VII. STATISTICĂ

7.1. INDICATORII TENDINŢEI CENTRALE

7.1.1. Mărimile medii

***Media** nivelurilor individuale ale unei variabile (caracteristici) statistice este expresia sintetizării într-un singur nivel reprezentativ a tot ceea ce este esenţial, tipic şi obiectiv în apariţia, manifestarea şi dezvoltarea acesteia.*

7.1.1.1 Media aritmetică

Media aritmetică se foloseşte atunci când fenomenul supus cercetării înregistrează modificări aproximativ constante, în progresie aritmetică, prezentând, deci, o tendinţă liniară.

Media aritmetică simplă se foloseşte pentru seriile simple, adică în cazul în care numărul variantelor caracteristicii studiate este egal cu numărul unităţilor sau când se cunoaşte nivelul totalizat al caracteristicii şi numărul unităţilor. Pentru o caracteristică statistică *X*, cu valorile $x_1, x_2, \ldots, x_n$, şi ţinând cont că funcţia determinantă pentru media aritmetică simplă este de tip adiţional, adică: $x_1 + x_2 + \ldots + x_n = \sum x_i$,

$$\overline{x} = \frac{\sum x_i}{n};\ i = \overline{1,n}.$$

Media aritmetică ponderată este întâlnită în cazul seriilor de distribuţie, când unele variante ale caracteristicii se înregistrează de mai multe ori. Dacă fiecare variantă x_i a caracteristicii are o frecvenţă de apariţie f_i în colectivitate, atunci suma simplă este înlocuită cu suma produsului $x_i \cdot f_i$, rezultând:

$$\left.\begin{aligned} x_1 \cdot f_1 + x_2 \cdot f_2 + \ldots + x_n \cdot f_n &= \sum x_i \cdot f_i \\ \overline{x} \cdot f_1 + \overline{x} \cdot f_2 + \ldots + \overline{x} \cdot f_n &= \overline{x} \cdot \sum f_i \end{aligned}\right\} \Rightarrow\ \overline{x} \cdot \sum f_i = \sum x_i \cdot f_i \Rightarrow\ \overline{x} = \frac{\sum x_i \cdot f_i}{\sum f_i};\ i = \overline{1,n}.$$

Proprietăţile mediei aritmetice

- Media aritmetică este cuprinsă între varianta minimă şi varianta maximă, adică: $x_{min} < \overline{x} < x_{max}$;
- Suma abaterilor variantelor caracteristicii de la media lor este egală cu zero:
- Dacă dintr-o serie X ($x_1, x_2, \ldots x_n$) construim seria X^* prin adăugarea sau scăderea unei constante **a** ($x_1 \pm a, x_2 \pm a, \ldots x_n \pm a$), atunci media seriei X^* va fi: $\overline{x}^* = \overline{x} \pm a$;
- Dacă dintr-o serie X ($x_1, x_2, \ldots x_n$) construim seria X^* prin mărirea sau micşorarea de k ori $\left(x_i \cdot k \ sau \ \frac{x_i}{k} \right)$, atunci media seriei X^* se va mări sau micşora de k ori: $\overline{x}^* = \overline{x} \cdot k \ sau \ \overline{x}^* = \frac{\overline{x}}{k}$;

Combinând ultimele două proprietăţi, se obţine **formula de calcul simplificat a mediei aritmetice**:

$$\overline{x} = \frac{\sum \frac{x_i - a}{k} \cdot f_i}{\sum f_i} \cdot k + a.$$

Media aritmetică a variabilei alternative

Variabila alternativă sau binară, cunoscută şi sub denumirea de variabilă aleatoare a lui Bernoulli, admite doar două variante posibile, variante care se exclud reciproc. În realitate există diverse astfel de situaţii: admis / respins (candidaţii la un concurs), rebut / nonrebut (piesele realizate într-o întreprindere), calificat / necalificat (sportivii într-o anumită competiţie) etc. Deci, avem două situaţii ce nu pot apărea concomitent (un candidat ori este admis ori este respins, nu poate să fie în acelaşi timp şi admis, şi respins).

Pentru prelucrarea şi analiza statistică se consideră următoarele convenţii şi notaţii:

- situaţiilor corespunzătoare răspunsurilor afirmative, cele care constituie varianta x_1, li se atribuie cifra **1**, având frecvenţa absolută f_1 şi frecvenţa relativă p;
- situaţiilor corespunzătoare răspunsurilor negative, cele care constituie varianta x_2, li se atribuie cifra **0**, având frecvenţa absolută f_2 şi frecvenţa relativă q.

Astfel, dacă vom însuma frecvenţele absolute f_1 şi f_2 vom obţine volumul colectivităţii generale. În plus, cunoscând modul de determinare al frecvenţelor relative, rezultă că:

p + q = 1 rezultă p = 1 – q şi q = 1 – p.

Media aritmetică în acest caz va fi:

$$\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2}{f_1 + f_2} = x_1 \cdot \frac{f_1}{f_1 + f_2} + x_2 \cdot \frac{f_2}{f_1 + f_2} = 1 \cdot p + 0 \cdot q = p$$

$$\bar{x} = p.$$

7.1.1.2. Media armonică

Media armonică se determină doar pentru variabile cantitative şi se aplică numai în cazuri speciale. În general, utilizarea acestui tip de medie este recomandat atunci când două variabile interdependente se află în raport de inversă proporţionalitate.

Media armonică are, în principiu, aceeaşi metodologie de calcul ca media aritmetică, funcţia determinantă fiind tot de tip adiţional; deosebirea constă în aceea că nu se folosesc variantele x_1, x_2, …, x_n, ci inversul acestora, adică $\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_n}$.

Media armonică simplă este specifică seriilor simple, determinându-se astfel:

$$\left.\begin{aligned} \frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n} &= \sum \frac{1}{x_i} \\ \frac{1}{\bar{x}_h} + \frac{1}{\bar{x}_h} + \ldots + \frac{1}{\bar{x}_h} &= \frac{n}{\bar{x}_h} \end{aligned}\right\} \Rightarrow \frac{n}{\bar{x}_h} = \sum \frac{1}{x_i} \Rightarrow \bar{x}_h = \frac{n}{\sum \frac{1}{x_i}}$$

Media armonică ponderată se utilizează în cazul seriilor de frecvenţe, determinându-se astfel:

$$\left.\begin{aligned} \frac{1}{x_1} \cdot f_1 + \frac{1}{x_2} \cdot f_2 + \ldots + \frac{1}{x_n} \cdot f_n &= \sum \frac{1}{x_i} \cdot f_i \\ \frac{1}{\bar{x}_h} \cdot f_1 + \frac{1}{\bar{x}_h} \cdot f_2 + \ldots + \frac{1}{\bar{x}_h} \cdot f_n &= \frac{\sum f_i}{\bar{x}_h} \end{aligned}\right\} \Rightarrow \frac{\sum f_i}{\bar{x}_h} = \sum \frac{1}{x_i} \cdot f_i \Rightarrow \bar{x}_h = \frac{\sum f_i}{\sum \frac{1}{x_i} \cdot f_i}$$

Observăm că $\bar{x}_h < \bar{x}$.

Frecvent utilizată este forma transformată a mediei aritmetice ponderate, care ia forma unei medii armonice cu ponderi compuse. Se foloseşte atunci când nu se cunosc frecvenţele

În cazul ***mediei armonice ca formă transformată*** a mediei aritmetice ponderate, relaţiile de calcul se obţin prin substituirea frecvenţelor din numitorul relaţiei mediei aritmetice ponderate astfel $f_i = \frac{1}{x_i} \cdot x_i f_i$, datorită faptului că x_i şi $x_i f_i$ sunt cunoscute.

Dacă $x_i f_i$ sunt egale ($x_1 f_1 = x_2 f_2 = \ldots = x_n f_n$), se obţine ***media armonică simplă***:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum x_i f_i}{\sum \frac{1}{x_i} x_i f_i} = \frac{n \cdot x_i f_i}{x_i f_i \cdot \sum \frac{1}{x_i}} = \frac{n}{\sum \frac{1}{x_i}} = \bar{x}_h$$

Dacă $x_i f_i$ sunt diferite ($x_1 f_1 \neq x_2 f_2 \neq \ldots \neq x_n f_n$), se obţine ***media armonică ponderată***:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum x_i f_i}{\sum \frac{1}{x_i} x_i f_i} = \bar{x}_h$$

7.1.1.3. Media pătratică

Media pătratică se foloseşte în cazul în care fenomenele înregistrează creşteri, aproximativ, în progresie exponenţială, adică atunci când creşterea este mai lentă la începutul seriei şi din ce în ce mai pronunţată spre sfârşitul acesteia, fiind utilizată, deci, în analiza tendinţelor neliniare, de tip exponenţial. Este folosită şi ca model matematic în calculul indicatorilor sintetici ai variaţiei (abaterea standard).

Media pătratică se determină în mod asemănător mediei aritmetice, funcţia determinantă fiind tot de tip adiţional, cu deosebirea că, în cazul mediei pătratice, se foloseşte pătratul caracteristicii.

Media pătratică simplă este utilizată pentru seriile simple şi se determină astfel:

$$\left.\begin{aligned} x_1^2+x_2^2+\ldots+x_n^2&=\sum x_i^2\\ \bar{x}_p^2+\bar{x}_p^2+\ldots+\bar{x}_p^2&=n\cdot\bar{x}_p^2\end{aligned}\right\}\Rightarrow n\cdot\bar{x}_p^2=\sum x_i^2\Rightarrow\bar{x}_p=\sqrt{\frac{\sum x_i^2}{n}}.$$

Media pătratică ponderată se utilizează pentru seriile de frecvenţe, obţinându-se astfel:

$$\left.\begin{aligned} x_1^2\cdot f_1+x_2^2\cdot f_2+\ldots+x_n^2\cdot f_n&=\sum x_i^2\cdot f_i\\ \bar{x}_p^2\cdot f_1+\bar{x}_p^2\cdot f_2+\ldots+\bar{x}_p^2\cdot f_n&=\bar{x}_p^2\cdot\sum f_i\end{aligned}\right\}\Rightarrow\bar{x}_p^2\cdot\sum f_i=\sum x_i^2\cdot f_i\Rightarrow\bar{x}_p=\sqrt{\frac{\sum x_i^2\cdot f_i}{\sum f_i}}.$$

Dacă pentru aceeaşi serie se calculează media aritmetică şi media pătratică, întotdeauna: $\bar{x}<\bar{x}_p$.Acesta este şi motivul pentru care această medie este indicată pentru analiza fenomenelor ce înregistrează tendinţe exponenţiale.

7.1.1.4. Media geometrică

Media geometrică se foloseşte în cazurile în care fenomenele înregistrează modificări, aproximativ, în progresie geometrică. Se utilizează mai frecvent în situaţia în care diferenţele dintre variantele caracteristicii sunt mai mari la începutul seriei şi din ce în ce mai mici către sfârşitul acesteia. Rezultă că, media geometrică este recomandată pentru analiza tendinţelor neliniare care evidenţiază creşteri la început şi o atenuare a acestora spre sfârşitul seriei.

Este folosită ca model matematic în calculul unuia dintre indicatorii sintetici ai seriilor cronologice (indicele mediu al dinamicii).

În cazul mediei geometrice funcţia determinantă este de tipul produsului.

Media geometrică simplă este specifică seriilor simple, determinându-se astfel:

$$\left.\begin{aligned} x_1\cdot x_2\cdot\ldots\cdot x_n&=\Pi x_i\\ \bar{x}_g\cdot\bar{x}_g\cdot\ldots\cdot\bar{x}_g&=\bar{x}_g^n\end{aligned}\right\}\Rightarrow\ \bar{x}_g^n=\Pi x_i\Rightarrow\ \bar{x}_g=\sqrt[n]{\Pi x_i}.$$

Media geometrică ponderată se determină pentru seriile de frecvenţe, astfel:

$$\left.\begin{aligned} x_1^{f_1}\cdot x_2^{f_2}\cdot\ldots\cdot x_n^{f_n}&=\Pi x_i^{f_i}\\ \bar{x}_g^{f_1}\cdot\bar{x}_g^{f_2}\cdot\ldots\cdot\bar{x}_g^{f_n}&=\bar{x}_g^{\sum f_i}\end{aligned}\right\}\Rightarrow\ \bar{x}_g^{\sum f_i}=\Pi x_i^{f_i}\Rightarrow\ \bar{x}_g=\sqrt[\sum f_i]{\Pi x_i^{f_i}}.$$

Dacă pentru aceleaşi date se calculează media aritmetică, pătratică şi geometrică, întotdeauna:

$$\bar{x}_g<\bar{x}<\bar{x}_p.$$

Din acest motiv media geometrică este recomandată pentru analiza seriilor în cadrul cărora se manifestă tendinţe de reducere a ritmului de creştere.

Dacă pentru aceeaşi serie de date calculăm cele patru tipuri de medie prezentate, între ele există următoarea relaţie de ordine (aşa cum rezultă şi din figura 3.1.):

$$\bar{x}_h\le\bar{x}_g\le\bar{x}_a\le\bar{x}_p.$$

Egalitatea dintre medii are loc numai atunci când valorile din cadrul seriei sunt constante.

7.1.2. Cuantilele

***Cuantilele** sunt indicatori de poziţie care împart seria de distribuţie într-un anumit număr de părţi cu efective egale.*

Fie *n* volumul unităţilor statistice analizate şi $z=\frac{k}{n}$ un număr raţional ($z\in(0,1)$, deci $k<n$). Se numeşte cuantila de ordinul *z*, valoarea x_z a variabilei aleatoare *X*, cu proprietatea: $F_n(x_z) = z$,

unde $F_n(x_z)$ este funcţia empirică de repartiţie (funcţia frecvenţelor relative cumulate). În mod uzual, *z* are una din valorile:

- $z=\frac{1}{2}\Rightarrow$ cuantila $x_{\frac{1}{2}}=Me$ se numeşte ***mediană*** şi împarte seria de variaţie în două părţi de efective egale cu $\frac{n}{2}$;
- $z\in\left\{\frac{1}{4},\frac{2}{4},\frac{3}{4}\right\}\Rightarrow$ cuantilele $x_{\frac{1}{4}}=x_{Q_1},x_{\frac{2}{4}}=x_{Q_2},x_{\frac{3}{4}}=x_{Q_3}$ se numesc ***cuartile*** şi împart seria de variaţie în patru părţi de efective egale cu $\frac{n}{4}$;
- $z\in\left\{\frac{1}{10},\frac{2}{10},...,\frac{9}{10}\right\}\Rightarrow$ cuantilele $x_{\frac{1}{10}}=x_{D_1},x_{\frac{2}{10}}=x_{D_2},...,x_{\frac{9}{10}}=x_{D_9}$ se numesc ***decile*** şi împart seria de variaţie în zece părţi de efective egale cu $\frac{n}{10}$;
- $z\in\left\{\frac{1}{100},\frac{2}{100},...,\frac{99}{100}\right\}\Rightarrow$ cuantilele $x_{\frac{1}{100}}=x_{P_1},x_{\frac{2}{100}}=x_{P_2},...,x_{\frac{99}{100}}=x_{P_{99}}$ se numesc ***percentile*** şi împart seria de variaţie în o sută părţi de efective egale cu $\frac{n}{100}$.

7.1.2.1. Mediana

Mediana reprezintă acea valoare care împarte seria (ordonată crescător sau descrescător) în două părţi egale.

Cum seria de date trebuie să fie ordonată, rezultă că această măsură a tendinţei centrale nu poate fi definită decât pentru serii ale căror valori sunt mărimi cantitative sau ordinale, neavând sens pentru o caracteristică nominală. Metodologia de calcul a medianei diferă după cum seria este simplă sau de frecvenţe.

- **Pentru o serie simplă** vom parcurge etapele:
  - se ordonează crescător sau descrescător elementele seriei;
  - se calculează valoarea mediană într-una din următoarele două variante:
  - dacă seria are un număr impar de termeni, atunci: $Me=x_{\frac{n+1}{2}}$;
  - dacă seria este formată dintr-un număr par de termeni, atunci mediana este semisuma termenilor de rang $\frac{n}{2}$ şi $\frac{n}{2}+1$, adică: $Me=\frac{x_{\frac{n}{2}}+x_{\frac{n}{2}+1}}{2}$.
- **Pentru seriile de distribuţie** se deosebesc două posibilităţi de calcul:

Pentru o serie de distribuţie după variante, determinarea medianei presupune parcurgerea următoarelor etape:

- se determină frecvenţele cumulate crescător sau descrescător (Fc_i);
- determinăm unitatea mediană după relaţia: $U_{Me}=\frac{n}{2}$;

- stabilim mediana, care este egală cu prima valoare din cadrul seriei de valori pentru care:

$$U_{Me} \le Fc_i .$$

- ***Pentru o serie de distribuţie pe intervale***, determinarea medianei se face parcurgând etapele următoare:

- se determină frecvenţele cumulate crescător sau descrescător (Fc_i);

- determinăm unitatea mediană după relaţia: $U_{Me} = \frac{n}{2}$;

- se stabileşte intervalul median $I_{Me} = (x_{Me}^{inf}, x_{Me}^{sup})$, respectiv intervalul pentru care este respectată relaţia: $U_{Me} \le Fc_i$;

- se calculează mediana cu ajutorul relaţiei: $Me = x_{Me}^{inf} + \left(\frac{n}{2} - S_n \right) \cdot \frac{k}{f_{Me}}$,

unde: x_{Me}^{inf} – reprezintă limita inferioară a intervalului median;

S_n – reprezintă suma frecvenţelor care preced intervalul median;

k – mărimea intervalului în care se plasează median;

f_{Me} – frecvenţa intervalului median.

Această relaţie are la bază ipoteza că, în interiorul intervalului de variaţie unităţile statistice sunt uniform distribuite.

7.1.2.2 Cuartilele

Există trei cuartile $(x_{Q_1}, x_{Q_2}, x_{Q_3})$ care împart seria de distribuţie în patru părţi cu efective egale. Cele trei cuartile sunt: x_{Q_1} - cuartila inferioară, x_{Q_2} - mediana şi x_{Q_3} - cuartila superioară.

Metodologia determinării cuartilelor este asemănătoare celei a medianei. Metoda de calcul algebric a cuartilelor presupune parcurgerea următoarelor etape:

se stabileşte intervalul cuartilic I_{Q_h} corespunzător cuartilei x_{Q_h}. Acest interval conţine unitatea cuartilică U_{Q_h}, unitate care se obţine astfel: $U_{Q_h} = \frac{h \cdot n}{4}$, $h=1,2,3$;

se calculează cuartilele pe baza relaţiei: $x_{Q_h} = x_{Q_h}^{inf} + \left(\frac{h \cdot n}{4} - S_{Q_h - 1} \right) \cdot \frac{k}{f_{Q_h}}$,

unde: $x_{Q_h}^{inf}$ – reprezintă limita inferioară a intervalului în care se plasează cuartila x_{Q_h};

$S_{Q_h - 1}$ – reprezintă suma frecvenţelor care preced intervalul în care se plasează cuartila x_{Q_h}: $S_{Q_h - 1} = f_1 + ... + f_{Q_h - 1}$;

k – mărimea intervalului în care se plasează cuartila x_{Q_h};

f_{Q_h} – frecvenţa intervalului în care se plasează cuartila x_{Q_h}.

7.1.3 Modul

Modul (dominanta) reprezintă valoarea caracteristicii care are frecvenţa cea mai mare. Din această definiţie rezultă că modul este un indicator specific seriilor de distribuţie.

Determinarea modului poate fi făcută astfel:

- Dacă valorile intervalului modal sunt uniform repartizate, atunci modul se determină pe baza relaţiei:

$$Mo = x_{Mo}^{inf} + k \frac{\Delta_1}{\Delta_1 + \Delta_2},$$

unde: k – reprezintă mărimea intervalului modal;

Δ_1 – reprezintă diferenţa dintre frecvenţa maximă şi frecvenţa intervalului precedent:

$\Delta_1 = f_{Mo} - f_{Mo-1}$;

Δ_2 – reprezintă diferenţa dintre frecvenţa maximă şi frecvenţa intervalului următor:

$\Delta_2 = f_{Mo} - f_{Mo+1}$;

7.2 INDICATORII VARIAŢIEI

7.2.1. Indicatorii simpli ai variaţiei

Indicatorii simpli sunt folosiţi pentru caracterizarea gradului de împrăştiere a unităţilor colectivităţii cercetate faţă de medie sau faţă de o anumită valoare din serie. Se pot exprima atât în unităţi absolute, aceleaşi ca şi cele ale caracteristicii studiate, cât şi în mărimi relative, calculate în raport cu media. Aceşti indicatori sunt amplitudinea variaţiei şi abaterile individuale ale fiecărui termen de la media lor.

- **Amplitudinea variaţiei (*A*)**

Amplitudinea variaţiei oferă posibilitatea delimitării câmpului de variaţie a unui fenomen şi se prezintă sub două forme:

- ***amplitudinea absolută (A_a)*** – se obţine ca diferenţă între valoarea maximă (X_{max}) şi valoarea minimă (X_{min}) a seriei, adică: $A_a = X_{max} - X_{min}$.

În cazul unor serii de distribuţie pe intervale, amplitudinea se determină ca diferenţă între limita superioară a ultimului interval şi limita inferioară a primului interval;

- ***amplitudinea relativă (A_r)*** – se calculează ca raport între amplitudinea absolută şi media aritmetică, exprimându-se procentual, astfel: $A_r = \frac{A_a}{\bar{x}} \cdot 100 = \frac{X_{max} - X_{min}}{\bar{x}} \cdot 100$.

- **Abaterile individuale (d_i)**

Abaterile individuale permit cunoaşterea structurii variaţiei la nivelul fiecărei unităţi statistice. Se prezintă sub două forme:

- ***abaterile individuale absolute (da_i)*** – se calculează ca diferenţă între fiecare valoare înregistrată şi media aritmetică a seriei:

$$da_i = x_i - \bar{x} \rightarrow \begin{cases} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \dots \\ x_n - \bar{x} \end{cases}$$

- ***abaterile individuale relative (dr_i)*** – se calculează ca raport între abaterile individuale absolute şi media aritmetică a caracteristicii studiate, exprimându-se procentual, astfel:

$$dr_i = \frac{da_i}{\bar{x}} \cdot 100 = \frac{x_i - \bar{x}}{\bar{x}} \cdot 100.$$

Abaterile individuale pot fi negative sau pozitive în funcţie de mărimea fiecărui termen faţă de media lor. În analizele statistice se urmăresc în mod deosebit abaterea individuală minimă şi abaterea individuală maximă, calculate în cifre absolute şi relative astfel:

$$da_{max-} = x_{min} - \bar{x} \text{ sau } dr_{max-} = \frac{da_{max-}}{\bar{x}} \cdot 100.$$

$$da_{max+} = x_{max} - \bar{x} \text{ sau } dr_{max+} = \frac{da_{max+}}{\bar{x}} \cdot 100.$$

În cazul unei distribuţii simetrice $|da_{max+}| = |da_{max-}|$, iar în interiorul seriei la abateri egale dar de semne contrare, le corespund frecvenţe egale de apariţie. Aceasta conduce la compensarea pe total (la nivelul întregului ansamblu) a abaterilor individuale.

Pentru determinarea abaterilor individuale în locul mediei se folosesc, mai rar, şi ceilalţi indicatori ai tendinţei centrale (mediana, modul).

7.2.2. Indicatorii sintetici ai variaţiei

Indicatorii simpli ai variaţiei nu pot exprima şi caracteriza întreaga variaţie a caracteristicii studiate, fiind necesară calcularea indicatorilor sintetici. Aceşti indicatori caracterizează gradul de variaţie, luând în considerare toţi termenii seriei. Indicatorii sintetici sunt: abaterea medie liniară, dispersia, abaterea standard şi coeficientul de variaţie.

- **Abaterea medie liniară ($\bar{d}$)**

Abaterea medie liniară se calculează ca o medie aritmetică simplă sau ponderată a abaterilor absolute ale termenilor seriei de la media lor, luate sub formă de modul, astfel:

- $\bar{d} = \dfrac{\sum |x_i - \bar{x}|}{n}$ - pentru o serie simplă;
- $\bar{d} = \dfrac{\sum |x_i - \bar{x}| f_i}{\sum f_i}$ - pentru o serie de frecvenţe;

Abaterea medie liniară arată, în medie, cu cât se abat termenii seriei de la media lor. Prezintă dezavantajul că nu ţine seama de semnul algebric (abaterea fiind calculată în modul), acordând aceeaşi importanţă atât abaterilor pozitive cât şi abaterilor negative. Abaterea medie liniară poate fi un indicator concludent numai dacă seria prezintă un grad mare de omogenitate. Aceste neajunsuri se înlătură prin calculul dispersiei.

Abaterea medie liniară se calculează şi se analizează nu numai pentru seriile de distribuţie, ci şi pentru seriile cronologice sau teritoriale.

Se foloseşte la determinarea intervalului mediu de variaţie: $\bar{x} \pm \bar{d} \Rightarrow \begin{cases} \bar{x} + \bar{d} \\ \bar{x} - \bar{d} \end{cases}$

- **Dispersia (σ^2)**

Cunoscută şi sub denumirea de varianţă, **dispersia** se calculează ca o medie aritmetică simplă sau ponderată a pătratelor abaterilor termenilor seriei de la tendinţa lor centrală. Aceasta înseamnă că în calculul dispersiei poate fi luată în considerare media sau alt indicator al tendinţei centrale (mediana, modul).

Relaţiile de calcul ale dispersiei sunt următoarele:

- $\sigma^2 = \dfrac{\sum (x_i - \bar{x})^2}{n}$ - pentru o serie simplă;
- $\sigma^2 = \dfrac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}$ - pentru o serie de frecvenţe;

Dispersia este un indicator abstract, nu are formă concretă de exprimare şi arată modul în care valorile caracteristicii gravitează în jurul mediei. Măsoară variaţia totală a caracteristicii studiate datorită cauzelor esenţiale şi întâmplătoare. Este un indicator util în verificări de ipoteze statistice, în calculul altor indicatori statistici etc.

Proprietăţile dispersiei

- Dispersia calculată din abaterile variantelor x_i de la o constantă **a**, este mai mare decât dispersia reală cu pătratul diferenţei dintre medie şi constanta **a**, astfel:

$$\sigma^2 = \frac{\sum (x_i - a)^2 f_i}{\sum f_i} - (\bar{x} - a)^2 .$$

- Dispersia calculată din abaterile variantelor x_i de la media lor, micşorate în prealabil prin împărţire la o constantă **k**, este mai mică decât dispersia reală de **k^2** ori, astfel:

$$\sigma^2 = \frac{\sum \left(\frac{x_i - \bar{x}}{k} \right)^2 f_i}{\sum f_i} \cdot k^2 .$$

Din combinarea ultimelor două proprietăţi rezultă **relaţia de calcul simplificat a dispersiei**:

$$\sigma^2 = \frac{\sum \left(\frac{x_i - a}{k} \right)^2 f_i}{\sum f_i} \cdot k^2 - (\bar{x} - a)^2 .$$

Dispersia variabilei alternative

Se foloseşte relaţia de calcul obişnuit a dispersiei, introducându-se elementele specifice variabilei alternative. Vom folosi notaţiile şi convenţiile utilizate la media aritmetică pentru variabila alternativă. De asemenea, luăm în considerare şi rezultatul obţinut pentru media aritmetică, $\bar{x} = p$. Dispersia va fi:

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2}{f_1 + f_2} = \frac{(x_1 - \bar{x})^2 f_1}{f_1 + f_2} + \frac{(x_2 - \bar{x})^2 f_2}{f_1 + f_2} =$$

$$= (1 - p)^2 p + (0 - p)^2 q = pq(p + q)^2$$

$$\sigma^2 = pq .$$

- **Abaterea standard (σ)**

Denumită şi *abatere medie pătratică*, ***abaterea standard*** *se calculează ca o medie pătratică simplă sau ponderată a abaterilor valorilor seriei faţă de media lor, respectiv rădăcina pătrată din dispersie*:

- $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$ - pentru serii simple;
- $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}$ - pentru serii de frecvenţe.

Abaterea standard este indicatorul cel mai frecvent folosit pentru analiza variaţiei unei serii statistice. O serie de date prezintă o omogenitate mare dacă □este mic.

Abaterea standard a variabilei alternative

Abaterea standard pentru variabila alternativă este: $\sigma = \sqrt{\sigma^2} = \sqrt{pq}$.

- **Coeficientul de variaţie (*Cv*)**

Deoarece atât media, cât şi abaterea standard sunt indicatori exprimaţi în unităţi de măsură concrete, ei nu pot fi folosiţi pentru compararea a două serii de date exprimate în unităţi de măsură diferite. Spre exemplu, nu putem compara mediile şi abaterile standard calculate pentru două serii referitoare la vânzarea unor produse pe o piaţă, cu valori exprimate fizic, dacă aceste produse se exprimă în unităţi de măsură diferite. Pentru înlăturarea acestui inconvenient se calculează parametrul adimensional denumit *coeficient de variaţie*.

Coeficientul de variaţie, propus de Pearson, se calculează ca raport între abaterea standard şi nivelul mediu, adică: $Cv = \frac{\sigma}{\bar{x}} \cdot 100$.

Cum $Cv < 35\%$, rezultă că seria analizată prezintă un grad de omogenitate ridicat, iar indicatorii tendinţei centrale sunt reprezentativi pentru această serie.

7.2.3. Abaterile intercuantilice

O altă categorie de indicatori ai variaţiei o reprezintă abaterile intercuantilice. Aceste abateri pot fi definite pentru variabile cantitative sau ordinale.

Într-o serie perfect simetrică, cuantilele se distribuie în mod simetric în ambele sensuri faţă de valoarea tendinţei centrale a seriei, calculată ca valoare mediană. Calculând abaterile dintre valorile mediilor de poziţie şi valoarea mediană se poate interpreta tendinţa de distribuţie a frecvenţelor de repartiţie ale variantelor caracteristicii.

- ***Abaterea intercuartilică (Q_c)***

Abaterea intercuartilică se calculează ca o medie a celor două abateri ale cuartilelor extreme faţă de cuartila centrală:

$$Q_c = \frac{\left(Me - x_{Q_1}\right) + \left(x_{Q_3} - Me\right)}{2} = \frac{x_{Q_3} - x_{Q_1}}{2}.$$

Datorită faptului că se bazează numai pe relaţia dintre cele două cuartile extreme, abaterea intercuartilică s-ar mai putea numi şi *amplitudine semi-intercuartilică*.

Ca orice indicator absolut, şi abaterea intercuartilică se exprimă în unităţile de măsură ale caracteristicii studiate şi nu poate fi supusă direct comparaţiei statistice a mai multor serii. De aceea, se calculează ***coeficientul de variaţie intercuartilică***, ca raport între abaterea intercuantilică şi valoarea mediană, astfel:

$$V_Q = \frac{Q_c}{Me} \cdot 100 = \frac{x_{Q_3} - x_{Q_1}}{2Me} \cdot 100.$$

7.2.4 Dispersia în analiza distribuţiilor bidimensionale

Analiza variabilităţii în cazul distribuţiilor bidimensionale de frecvenţe este un proces mai complex ce necesită o atenţie suplimentară, întrucât variabilitatea, de această dată, este provocată de două categorii de factori: esenţiali şi întâmplători. Ca atare, variaţia trebuie descompusă pe cele două surse de factori care o generează, fiind necesar ca studiul acesteia pe întreaga colectivitate să fie completat cu studiul ei în cadrul fiecărei grupe şi între grupe.

Presupunem că avem două caracteristici X_i şi Y_j şi unităţile au fost împărţite în n grupe după variaţia lui X_i, obţinându-se următoarele distribuţii condiţionate de factorul de grupare:

Grupare după X	**Grupare după Y**						**Total f_x**	**Medii de grupă $\bar{y}_i$**	**Dispersii de grupă σ_i^2**
	y_1	**y_2**	**…**	**y_j**	**…**	**y_m**			
x_1	f_{11}	f_{12}	…	f_{1j}	…	f_{1m}	f_1	$\bar{y}_1$	σ_1^2
x_2	f_{21}	f_{22}	…	f_{2j}	…	f_{2m}	f_2	$\bar{y}_2$	σ_2^2
⋮	⋮	⋮	…	⋮	…	⋮	⋮	⋮	⋮
x_i	f_{i1}	f_{i2}	…	f_{ij}	…	f_{im}	f_i	$\bar{y}_i$	σ_i^2
⋮	⋮	⋮	…	⋮	…	⋮	⋮	⋮	⋮
x_n	f_{n1}	f_{n2}	…	f_{nj}	…	f_{nm}	f_n	$\bar{y}_n$	σ_n^2
Total f_y	f_1	f_2	…	f_j	…	f_m	n	$\bar{y}_0$	σ_0^2

Tabelul poate fi considerat cu dublă intrare, în care prima intrare se referă la frecvenţele variabilei principale X_i, iar cea de-a doua intrare la frecvenţele variabilei secundare Y_j. Din întretăierea celor două variabile rezultă frecvenţele f_{ij}.

Pentru analiza variaţiei caracteristicii Y_j, în funcţie de variaţia caracteristicii de grupare X_i, precum şi a interdependenţei dintre ele, se pot calcula medii şi dispersii condiţionate pentru fiecare grupă. Frecvenţele pe fiecare grupă se obţin prin însumarea frecvenţelor din interiorul grupelor, pentru grupa i având: $\sum_{j=1}^{m} f_{ij} = f_{i1} + f_{i2} + \ldots + f_{ij} + \ldots + f_{im} = f_i$.

Se poate calcula, în acest caz, o medie generală ($\bar{y}_0$) care sintetizează variaţia valorilor individuale ale colectivităţii totale şi valorile mediilor de grupă. Pentru caracteristica Y_j se pot calcula 3 feluri de indicatori, care să descrie:

- variaţia valorilor y_j în jurul mediei lor de grupă $\left(y_j - \bar{y}_i\right)$ datorată acţiunii cauzelor întâmplătoare (pe fiecare grupă);
- variaţia mediilor de grupă în jurul mediei colectivităţii totale $\left(\bar{y}_i - \bar{y}_0\right)$ datorată acţiunii cauzelor esenţiale (factorul principal de grupare);

- variaţia valorilor y_j în jurul mediei colectivităţii totale $\left(y_j - \bar{y}_0\right)$ datorată atât influenţei cauzelor esenţiale, cât şi influenţei cauzelor întâmplătoare.

Având în vedere cei 3 indicatori de mai sus (inclusiv modul lor de definire), se poate scrie:

$$\left(y_j - \bar{y}_0\right) = \left(y_j - \bar{y}_i\right) + \left(\bar{y}_i - \bar{y}_0\right).$$

Pornind de la această relaţie se pot determina dispersiile caracteristice distribuţiilor bidimensionale, dispersii pe baza cărora se face analiza variaţiei în cadrul acestor serii. Aceste dispersii sunt: dispersia de grupă; media dispersiilor de grupă; dispersia dintre grupe; dispersia generală.

- **Dispersia de grupă** $\left(\sigma_i^2\right)$ – cunoscută şi sub denumirea de dispersie parţială, se determină ca o medie aritmetică ponderată a pătratelor abaterilor variantelor caracteristicii de la media grupei, pe baza relaţiei următoare:

$$\sigma_i^2 = \frac{\sum_{j=1}^{m}\left(y_j - \bar{y}_i\right)^2 f_{ij}}{\sum_{j=1}^{m} f_{ij}} = \frac{\sum_{j=1}^{m}\left(y_j - \bar{y}_i\right)^2 f_{ij}}{f_i},$$

unde: $\bar{y}_i$ – mediile de grupă determinate ca medii aritmetice ponderate, astfel:

$$\bar{y}_i = \frac{\sum_{j=1}^{m} y_j f_{ij}}{\sum_{j=1}^{m} f_{ij}} = \frac{\sum_{j=1}^{m} y_j f_{ij}}{f_i}.$$

Dispersia de grupă măsoară variaţia caracteristicii Y_j determinată de acţiunea cauzelor întâmplătoare la nivelul fiecărei grupe. Se vor calcula atâtea dispersii de grupă câte grupe are colectivitatea cercetată, cu valori mai mici sau mai mari în funcţie de gradul de omogenitate sau eterogenitate a grupelor.

Spre exemplu, considerăm o distribuţie bidimensională a unei echipe de muncitori în funcţie de vechimea în muncă şi salariul realizat de muncitori. Dacă vechimea în muncă ar fi unicul factor de influenţă asupra salariului, atunci pentru fiecare grupă de vechime am avea un singur nivel al salariului. Cum, în general, avem mai multe niveluri ale salariului pentru o grupă de vechime în muncă, deducem că la nivelul fiecărei grupe îşi exercită influenţa şi alţi factori. Într-adevăr, în realitate, salariul este condiţionat şi de alţi factori, cum ar fi: productivitatea muncii, nivelul de calificare al muncitorilor, dotarea tehnică etc. Toţi ceilalţi factori, în afara vechimii în muncă, sunt consideraţi factori întâmplători, şi, ca atare, dispersia de grupă va cuantifica influenţa acestor factori la nivelul fiecărei grupe.

- **Media dispersiilor de grupă** $\left(\overline{\sigma}^2\right)$ – sintetizează influenţa factorilor întâmplători la nivelul întregii colectivităţi şi se calculează ca o medie aritmetică ponderată a dispersiilor de grupă, cu ajutorul relaţiei: $\overline{\sigma}^2 = \dfrac{\sum_{i=1}^{n}\sigma_i^2 f_i}{\sum_{i=1}^{n} f_i}$.

- **Dispersia dintre grupe** $\left(\delta^2\right)$ – reflectă variaţia caracteristicii secundare datorată acţiunii cauzelor esenţiale la nivelul întregii colectivităţi şi se calculează ca o medie aritmetică ponderată a pătratelor abaterilor mediilor de grupă de la media generală, pe baza relaţiei:

$$\delta^2 = \frac{\sum_{i=1}^{n}\left(\bar{y}_i - \bar{y}_0\right)^2 f_i}{\sum_{i=1}^{n} f_i},$$

unde: $\bar{y}_0$ – media generală şi se determină fie ca o medie aritmetică ponderată a distribuţiei marginale, fie ca o medie generală a mediilor de grupă, astfel: $\bar{y}_0 = \frac{\sum_{j=1}^{m} y_j f_j}{\sum_{j=1}^{m} f_j} = \frac{\sum_{i=1}^{n} \bar{y}_i f_i}{\sum_{i=1}^{n} f_i}$.

Din cele trei tipuri de dispersii prezentate, reţinem faptul că media dispersiilor de grupă şi dispersia dintre grupe pot fi comparate (pentru că ele caracterizează întreaga colectivitate). Putem, astfel, determina care dintre factori (esenţiali sau întâmplători) au avut o influenţă mai puternică asupra caracteristicii studiate.

O atenţie deosebită se cuvine să acordăm influenţei factorilor întâmplători pentru a cunoaşte cauzele care au condus la dispersarea unităţilor statistice din cadrul grupelor. Putem determina în acest fel cauzele obiective, dar şi subiective, care au determinat deplasarea frecvenţelor f_{ij} din cadrul grupei i.

- **Dispersia generală** (σ_0^2) – se calculează ca o medie aritmetică ponderată a pătratelor abaterilor termenilor faţă de media generală, pe baza relaţiei următoare: $\sigma_0^2 = \frac{\sum_{j=1}^{m} (y_j - \bar{y}_0)^2 f_j}{\sum_{j=1}^{m} f_j}$.

Dispersia generală măsoară variaţia totală a caracteristicii secundare (Y_j), variaţie determinată atât de acţiunea factorilor întâmplători, cât şi de cea a factorilor esenţiali, la nivelul colectivităţii generale. Această dispersie va avea o valoare mai mare în colectivităţile eterogene influenţate de un număr mare de factori (întâmplători sau esenţiali) şi o valoare mai mică în cazul colectivităţilor omogene.

Având în vedere conţinutul dispersiilor calculate, rezultă **regula de adunare a dispersiilor:**

$$\sigma_0^2 = \bar{\sigma}^2 + \delta^2.$$

7.3. INDICATORII FORMEI

Pentru caracterizarea seriilor de distribuţie se utilizează, alături de indicatorii tendinţei centrale şi ai gradului de dispersare, şi măsuri pentru asimetrie şi boltire. Măsurarea asimetriei şi a boltirii unei serii de distribuţie poate fi făcută atât prin intermediul unor parametri specifici, cât şi pe cale grafică. Dacă metoda grafică poate fi utilizată şi în cazul variabilelor calitative, indicatorii de asimetrie şi boltire sunt calculaţi numai pentru caracteristici numerice. Ambele metode au, însă, ca scop verificarea caracterului normal al distribuţiei.

7.3.1. Asimetria

În urma prelucrării primare a datelor, se obţin repartiţii de frecvenţe empirice, care se pot compara cu repartiţiile teoretice, pentru care s-au calculat indicatorii tendinţei centrale şi variaţiei, şi este cunoscută forma lor de repartiţie. Cea mai frecventă repartiţie teoretică cu care se compară seriile empirice este distribuţia normală sau funcţia Gauss-Laplace, ale cărei frecvenţe se distribuie simetric de o parte şi de alta a frecvenţei maxime plasate în centrul seriei, iar graficul acesteia are forma de clopot (clopotul Gauss-Laplace).

În practica statisticii economico-sociale se pot întâlni serii de repartiţie de frecvenţe simetrice, uşor asimetrice sau cu tendinţă pronunţată de asimetrie.

- ***Coeficientul de asimetrie al lui Pearson*** – este cel mai frecvent folosit indicator pentru determinarea asimetriei şi se obţine pe baza relaţiei următoare: $Cas = \frac{\bar{x} - Mo}{\sigma}$.

Acest indicator are o valoare abstractă, dar nu şi lipsită de semnificaţie. El oferă informaţii atât asupra sensului asimetriei, cât şi asupra intensităţii acesteia. Valorile pe care le ia sunt cuprinse în intervalul (–1,1). Pentru seriile de repartiţie moderat asimetrice, coeficientul de asimetrie ia valori în intervalul [-0,3;0,3]. Semnul indicatorului arată sensul asimetriei, astfel:

- $Cas < 0$ - serie cu asimetrie spre stânga (negativă);
- $Cas = 0$ - serie simetrică;
- $Cas > 0$ - serie cu asimetrie spre dreapta (pozitivă).

7.3.2. Boltirea

Boltirea (aplatizarea) *apare atunci când distribuţia prezintă o variaţie slabă a variabilei X şi o variaţie puternică a frecvenţei absolute (şi invers), în comparaţie cu o distribuţie normală, de aceeaşi medie şi dispersie.*

Deci, boltirea unei serii de repartiţie se defineşte prin raportarea la repartiţia normală sub aspectul variaţiei variabilei X şi a frecvenţelor absolute f_i. Boltirea se poate evalua fie pe cale grafică, fie pe calea calculelor algebrice.

- ***coeficientul de boltire Pearson (β_2)*** – se calculează pe baza momentelor centrate de ordinul 2 şi 4, cu ajutorul relaţiei: $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}$.

unde μ_2 şi μ_4 reprezintă momentele centrate de ordinul 2 şi 4.

Acest coeficient ia valoarea 3 ($\beta_2 = 3$) pentru o distribuţie normală – curba mezocurtică. Pentru $\beta_2 > 3$ avem o curbă leptocurtică (mai ascuţită decât curba normală), iar pentru $\beta_2 < 3$ avem o curbă platicurtică (mai plată decât curba normală).

- ***coeficientul de boltire Fisher (γ_2)*** – mai este cunoscut şi sub denumirea de *coeficient al excesului*, deoarece măsoară excesul faţă de boltirea unei distribuţii normale Gauss-Laplace. Se determină pornind de la coeficientul de boltire al lui Pearson, ţinând cont şi de faptul că acest indicator pentru distribuţia normală ia valoarea 3, astfel: $\gamma_2 = \beta_2 - 3$.

Pentru $\gamma_2 = 0$ avem o curbă mezocurtică, pentru $\gamma_2 > 0$ (avem un exces de frecvenţe în zona centrală) curba este leptocurtică, iar pentru $\gamma_2 < 0$ avem o curbă platicurtică.

7.4. SONDAJUL STATISTIC

În cercetarea oricărui fenomen sau proces se pot utiliza două categorii de cercetări statistice:

1. cercetări statistice totale - în care sunt supuse studiului toate unităţile statistice din populaţia vizată;
2. cercetări statistice parţiale - în care sunt supuse studiului doar o parte din unităţile statistice din populaţia vizată.

Sondajul statistic face parte din categoria cercetărilor statistice parţiale. Se utilizează, de obicei, în cazurile în care se doreşte caracterizarea unor fenomene şi procese social-economice pentru care nu dispunem de date suficiente, care nu pot fi studiate în totalitate sau pentru care nu este recomandabilă o cercetare totală datorită faptului că unităţile statistice supuse studiului sunt distruse în urma cercetării.

Sondajul statistic reprezintă o formă de cercetare statistică parţială care îşi propune estimarea parametrilor populaţiei de bază pe baza datelor culese de la nivelul unui eşantion reprezentativ.

Principalele etape ale sondajului statistic sunt:

1. **Stabilirea obiectivelor cercetării**. Această etapă presupune:

- definirea clară a obiectivelor ce vor fi urmărite în cadrul cercetării prin sondaj
- identificarea şi delimitarea spaţio-temporală a populaţiei ce va fi supusă studiului
- identificarea principalelor surse de date aferente populaţiei de bază, posibil de utilizat în cercetare

- stabilirea variabilelor (caracteristicilor) ce vor fi supuse observării şi a modalităţii de observare şi înregistrare a acestora
- estimarea costurilor şi stabilirea bugetului cercetării
- stabilirea necesarului de personal şi alcătuirea echipei ce va realiza cercetarea

2. **Extragerea eşantionului**. Luând în considerare caracteristicile populaţiei de bază dar şi cerinţele de precizie în estimarea parametrilor în urma cercetării prin sondaj, se stabileşte modul în care va fi extras eşantionul din populaţia de bază. Într-o cercetare prin sondaj, aceasta reprezintă problema esenţială. Calitatea eşantionului determină precizia estimărilor şi gradul de realism al rezultatelor obţinute şi interpretărilor acestora.

3. **Elaborarea chestionarului**. Instrumentul utilizat de cele mai multe ori pentru culegerea datelor în cadrul sondajului statistic este reprezentat de chestionar. Calitatea acestuia este determinantă în reducerea erorilor de înregistrare şi implicit calitatea datelor.

4. **Culegerea datelor**. Este etapa cea mai laborioasă a cercetării, sub aspectul efortului dar şi resurselor necesare. În această etapă se realizează efectiv înregistrarea valorilor pentru variabilele cuprinse în planul cercetării, stabilite în etapa 1. Culegerea datelor se poate realiza cu ajutorul mai multor tehnici, dintre care menţionăm:

- culegere directă, presupune deplasarea până la fiecare unitate statistică a populaţiei de bază care a fost inclusă în eşantion şi înregistrarea datelor prin observarea, măsurarea directă a acestora sau prin interviu direct dacă este posibil;
- completarea unui chestionar;
- interviu prin telefon;
- culegerea datelor prin corespondenţă, etc.

5. **Codificarea şi prelucrarea primară a datelor**. Este etapa în care se identifică şi se elimină datele eronate.

6. **Prelucrarea propriu-zisă a datelor**. Această etapă presupune:

- evidenţierea erorilor apărute în procesul de realizare a cercetării pin sondaj;
- calculul parametrilor la nivelul eşantionului;
- estimarea parametrilor la nivelul populaţiei de bază;
- ameliorarea estimatorilor obţinuţi prin utilizarea de date şi informaţii auxiliare.

7. **Analiza şi interpretarea rezultatelor**. În această etapă, rezultatele obţinute în urma cercetării sunt analizate şi sintetizate în concluzii.

7.4.1. Procedee de extragere a eşantionului

În organizarea unui sondaj cea mai importantă problemă o reprezintă formarea eşantionului.

Pentru ca rezultatele sondajului să aibă o precizie cât mai mare este necesar ca eşantionul să respecte condiţia de reprezentativitate, adică să reproducă pe cât posibil structura lotului de bază din care a fost extras.

Pentru a extrage un eşantion reprezentativ trebuie respectate următoarele condiţii:

a) lotul de bază să fie cât mai omogen;

b) extragerea unităţilor statistice, din lotul de bază, să se facă absolut întâmplător, astfel încât toate unităţile să aibă şanse egale de a fi extrase.

La extragerea eşantionului se pot folosi următoarele procedee:

- **procedee aleatoare** (probabilistice) şi
- **procedee subiective** (dirijate)

7.4.1.1 Procedee aleatoare

a) Procedeul tragerii la sorţi, care poate fi realizat în două variante: cu repetare; fără repetare

Procedeul tragerii la sorţi în varianta „cu repetare” presupune formarea eşantionului prin extragerea unitate cu unitate din populaţia de bază astfel încât, după fiecare extragere, unitatea se reintroduce în populaţie. Astfel, volumul populaţiei de bază rămâne constant pe toată durata extragerii.

Procedeul tragerii la sorţi în varianta „fără repetare” presupune formarea eşantionului în acelaşi mod ca la procedeul tragerii la sorţi în varianta cu repetare, cu deosebirea că odată extrase unităţile statistice nu se reintroduc în populaţia de bază.

b) Procedeul mecanic (sistematic)

Se utilizează în cazul în care populaţia de bază din care urmează să extragem eşantionul este formată dintr-o populaţie statistică deja organizată după un anumit criteriu. (*de exemplu: studenţii unei facultăţi ordonaţi după numărul matricol, pomii dintr-o livadă sunt plantaţi după un anumit model*).

Utilizarea acestui procedeu presupune extragearea aleatoare doar a primei unităţi care va fi inclusă în eşantion. Restul unităţilor ce vor forma eşantionul se determină mecanic (sistematic) pornind de la prima extrasă pe baza unei relaţii.

Pentru aplicarea acestui procedeu se calculează mai întâi un pas de numărare:

$$k = \frac{N}{n}, \quad n \text{ - volumul eşantionului}$$

Se introduc apoi într-o urnă, bilete (sau jetoane, cartonaşe, etc.) numerotate de la 1 la k din care se extrage unul singur.

Numărul înscris pe biletul extras va indica numărul de ordine al unităţii statistice din lotul de bază care va fi prima extrasă în eşantion. Restul unităţilor care intră în eşantion se determină adăugând pasul de numărare la numărul de ordine al ultimei unităţi extrase din lotul de bază.

Datorită faptului că nu se realizează o extragere a eşantionului complet întâmplătoare, rezultatele acestui tip de sondaj sunt mai puţin exacte decât în cazul utilizării procedeului tragerii la sorţi pe baza schemei bilei nerevenite.

7.4.2. Principalele tipuri de sondaje

Tipul sondajului este determinat de următorii factori:

a) Modul de organizare al lotului de bază în momentul extragerii:
- lotul de bază neorganizat;
- lotul de bază organizat în grupe tipice.

b) Procedeul de eşantionare folosit. Cel mai utilizat este procedeul tip loterie în variantele:
- repetat;
- nerepetat.

c) Numărul de unităţi extrase deodată din lotul de bază:
- unitate cu unitate;
- un grup de unităţi (serie).

Combinând cei trei factori rezultă următoarele tipuri importante de sondaj:

a) sondaj simplu întâmplător: -repetat
-nerepetat

b) sondaj tipic (stratificat): -repetat
-nerepetat

c) sondaj de serii, se organizat în practică numai în varianta nerepetat, pentru că se operează cu un număr mic de serii.

7.4.2.1. Sondajul simplu întâmplător

Acest tip de sondaj se utilizează la cercetarea populaţiilor statistice care prezintă un grad de omogenitate ridicat.

În cazul acestui tip de sondaj se utilizează următorii indicatori:

Tip sondaj / Indicatori	repetat	nerepetat
Eroarea medie de reprezentativitate	$\mu_{rep}=\sqrt{\frac{\sigma_s^2}{n}}$	$\mu_{nrep}=\sqrt{\frac{\sigma_0^2}{n}\cdot\left(1-\frac{n}{N}\right)}$
Eroarea limită admisă	$\pm\Delta x_{rep}=z\sqrt{\frac{\sigma_0^2}{n}}$	$\pm\Delta x_{nrep}=z\mu_{nrep}=z\sqrt{\frac{\sigma_0^2}{n}\left(1-\frac{n}{N}\right)}$
Volumul eşantionului	$n=\frac{z^2\sigma_0^2}{\Delta x_{rep}^2}$	$n_{nrep}=\frac{z^2\sigma_0^2}{\Delta x_{nrep}^2+\frac{z^2\sigma_0^2}{N}}$

Aceeaşi metodologie de calcul pentru indicatorii sondajului se poate utiliza şi pentru cazul în care **variabila studiată** prin sondaj **este de tip alternativ** ţinând cont de modul de determinare al mediei şi dispersiei pentru acest tip de variabilă.

Se folosesc notaţiile:

- pentru lotul de bază:

p - media generală;

σ_p^2 - dispersia generală ($\sigma_p^2=p(1-p)$);

- pentru eşantion:

w -media ;

σ_w^2 - dispersia ($\sigma_w^2=w(1-w)$).

7.4.2.2. Sondajul tipic (stratificat)

Se utilizează în cazul populaţiilor statistice care prezintă un grad de omogenitate scăzut. În astfel de situaţii populaţia din lotul de bază se organizează în prealabil în grupe omogene.

Pentru a respecta condiţia de reprezentativitate eşantionul trebuie format extrăgând un număr de unităţi direct proporţional cu volumul fiecărei grupe.

Pentru calculul indicatorilor sondajului, se foloseşte media dispersiilor de grupă:

$$\overline{\sigma}^2=\frac{\sum\sigma_i^2 f_i}{\sum f_i}$$ - media dispersiilor de grupă din lotul de bază

unde

$$\sigma_i^2=\frac{\sum\left(x_i-\overline{x}_i\right)^2 f_i}{\sum f_i}$$ - dispersia grupei *i* din lotul de bază

unde:

x_i - variabila urmărită în grupa *i*;

$\overline{x}_i$ - media grupei *i*;

f_i - volumul grupei *i*;

$\sum f_i=N$ - volumul lotului de bază.

În cazul acestui tip de sondaj se utilizează următorii indicatori:

Tip sondaj / Indicatori	repetat	nerepetat

Eroarea medie de reprezentativitate	$\mu_{rep} = \sqrt{\frac{\overline{\sigma}^2}{n}}$	$\mu_{nrep} = \sqrt{\frac{\overline{\sigma}^2}{n}\left(1-\frac{n}{N}\right)}$
Eroarea limită admisă	$\pm \Delta x_{rep} = z\mu = z\sqrt{\frac{\overline{\sigma}^2}{n}}$	$\pm \Delta x_{nrep} = z\mu = z\sqrt{\frac{\overline{\sigma}^2}{n}\left(1-\frac{n}{N}\right)}$
Volumul eşantionului	$n_{rep} = \frac{z^2\overline{\sigma}^2}{\Delta x_{rep}^2}$	$n_{nrep} = \frac{z^2\overline{\sigma}^2}{\Delta x_{nrep}^2 + \frac{z^2\overline{\sigma}^2}{N}}$

Dacă volumul eşantionului este suficient de mare (*n > 100*) se poate folosi în locul mediei dispersiilor de grupă din lotul de bază ($\overline{\sigma}^2$), media dispersiilor de grupă din eşantion ($\overline{\sigma}_s^2$).

Dacă variabila urmărită este de tip alternativ, atunci indicatorii sondajului se vor determina ţinând cont de modul de calcul al mediei şi dispersiei pentru acest tip de variabilă:
În varianta **cu revenire**:

- eroarea medie de reprezentativitate (μ): $\mu_{rep} = \sqrt{\frac{\overline{p(1-p)}}{n}} = \sqrt{\frac{\overline{\sigma}_p^2}{n}}$
- eroarea-limită admisă (Δx): $\pm \Delta x_{rep} = z\mu_{rep}$
- volumul eşantionului (*n*): $n_{rep} = \frac{z^2\overline{\sigma}_p^2}{\Delta x_{rep}^2}$

În varianta **fără revenire**:

- eroarea medie de reprezentativitate (μ): $\mu_{nrep} = \sqrt{\frac{\overline{p(1-p)}}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{\overline{\sigma}_p^2}{n}\left(1-\frac{n}{N}\right)}$
- eroarea-limită admisă (Δx): $\pm \Delta x_{nrep} = z\mu_{nrep}$
- volumul eşantionului (*n*): $n_{nrep} = \frac{z^2\overline{\sigma}_p^2}{\Delta x_{nrep}^2 + \frac{z^2\overline{\sigma}_p^2}{N}}$

7.4.2.3. Sondajul de serii

Se utilizează atunci când populaţia statistică din lotul de bază este alcătuită nu din unităţi simple ci din unităţi complexe (ex. numărul de muncitori ai unei firme organizaţi în echipe, produse ambalate în seturi etc). În acest caz eşantionul se formează prin extragerea de unităţi complexe (serii).

Pentru calculul indicatorilor sondajului se utilizează dispersia mediilor seriilor de la media generală (δ^2), astfel:

$$\delta^2 = \frac{\sum\left(\overline{x}_i - \overline{x}_0\right)^2}{r}$$

Tip sondaj / Indicatori	nerepetat
Eroarea medie de reprezentativitate	$\mu_{nrep} = \sqrt{\frac{\delta^2}{r-1}\left(\frac{R-r}{R-1}\right)}$
Eroarea limită	$\pm \Delta x_{nrep} = z\mu_{nrep}$

admisă	
Volumul eşantionului	$r_{nrep} = \dfrac{Rt^2\delta^2}{(R-1)\Delta x^2_{nrep} + z^2\delta_0^2}$

unde:

r - numărul de serii ce formează eşantionul;

R - numărul total de serii din care este alcătuit lotul de bază.

Se poate utiliza în locul dispersiei mediilor seriilor de la media generală a lotului de bază (δ^2), dispersia mediilor seriilor de la media eşantionului (δ_s^2).

Dacă variabila studiată prin sondaj este de tip alternativ, indicatorii sondajului se vor calcula cu relaţiile:

- eroarea medie de reprezentativitate (μ_w): $\mu_{wnrep} = \sqrt{\dfrac{\delta_w^2}{r-1}\left(\dfrac{R-r}{R-1}\right)}$

- eroarea-limită admisă (Δw): $\pm\Delta w_{nrep} = z\mu_{wnrep}$

- numărul de serii ce formează eşantionul (r): $r_{nrep} = \dfrac{Rz^2\delta_w^2}{(R-1)\Delta w^2_{nrep} + z^2\delta_w^2}$

7.5. CORELAŢIE ŞI REGRESIE

De foarte multe ori, fenomenele şi procesele, indiferent de domeniul în care se manifestă şi de natura lor, sunt legate între ele prin conexiuni care nu sunt cunoscute şi observate de la bun început, ci de regulă sunt descoperite pe parcursul studierii lor. Efectele generate de manifestarea unuia pot provoca apariţia, încetarea sau modificarea altuia, determinând relaţii de interdependenţă sau cauzalitate.

Expresia sintetică a intensităţii legăturii cauzale dintre fenomene - poartă denumirea de ***corelaţie***.

După cum am precizat anterior fenomenele între care există o legătură se pot găsi în una din situaţiile:

- **cauză** - atunci când determină apariţia sau modificarea unui alt fenomen;
- **efect** - atunci când sunt rezultatul efectelor generate de manifestarea unor alte fenomene.

Variabilele care caracterizează cele două categorii de fenomene pot fi: **variabile cauză** (independente, factoriale) - atunci când caracterizează un fenomen cauză, respectiv **variabile efect** (dependente, rezultative) - atunci când caracterizează un fenomen efect.

Cuplul corelativ poate cuprinde doar două variabile, una factorială şi una rezultativă sau poate cuprinde mai multe variabile dintre care doar una este variabilă rezultativă şi restul factoriale.

Având în vedere elementele urmărite în studiul corelaţiei se pot determina mai multe tipuri de corelaţie:

1. **După numărul variabilelor din cuplul corelativ**, avem:

- **corelaţie simplă** - atunci când cuplul corelativ cuprinde o singură variabilă rezultativă şi o singură variabilă cauză;
- **corelaţie multiplă** - atunci când cuplul corelativ cuprinde o singură varabilă rezultativă şi mai multe variabile cauză.

2. **După sensul legăturii dintre variabilele cuplului corelativ**, avem:

- **corelaţie directă**- atunci când variabila rezultativă şi cele cauzale urmează acelaşi sens al modificării - valorile lor cresc sau descresc simultan.

- **corelaţie inversă** - atunci când variabila rezultativă urmează un sens al modificării, iar cele cauză sensul opus - în timp ce valorile rezultativei cresc cele ale variabilelor cauzale scad şi invers.

3. **După forma legăturii dintre variabilele cuplului corelativ**, avem:

- **corelaţie liniară** - atunci când variabila rezultativă urmează o tendinţă liniară, generată de influenţele variabilelor factoriale;
- **corelaţie neliniară** - atunci când variabila efect urmează o tendinţă diferită de cea liniară, generată de influenţele variabilelor factoriale.

Metodele statistice utilizate pentru studiul legăturii dintre două sau mai multe fenomene se pot grupa în două mari categorii:

- **metode elementare** - prin care se poate determina existenţa legăturii dintre fenomene, a tăriei, a sensului şi a formei acesteia dar nu cu o precizie foarte mare, ele fiind de obicei folosite pentru orientarea către metode de altă natură, mai rafinate pentru determinarea elementelor de mai sus foarte precis.
- **metode analitice** - prin care se pot determina aceleaşi elemente ca şi prin metodele elementare, dar cu o precizie mult mai mare, ele permiţând de asemenea şi studiul legăturii dintre un fenomen efect şi mai multe fenomene cauză simultan.

7.5.1 Metode elementare

Dintre metodele elementare, mai des, sunt folosite următoarele: metoda tabelulului de corelaţie, metoda grafică.

7.5.1.1 Metoda tabelului de corelaţie

Permite evidenţierea tuturor elementelor necesare pentru confirmarea existenţei unei legături dintre două fenomene, pe baza observaţiei modului de manifestare prin utilizarea măsurătorilor unor variabile care caracterizează fenomenele supuse studiului.

Pentru utilizarea este necesară distribuţia bidimensională obţinută prin prelucrarea perechilor de valori determinate prin măsurarea ceor două variabile care caracterizează fenomenul cauză, respectiv fenomenul efect.

Modul în care se distribuie frecvenţele în interiorul acestei distribuţii (figura 9.1) oferă toate elementele pentru evidenţierea unei eventuale legături între cele două fenomene.

Y \ X	y_1	y_2	y_3	$\dots y_j \dots$	y_n	Fx
x_1	f_{11}	f_{12}	f_{13}	$\dots f_{1j} \dots$	f_{1n}	Fx_1
x_2	f_{21}	f_{22}	f_{23}	$\dots f_{2j} \dots$	f_{2n}	Fx_2
x_3	f_{31}	f_{32}	f_{33}	$\dots f_{3j} \dots$	f_{3n}	Fx_3
: x_i :	: f_{i1} :	: f_{i2} :	: f_{i1} :	... f_{ij} ...	: f_{ni} :	: Fx_i :
x_n	f_{n1}	f_{n2}	f_{n3}	$\dots f_{nj} \dots$	f_{nn}	Fx_n
Fy	Fy_1	Fy_2	Fy_3	$\dots Fy_j \dots$	Fy_n	F

X - variabila cauză;
Y - variabila efect;
$x_1 \dots x_n$ - valorile variabilei cauză;
$y_1 \dots y_n$ - valorile variabilei efect;
f_{ij} - frecvenţa de apariţie a perechii de valori (x_i, y_j);

Fx_i - frecvenţa de apariţie a valorii x_i;
Fy_i - frecvenţa de apariţie a valorii y_i;
F - numărul total de perechi de valori (x_i, y_j);
Elementele care pot fi evidenţiate cu ajutorul acestei metode:

1. **Existenţa legăturii dintre variabila X factorială şi Y rezultativă:**

Dacă frecvenţele f_{ij} se distribuie într-o bandă grupată de-a lungul unei diagonale a tabelului

Sensul legăturii:

Dacă banda în care sunt grupate frecvenţele f_{ij} se află pe diagonala tabelului care corespunde aceluiaşi sens de variaţie a valorilor corespunzătoare celor două variabile X,Y înseamnă că între cele două variabile există o legătură directă. Dacă se află pe cealaltă diagonală care corespunde sensului diferit de variaţie a celor două variabile X,Y atunci legătura dintre cele două variabile este inversă.

2. **Intensitatea legăturii:**

   Este dată de lăţimea benzii în care sunt grupate frecvenţele f_{ij}. Cu cât este mai îngustă intensitatea legăturii creşte.

3. **Forma legăturii:**

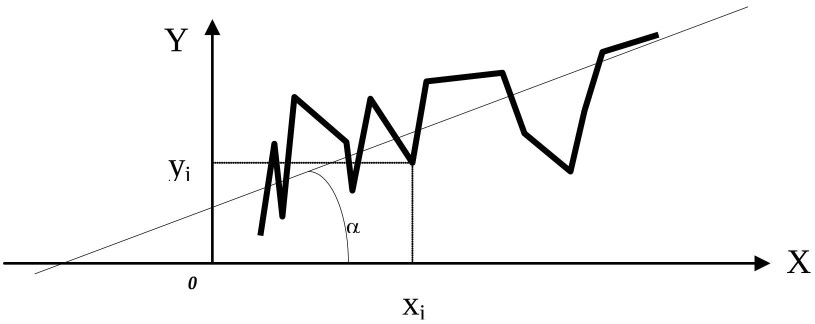
Este dată de forma benzii, putând fi liniară dacă banda este liniară sau neliniară dacă banda are altă formă decât cea liniară.

7.5.1.2. Metoda grafică

Ca şi metoda precedentă permite evidenţierea prin apreciere vizuală a elementelor ce caracterizează legătura dintre două variabile.

În acest caz este necesară construirea corelogramei. Pe abscisă se trec valorile scării de reprezentare corespunzătoare variabilei cauză X, iar pe ordonată, valorile scării de reprezentare corespunzătoare variabilei Y.

Prin unirea cu segmente de dreaptă a punctelor obţinute reprezentând grafic perechile de valori (x_i, y_j) se obţine corelograma.



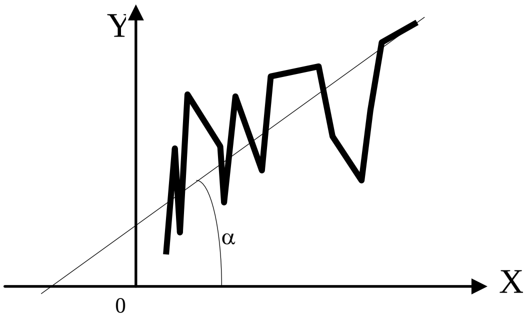
Corelograma

1. **Existenţa legăturii:**

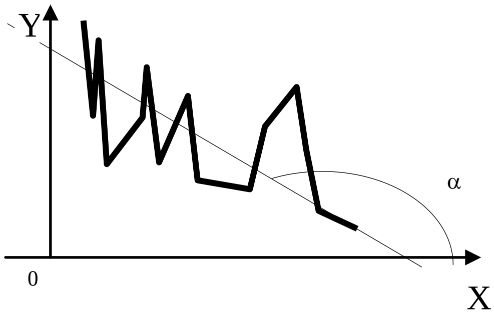
Se determină prin existenţa unghiului (α) realizat de linia de tendinţa cu orizontala diferit de 0.

2. **Sensul legăturii:**

   - legătură directă - atunci când linia de tendinţă este ascendentă;
   - legătură inversă - atunci când linia de tendinţă este descendentă;



Corelaţie directă



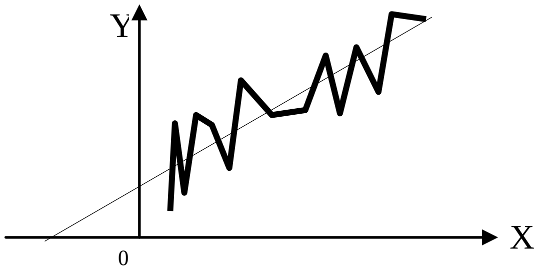
Corelaţie inversă

Intensitatea legăturii:

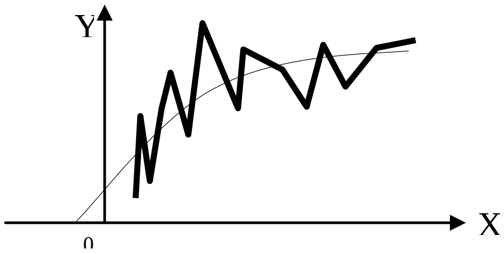
Dată de apropierea valorii unghiului α de 45°.

3. **Forma legăturii:**

Este dată de forma corelogramei.



Corelație liniară



Corelatie neliniară

7.5.2. Metode analitice

Determinarea precisă a legăturii dintre două sau mai multe variabile se realizează cu ajutorul metodelor analitice.

Sunt urmărite în special două aspecte:

a) **regresia** - cu ajutorul căreia se detemină contribuţia variabilelor factoriale la modificarea variabilei rezultative;

b) **intensitatea legăturii** - determinată cu ajutorul coeficienţilor de corelaţie corespunzători tipului de corelaţie existent liniar sau neliniar.

7.5.2.1. Corelaţia simplă liniară

Are la bază utilizarea funcţiei liniare pentru analiza regresiei:

$$Y_x = a + bx$$

Graficul funcţiei liniare

în care:

Y_x - valorile calculate (teoretice) ale variabilei rezultative Y prin funcţia de regresie;

a - valoarea pe care o ia variabila rezultativă atunci când variabila factorială nu o influenţează;

b - coeficientul de regresie - arată contribuţia variabilei factoriale la modificarea cu o unitate a valorii variabilei rezultative;

x - valorile x_1, x_2, x_3,... x_n, ale variabilei factoriale X.

Intensitatea corelaţiei liniare

Determinarea intensităţii corelaţiei liniare se realizează cu ajutorul coeficientului de corelaţie liniară al lui Pearson care se determină cu ajutorul uneia din relaţiile următoare:

$$r_{y,x} = \frac{\sum\left(\frac{x_i - \bar{x}}{\sigma_x}\right)\left(\frac{y_i - \bar{y}}{\sigma_y}\right)}{n}$$

După mai multe prelucrări se obţine o formă aplicabilă mult mai uşor:

$$r_{y,x} = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{\sqrt{\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right]\left[n\sum y_i^2 - \left(\sum y_i\right)^2\right]}}$$

sau

$$r_{y,x} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \bar{y})^2}}$$

iar pentru serii de distribuţie bidimensionale se poate utiliza relaţia:

$$r_{y,x} = \frac{\sum f \cdot \sum\left(x_i \sum y_j f_{ij}\right) - \left(\sum x_i F_{xi}\right)\left(\sum y_j F_{yj}\right)}{\sqrt{\left[\sum f \cdot \sum x_i^2 F_{xi} - \left(\sum x_i F_{xi}\right)^2\right]\left[\sum f \cdot \sum y_j^2 F_{yj} - \left(\sum y_j F_{yj}\right)^2\right]}}$$

unde $\sum f = \sum\sum f_{ij} = \sum F_{xi} = \sum F_{y_j}$

Coeficientul de corelaţie ***r*** ia valori în intervalul **[-1;1]**, intensitatea legăturii crescând pe măsură ce se apropie de extremele intervalului.

Valorile negative semnifică existenţa unei corelaţii *inverse*, cele pozitive o corelaţii *directe* între variabila rezultativă Y şi cea factorială X.

Regresia simplă liniară

Pentru determinarea coeficienţilor *a*, *b* se foloseşte drept criteriu de apreciere al calităţii funcţiei de regresie expresia celor mai mici pătrate:

$$\sum(y - Y_x)^2 = \min$$

în care:

y - valorile observate (empirice) ale variabilei rezultative Y.

Prin aplicarea acestui criteriu asupra funcţiei de regresie liniare se obţine un sistem cu două ecuaţii din care se pot determina uşor valorile corespunzătoare coeficienţilor *a* şi *b*.

$$\begin{cases} \sum y_i = na + b\sum x_i \\ \sum x_i y_i = a\sum x_i + b\sum x_i^2 \end{cases}$$

O metodă de a rezolva acest sistem este cea folosind determinanţi:

$$a = \frac{\Delta_a}{\Delta};\ b = \frac{\Delta_b}{\Delta}$$

unde, din sistemul iniţial se pot determina:

$$\Delta = \begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix};\ \Delta_a = \begin{vmatrix} \sum y & \sum x \\ \sum xy & \sum x^2 \end{vmatrix};\ \Delta_b = \begin{vmatrix} n & \sum y \\ \sum x & \sum xy \end{vmatrix}$$

După cum se observă din determinantul iniţial Δ se determină Δ_a respectiv Δ_b înlocuind coloana corespunzătoare parametrului a sau b cu coloana termenilor liberi din sistem.

7.6. ANALIZA SERIILOR CRONOLOGICE

Orice fenomen sau proces al activităţii umane poate fi studiat atât în timp, cât şi în spaţiu. Analiza în timp presupune, în principal, o cercetare cu ajutorul unor indicatori statistici specifici de-a lungul diferitelor perioade.

7.6.1. Indicatorii dinamicii

Pentru a caracteriza dinamica fenomenelor economico-sociale, prelucrarea unor serii dinamice conduce la obţinerea unei varietăţi de indicatori. După modul de calcul şi exprimare aceştia pot fi grupaţi în trei categorii: indicatori absoluţi; indicatori relativi; indicatori medii.

Toţi indicatorii dinamicii se pot calcula în două variante:

- cu bază fixă: când se compară valorile din oricare perioadă (t) aferente indicatorilor cuprinşi în serie cu valorile acestora aferente unei singure perioade (1);
- cu bază în lanţ: când se compară valorile din oricare perioadă (t) aferente indicatorilor cuprinşi în serie cu valorile acestora din perioada precedentă (t-1);

7.6.1.1 Indicatorii absoluţi

Indicatorii absoluţi se exprimă în aceeaşi unitate de măsură cu fenomenul supus cercetării. În cadrul lor întâlnim două categorii:

- ***nivelul absolut*** – este dat de şirul nivelurilor fenomenului a cărui evoluţie se urmăreşte. Dacă seria este simplă, atunci nivelurile absolute pentru variabila *Y* sunt $y_1, y_2, \ldots, y_n$;
- ***modificarea absolută*** – se determină ca diferenţă între nivelurile absolute ale uneia dintre variabilele seriei, luate succesiv, şi un nivel oarecare considerat bază de comparaţie (această bază trebuie să fie un moment sau interval de timp considerat reprezentativ pentru seria supusă cercetării). Modificarea absolută exprimă, în valori absolute, cu cât a crescut sau a scăzut nivelul fenomenului cercetat în perioada de timp considerată. În funcţie de baza de comparaţie aleasă, modificarea absolută poate fi:

  - *cu baza fixă*: $\Delta_{t/1} = y_t - y_1$;

  - *cu baza în lanţ*: $\Delta_{t/t-1} = y_t - y_{t-1}$,

  unde: y_1 - nivelul indicatorului în perioada de referinţă;
  y_t - nivelul indicatorului în perioada *t*;

y_{t-1} - nivelul indicatorului în perioada t-1.

Comparând relaţiile de calcul ale celor două variante, rezultă că: $\Delta_{t/1} = \sum \Delta_{t/t-1}$.

7.6.1.2 Indicatorii relativi

Indicatorii relativi se calculează ca raport între doi indicatori absoluţi ai aceluiaşi fenomen şi se exprimă, de regulă, sub formă de coeficienţi sau în procente. În cadrul lor întâlnim indicele dinamicii, ritmul dinamicii şi valoarea absolută a unui procent de creştere (reducere).

- ***Indicele dinamicii*** – se calculează ca raport între nivelul indicatorului de comparat şi nivelul indicatorului folosit ca bază de comparaţie. Acesta exprimă de câte ori sau în ce proporţie s-a modificat fenomenul y în perioada considerată. În funcţie de baza de comparaţie aleasă, indicele dinamicii poate fi de două feluri:

  - *cu baza*: $I_{t/1} = \dfrac{y_t}{y_1}$;

  - *cu baza în lanţ*: $I_{t/t-1} = \dfrac{y_t}{y_{t-1}}$.

  Comparând relaţiile celor două variante de indici, rezultă că: $I_{t/1} = \prod I_{t/t-1}$.

  Indicii dinamicii se pot exprima şi în procente.

- ***Ritmul dinamicii*** – exprimă, în mărimi relative, cu cât a crescut sau a scăzut nivelul fenomenului cercetat în perioada de timp considerată. Se poate calcula în trei moduri: pe baza nivelurilor absolute, pe baza modificărilor absolute sau pe baza indicilor. În funcţie de baza de comparaţie aleasă, ritmul dinamicii poate fi:

  - *cu baza fixă*: $R_{t/1} = \dfrac{y_t - y_1}{y_1} \cdot 100 = \dfrac{\Delta_{t/1}}{y_1} \cdot 100 = \left(I_{t/1} - 1\right) \cdot 100$;

  - *cu baza în lanţ*: $R_{t/t-1} = \dfrac{y_t - y_{t-1}}{y_{t-1}} \cdot 100 = \dfrac{\Delta_{t/t-1}}{y_{t-1}} \cdot 100 = \left(I_{t/t-1} - 1\right) \cdot 100$.

  Ritmul dinamicii se exprimă numai în procente.

7.6.1.3 Indicatorii medii

Indicatorii medii sunt indicatori calculaţi pe baza tuturor termenilor seriei cronologice. Astfel, în timp ce indicatorii absoluţi şi relativi ne arată nivelurile individuale înregistrate de-a lungul perioadei, indicatorii medii reunesc aceste valori individuale într-una singură. În această categorie de indicatori regăsim: nivelul mediu, modificarea medie, indicele mediu, ritmul mediu şi valoarea medie absolută a unui procent de creştere.

- ***nivelul mediu*** – se calculează în mod diferit după cum seria dinamică este de *intervale* sau de *momente*:
  - *dacă seria cronologică este de intervale,* nivelul mediu se calculează folosind:
    - media aritmetică – dacă valorile $\Delta_{t/t-1}$ sunt aproximativ constante;
    - media pătratică – dacă valorile $\Delta_{t/t-1}$ sunt mai mici la începutul seriei şi din ce în ce mai mari spre sfârşitul acesteia;
    - media geometrică – dacă valorile $\Delta_{t/t-1}$ sunt mai mari la începutul seriei şi din ce în ce mai mici spre sfârşitul acesteia.
  - *dacă seria cronologică este de momente,* nivelul mediu se determină ca o medie cronologică. Media cronologică este, în principiu, o medie aritmetică, şi se determină în două etape: a) *calculul mediilor mobile* – acum are loc transformarea seriei de momente în serie de intervale, mediile mobile nefiind altceva decât medii aritmetice simple calculate din câte doi, trei sau mai mulţi termeni ai seriei, în cadrul cărora unul sau mai mulţi termeni se repetă; b) *calculul mediei cronologice* – se obţine ca medie aritmetică a

mediilor mobile. Intervalele dintre momentele seriei pot fi egale sau inegale, rezultând medii cronologice simple sau ponderate.

Media cronologică simplă este utilizată în cazul în care intervalele dintre momente sunt egale ($t_1 = t_2 = \ldots = t_k$, unde k – numărul mediilor mobile sau numărul intervalelor dintre momente, $k = n - 1$). Determinarea mediei cronologice simple se face după etapele precizate anterior, astfel:

- calculul mediilor mobile: $\bar{x}_i = \frac{x_i + x_{i+1}}{2}$;
- calculul mediei cronologice simple: $\bar{x}_c = \frac{\sum_{i=1}^{k} \bar{x}_i}{k}$.

Media cronologică ponderată se foloseşte atunci când intervalele dintre momente sunt inegale ($t_1 \neq t_2 \neq \ldots \neq t_k$). Ca şi în cazul mediei cronologice simple, media cronologică ponderată se determină urmând cele două etape:

- calculul mediilor mobile: $\bar{x}_i = \frac{x_i + x_{i+1}}{2}$;
- calculul mediei cronologice ponderate: $\bar{x}_c = \frac{\sum_{i=1}^{k} \bar{x}_i t_i}{\sum_{i=1}^{k} t_i}$.

- ***modificarea medie*** – exprimă, sub formă de medie, modificarea înregistrată în fiecare perioadă a seriei cronologice. Se calculează ca o medie aritmetică simplă a modificărilor cu baza în lanţ, pe baza relaţiei următoare:

$$\bar{\Delta} = \frac{\sum \Delta_{t/t-1}}{n-1} = \frac{\Delta_{t/1}}{n-1}.$$

  Modificarea medie prezintă importanţă pentru stabilirea tendinţei (trendului) unui fenomen, astfel:
  - dacă $\bar{\Delta} > 0 \Rightarrow$ tendinţă evolutivă (crescătoare);
  - dacă $\bar{\Delta} < 0 \Rightarrow$ tendinţă involutivă (descrescătoare).

  De asemenea, acest indicator permite ajustarea seriei dinamice şi elaborarea de prognoze privind evoluţia viitoare a fenomenului analizat.
- ***indicele mediu*** – reuneşte într-un singur indicator nivelurile individuale ale indicilor cu baza în lanţ calculaţi pentru o serie dinamică. Se determină ca o medie geometrică simplă a indicilor cu baza în lanţ, pe baza relaţiei: $\bar{I} = \sqrt[n-1]{\prod I_{t/t-1}} = \sqrt[n-1]{I_{t/1}}$.

  Indicele mediu se foloseşte la ajustarea seriei dinamice, precum şi la determinarea ritmului mediu.
- ***ritmul mediu*** – arată cu cât a crescut sau a scăzut în medie, pe fiecare perioadă, fenomenul analizat şi se exprimă în procente. Se calculează pe baza relaţiei: $\bar{R} = (\bar{I} - 1) \cdot 100$.
- ***valoarea medie absolută a unui procent de creştere*** – exprimă cât din modificarea medie a unui fenomen revine la un procent din ritmul mediu şi se determină pe baza relaţiei: $\bar{A} = \frac{\bar{\Delta}}{\bar{R}}$.

7.6.2. Ajustarea mecanică a seriilor cronologice

Ajustarea seriilor cronologice constă în aplicarea unor metode statistico-matematice adecvate asupra unor serii de timp în dorinţa de a extrage ceea ce este esenţial şi tipic în evoluţia fenomenului sau procesului analizat şi care prezintă caracter de lege.

În teoria şi practica statistică sunt utilizate următoarele metode de ajustare:

- ajustarea grafică;

- ajustarea mecanică;
- ajustarea analitică.

- **Ajustarea grafică** – acest procedeu presupune trasarea liberă şi aproximativă a unei drepte sau curbe asupra unei serii cronologice empirice. O asemenea ajustare are un caracter orientativ şi oferă informaţii asupra tendinţei generale a evoluţiei fenomenului sau procesului supus cercetării. Ajustarea grafică este, însă, subiectivă putând conduce la determinări diferite. Acesta este şi motivul pentru care este folosită mai rar.
- **Ajustarea mecanică** – acest procedeu constă în aplicarea succesivă, în mod mecanic, a unor formule de calcul stabilite dinainte, pentru toţi termenii seriei. În cadrul ajustării mecanice întâlnim următoarele metode: metoda mediilor eşalonate, metoda mediilor mobile, metoda sporului mediu şi metoda indicelui mediu.

- ***Metoda mediilor eşalonate*** – constă în calculul mediilor eşalonate, ca medii aritmetice simple din câte doi, trei sau mai mulţi termeni (în cadrul cărora nu se repetă nici un termen) şi aprecierea tendinţei evolutive cu ajutorul seriei formate din aceste medii. Considerând y_1 , y_2 , …, y_n nivelurile absolute dintr-o serie dată, mediile eşalonate, calculate din câte doi termeni, sunt: $\overline{y}_1 = \frac{y_1 + y_2}{2}, \overline{y}_2 = \frac{y_3 + y_4}{2}, ..., \overline{y}_{n/2} = \frac{y_{n-1} + y_n}{2}$.

  Seria mediilor eşalonate va fi: $\overline{y}_1, \overline{y}_2, ..., \overline{y}_{n/2}$.

  Pentru seriile cu un număr mare de termeni se poate continua calculul mediilor eşalonate, folosindu-se ca bază de calcul mediile deja calculate. Se obţin astfel medii de rang superior, putându-se aprecia mai exact tendinţa evolutivă. Deşi prin determinarea mediilor de rang superior sunt atenuate într-o anumită măsură fluctuaţiile evolutive generate de acţiunea factorilor întâmplători, nu este posibilă înlăturarea lor în totalitate.
- ***Metoda mediilor mobile*** – constă în determinarea tendinţei evolutive după procedeul prezentat la metoda anterioară, cu deosebirea că, în calculul mediilor, unul, doi sau mai mulţi termeni se repetă. Mediile mobile, calculate din câte doi termeni, sunt:

$$\overline{y}_1 = \frac{y_1 + y_2}{2},\ \overline{y}_2 = \frac{y_2 + y_3}{2},\ ...\ \overline{y}_{n-1} = \frac{y_{n-1} + y_n}{2}.$$

Seria mediilor mobile va fi: $\overline{y}_1, \overline{y}_2, ..., \overline{y}_{n-1}$.

Şi în acest caz pot fi determinate medii de rang superior. Nici prin această metodă nu sunt eliminate în totalitate fluctuaţiile întâmplătoare.

- ***Metoda sporului mediu*** – este o metodă mecanică de ajustare care are la bază relaţia dintre primul termen al seriei, sporul mediu şi un termen oarecare al seriei. Se foloseşte, de regulă, atunci când se obţin sporuri cu baza în lanţ cu valori apropiate. Aceasta corespunde unei creşteri a nivelurilor caracteristicii studiate sub forma unei progresii aritmetice cu raţia egală cu modificarea medie absolută. Relaţia care stă la baza ajustării prin procedeul modificării medii absolute este: $Y_i = y_1 + k \cdot \overline{\Delta}$,

  unde $i = 1, 2, 3, …, n$;

  $k = 0, 1, 2, …, n-1$;

  y_1 – reprezintă termenul luat ca bază de ajustare.

  Observăm că:

  $$Y_1 = y_1 + 0 \cdot \overline{\Delta} = y_1; \quad Y_2 = y_1 + 1 \cdot \overline{\Delta};\ Y_3 = y_1 + 2 \cdot \overline{\Delta};\ ...;\ Y_n = y_1 + (n-1) \cdot \overline{\Delta} = y_n.$$

  În cadrul acestei metode, primul şi ultimul termen ai seriei teoretice, respectiv Y_1 şi Y_n sunt identici cu primul şi ultimul termen ai seriei empirice, adică y_1 şi y_n ; această proprietate este folosită ca mijloc de control ($Y_1 = y_1$, $Y_n = y_n$).

  Cu ajutorul acestei metode sunt eliminate toate fluctuaţiile evolutive întâmplătoare, valorile teoretice Y_i înscriindu-se pe o linie dreaptă.
- ***Metoda indicelui mediu*** – este tot o metodă mecanică, uşor de aplicat, care se bazează pe relaţia existentă între primul termen al seriei, indicele mediu şi un termen oarecare al seriei. Se

folosește atunci când termenii seriei au tendința unei progresii geometrice, în care rația poate fi considerată egală cu indicele mediu al dinamicii. Relația care stă la baza ajustării prin procedeul modificării medii absolute este: $Y_i = y_1 \cdot \bar{I}^k$.

În acest caz vom avea: $Y_1 = y_1 \cdot \bar{I}^0 = y_1$; $Y_2 = y_1 \cdot \bar{I}^1$; $Y_3 = y_1 \cdot \bar{I}^2$; …; $Y_n = y_1 \cdot \bar{I}^{n-1} = y_n$.

Și în cadrul acestei metode $Y_1 = y_1$ și $Y_n = y_n$. Și cu ajutorul acestei metode sunt eliminate toate fluctuațiile evolutive întâmplătoare, valorile teoretice Y_i înscriindu-se pe o linie curbă.

BIBLIOGRAFIE

Carmen Radu, Costel Ionașcu, Murărița Ilie, *Statistică teoretică*, Editura Universitaria Craiova, 2009

Georgescu V., *Statistică descriptivă și inferențială*, Editura Universitaria, Craiova, 2006

Georgescu V., *Bazele statisticii*, Editura Reprograph, Craiova, 2001

Andrei T., *Statistică și econometrie*, Editura Economică, București, 2003