către nivelul precedent se numește ritm de creștere.

1.7.2. SERIA VARIABILĂ. VALORILE MEDII

În studiul sistemic al populației, rezultatele observațiilor trebuie ordonate în raport cu o variabilă statistică.

Variabila statistică este însușirea comună a unităților unei populații care își poate schimba nivelul de dezvoltare în timp și/sau spațiu, sub influența diferitor factori. Variabila poate fi calitativă și cantitativă.

Variabila calitativă este reprezentată printr-o literă și nu numeric (sexul; ocupația; naționalitatea etc.). Ea poate fi recunoscută cantitativ prin numărarea unităților de observație.

Variabila cantitativă este o caracteristică măsurabilă exprimată prin cifre, cu proprietăți de ierarhizare, mărimile cifrelor exprimând și intensitatea fenomenului.

Variabila continuă este exprimată în valori relative (natalitate, mortalitate, morbiditate).

În general în toate cercetările, dar în special în cele care privesc starea de sănătate a populației, un interes deosebit prezintă cunoașterea comparativă a fenomenelor studiate față de un etalon – valoare medie.

Valorile medii sunt indicatori ai însușirii calitative esențiale a fenomenelor colectivității studiate, măsura tendinței centrale a repartițiilor de frecvențe ale variabilei.

Pentru o prelucrare corectă a caracteristicilor unităților de studiu redată în cifre este nevoie să construim o serie de variații (*tabelul 1*) ale variabilelor supuse analitic.

Durata medie a incapacității de muncă (în zile)

Tabelul 1

Numărul zilelor incapacității, V	Numărul bolnavilor, f	Produsul $V \cdot f$
12	24	288
13	25	325
14	26	364
15	22	330
16	21	336
17	18	306
18	17	216

Seria variabilă constă din unități (variante – „ V ”) ale unei și aceleiași caracteristici aranjate în ordine crescândă sau descrescândă. Dacă o variantă se întâlnește o singură dată este o serie simplă, iar dacă se repetă de mai multe ori – seria este ponderată. În acest caz fiecare variantă „ V ” are frecvența sa – „ f ”.

Valoarea medie calculată are aceleași dimensiuni concrete cu ale variabilei a cărei repartiție de frecvențe o caracterizează. Astfel, dacă variabila privește greutatea în kg și valoarea medie va fi redată în kg.

Valorile medii pot fi exprimate prin cifre absolute sau prin indici relativi. De exemplu, în cazul cunoașterii TA sau înălțimii, sau greutății medii la diferite vârste

și sexe, valorile medii se exprimă în cifre absolute (mmHg, cm, kg). În cazul morbidității sau mortalității generale sau specifice pe vârste sau sexe, valorile medii se exprimă prin cifre relative (la 1 000, 100 000 locuitori etc.).

Valorile medii mai frecvent utilizate în studiile stării de sănătate sunt:

1. Mediana – Me.
2. Modulul (dominanta) – Mo.
3. Media aritmetică simplă – M.
4. Media aritmetică ponderată – Mp.
5. Media armonică – Marm.
6. Media cronologică – Mcr.
7. Media geometrică – Mge.

Mediana – o valoare medie a cărei mărime corespunde variabilei ce se găsește la mijlocul seriei de variații. Dacă în seria de variații este un număr par de variante se ia media aritmetică a celor două variante din mijloc.

Modulul – o valoare medie, care corespunde variantei cu o frecvență maximă în seria de variații.

Media aritmetică simplă – suma variabilelor împărțită la numărul de variante.

$$M = \frac{\sum V}{n}$$

Media aritmetică ponderată:

$$M = \frac{\sum V \cdot f}{n}$$

Dacă calculăm media aritmetică ponderată a exemplului propus în *tabelul 1* obținem:

$$M = \frac{(12 \times 24) + (13 \times 25) + (14 \times 26) + (15 \times 22) + (16 \times 21) + (17 \times 18) + (18 \times 17)}{153} = 14,74 \text{ zile}$$

Însă, apare întrebarea: este oare reprezentativă această medie? Răspunsul ni-l dă deviația standard.

Deviația standard (σ) – permite să apreciem gradul de omogenitate a variabilelor din seria de variație și, de asemenea, prin deviația standard se poate aprecia dacă media aritmetică este reprezentativă. Deviația standard arată gradul de dispersie a valorilor seriei de variație față de media aritmetică.

Pentru seriile de variație simple deviația standard se calculează după formula (dacă eșantionul constă dintr-un număr mai mic de 30 de unități, la numitor, în formulă va fi „n-1” :

$$\sigma = \pm \sqrt{\frac{\sum d^2}{n}} \text{ . Pentru seriile de variație ponderate: } \sigma = \pm \sqrt{\frac{\sum d^2 \times f}{n}} \text{ , unde}$$

$$d = M - V$$

M – media aritmetică

V – fiecare variabilă

n – numărul de variante

f – numărul de frecvențe

Pentru a vedea care-i mărimea relativă a deviației standard în statistică se calculează coeficientul de variație ce reflectă în % gradul de dispersie a variantelor față de media aritmetică: $Cv = \pm \frac{\sigma}{M} \times 100$

– Dacă Cv este cuprins între 0 și 10 %, gradul de dispersie a valorilor față de media aritmetică este mic, ceea ce înseamnă că omogenitatea seriei de variație este foarte înaltă.

– Dacă Cv = 10 – 20 %, gradul de dispersie este mediu, deci omogenitatea seriei de variație este medie.

– Dacă Cv este mai mare de 20%, omogenitatea seriei de variație este foarte slabă, deci este necesar de a mări numărul de observări.

Media armonică se utilizează în calcularea valorii medii, în cazul mai multor valori relative. Se utilizează rar, în special în stabilirea prețurilor medii. Este, de obicei, mai mare decât media aritmetică.

$$M \text{ arm.} = \pm \frac{\sum^n}{\sum \frac{1}{V}^n}, \text{ unde:}$$

V – valoarea relativă

n – numărul populației la care s-a calculat valoarea relativă

1/V – inversul mărimilor relative

Media cronologică – seriile cronologice.

Media geometrică – este valoarea medie, care se utilizează în cazul coeficienților de creștere medie a valorilor unei serii cronologice logaritmice.

1.7.3. SERIA CRONOLOGICĂ ȘI ANALIZA EI

O sarcină importantă a medicinei și ocrotirii sănătății este studiul sănătății publice, privind caracterul, volumul și calitatea asistenței medicale acordate și a activității instituțiilor sanitare, ținând cont de schimbările fenomenelor și proceselor în dinamică. Pentru analiza în dinamică a proceselor și fenomenelor este necesar de a compara serii cronologice de diferit tip, de a cunoaște modul de ajustare și analiză a lor.

Serie cronologică se numește șirul alcătuit din valori omogene comparabile, care caracterizează schimbările unui fenomen oarecare într-o perioadă de timp.

Valorile numerice ale elementelor constitutive ale șirului analizat se numesc **niveluri** ale seriei cronologice. Nivelurile seriei cronologice pot fi reprezentate prin valori absolute, relative și medii.

Tipurile seriilor cronologice. Seriile cronologice pot fi **simple** (alcătuite din valori absolute) și **compuse** (alcătuite din valori relative sau medii). Seria cronologică simplă, la rândul ei, poate fi de 2 tipuri: **de moment** și **de interval**. Seria cronologică simplă de moment este alcătuită din valori ce caracterizează nivelul fenomenului la un moment dat. Nivelurile seriei cronologice simple de moment nu pot fi fracționate sau divizate (*tabelul 2*).

Tabelul 2

Numărul de paturi (la sfârșit de an) în staționarul raionului «C» în perioada anilor 2001–2005					
Anii	2001	2002	2003	2004	2005
Numărul de paturi	122	120	114	109	98

Seria cronologică simplă de interval este o serie de valori ce caracterizează unele totaluri pentru un interval de timp (zi, săptămână, lună, an) (*tabelul 3*).

Tabelul 3

Numărul de nou-născuți în județul «N» în perioada anilor 2001–2005					
Anii	2001	2002	2003	2004	2005
Numărul de nou-născuți	690	760	640	620	600

Fenomenele seriei cronologice simple de interval, spre deosebire de cea de moment, pot fi divizate în perioade de timp mai mici sau pot fi comasate în intervale de timp mai mari. Astfel, în anul 2002 numărul nou-născuților a constituit 760, iar într-un trimestru al acestui an, de 4 ori mai mic – $760 : 4 = 190$ de copii.

Seriile de interval pot fi alcătuite nu numai din numărul nou-născuților, dar și din numărul celor decedați, din numărul bolnavilor etc., cu alte cuvinte, ele reprezintă datele fenomenelor, care se cumulează în timp.

Alegerea perioadei pentru seria de interval (an, lună, săptămână, zi, oră etc.) este determinată într-o oarecare măsură de gradul de variabilitate a fenomenului (mortalitate, morbiditate–natalitate etc.). Cu cât mai lent se schimbă fenomenul în timp, cu atât mai mari pot fi perioadele de analiză.

Seriile simple (atât de moment, cât și de interval) sunt inițiale (primare) în construirea seriilor compuse. Seriile compuse sunt alcătuite din valori medii (durata medie de tratament, numărul mediu de paturi pentru câțiva ani etc.) sau din valori relative (morbiditate, mortalitate, natalitate).

Ajustarea seriei cronologice. Seria cronologică nu întotdeauna este alcătuită din niveluri care se schimbă continuu spre diminuare sau creștere. Uneori nivelurile reprezintă o diversitate de oscilații, care fac imposibilă identificarea legităților de bază caracteristice fenomenului studiat. În astfel de cazuri, pentru a depista tendința dinamică sau legitățile de manifestare a fenomenului, este nevoie de ajustarea seriei cronologice.

Există câteva metode de ajustare a seriilor cronologice: *majorarea intervalului, nivelarea seriei cu ajutorul mediei de grup și nivelarea seriei cu ajutorul mediei glisante*. Ajustarea nivelurilor seriilor cronologice se face, însă, numai după analiza cauzelor, care se explică prin oscilarea acestor nivele.

Majorarea intervalului se face prin sumarea datelor pentru un șir de perioade megieșe (tabelul 4). Precum se vede din tabel, numărul de îmbolnăviri de anghină în fiecare lună oscilează, mărindu-se sau micșorându-se. Efectuând majorarea intervalelor prin sumarea nivelurilor după trimestrele anului, observăm o legitate sezonieră determinantă: cel mai mare nivel de îmbolnăviri se înregistrează în perioada de vară spre toamnă.

Tabelul 4

Oscilațiile sezoniere ale cazurilor de anghină în orașul «B», pe parcursul anului 2005													
Luna	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
	Numărul de cazuri												
Pe lună	129	193	133	387	230	288	530	270	380	231	137	260	3 268
Pe trimestru			455			950			1280			628	3 268

Calcularea mediei de grup pentru fiecare perioadă majorată se face în felul următor: suma nivelurilor perioadelor învecinate se raportează la numărul de niveluri sumate (tabelul 5).

Tabelul 5

**Dinamica procentului de divergență
a diagnosticelor clinice și patomorfologice după datele spitalului «C»**

Anii	2000	2001	2002	2003	2004	2005	2006	2007
% divergenței diagnosticelor	11,0	9,8	8,0	9,2	8,2	8,6	8,5	7,9
Media de grupă	–	10,4	–	8,6	–	8,4	–	–

Pentru nivelurile prezentate în tabel sunt caracteristice oscilații ondulatorii. Ajustarea seriei cu ajutorul mediei de grup ne permite să obținem datele ce ilustrează o tendință clară spre scăderea treptată a procentului de divergență a diagnosticelor clinice și patomorfologice în spitalul studiat.

Ajustarea seriei cronologice prin calcularea mediei glisante înlocuiește nivelurile cu o valoare medie obținută din nivelul dat și două învecinate (tabelul 6).

Seria cronologică ajustată cu ajutorul mediei glisante prezintă o tendință consecutivă de scădere a procentului de divergență a diagnosticelor. Astfel, calcularea mediei glisante este o metodă simplă de ajustare a seriei cronologice. Cu ajutorul acestei metode are loc nivelarea și eliminarea oscilațiilor extremale ale seriei cronologice.

Tabelul 6

**Dinamica procentului de divergență
a diagnosticelor clinice și patomorfologice după datele spitalului «C»**

Anii	2000	2001	2002	2003	2004	2005	2006	2007
% divergenței diagnosticelor	11,0	9,8	8,0	9,2	8,2	8,6	8,5	7,9
Media de grupă	–	9,6	9,0	8,7	8,6	8,9	8,3	–

Iată un exemplu de calcul pentru anul 2001 $(11,0 + 9,8 + 8,0) : 3 = 9,6$, pentru anul 2002 $(9,8 + 8,0 + 9,2) : 3 = 9$ și așa mai departe. Rezultatele obținute după ajustare nu pot fi considerate ca indici reali de analiză a seriei cronologice, dar se folosesc cu scop auxiliar.

Analiza seriei cronologice. Pentru o analiză mai detaliată se recomandă mai întâi reprezentarea grafică a seriei cronologice. Graficul arată în mod intuitiv succesivitatea schimbării fenomenului studiat. Seria, ajustată cu ajutorul mediei glisante, demonstrează o tendință clară: scăderea treptată a indicilor de frecvență a divergențelor de diagnostice.

Pentru analiza seriei cronologice utilizăm următorii indici: sporul absolut, ritmul de creștere, ritmul sporului, valoarea absolută a 1% de spor (tabelul 7).

Tabelul 7

**Incapacitatea temporară de muncă în legătură cu bolile sistemului nervos
ale lucrătorilor de la uzina «C» în perioada anilor 2002–2005**

Indicii	2002	2003	2004	2005	Total
Numărul de zile la 100 de lucrători	39,8	44,6	55,5	59,7	–
Sporul absolut	–	+ 4,8	+ 10,9	+ 4,2	+ 19,9
Ritmul sporului	–	+ 12,1	+ 24,4	+ 7,5	+ 50,0
Ritmul de creștere	–	112,1	124,4	107,5	150,0
Valoarea absolută a unui % de spor	–	0,39	0,44	0,56	

Metodele de calcul a indicilor.

1. **Sporul absolut** este diferența dintre nivelurile anului curent și precedent.

Exemplu: pentru anul 2003 el va fi: $44,6 - 39,8 = +4,8$.

2. **Ritmul sporului** este raportul procentual al sporului absolut către nivelul precedent. *Exemplu:* pentru anul 2003 el va fi: $(4,8 : 39,8) \times 100 = +12,1\%$.

Ritmul sporului denotă cu câte procente a crescut frecvența fenomenului studiat într-o perioadă de timp.

Dacă nivelul fenomenului scade, ritmul sporului va fi cu semn negativ (-) și va caracteriza scăderea relativă a fenomenului.

3. **Ritmul de creștere** este raportul procentual al nivelului următor către nivelul precedent. *Exemplu:* pentru anul 2003 el va fi: $(44,6 : 39,8) \times 100 = 112,1\%$.

4. **Valoarea absolută a unui procent de spor** este raportul dintre sporul absolut și ritmul sporului. *Exemplu:* pentru anul 2003 ea va fi: $4,8 : 12,1 = 0,39$.

Analiza statistică a seriilor cronologice trebuie efectuată reieșind din analiza calitativă preliminară a fenomenului studiat.

1.8. TESTAREA CERTITUDINII VALORILOR RELATIVE ȘI MEDII

1.8.1. EROAREA STANDARD

În cadrul cercetărilor este foarte greu să culegi o informație primară despre toate unitățile de lucru. De obicei, când calculăm valorile relative sau medii, le calculăm dintr-o parte a colectivității. Valorile calculate din colectivități parțiale se numesc valori empirice, care au o eroare, numită standard.

1.8.2. EROAREA STANDARD A VALORILOR RELATIVE

Când se testează certitudinea valorilor relative empirice se stabilește cu cât a greșit cercetătorul calculând indicatorii relativi într-o cercetare parțială față de rezultatele pe care el le-ar fi obținut într-o cercetare integrală.

$$eP = \pm \sqrt{\frac{p \times q}{n}}, \text{ unde:}$$

p – indicatorul relativ

q – diferența dintre înmulțitor și indicatorul relativ

n – numărul unităților de studiu cercetate

De exemplu: Într-o cercetare parțială, care avea drept scop de a determina nivelul incidenței prin HTA din localitatea X, s-au examinat 800 de persoane. La 112 din ele tensiunea arterială depășea limitele normei. Inițial calculăm indicele intensiv, adică incidența prin HTA:

$$P = \frac{112}{800} \times 1000 = 140,00\%$$

$$eP = \pm \sqrt{\frac{140 \times 860}{800}} = \pm \sqrt{150,5} = \pm 12,27$$

Dacă înmulțim eroarea cu 10 și ea depășește indicatorul, atunci această eroare nu este certă și deci indicatorul este calculat dintr-o colectivitate nesemnificativă: $(12,27 \times 10 = 122,7)$ – este mai mic ca 140, indicele este cert.

Această formulă se folosește când indicatorul nu depășește înmulțitorul. Dacă indicatorul depășește înmulțitorul, atunci $p - 1000 =$ cu o valoare negativă și nu putem efectua calcule, deoarece ar contraveni legilor matematice. Atunci utilizăm o altă formulă:

$$eP = \pm \sqrt{\frac{X}{n}}, \text{ unde:}$$

X – numărul mediu de cazuri pentru o unitate de studiu.

De exemplu: Unei cercetări ce avea drept scop studierea sănătății muncitorilor au fost supuși 800 de angajați ai întreprinderii Y. Pe parcursul unui an la ei au fost înregistrate 940 de cazuri de incapacitate temporară de muncă. De asemenea, inițial se calculează indicele intensiv:

$$P = \frac{940}{800} \times 1000 = 1175,00\%$$

1 175 1 000 angajați
1,175 1 angajat

$$ep = \pm \sqrt{\frac{1,175}{800}} = 0,04$$

0,04 1 angajat
4,0 100 angajați

Indicele incapacității temporare de muncă = $1\ 175 \pm 4,0\%$. Similar, dacă înmulțim eroarea cu 10, aflăm certitudinea indicatorului relativ. Metodele de testare a certitudinii valorilor relative și medii sunt utile atât pentru o analiză corectă a rezultatelor cercetărilor științifice, cât și pentru o evaluare a oricărui fenomen din activitatea curentă a organizatorilor ocrotirii sănătății.

1.8.3. EROAREA STANDARD A MEDIEI ARITMETICE

Eroarea medie a mediei aritmetice arată care este gradul de eroare a rezultatelor obținute într-o cercetare parțială față de rezultatele pe care cercetătorul le-ar fi obținut într-o cercetare integrală. Acest grad de eroare poate fi calculat cu ajutorul erorii medii care pentru media aritmetică are următoarea formulă:

$$e_m = \pm \frac{\sigma}{\sqrt{n}}$$

Nu se permite publicarea rezultatului fără a prezenta eroarea lui.

De exemplu: Să calculăm eroarea medie a unui eșantion privind lungimea la 262 de nou-născuți care au: $n = 262$; $M = 50,33$ cm; $\sigma = \pm 1,37$ cm.

$$eM = \pm \frac{1,37}{\sqrt{262}} = \pm 0,08$$

Indicele erorii medii ne arată că media absolută a înălțimii n/n observați ar fi cuprinsă între media practică și plus sau minus eroarea ei medie:

$M = 50,33 \pm 0,08$, adică între 50,25 și 50,47 cm.

Pentru o mai mare exactitate se calculează eroarea medie dublă sau triplă.

$$2eM = \pm 0,08 \times 2 = \pm 0,16.$$

În acest caz probabilitatea că media absolută se află între 50,17 și 50,49 este de 95,45%.

$$3eM = \pm 0,08 \times 3 = \pm 0,24.$$

Probabilitatea că media absolută se află între 50,09 și 50,57 este de 99,73%.

Ca și în cazul deviației standard (σ) și eroarea medie se calculează în cifre absolute: cm, grame, litri, indici etc. Pentru a putea compara eroarea medie la colectivități se calculează valoarea procentuală a erorii medii, care permite să aflăm certitudinea ei.

$$e_m \% = \pm \frac{e_m}{M} \times 100$$

- O proporție sub 1% a e_m % indică faptul că media aritmetică poate fi luată în considerare cu o certitudine foarte înaltă. Cu cât proporția va fi mai mică ca 1%, cu atât certitudinea va fi mai mare.
- O proporție mai mare de 3% indică faptul că media aritmetică nu poate fi luată în considerare sau numărul de observări nu-i suficient de mare.

1.8.4. INTERVALUL DE ÎNCREDERE

Dacă o repartiție de selecție este mai mare de 30 de unități, 68,27% din cazuri se vor găsi în intervalul dintre medie $\pm 1\sigma$; 95,45% – între medie $\pm 2\sigma$ și 99,73% – între medie $\pm 3\sigma$.

În același fel și media poate fluctua între aceste intervale, în cazul acesta calculându-se eroarea medie a mediei „ eM ”, deci media se poate găsi în intervalul „ $M \pm eM$ ” sau „ $M \pm 2eM$ ”, sau „ $M \pm 3eM$ ” cu probabilitatea respectivă de 68,27%, 95,45% și 99,73%. Intervalele de fluctuație a valorilor relative vor fi: „ $P \pm eP$ ” sau „ $P \pm 2eP$ ”, sau „ $P \pm 3eP$ ” pentru aceleași probabilități.

Intervalele „ $M \pm eM$ ” și „ $P \pm eP$ ” se numesc **interval de încredere** (sau regiuni de acceptare) a mediei sau indicatorului relativ, iar limitele lor se numesc limite de încredere. Peste limita de încredere vorbim de intervalul de respingere.

Unei probabilități de 95% sau 99% (sau 0,95 respectiv 0,99) denumită nivel de încredere, îi corespunde un coeficient de încredere sau *valoare critică*, notată „ z ”. Aceste corespondente sunt redate în *tabelul 8*.

Tabelul 8

Dependența valorii critice „ z ” de nivelul de încredere a prognosticului cert					
Nivelul de încredere	99,73%	99%	95,45%	95%	90%
Valoarea critică (z)	3,00	2,58	2,00	1,96	1,65

Deci, dacă media de selecție este „ M ”, atunci, cu un nivel de încredere de 95% sau 99%, ea se găsește în intervalul de încredere „ $M \pm 1,96 eM$ ”, respectiv „ $M \pm 2,58eM$ ”.

1.8.5. DIFERENȚĂ SEMNIFICATIVĂ DINTRE DOUĂ VALORI MEDII SAU RELATIVE

Testarea diferenței semnificative dintre două sau mai multe valori (medii sau relative) este o parte componentă a ultimei etape de cercetare – sinteza și analiza rezultatelor obținute.

Pentru testarea diferenței semnificative între valorile relative sau mediile obținute în loturile experimentale și martor se utilizează testul t-Student.

$$t = \frac{P_1 - P_2}{\sqrt{ep_1^2 + ep_2^2}} \text{ sau } t = \frac{M_1 - M_2}{\sqrt{em_1^2 + em_2^2}}, \text{ unde:}$$

P_1 și P_2 – indicatorii relativi din lotul experimental și lotul martor;
 ep_1 și ep_2 – erorile medii ale indicatorilor supuși comparației; și, respectiv:
 M_1 și M_2 – mediile aritmetice din lotul experimental și lotul martor;
 em_1 și em_2 – erorile medii ale mediilor aritmetice supuse comparației.

- Dacă nr. de observații depășește 120 și valoarea testului „ $t \geq 1,96$ ”, atunci între valorile comparate există o diferență semnificativă, care se apreciază cu o probabilitate de 95% ($p < 0,05$).
- Dacă „ $t \geq 3$ ”, atunci între valorile comparate, de asemenea, există o diferență semnificativă care se apreciază cu o probabilitate de 99% ($p < 0,01$).
- Dacă „ $t < 1,96$ ” – între valorile comparate nu există diferență semnificativă, deci această diferență poate fi considerată întâmplătoare.

Pentru un număr mai mic de observații, pentru a testa diferența semnificativă după valoarea testului „ t ” este nevoie să calculăm gradul de libertate după formula: $\gamma = n_1 + n_2 - 2$, unde: n_1 și n_2 – numerele de observații în loturile care se compară.

Utilizând tabelul valorilor lui „ t ” (tabelul 9) în funcție de numărul gradelor de libertate și valoarea reală a testului „ t ”, obținem gradul de probabilitate a diferenței dintre două sau mai multe valori.

Deci, dacă $n_1 + n_2 \leq 120$, atunci, în funcție de gradul de libertate „ γ ”, se compară valorile lui t_{calculat} și t_{tabelar} :

$t_{\text{calculat}} \geq t_{\text{tabelar}}$ – diferență statistică semnificativă;

$t_{\text{calculat}} < t_{\text{tabelar}}$ – diferență statistică nesemnificativă.

Metodele de testare a certitudinii valorilor relative și medii sunt utile atât pentru o analiză corectă a rezultatelor cercetării, cât și pentru o evaluare a oricărui fenomen din activitatea curentă a organizatorilor ocrotirii sănătății.

1.8.6. COMPARAREA VALORILOR ABSOLUTE SAU A DISTRIBUȚIILOR DE FRECVENȚĂ – TESTUL χ^2

Când avem de comparat între ele valori absolute sau distribuții de frecvențe, folosim **testul χ^2 (chi pătrat) al lui Pearson**. Testul χ^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 este:

$$\chi^2 = \sum \frac{(FT - FO)^2}{FT}, \text{ unde:}$$

χ^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.

Tabelul 9

Valorile lui „t” în funcție de nivelul de probabilitate și numărul gradelor de libertate

γ	p	Valorile lui „t” în funcție de nivelul de probabilitate și numărul gradelor de libertate			
		0,05 (95%)	0,02 (98%)	0,01 (99%)	0,001 (99,9%)
1		12,706	31,821	63,657	636,619
2		4,303	6,96,	9,925	31,598
3		3,182	4,541	5,841	12,941
4		2,776	3,747	4,614	8,610
5		2,571	3,365	4,032	6,859
6		2,447	3,143	3,707	5,959
7		2,365	2,998	3,499	5,405
8		2,306	2,896	3,355	5,041
9		2,262	2,821	3,250	4,781
10		2,228	2,764	3,169	4,587
11		2,201	2,718	3,106	4,437
12		2,179	2,681	3,005	4,318
13		2,160	2,650	3,012	4,221
14		2,145	2,624	2,977	4,140
15		2,131	2,602	2,947	4,073
16		2,120	2,583	2,921	4,015
17		2,110	2,567	2,898	3,995
18		2,101	2,550	2,878	3,992
19		2,093	2,539	2,861	3,883
20		2,086	2,528	2,845	3,850
21		2,080	2,518	2,831	3,819
22		2,074	2,508	2,819	3,792
23		2,069	2,500	2,807	3,767
24		2,064	2,492	2,797	3,745
25		2,060	2,485	2,787	3,725
26		2,056	2,479	2,779	3,707
27		2,052	2,473	2,771	3,690
28		2,048	2,467	2,763	3,674
29		2,042	2,457	2,756	3,646
30		2,042	2,457	2,756	3,646
40		2,021	2,423	2,704	3,551
60		2,000	2,390	2,660	3,460
120		1,980	2,358	2,617	2,373
∞		1,960	2,326	2,576	3,291

În situația în care avem o distribuție a colectivității **generale**, pe care o comparăm cu **distribuția unui eșantion**, introducem datele respective în formulă și calculăm valoarea lui χ^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 χ^2 , obținută pe baza formulei de mai sus, se numește valoarea lui „ χ^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 χ^2 tabelar. Această valoare o luăm din **tabelul testului χ^2 (tabelul 10)**, care este asemănător cu tabelul testului t. Valoarea lui χ^2 tabelar se citește în tabelul testului χ^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 gradul de libertate în tabelul testului χ^2 înțelegem produsul dintre numărul rândurilor tabelului (în care am introdus distribuțiile de frecvență) minus 1 și numărul coloanelor acestui tabel minus 1.

$$GL = (Nr.R-1) - (Nr.C-1)$$

Tabelul 10

Pragul de semnificație a diferitor valori pentru 1 – 100 grade de libertate			
Grad de libertate	Pragul 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	23,7	38,9	46,8
22	33,9	40,3	48,3
23	35,2	41,6	49,7
24	36,4	43,6	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

Exemplul 1. Examinând întreaga populație adultă a unei colectivități de 1 337 locuitori sub aspectul valorilor tensiunii arteriale sistolice și a unui eșantion de 120 de muncitori, am obținut următoarele distribuții de frecvențe pe grupe de valori ale tensiunii arteriale sistolice (tabelul 11).

Tabelul 11

Distribuțiile de frecvențe pe grupe de valori ale tensiunii arteriale sistolice

x_i = variabila TA sistolică	f_0 = frecvențe observate	f_i = 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
	$\Sigma=120$	$\Sigma=1\ 337$

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 χ^2 tabelar o vom citi din tabelul χ^2 la gradul de libertate: $GL = (\text{nr. de rânduri} - 1) \times (\text{nr. de coloane} - 1) = (7 - 1) \times (2 - 1) = 6$. Valoarea tabelară a lui χ^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 $\alpha = 0,1\%$ este de 22,5. 91% χ^2 calculat se obține:

$$\chi^2 = \frac{(180-6)^2}{180} + \frac{(210-17)^2}{210} + \frac{(290-31)^2}{290} + \frac{(320-42)^2}{320} + \frac{(230-17)^2}{230} + \frac{(65-5)^2}{65} + \frac{(42-2)^2}{42} + \frac{30276}{180} + \frac{37249}{210} + \frac{67081}{290} + \frac{77284}{320} + \frac{45369}{230} + \frac{3600}{65} + \frac{1600}{42} = 168,2 + 177,3 + 231,3 + 241,5 + 197,2 + 55,3 + 38 = 1108,8$$

Deci, valoarea lui χ^2 calculat (1 108,8) este mult mai mare decât valorile găsite în tabelul χ^2 pentru oricare din cele trei riscuri asumate (5%, 1%, 0,1%): deci, diferența dintre cele două distribuții de frecvențe teoretice și observate este semnifica-

tivă 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 χ^2 calculat ar fi fost mai mică decât valoarea lui χ^2 tabelar, atunci diferența dintre cele două distribuții de frecvențe ar fi fost ne semnificativă din punct de vedere statistic, eșantionul provenind din aceeași colectivitate generală, reflectând condițiile acesteia.

Exemplul 2. Dacă am stabilit că în teritoriul deservit de spitalul R populația adultă (39 620 de locuitori) se distribuie în raport cu dinții extrași după modelul prezentat în *tabelul 12* 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 χ^2 tabelar o vom citi din tabelul testului χ^2 la gradul de libertate 5. $GL = (6-1) \times (2-1) = 5$

Tabelul 12

Distribuția persoanelor din teritoriul deservit de spitalul R în raport cu dinții extrași		
Nr. de dinți absenți	Nr. pers. cot. gen.	Nr. pers. eșantion
X	FT	FO
1	2 850	860
2	5 250	920
3	10 320	1 130
4	15 810	1 620
5	3 860	650
6	1 530	320

La acest grad de libertate, valoarea lui χ^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 χ^2 calculat o obținem prin calcul. Introducând datele în formula lui χ^2 calculat, avem:

$$\chi^2 = (2\ 850-860)^2:2\ 850 + (5\ 250-920)^2:5\ 250 + (10\ 320-1\ 130)^2:10\ 320 + (15\ 810-1\ 620)^2:15\ 810 + (3\ 860-650)^2:3\ 860 + (1\ 530-320)^2:1\ 530 = 1\ 990^2:2\ 850 + 4\ 330^2:5\ 250 + 9\ 190^2:10\ 320 + 14\ 190^2:15\ 810 + 3\ 210^2:3\ 860 + 1\ 210^2:1\ 530 = 3\ 960\ 100:2\ 850 + 18\ 748\ 900:5\ 250 + 84\ 456\ 100:10\ 320 + 201\ 356\ 100:15\ 810 + 10\ 304\ 100:3\ 860 + 1\ 464\ 100:1\ 530 = 1\ 389 + 3\ 571 + 8\ 183 + 12\ 735 + 2\ 669 + 956 = 29\ 503.$$

Deci χ^2 calculat are o valoare de 29 503.

Interpretarea diferenței dintre cele două distribuții de frecvențe (a colectivității generale și a eșantionului nostru) se face comparând cele două valori ale lui χ^2 (obținute prin calcul și din tabele). Dacă valoarea lui χ^2 calculat este mai mare decât a lui χ^2 tabelar, atunci între cele două distribuții de frecvențe este o diferență semnificativă din punct de vedere statistic. Dacă, din contră, valoarea lui χ^2 calculat este mai mică decât valoarea lui χ^2 tabelar, diferența dintre distribuții este ne semnificativă, în exemplul nostru χ^2 calculat, având valoarea de 38 116, adică mai mare decât a lui χ^2 tabelar, testează o diferență semnificativă din punct de vedere statistic dintre 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 oncologică 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 de 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 tabelul 12 a.

Tabelul 12 a

Repartizarea bolnavilor de cancer după sex între fumători și nefumători					
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

Adăunțând de la început că între fumători și nefumători nu există diferențe statistice, deci admitând ipoteza nulă (H_0), în baza regulii a 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 \times fumători

$$X = \frac{60 \times 45}{78} = 39,8 (A)$$

Similar vom obține prin calcul și celelalte valori (B, C, D), pe care le vom introduce în tabelul corespunzător frecvențelor teoretice. Disponând în acest fel atât de valorile frecvențelor observate, cât și de frecvențele teoretice, putem obține prin formula menționată anterior:

$$\chi_c^2 = \sum \frac{(F.O - F.T)^2}{F.T} \quad \text{valoarea lui } \chi^2 \text{ calculat astfel:}$$

$$\chi_c^2 = \sum \frac{(39 - 39,8)^2}{39,8} + \frac{(30 - 29,2)^2}{29,2} + \frac{(6 - 5,2)^2}{5,2} + \frac{(3 - 3,8)^2}{3,8} =$$

$$\frac{0,64}{39,8} + \frac{0,64}{29,2} + \frac{0,64}{5,2} + \frac{0,64}{3,8} = 0,016 + 0,0219 + 0,123 + 0,168 + 0,3289 = 0,33$$

deci χ^2 calculat = 0,33.

Pentru a vedea dacă diferența dintre bărbații și femeile fumătoare sau dacă între bărbații și femeile nefumătoare este sau nu o diferență semnificativă din punct de vedere statistic, comparăm valorile $\chi_c^2=0,33$ cu valoarea χ_t^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 χ^2 calculat este mai mic decât χ^2 tabelar pentru oricare din riscurile α de mai sus asumate, deci diferența este ne semnificativă statistic. Dacă χ^2 calculat

ar fi fost mai mare decât χ^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 eșantion de 810 elevi (470 băieți și 340 fete) dintr-o colectivitate școlară, am constatat următoarele:

=> dintre cei 470 de băieți examinați 250 prezentau carii dentare, iar 220 erau fără carii;

=> dintre 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 contra sexul nu are nici o influență în determinarea ei.

Pentru a putea calcula valoarea lui χ^2 avem nevoie de frecvențele teoretice. Cum le obținem? Introducem datele cunoscute, deci frecvențele observate, într-un tabel cvadruplu (tabelul 13).

Tabelul 13

Repartizarea elevilor după sex în dependență de prezența sau absența cariei dentare

Lotul de elevi	Elevi cu carie		Elevi fără carie		TOTAL ELEVI
	FO	FT	FO	FT	
Băieți	250	249,50	220	220,50	470
Fete	180	180,50	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 a trei simple obținem frecvențele teoretice, făcând raționamentul: dacă la cei 810 elevi examinați corespund 430 de elevi cu carii, dintre cei 470 de 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 dintre 810 elevi au fost găsiți 380 de elevi fără carii, dintre cei 470 de băieți câți vor fi fără carie?

$$X = 470 \times 380 : 810 = 220,50$$

Același raționament îl facem și pentru obținerea frecvențelor teoretice ale fetelor cu și fără carie:

$$X = 340 \times 430 : 810 = 180,50$$

Deci, frecvențele teoretice ale fetelor cu carie – 180,50, le introducem în căsuța corespunzătoare a tabelului:

$$X = 340 \times 380 : 810 = 159,50$$

De asemenea și frecvențele teoretice ale fetelor fără carie le introducem în căsuța corespunzătoare a tabelului nostru.

Având acum frecvențele teoretice și pe cele observate, le introducem în formula lui χ^2 și obținem valoarea lui χ^2 calculat:

$$\chi^2 = (249,50-250):249,50 + (220,50-220):220,50 + (180,50-180):180,50 + (159,50-160) : 159,50 = (-0,50) : 249,50 + 0,50:220,50 + 0,50:180,50 + (-0,50) : 159,50 = 0,25:249,50 + 0,25:220,50 + 0,25 : 180,50 + 0,25 : 159,50 = 0,0010 + 0,0011 + 0,0015 = 0,0049, \text{ deci } \chi^2 \text{ calculat} = 0,0049.$$

Având în vedere că datele noastre le-am înscris într-un tabel cu două rânduri și cu două coloane, χ^2 tabelar îl vom citi din tabelul testului χ^2 la gradul de libertate 1 și el 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%). χ^2 calculat având o valoare mult mai mică (0,0049) decât valoarea lui χ^2 tabelar (3,84; 6,63; 10,80) testează lipsa de semnificație a diferenței dintre frecvența cariei dentare la băieți și la fete.

Prin urmare, 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 χ^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:

$$\chi^2 = \frac{(ad - bc)^2}{n_1 n_2 n_3 n_4} \times N(-1)$$

Această formulă rezultă din desfacerea binomului lui Newton $(a+b)^n$.

Exemplul 5. Din exemplul 3 putem repartiza datele în tabelul 14.

Tabelul 14

Repartizarea bolnavilor
de cancer pulmonar după sex între fumători și nefumători

Cancer pulmonar	Fumători	Nefumători	Total
Bărbați	39=a	30=b	69=a+b=n ₁
Femei	6=c	3=d	9=c+d=n ₂
Total	45=a+c=n ₃	33=b+d=n ₄	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:

$$\begin{aligned} \chi_c^2 &= \frac{(a \times d - b \times c)^2 \times N(-1)}{n_1 \times n_2 \times n_3 \times n_4} = \frac{(39 \times 3 - 30 \times 6)^2 \times (78 - 1)}{69 \times 9 \times 45 \times 33} = \\ &= \frac{(117 - 180)^2 \times 77}{922185} = \frac{2045813}{922185} = 2,21 \end{aligned}$$

χ^2 pentru un risc $\alpha=5\%$ și pentru GL=(numărul de rânduri-1) x (număr de coloane - 1)=(2-1) x (2-1)=1 are valoarea de 3,8. Valoarea calculată este mai mică

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 ₁ =20
Tratament cu streptomycină	c=16	d=4	c+d=n ₂ =20
Total	a+c=n ₃ =30	b+d=n ₄ =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