2

ANALIZA SERIILOR DE REPARTIŢIE

2.1. Indicatorii tendinţei centrale

Luarea unei decizii, în orice tip de activitate, implică necesitatea cunoaşterii acelui domeniu, respectiv a fenomenelor de masă manifestate în acel domeniu. Cu cât această cunoaştere este mai profundă, cu atât riscurile acţiunilor întreprinse sunt mai mici. Fenomenele de masă prezintă o variabilitate însemnată la nivelul formelor de manifestare, variabilitate determinată de acţiunea combinată a unui complex de factori, esenţiali sau neesenţiali, obiectivi sau subiectivi, sistematici sau întâmplători. Însă, importantă în cunoaşterea fenomenelor de masă nu este situaţia fiecărei unităţi din colectivitate, ci tendinţa manifestată de întreaga colectivitate. O posibilitate de cunoaştere a mediului economico-social o reprezintă determinarea diferiţilor indicatori statistici, dintre care un rol de seamă îl au indicatorii tendinţei centrale.

Indicatorii tendinţei centrale se determină ca indicatori medii sau indicatori de poziţie, în funcţie de natura variabilelor urmărite în colectivitatea analizată, de scopul analizei etc. Indicatorii tendinţei centrale folosiţi mai frecvent sunt mărimile medii şi indicatorii de poziţie.

2.1.1. Mărimile medii

Primul contact îl vom avea cu mărimile medii care sunt utilizate frecvent atât în activitatea de planificare şi conducere, cât şi în diversele cercetări statistice. Mărimile medii au un mare grad de aplicabilitate în activitatea practică, reprezentând, totodată, şi principale instrumente de cunoaştere a fenomenelor de masă. Aceste mărimi redau ceea ce este tipic, comun şi general, în evoluţia fenomenelor.

Aplicarea corectă a metodei mediilor necesită respectarea următoarelor condiţii:

- calcularea mediilor trebuie să se bazeze pe folosirea unui număr mare de cazuri individuale diferite sub care s-a înregistrat caracteristica, a căror variaţie este întâmplătoare în raport cu fenomenul în totalitatea lui;
- valorile din care se va calcula media să fie omogene;
- alegerea acelui tip de medie care corespunde cel mai bine formei de variaţie a caracteristicii cercetate şi informaţiilor de care dispunem.

Spre exemplu, dacă am avea următoarea situaţie a notelor studenţilor unei grupe la un examen:

Nota	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Număr studenţi	9	10	10	2	2	1	1

Dacă am calcula media obţinută de studenţi la acest examen folosind metodologia mediei aritmetice simple am obţine următorul rezultat:

$$m = \frac{4+5+6+7+8+9+10}{7} = 7 \,.$$

Rezultatul acesta ar fi corect în situaţia în care pentru fiecare notă am fi avut acelaşi număr de studenţi (5). Având în vedere că realitatea este alta, calculul corect al mediei notelor obţinute de studenţi este următorul:

$$m = \frac{4 \cdot 9 + 5 \cdot 10 + 6 \cdot 10 + 7 \cdot 2 + 8 \cdot 2 + 9 \cdot 1 + 10 \cdot 1}{35} = 5{,}57 \,.$$

Astfel, în primul caz am putea spune că nivelul de pregătire al studenţilor la acest examen a fost mediu (media este 7), în timp ce în realitate acest nivel a fost foarte scăzut (puţin peste nota de promovare – 5,57).

Din acest exemplu ne putem da seama, destul de uşor, de importanţa alegerii corecte a tipului de medie.

***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.*

Având în vedere că media este o valoare reprezentativă pentru toate nivelurile pe care le sintetizează, înseamnă că ea le poate substitui. Această substituire poate fi privită sub două aspecte:

- unul *cantitativ*, care constă în faptul că nivelul total al caracteristicii supuse cercetării, calculat prin totalizarea nivelurilor individuale nu trebuie să se schimbe atunci când aceste niveluri sunt substituite cu media lor;
- unul *calitativ*, legat de semnificaţia şi conţinutul mediei calculate, conţinut care este asigurat atunci când unităţile statistice au un grad înalt de omogenitate.

Rezultă că media cuantifică influenţa cauzelor esenţiale, făcând abstracţie de cauzele întâmplătoare. În statistică, media poate fi interpretată ca nivelul la care ar fi ajuns caracteristica înregistrată, dacă, în toate cazurile, toţi factorii esenţiali şi neesenţiali ar fi acţionat constant, deci s-ar fi obţinut o valoare identică. Ca atare, putem aprecia că media este „speranţa matematică” spre care tind toate valorile, variaţia dintre ele nefiind altceva decât influenţa factorilor aleatori. Într-adevăr, dacă fenomenele sunt de acelaşi tip calitativ, variaţia dintre ele este minimă şi ar putea fi considerată aleatoare, iar dacă sunt de tipuri diferite, atunci colectivitatea se împarte pe grupe omogene. Atunci se operează cu două tipuri de variaţie: *variaţia din interiorul grupelor*, care este influenţa factorilor aleatori (neesenţiali), şi *variaţia dintre grupe*, care este influenţa unor factori esenţiali / sistematici care structurează obiectiv întregul ansamblu pe tipuri calitative. În primul caz este o singură medie, în al doilea caz, pe lângă media ansamblului, sunt şi medii condiţionate de factorii esenţiali care structurează colectivitatea. Pentru a verifica gradul de semnificaţie a mediei este necesar să se continue cu studiul variaţiei (studiu realizat în paragraful 2.2., *Indicatorii variaţiei*).

Dată fiind marea diversitate a fenomenelor economico-sociale, precum şi complexitatea variabilităţii acestor fenomene, în practică trebuie să se aleagă tipul de medie adecvat. Mediile cel mai frecvent întâlnite sunt: aritmetică, armonică, pătratică şi geometrică, calculate ca medii simple sau ponderate în funcţie de tipul de serie asupra căreia se aplică.

2.1.1.1. 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 $x_1, x_2, ..., x_n$, şi ţinând cont că funcţia determinantă pentru media aritmetică simplă este de tip adiţional, adică:

$$x_1 + x_2 + ... + x_n = \sum x_i \ ,$$

înlocuind variantele caracteristicii cu media lor, atunci:

$$\bar{x} + \bar{x} + ... + \bar{x} = \sum x_i \Rightarrow$$

$$n \cdot \bar{x} = \sum x_i \Rightarrow$$

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

Exemplul 2.1.

Producţia obţinută de 5 firme din oraşul Craiova, în luna decembrie 2006, se prezintă astfel:

Tabelul 2.1.

Firma	**1**	**2**	**3**	**4**	**5**
Producţia realizată (mii lei)	50	65	42	74	87

Date convenţionale

Să se determine producţia medie a celor 5 firme.

Rezolvare

$$\bar{x} = \frac{\sum x_i}{n} = \frac{50 + 65 + 42 + 74 + 87}{5} = \frac{318}{5} = 63{,}6 \text{ mii lei}$$

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:

$$\left.\begin{array}{l} x_1 \cdot f_1 + x_2 \cdot f_2 + ... + x_n \cdot f_n = \sum x_i \cdot f_i \\ \bar{x} \cdot f_1 + \bar{x} \cdot f_2 + ... + \bar{x} \cdot f_n = \bar{x} \cdot \sum f_i \end{array}\right\} \Rightarrow$$

$$\bar{x} \cdot \sum f_i = \sum x_i \cdot f_i \Rightarrow$$

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

Observaţie: în cazul seriilor de distribuţie după intervale, variantele x_i vor fi date de centrele intervalelor.

Dacă în locul frecvenţelor absolute *(f_i)* se folosesc frecvenţele relative *(p_i)*, relaţia de calcul devine:

$$\bar{x} = \frac{\sum x_i \cdot p_i}{\sum p_i}$$

şi se poate scrie în următoarele două variante:

$\bar{x} = \frac{\sum x_i \cdot p_i}{100}$ - dacă p_i este exprimat în procente ($\sum p_i = 100$);

$\bar{x} = \sum x_i \cdot p_i$ - dacă p_i este exprimat în coeficienţi ($\sum p_i = 1$).

Exemplul 2.2.

Situaţia salariului lunar obţinut de angajaţii unei întreprinderi din oraşul Craiova în luna decembrie 2006 este prezentată în tabelul 2.2.

Tabelul 2.2.

Salariul lunar realizat (lei)	**Numărul de muncitori (f_i)**	**x_i**
→ 450	50	400
450 – 550	150	500
550 – 650	350	600
650 – 750	300	700
750 – 850	100	800
850 →	50	900
Total	1000	-

Date convenţionale

Să se determine salariul mediu realizat de cei 1000 angajaţi ai acestei întreprinderi.

Rezolvare

$$\bar{x} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{400 \cdot 50 + 500 \cdot 150 + 600 \cdot 350 + 700 \cdot 300 + 800 \cdot 100 + 900 \cdot 50}{1000}$$

$\bar{x} = 640$ lei

Proprietăţile mediei aritmetice

- *Media aritmetică este cuprinsă între varianta minimă şi varianta maximă, adică:* $x_{min} < \bar{x} < x_{max}$;
- *Suma abaterilor variantelor caracteristicii de la media lor este egală cu zero:*
  - $\sum (x_i - \bar{x}) = 0$ - pentru media aritmetică simplă;

  Demonstraţie:

  $$\sum (x_i - \bar{x}) = \sum x_i - \sum \bar{x} = \sum x_i - n\bar{x} = \sum x_i - n\frac{\sum x_i}{n} = 0.$$

  - $\sum (x_i - \bar{x}) f_i = 0$ - pentru media aritmetică ponderată;

  Demonstraţie:

  $$\sum (x_i - \bar{x}) f_i = \sum x_i f_i - \sum \bar{x} f_i = \sum x_i f_i - \frac{\sum x_i f_i}{\sum f_i} \sum f_i = 0.$$

- *Media aritmetică a unei variabile aleatoare X care are valorile individuale egale între ele este egală cu valoarea lor:*

$\bar{x} = x_1 = x_2 = ... = x_n$;

- *Dacă dintr-o serie X (x_1, x_2, ... x_n) construim seria X^* prin adăugarea sau scăderea unei constante **a** ($x_1 \pm a$, $x_2 \pm a$, ... $x_n \pm a$), atunci media seriei X^* va fi:*

  $\bar{x}^* = \bar{x} \pm a$;

  Demonstraţie:

$$\bar{x}^* = \frac{\sum x_i^*}{n} = \frac{\sum (x_i \pm a)}{n} = \frac{\sum x_i}{n} \pm a = \bar{x} \pm a .$$

- *Dacă dintr-o serie X (x_1, x_2, ... x_n) construim seria X^* prin mărirea sau micşorarea de k ori $\left(x_i \cdot k \text{ sau } \frac{x_i}{k} \right)$, atunci media seriei X^* se va mări sau micşora de k ori:*

  $\bar{x}^* = \bar{x} \cdot k \text{ sau } \bar{x}^* = \frac{\bar{x}}{k}$;

  Demonstraţie:

$$\bar{x}^* = \frac{\sum x_i^*}{n} = \frac{\sum x_i \cdot k}{n} = \frac{\sum x_i}{n} \cdot k = \bar{x} \cdot k .$$

$$\bar{x}^* = \frac{\sum x_i^*}{n} = \frac{\sum \frac{x_i}{k}}{n} = \frac{\sum x_i}{n \cdot k} = \frac{\bar{x}}{k} .$$

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

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

Evident, la prima vedere pare mai complicată această nouă relaţie de calcul a mediei aritmetice, însă dacă pentru o serie de distribuţie vom considera constanta ***a*** ca fiind varianta caracteristicii cu frecvenţa cea mai mare şi constanta ***k*** mărimea intervalului de variaţie, atunci valorile raportului $\frac{x_i - a}{k}$ vor fi 0 pentru varianta corespunzătoare lui ***a***, -1, -2, -3 … deasupra lui ***a*** şi 1, 2, 3 ... sub ***a***.

- *Dacă dintr-o serie X (x_1, x_2, ... x_n) construim seria X^* prin mărirea sau micşorarea de k ori a frecvenţelor corespunzătoare valorilor individuale, atunci media seriei X^* va fi egală cu cea a seriei X:*

  $\bar{x}^* = \bar{x}$;

  Demonstraţie:

$$\bar{x}^* = \frac{\sum x_i f_i^*}{\sum f_i^*} = \frac{\sum x_i \frac{f_i}{k}}{\sum \frac{f_i}{k}} = \frac{\frac{1}{k} \sum x_i f_i}{\frac{1}{k} \sum f_i} = \frac{\sum x_i f_i}{\sum f_i} = \bar{x} .$$

$$\bar{x}^* = \frac{\sum x_i f_i^*}{\sum f_i^*} = \frac{\sum x_i f_i k}{\sum f_i k} = \frac{k \sum x_i f_i}{k \sum f_i} = \frac{\sum x_i f_i}{\sum f_i} = \bar{x} .$$

- *Pentru o serie de distribuţie X (x_1, x_2, ... x_n) dacă frecvenţele sunt constante ($f_1 = f_2 = ... = f_n = r$) avem:*

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum x_i r}{\sum r} = \frac{r \sum x_i}{n \cdot r} = \frac{\sum x_i}{n};$$

- *Media aritmetică a unei variabile Z, definită ca sumă a două variabile aleatoare independente X şi Y (Z = X + Y), este egală cu suma mediilor celor două variabile:*

  $\overline{x+y} = \bar{x} + \bar{y}$;

- *Media aritmetică a unei variabile Z, definită ca produs a două variabile aleatoare independente X şi Y (Z = X · Y), este egală cu produsul mediilor celor două variabile:*

  $\overline{x \cdot y} = \bar{x} \cdot \bar{y}$;

- *În cazul în care colectivitatea generală este structurată, valoarea medie a caracteristicii studiate se calculează ca medie aritmetică ponderată a mediilor parţiale. Astfel, pentru o serie X (x_1, x_2, ... x_r, x_{r+1}... x_n) împărţită în două clase omogene de mărime f_a $\left(f_a = \sum_{i=1}^{r} f_i\right)$ şi f_b $\left(f_b = \sum_{i=r+1}^{n} f_i\right)$, pentru care vom avea mediile parţiale $\bar{x}_a$ şi $\bar{x}_b$, media va fi:*

  $$\bar{x} = \frac{f_a \bar{x}_a + f_b \bar{x}_b}{f_a + f_b};$$

  Demonstraţie:

  $$\frac{f_a \bar{x}_a + f_b \bar{x}_b}{f_a + f_b} = \frac{f_a \dfrac{\sum_{i=1}^{r} x_i f_i}{\sum_{i=1}^{r} f_i} + f_b \dfrac{\sum_{i=r+1}^{n} x_i f_i}{\sum_{i=r+1}^{n} f_i}}{f_a + f_b} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \bar{x}.$$

Exemplul 2.3.

Considerăm datele de la exemplul 2.2. Pentru determinarea mediei aritmetice, pe baza calculului simplificat, vom construi tabelul 2.3.

Tabelul 2.3.

Salariul lunar realizat (lei)	**Numărul de muncitori (f_i)**	x_i	$\frac{x_i - a}{k}$	$\frac{x_i - a}{k} \cdot f_i$
→ 450	50	400	-2	-100
450 – 550	150	500	-1	-150
550 – 650	350	600	0	0
650 – 750	300	700	1	300
750 – 850	100	800	2	200
850 →	50	900	3	150
Total	1000	-	-	400

a=600; k=100

Rezolvare

$\bar{x} = \frac{400}{1000} \cdot 100 + 600 = 640$ lei.

Principalul dezavantaj al folosirii mediei aritmetice îl constituie sensibilitatea sa faţă de valorile extreme. Ea devine nereprezentativă dacă termenii seriei sunt prea dispersaţi, iar dacă în colectivitatea statistică se observă manifestări distincte, din punct de vedere calitativ, media riscă să devină o mărime lipsită de conţinut. În acest caz, este indicat să se calculeze medii parţiale pentru fiecare tip calitativ al colectivităţii şi, în final, să se determine media generală. Omogenitatea colectivităţii pentru care se calculează media este, de fapt, o condiţie a reprezentativităţii pentru orice tip de mărime medie.

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:

$p + q = 1 \Rightarrow p = 1 - q$ şi $q = 1 - p$.

Media aritmetică în acest caz va fi:

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

$\bar{x} = p$.

Exemplul 2.4.

Dacă analizăm salariul muncitorilor din această unitate prin prisma nivelului de trai şi considerăm că un salariu sub 550 lei este necorespunzător din acest punct de vedere, iar unul peste 550 lei corespunzător, putem regrupa datele din exemplul 2.2. ca în tabelul 2.4.

Tabelul 2.4.

Salariul lunar realizat	**Numărul de muncitori (f_i)**	**Frecvenţe relative (p_i)**
necorespunzător	200	0,2
corespunzător	800	0,8
Total	1000	1

Să se determine media salariilor „necorespunzătoare”.

Rezolvare

$\bar{x} = p = 0{,}2$ (20%).

2.1.1.2. 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 $x_1, x_2, \ldots, x_n$, 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:

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

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

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

Exemplul 2.5.

Considerăm datele de la exemplul 2.2. Să se determine salariul mediu aplicând media armonică. Pentru aceasta, vom construi tabelul următor (tabelul 2.5.):

Tabelul 2.5.

Salariul lunar realizat (lei)	**Numărul de muncitori (f_i)**	x_i	$\frac{1}{x_i}$	$\frac{1}{x_i}\cdot f_i$
→ 450	50	400	0,002500	0,125000
450 – 550	150	500	0,002000	0,300000
550 – 650	350	600	0,001667	0,583333
650 – 750	300	700	0,001429	0,428571
750 – 850	100	800	0,001250	0,125000
850 →	50	900	0,001111	0,055556
Total	1000	-	-	1,617460

$$\bar{x}_h = \frac{\sum f_i}{\sum \frac{1}{x_i}\cdot f_i} = \frac{1000}{1,61746} = 618,25 \text{ lei.}$$

Observăm că pentru aceste date $\bar{x}_h < \bar{x}$.

Media armonică este mai rar folosită în practică. În schimb, mult mai 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. De asemenea, mai este folosită şi ca model matematic în calculul unor indicatori statistici, cum ar fi indicele mediu armonic al preţurilor (cazul tipic îl constituie determinarea preţului mediu al bunurilor de consum ce compun coşul zilnic, determinat pe baza bugetelor de familie ale unui eşantion reprezentativ de consumatori; de regulă, aceştia nu declară cantităţile cumpărate din fiecare produs, ci doar valoarea bunurilor consumate).

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***:

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

Proprietăţile mediei armonice

- Pentru aceeaşi serie de valori, între media aritmetică şi media armonică se verifică relaţia de ordine:

  $\bar{x}_h \leq \bar{x}$.

  Egalitatea între cele două medii are loc numai pentru serii cu valori egale.

- Dacă între două variabile există raportul de inversă proporţionalitate, $\frac{y}{x} = 1$, atunci acelaşi raport se păstrează şi între mediile calculate pentru cele două variabile. Dacă în cazul primei variabile utilizăm media aritmetică, atunci pentru cealaltă variabilă se impune folosirea mediei armonice;
- Dacă pentru o caracteristică numerică se cunoaşte seria de valori (x_i, f_i), $i=1,n$, atunci pentru determinarea nivelului mediu se va utiliza media aritmetică, iar dacă avem valorile (x_i, $x_i \cdot f_i$), $i=1,n$, se va utiliza media armonică. Mediile calculate în cele două cazuri sunt egale:

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

Exemplul 2.6.

Pentru cinci produse din aceeaşi grupă sortimentală, vândute de o întreprindere în luna decembrie 2006, s-a încasat suma de 10.000 lei, constatându-se faptul că sumele încasate la fiecare produs au fost egale. Să se determine preţul mediu de vânzare, cunoscând că preţurile de vânzare ale celor cinci produse au fost următoarele (tabelul 2.6.):

Tabelul 2.6.

Produsul	Preţul (lei/bucată)
A	5
B	4
C	1
D	2
E	3

Rezolvare

Ştiind că sumele încasate pentru cele cinci produse sunt egale, dar neavând la dispoziţie date despre cantităţile vândute, putem aplica media armonică simplă ca formă transformată a mediei aritmetice ponderate:

$$\bar{x}_h = \frac{5}{\frac{1}{5} + \frac{1}{4} + \frac{1}{1} + \frac{1}{2} + \frac{1}{3}} = 2{,}19 \text{ lei/buc.}$$

2.1.1.3. 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:

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

$$n \cdot \overline{x}_p^2 = \sum x_i^2 \Rightarrow$$

$$\overline{x}_p = \sqrt{\frac{\sum x_i^2}{n}}\,.$$

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

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

$$\overline{x}_p^2 \cdot \sum f_i = \sum x_i^2 \cdot f_i \Rightarrow$$

$$\overline{x}_p = \sqrt{\frac{\sum x_i^2 \cdot f_i}{\sum f_i}}\,.$$

Dacă pentru aceeaşi serie se calculează media aritmetică şi media pătratică, întotdeauna:

$$\overline{x} < \overline{x}_p\,.$$

Această proprietate este determinată de faptul că, în cazul mediei pătratice, variantele caracteristicii participă, prin ridicare la pătrat, la calculul mediei în mod diferenţiat, pătratul lor îndeplinind rolul de frecvenţă. Acesta este şi motivul pentru care această medie este indicată pentru analiza fenomenelor ce înregistrează tendinţe exponenţiale.

Exemplul 2.7.

Considerăm datele de la exemplul 2.2. Să se determine salariul mediu aplicând media pătratică. Pentru aceasta, vom construi tabelul următor (tabelul 2.7.):

Tabelul 2.7.

Salariul lunar realizat (lei)	**Numărul de muncitori (f_i)**	x_i	x_i^2	$\sum x_i^2 \cdot f_i$
→ 450	50	400	160000	8000000
450 – 550	150	500	250000	37500000
550 – 650	350	600	360000	126000000
650 – 750	300	700	490000	147000000
750 – 850	100	800	640000	64000000
850 →	50	900	810000	40500000
Total	1000	-	-	423000000

Date convenţionale

Rezolvare

$$\overline{x}_p = \sqrt{\frac{\sum x_i^2 \cdot f_i}{\sum f_i}} = \sqrt{\frac{423000000}{1000}} = \sqrt{423000} = 650{,}38 \text{ lei}$$

2.1.1.4. 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:

$$\left.\begin{array}{l} x_1 \cdot x_2 \cdot \ldots \cdot x_n = \Pi x_i \\ \bar{x}_g \cdot \bar{x}_g \cdot \ldots \cdot \bar{x}_g = \bar{x}_g^n \end{array}\right\} \Rightarrow$$

$$\bar{x}_g^n = \Pi x_i \Rightarrow$$

$$\bar{x}_g = \sqrt[n]{\Pi x_i} .$$

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

$$\left.\begin{array}{l} x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots \cdot x_n^{f_n} = \Pi x_i^{f_i} \\ \bar{x}_g^{f_1} \cdot \bar{x}_g^{f_2} \cdot \ldots \cdot \bar{x}_g^{f_n} = \bar{x}_g^{\sum f_i} \end{array}\right\} \Rightarrow$$

$$\bar{x}_g^{\sum f_i} = \Pi x_i^{f_i} \Rightarrow$$

$$\bar{x}_g = \sqrt[\sum f_i]{\Pi x_i^{f_i}} .$$

Prin logaritmare, statistica mediei geometrice capătă o formă similară celei a mediei aritmetice, cu deosebirea că nu se aplică termenilor seriei ca atare, ci logaritmilor acestora:

- $log\, \bar{x}_g = \dfrac{\sum log\, x_i}{n}$ - pentru serii simple şi
- $log\, \bar{x}_g = \dfrac{\sum f_i \cdot log\, x_i}{\sum f_i}$ - pentru serii de frecvenţe.

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.

Exemplul 2.8.

Considerăm datele de la exemplul 2.2. Să se determine salariul mediu aplicând media geometrică.

Rezolvare

$$\bar{x}_g = \sqrt[1000]{400^{50} \cdot 500^{150} \cdot 600^{350} \cdot 700^{300} \cdot 800^{100} \cdot 900^{50}} \Rightarrow$$

$$lg\, \bar{x}_g = \frac{50 lg\, 400 + 150 lg\, 500 + 350 lg\, 600 + 300 lg\, 700 + 100 lg\, 800 + 50 lg\, 900}{1000}$$

$lg\,\bar{x}_g = 2{,}79885 \Rightarrow \bar{x}_g = 629{,}29$ lei

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

2.1.2. 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 $\frac{n}{4}$;
- $z \in \left\{\frac{1}{10}, \frac{2}{10}, \ldots, \frac{9}{10}\right\} \Rightarrow$ cuantilele $x_{\frac{1}{10}} = x_{D_1}, x_{\frac{2}{10}} = x_{D_2}, \ldots, x_{\frac{9}{10}} = x_{D_9}$ se numesc ***decile*** şi împart seria de variaţie în zece părţi de efective egale cu $\frac{n}{10}$;
- $z \in \left\{\frac{1}{100}, \frac{2}{100}, \ldots, \frac{99}{100}\right\} \Rightarrow$ cuantilele $x_{\frac{1}{100}} = x_{P_1}, x_{\frac{2}{100}} = x_{P_2}, \ldots, x_{\frac{99}{100}} = x_{P_{99}}$ se numesc ***percentile*** şi împart seria de variaţie în o sută părţi de efective egale cu $\frac{n}{100}$.

2.1.2.1. 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:

1) se ordonează crescător sau descrescător elementele seriei;
2) 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:

$$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}.$$

Exemplul 2.9.

Fie seria de date *X={18, 27, 16, 35, 38, 44, 13}*, reprezentând numărul de puncte obţinute de 7 candidaţi la un examen. Să se determine mediana.

Rezolvare

Mai întâi ordonăm crescător seria: *X={13, 16, 18, 27, 35, 38, 44}*. Cum seria este formată dintr-un număr impar de termeni, vom avea:

Me = 27.

Dacă la seria iniţială mai adăugăm o valoare: *X={18, 27, 16, 35, 38, 44, 13, 30}*, atunci numărul termenilor seriei va deveni par şi vom avea o altă mediană. Seria ordonată crescător va fi: *X={13, 16, 18, 27, 30, 35, 38, 44}*. În acest caz mediana va fi:

$$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n+1}{2}}}{2} = \frac{27+30}{2} = 28,5$$

♦ **Pentru seriile de distribuţie** se deosebesc două posibilităţi de calcul:

A. Calculul algebric

- ***Pentru o serie de distribuţie după variante***, determinarea medianei presupune parcurgerea următoarelor etape:
  1) se determină frecvenţele cumulate crescător sau descrescător (Fc_i);
  2) determinăm unitatea mediană după relaţia:

     $$U_{Me} = \frac{n}{2};$$

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

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

Exemplul 2.10.

Considerăm notele obţinute de studenţii unei grupe la examenul de *Statistică* (tabelul 2.8.):

Tabelul 2.8.

Nota obţinută (x_i)	**Număr de studenţi (f_i)**	Fc_i
3	5	5
4	4	9
5	2	11
6	3	14
7	6	20
8	4	24
9	2	26
10	2	28
Total	**28**	**-**

$$U_{Me} = \frac{n}{2} = \frac{28}{2} = 14 \Rightarrow Me = 6.$$

- ***Pentru o serie de distribuţie pe intervale***, determinarea medianei se face parcurgând etapele următoare:
  1) se determină frecvenţele cumulate crescător sau descrescător (Fc_i);
  2) determinăm unitatea mediană după relaţia:
  $$U_{Me} = \frac{n}{2};$$
  3) se stabileşte intervalul median $I_{Me} = (x_{Me}^{inf}, x_{Me}^{sup})$, respectiv intervalul pentru care este respectată relaţia:
  $$U_{Me} \leq Fc_i;$$
  4) 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.

Exemplul 2.11.

Considerăm datele de la exemplul 2.2. Să se determine nivelul mediu cu ajutorul medianei. La tabelul iniţial mai adăugăm o coloană cu frecvenţele cumulate (tabelul 2.9.).

Tabelul 2.9.

Salariul lunar realizat (lei)	**Numărul de muncitori (f_i)**	**Frecvenţe cumulate (Fc_i)**
$\rightarrow$ 450	50	50
450 – 550	150	200
550 – 650	350	550
650 – 750	300	850
750 – 850	100	950
850 $\rightarrow$	50	1000
Total	1000	-

Rezolvare

$$U_{Me} = \frac{n}{2} = \frac{1000}{2} = 500$$

Primul interval pentru care $U_{Me} \le Fc_i$ este $I_{Me} = [550, 650]$.

$$Me = 550 + \left(\frac{1000}{2} - 200 \right) \cdot \frac{100}{350} = 550 + 85{,}71 = 635{,}71 \text{ lei}$$

Rezultă că jumătate din angajaţi obţine salarii de până la 635,71 lei, în timp ce jumătatea cealaltă obţine salarii de peste 635,71 lei.

B. Calculul grafic

Pentru determinarea medianei pe cale grafică se foloseşte ogiva (curba frecvenţelor cumulate). De pe ordonată, din dreptul lui $\frac{n}{2}$, se duce o paralelă la abscisă şi din intersecţia acesteia cu ogiva, se coboară o perpendiculară pe abscisă; punctul de întâlnire a perpendicularei cu abscisa corespunde valorii medianei.

Exemplul 2.12.

Considerăm datele de la exemplul 2.2. Să se determine grafic mediana. Acest lucru este realizat în figura 2.1. Se observă că mediana se plasează pe intervalul [70, 90].

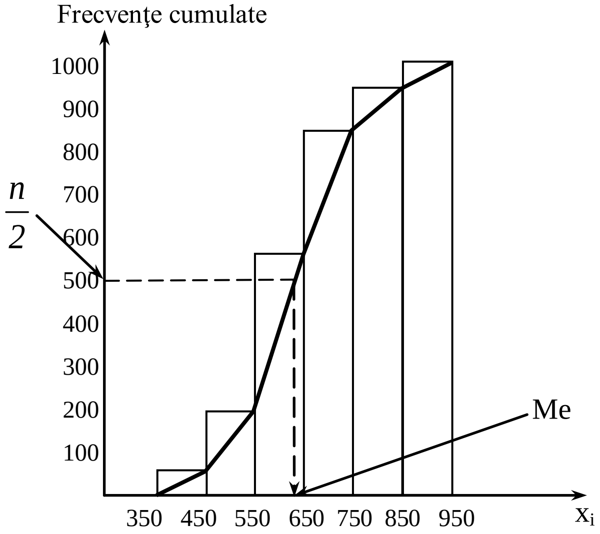


Figura 2.1. Calculul grafic al medianei.

În privinţa principalelor utilizări ale medianei menţionăm că poate fi folosită în locul mediei în aprecierea nivelului mediu al unor serii statistice, este folosită ca bază de calcul în determinarea unor indicatori ai asimetriei, sau poate fi folosită ca etalon în aprecierea gradului de semnificaţie a mediei.

2.1.2.2. Cuartilele

Există trei cuartile $\left(x_{Q_1}, x_{Q_2}, x_{Q_3}\right)$ 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:

1) 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:

$$U_{Q_h} = \frac{h \cdot n}{4}, \qquad h=1,2,3;$$

2) 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}.

Exemplul 2.13.

Considerând datele de la exemplul 2.2, să se determine cuartilele.

Rezolvare

Unităţile cuartilice sunt:

$$U_{Q_1} = \frac{1 \cdot 1000}{4} = 250;$$

$$U_{Q_2} = \frac{2 \cdot 1000}{4} = 500;$$

$$U_{Q_3} = \frac{3 \cdot 1000}{4} = 750.$$

Corespunzător acestor unităţi cuartilice vom avea intervalele:

I_{Q_1} =[550, 650];

I_{Q_2} =[550, 650];

I_{Q_3} =[650, 750].

Cele trei cuartile vor fi:

$$x_{Q_1} = 550 + \left(\frac{1 \cdot 1000}{4} - 200 \right) \cdot \frac{100}{350} = 550 + 14{,}28 = 564{,}28 \text{ lei;}$$

$$x_{Q_2} = 550 + \left(\frac{2 \cdot 1000}{4} - 200 \right) \cdot \frac{100}{350} = 550 + 85{,}71 = 635{,}71 \text{ lei;}$$

$$x_{Q_3} = 650 + \left(\frac{3 \cdot 1000}{4} - 550 \right) \cdot \frac{100}{300} = 650 + 66{,}67 = 716{,}67 \text{ lei.}$$

În concluzie, 25% dintre angajaţi au salarii sub 564,28 lei în luna decembrie 2006, 25% au avut salarii între 564,28 şi 635,71 lei, 25% între 635,71 şi 716,67 lei şi ceilalţi 25% au obţinut salarii peste 716,67 lei.

2.1.2.3. Decilele

Decilele sunt în număr de nouă ($x_{D_1}, x_{D_2}, \dots, x_{D_9}$) şi reprezintă acele valori care împart o serie de distribuţie în zece părţi cu efective egale. Observăm că decila x_{D_5} este chiar mediana, datorită faptului că $x_{D_5} = x_{\frac{5}{10}} = x_{\frac{1}{2}} = Me$.

Metodologia de calcul a decilelor este similară cu cea pe care am întâlnit-o în cazul medianei şi cuartilelor, astfel:

1) se stabileşte intervalul decilic I_{D_h} corespunzător decilei x_{D_h}. Acest interval conţine unitatea decilică U_{D_h}, unitate care se obţine astfel:

$$U_{D_h} = \frac{h \cdot n}{10};$$

2) se calculează decilele pe baza relaţiei:

$$x_{D_h} = x_{D_h}^{inf} + \left(\frac{h \cdot n}{10} - S_{D_h - 1} \right) \cdot \frac{k}{f_{D_h}}, \qquad h = \overline{1,9},$$

unde: $x_{D_h}^{inf}$ - reprezintă limita inferioară a intervalului în care se plasează decila x_{D_h};

$S_{D_h - 1}$ – reprezintă suma frecvenţelor care preced intervalul în care se plasează decila x_{D_h}: $S_{D_h - 1} = f_1 + \dots + f_{D_h - 1}$;

k – mărimea intervalului corespunzător decilei x_{D_h};

f_{D_h} – frecvenţa intervalului în care se plasează decila x_{D_h}.

Exemplul 2.14.

Considerând datele de la exemplul 2.2, să se determine decilele.

Rezolvare

Unităţile decilice sunt:

$$U_{D_1} = \frac{1 \cdot 1000}{10} = 100;$$

$$U_{D_2} = \frac{2 \cdot 1000}{10} = 200;$$

$$U_{D_3} = \frac{3 \cdot 1000}{10} = 300;$$

...

$$U_{D_9} = \frac{9 \cdot 1000}{10} = 900.$$

Corespunzător acestor unităţi decilice vom avea intervalele:

$I_{D_1} = I_{D_2} = [450, 550]$;

$I_{D_3} = I_{D_4} = I_{D_5} = [550, 650]$;

$I_{D_6} = I_{D_7} = I_{D_8} = [650, 750]$;

$I_{D_9} = [750, 850]$.

Decilele vor fi:

$$x_{D_1} = 450 + \left(\frac{1 \cdot 1000}{10} - 50 \right) \cdot \frac{100}{150} = 450 + 33{,}33 = 483{,}33 \text{ lei;}$$

$$x_{D_2} = 450 + \left(\frac{2 \cdot 1000}{10} - 50 \right) \cdot \frac{100}{150} = 450 + 100 = 550 \text{ lei;}$$

$x_{D_3} = 578{,}57$; $\quad x_{D_4} = 607{,}14$;

$x_{D_5} = 635{,}71$; $\quad x_{D_6} = 666{,}67$;

$x_{D_7} = 700$; $\quad x_{D_8} = 733{,}33$;

$x_{D_9} = 800$ lei.

2.1.2.4. Percentilele

Percentilele reprezintă acele valori care împart repartiţia în o sută de părţi cu efective egale şi sunt în număr de 99 ($x_{P_1}, x_{P_2}, \ldots, x_{P_{99}}$). Ele se folosesc numai pentru serii formate dintr-un număr foarte mare de unităţi statistice.

Referitor la percentile putem să facem următoarele observaţii:

- percentilele $x_{P_{25}}, x_{P_{50}}, x_{P_{100}}$ coincid cu cuartilele $x_{Q_1}, x_{Q_2}, x_{Q_3}$, deoarece $x_{P_{25}} = x_{\frac{25}{100}} = x_{\frac{1}{4}} = x_{Q_1}$, $x_{P_{50}} = x_{\frac{50}{100}} = x_{\frac{2}{4}} = x_{Q_2}$ şi $x_{P_{75}} = x_{\frac{75}{100}} = x_{\frac{3}{4}} = x_{Q_3}$;
- percentilele $x_{P_{10}}, x_{P_{20}}, x_{P_{30}}, x_{P_{40}}, x_{P_{50}}, x_{P_{60}}, x_{P_{70}}, x_{P_{80}}, x_{P_{90}}$ coincid cu decilele $x_{D_1}, x_{D_2}, x_{D_3}, x_{D_4}, x_{D_5}, x_{D_6}, x_{D_7}, x_{D_8}, x_{D_9}$, deoarece $x_{P_{10}} = x_{\frac{10}{100}} = x_{\frac{1}{10}} = x_{D_1}$, ... $x_{P_{90}} = x_{\frac{90}{100}} = x_{\frac{9}{10}} = x_{D_9}$;
- percentila $x_{P_{50}}$ coincide cu decila x_{D_5}, cu cuartila x_{Q_2}, respectiv cu mediana, deoarece $x_{P_{50}} = x_{\frac{50}{100}} = x_{\frac{5}{10}} = x_{\frac{2}{4}} = x_{\frac{1}{2}}$.

Metodologia de calcul a percentilelor este asemănătoare cu a celorlalte cuantile:

1) se stabilesc intervalele repartiţiei în care se găsesc variantele de rang $\frac{n}{100}$, $\frac{2n}{100}$, …, $\frac{99n}{100}$ pentru percentilele x_{P_1}, x_{D_2}, …, x_{D_9};
2) se calculează percentilele pe baza relaţiei:

$$x_{P_h} = x_{P_h}^{inf} + \left(\frac{h \cdot n}{100} - S_{P_h - 1}\right) \cdot \frac{k}{f_{P_h}}, \qquad h = \overline{1,99},$$

unde: $x_{P_h}^{inf}$ – limita inferioară a intervalului în care se plasează percentila x_{P_h};

S_{P_h-1} – suma frecvenţelor care preced intervalul în care se plasează percentila x_{P_h}:

$S_{P_h-1} = f_1 + ... + f_{P_h-1}$;

k – mărimea intervalului în care se plasează percentila x_{P_h};

f_{P_h} – frecvenţa intervalului în care se plasează percentila x_{P_h}.

În cazul distribuţiilor simetrice, între medie şi cuantile se verifică următoarele egalităţi:

$\bar{x} = Me$;

$\bar{x} - x_{Q_1} = x_{Q_3} - \bar{x}$;

$\bar{x} - x_{D_1} = x_{D_9} - \bar{x},\ \bar{x} - x_{D_2} = x_{D_8} - \bar{x}, ..., \bar{x} - x_{D_4} = x_{D_6} - \bar{x}$;

$\bar{x} - x_{P_1} = x_{P_{99}} - \bar{x},\ \bar{x} - x_{P_2} = x_{P_{98}} - \bar{x}, ..., \bar{x} - x_{P_{44}} = x_{P_{46}} - \bar{x}$.

2.1.3. Mediala

***Mediala (Md)** este un indicator de poziţie egal cu acel nivel al caracteristicii (x_i) care împarte suma termenilor seriei $\left(\sum x_i f_i\right)$ în două părţi egale.*

Mediala nu se confundă cu mediana, care reprezintă acel nivel al caracteristicii ce împarte efectivul total $\left(\sum f_i\right)$ al unei serii în două părţi egale.

Pentru aceeaşi serie de date mediala este mai mare decât mediana, cu excepţia cazului unei repartiţii simetrice, situaţie în care *Me = Md*. Compararea celor doi indicatori ne ajută la aprecierea fenomenului de concentrare.

Mediala se determină diferit în raport cu tipul seriei statistice.

♦ **Pentru o serie simplă** vom parcurge paşii următori:
1) se ordonează crescător termenii seriei;
2) se determină şirul valorilor individuale cumulate ale caracteristicii $(C_i(x_i))$;
3) determinăm unitatea medială după relaţia:

$$U_{Md} = \frac{\sum x_i}{2};$$

4) stabilim mediala, care este egală cu prima valoare din cadrul seriei de valori pentru care:

$$U_{Md} \le C_i(x_i).$$

♦ **Pentru seriile de distribuţie** se deosebesc două posibilităţi de calcul:

- ***Pentru o serie de distribuţie după variante***, pentru determinarea medialei se parcurg următoarele etape:
  1) se determină şirul produselor ($x_i f_i$) cumulate ($C_i(x_i f_i)$);
  2) determinăm unitatea medială după relaţia:

     $$U_{Md} = \frac{\sum x_i f_i}{2};$$
  3) stabilim mediala, care este egală cu prima valoare din cadrul seriei de valori pentru care:

     $$U_{Md} \le C_i(x_i f_i).$$

- ***Pentru o serie de distribuţie pe intervale de variaţie***, determinarea medialei se face parcurgând etapele următoare:
  1) se determină şirul produselor ($x_i f_i$) cumulate ($C_i(x_i f_i)$);
  2) determinăm unitatea medială după relaţia:

     $$U_{Md} = \frac{\sum x_i f_i}{2};$$
  3) stabilirea intervalului medial, respectiv a intervalului pentru care:

     $$U_{Md} \le C_i(x_i f_i);$$

     $$I_{Md} = (x_{Md}^{inf}, x_{Md}^{sup});$$
  4) se calculează mediala cu ajutorul relaţiei:

     $$Md = x_{Md}^{inf} + (U_{Md} - S_n) \cdot \frac{k}{P_{Md}},$$

     unde: x_{Md}^{inf} - reprezintă limita inferioară a intervalului în care se plasează mediala;
     S_n – reprezintă suma produselor ($x_i f_i$) care preced intervalul în care se plasează mediala;
     k – mărimea intervalului în care se plasează mediala;
     P_{Md} – produsul ($x_i f_i$) corespunzător intervalului medial.

Exemplul 2.15.

Se consideră datele de la exemplul 2.2. Să se determine mediala. Datele necesare calculului se găsesc în tabelul 2.10.

Tabelul 2.10.

Salariul lunar realizat (lei)	**Numărul de muncitori (f_i)**	x_i	$x_i f_i$	$C_i(x_i f_i)$
→ 450	50	400	20000	20000
450 – 550	150	500	75000	95000
550 – 650	350	600	210000	305000
650 – 750	300	700	210000	515000
750 – 850	100	800	80000	595000
850 →	50	900	45000	640000
Total	1000	-	640000	-

Rezolvare

$$U_{Md} = \frac{\sum x_i f_i}{2} = \frac{640000}{2} = 320000 \le C_i(x_i f_i) = 515000 \Rightarrow I_{Md} = [650;\ 750]$$

$$Md = x_{Me}^{inf} + (U_{Md} - S_n) \cdot \frac{k}{P_{Md}} = 650 + (320000 - 305000) \cdot \frac{100}{210000} = 657{,}14 \text{ lei.}$$

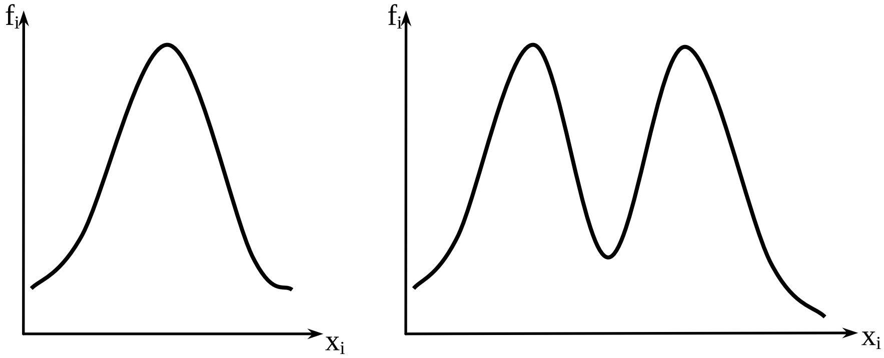
Se observă că *Md >Me (= 635,71).*

2.1.4. 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.

Pentru o serie de distribuţie putem avea una din următoarele situaţii:

- *seria de date are o singură valoare modală* – cazul în care există o singură valoare care are frecvenţa cea mai mare, iar seria se va numi unimodală (figura 2.2.a);
- *seria de date are mai multe valori modale* – există două sau mai multe valori dominante, adică frecvenţa cea mai mare corespunde la două sau mai multe variante din cadrul seriei. Seria se va numi multimodală (figura 2.2.b);
- *seria de date nu conţine valori modale* – cazul în care toate variantele au aceeaşi frecvenţă de apariţie.



a) **b)**

Figura 2.2. Tipuri de serii de repartiţie: a) unimodală; b) multimodală.

Dacă în ceea ce priveşte determinarea modului pentru seriile de distribuţie după variante lucrurile sunt clare (conform definiţiei, modul este dat de valoarea caracteristicii care are frecvenţa cea mai mare), pentru obţinerea în cazul seriilor de distribuţie după intervale trebuie să luăm în calcul şi alte elemente. Astfel, pentru seriile de intervale există mai multe posibilităţi de determinare a modului:

- pentru seriile de distribuţie cu intervale egale parcurgem etapele:
  1) determinarea intervalului modal, respectiv intervalul cu frecvenţa cea mai mare: $I_{Mo} = \left(x_{Mo}^{inf}, x_{Mo}^{sup}\right)$;
  2) determinarea modului.

Determinarea modului poate fi făcută în mai multe variante:

- *Varianta 1.* Valoarea modală se alege ca fiind centrul intervalului modal, astfel:

  $$Mo = \frac{x_{Mo}^{inf} + x_{Mo}^{sup}}{2};$$

  Deşi modul se obţine mai rapid astfel, rezultatul obţinut este aproximativ, motiv pentru care această variantă este mai rar utilizată;
- *Varianta 2.* 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: $\Delta_2 = f_{Mo} - f_{Mo+1}$;

  Această variantă este cea mai des utilizată dintre variantele de calcul algebric (primele 3 prezentate);
- *Varianta 3.* În cazul distribuţiilor unimodale simetrice, cei trei indicatori ai tendinţei centrale (media, mediana şi modul) coincid. În absenţa simetriei, însă, aceste trei valori sunt distincte. Totuşi, dacă asimetria nu este mare, cele trei puncte se găsesc într-un raport relativ constant. Distanţa dintre mod şi media aritmetică este relativ mare, în timp ce mediana se depărtează de medie cu a treia parte din distanţa care desparte media de mod (în cazul în care asimetria este pronunţată, acest raport nu se mai păstrează). Rezultă că, în cazul unei distribuţii uşor asimetrice, în care se cunosc valorile a două din cele trei mărimi, cealaltă se poate determina cu o oarecare aproximaţie. Desigur, determinarea modului pe această cale se face numai atunci când nici o altă metodă nu se poate aplica. Relaţia care există între medie, mediană şi mod este următoarea:

  $$\overline{x} - Me = \frac{1}{3}\left(\overline{x} - Mo\right),$$

  iar relaţia pe baza căreia se determină modul este:

  $$Mo = 3Me - 2\overline{x};$$
- *Varianta 4.* Constă în metoda grafică, respectiv utilizarea histogramei prin dreptunghiuri. Se unesc vârfurile coloanei maxime cu punctele de incidenţă ale acesteia cu coloanele adiacente şi din intersecţia segmentelor respective, se coboară o perpendiculară pe abscisă; valoarea corespunzătoare punctului de intersecţie al acestei perpendiculare cu abscisa reprezintă nivelul modului.

• dacă seriile de distribuţie au intervale inegale, trebuie să parcurgem următoarele etape:

1) se calculează mărimea fiecărui interval: $I_i = \left(x_i^{inf}, x_i^{sup}\right)$. Se va alege un interval etalon pentru seria de date, având lungimea intervalului de valori h;
2) se calculează pentru fiecare interval factorul de ajustare:

   $$k_i = \frac{h_i}{h};$$
3) se determină seria frecvenţelor ajustate:

   $$f_i^* = \frac{f_i}{k_i};$$

4) se determină modul prin diferite metode, precum în cazul seriei cu intervale egale.

Exemplul 2.16.

Considerăm datele de la exemplul 2.2. Să se determine, pentru aceste date, modul.

Rezolvare

Intervalul modal va fi $I_{Mo} = [550, 650]$, iar modul:

$$Mo = 550 + 100 \cdot \frac{350 - 150}{(350 - 150) + (350 - 300)} = 550 + 100 \cdot \frac{200}{250} = 630 \text{ lei.}$$

Aplicând varianta grafică se obţine figura 2.3.

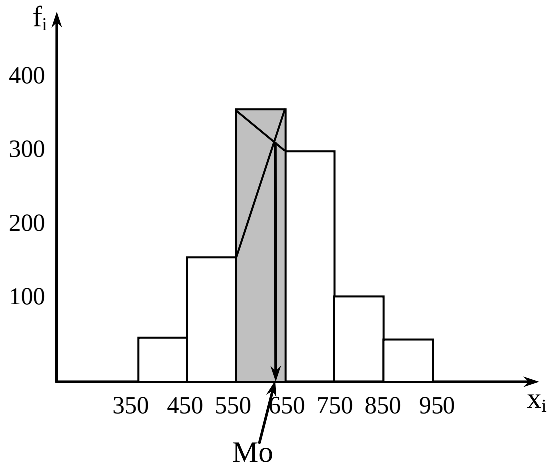


Figura 2.3. *Calculul grafic al modului.*

Modul are aceleaşi utilizări ca şi mediana; este folosit mai mult decât mediana în calculul unor indicatori ai asimetriei. Modul poate înlocui media atunci când aceasta nu se poate calcula sau nu are sens a fi calculată, ca de exemplu: numărul mediu la încălţăminte, talia medie în industria confecţiilor etc. În aceste cazuri se stabilesc ca valori modale numărul la pantofi cel mai căutat şi talia cea mai des solicitată. De asemenea modul este util când seria de date este asimetrică.

2.2. Indicatorii variaţiei

Studiul variaţiei fenomenelor economico-sociale ocupă un loc foarte important în cadrul analizei statistice. Indicatorii tendinţei centrale nu dau nici o indicaţie asupra împrăştierii, respectiv a modului în care termenii seriei se abat între ei sau de la medie (poziţia centrului de grupare). Centrul de grupare poate fi acelaşi pentru două sau mai multe serii de date, dar gradul de împrăştiere să fie diferit în jurul centrului de grupare. Spre exemplu, dacă am avea trei serii de repartiţie simetrice *X, Y* şi *Z* (figura 2.4.), ele pot avea aceeaşi medie, însă repartiţiile lor sunt diferite. Astfel, variabila *X* are o împrăştiere mai mică decât variabila *Y,* iar variabila *Y* are o împrăştiere mai mică decât variabila *Z.*

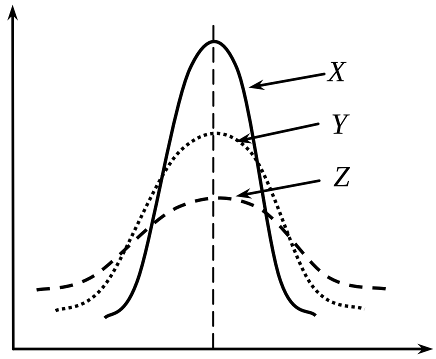


Figura 2.4. Variabile cu medie identică, dar împrăştiere diferită.

Media, prin modul său de determinare, ascunde structura colectivităţii pe grupe şi nu permite cunoaşterea abaterilor termenilor seriei (datorate cauzelor întâmplătoare) faţă de media lor. Nu este suficient să calculăm valorile tipice (indicatorii tendinţei centrale) ale unei serii, ci este necesar să verificăm şi gradul lor de reprezentativitate. Din cele prezentate rezultă necesitatea studierii variaţiei fenomenelor social-economice.

Indicatorii variaţiei utilizaţi în analizele statistice pot fi grupaţi după mai multe **criterii**:

- *după numărul variantelor luate în calcul* (sau după gradul lor de sinteză) avem indicatori simpli şi indicatori sintetici;
- *după modul de sistematizare a datelor primare*, există indicatori ai variaţiei calculaţi pentru serii de distribuţie unidimensionale şi indicatori ai variaţiei calculaţi pentru serii multidimensionale;
- *după modul de calcul şi exprimare*, există indicatori ai variaţiei calculaţi ca mărimi absolute şi ca mărimi relative.

Indiferent de natura lor, indicatorii variaţiei oferă informaţii necesare nu numai pentru cunoaşterea variabilităţii seriilor statistice analizate, ci şi pentru aprecierea „calităţii" valorilor tipice utilizate în procesul decizional.

Pentru caracterizarea variaţiei există o mare diversitate de indicatori, fiecare dintre aceştia prezentând o semnificaţie şi o metodologie de calcul specifice. Ţinând cont de gradul de sinteză a indicatorilor variaţiei, distingem, după cum arătam anterior, două categorii (indicatori simpli şi indicatori sintetici), pe care îi vom prezenta în cele ce urmează.

2.2.1. 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:

  $$A_r = \frac{A_a}{\bar{x}} \cdot 100 = \frac{X_{max} - X_{min}}{\bar{x}} \cdot 100.$$

  Amplitudinea relativă permite aprecierea şi compararea gradului de variaţie pentru două colectivităţi statistice în care caracteristica de grupare se exprimă în unităţi de măsură diferite.

Amplitudinea, fiind calculată numai pe baza valori extreme (X_{max} şi X_{min}) ale seriei, nu oferă posibilitatea cunoaşterii structurii interioare a colectivităţii (figura 2.5.). În plus, în cazul în care valorile extreme sunt neobişnuite, rezultatul la care ajungem conduce la concluzii greşite. Practic, acest indicator este folosit în prelucrarea informaţiilor – la alegerea numărului de grupe şi a mărimii intervalului de grupare (vezi capitolul 1).

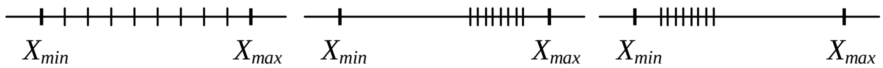


Figura 2.5. Exemple de serii cu aceeaşi amplitudine, dar cu o structură internă diferită.

- **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:

  $$da_i = x_i - \bar{x} \rightarrow \begin{cases} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \ldots \\ x_n - \bar{x} \end{cases}$$
- ***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).

Exemplul 2.17.

Considerăm datele de la exemplul 2.2. Să se determine, pentru aceste date, indicatorii simpli ai variaţiei.

Rezolvare

a) Amplitudinea variaţiei:

- Amplitudinea absolută:

$A_a = X_{max} - X_{min}$ = 950 – 350 = 600 lei

- Amplitudinea relativă:

$$A_r = \frac{A_a}{\bar{x}} \cdot 100 = \frac{600}{640} \cdot 100 = 93{,}75\%$$

Câmpul de variaţie al salariului muncitorilor este de 600 lei, aceasta reprezentând 93,75% din salariul mediu al muncitorilor.

b) Abaterile individuale:

- Abaterile individuale absolute:

$$da_i = x_i - \bar{x} \rightarrow \begin{cases} x_1 - \bar{x} = 400 - 640 = -240 \\ x_2 - \bar{x} = 500 - 640 = -140 \\ x_3 - \bar{x} = 600 - 640 = -40 \\ x_4 - \bar{x} = 700 - 640 = 60 \\ x_5 - \bar{x} = 800 - 640 = 160 \\ x_6 - \bar{x} = 900 - 640 = 260 \end{cases}$$

- Abaterile individuale relative:

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

$$dr_1 = \frac{da_1}{\bar{x}} \cdot 100 = \frac{-240}{640} \cdot 100 = -37{,}5\%;$$

$$dr_2 = \frac{da_2}{\bar{x}} \cdot 100 = \frac{-140}{640} \cdot 100 = -21{,}87\%;$$

dr_3 = - 6,25%;

dr_4 = 9,37%;

dr_5 = 25%;

dr_6 = 40,62%.

Remarcăm situaţia muncitorilor cu salarii mici (sub 450 lei) şi a celor cu salarii mari (peste 850 lei) care înregistrează cele mai mari abateri de la medie.

2.2.2. 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:

$$\bar{x} \pm \bar{d} \Rightarrow \begin{cases} \bar{x} + \bar{d} \\ \bar{x} - \bar{d} \end{cases}$$

Se poate face, însă, o distincţie între abaterile pozitive şi cele negative, astfel:

- $\bar{d}_1 = \dfrac{\sum (x_i - \bar{x})}{n_1}$ - abateri pozitive şi
- $\bar{d}_2 = \dfrac{\sum (x_j - \bar{x})}{n_2}$ - abateri negative,

unde: n_1 – numărul termenilor mai mari decât media;
n_2 – numărul termenilor mai mici decât media;
$n = n_1 + n_2$;
x_i - termeni mai mari decât media; $i=1,2,\ldots,n_1$;
x_j - termeni mai mici decât media; $j=1,2,\ldots,n_2$.

Abaterea medie liniară generală va fi:

$$\bar{d} = \frac{n_1\bar{d}_1}{n} + \left(-\frac{n_2\bar{d}_2}{n}\right) = \frac{\sum(x_i - \bar{x}) - \sum(x_j - \bar{x})}{n}.$$

Acest indicator are o valoare informaţională importantă, deoarece nivelul său avertizează asupra tendinţei evolutive a fenomenului supus analizei. Această metodă poate fi folosită pentru analiza variabilităţii unor indicatori, precum: producţia fabricată, volumul vânzărilor, volumul stocurilor etc.

Exemplul 2.18.

Considerăm datele de la exemplul 2.2. Să se determine abaterea medie liniară. Calculele intermediare necesare determinării abaterii medii liniare sunt prezentate în tabelul 2.11.

Tabelul 2.11.

Salariul lunar realizat (lei)	**Numărul de muncitori (f_i)**	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert f_i$
→ 450	50	240	12000
450 – 550	150	140	21000
550 – 650	350	40	14000
650 – 750	300	60	18000
750 – 850	100	160	16000
850 →	50	260	13000
Total	1000	900	94000

Date convenţionale

Rezolvare

$$\bar{d} = \frac{\sum \lvert x_i - \bar{x} \rvert f_i}{\sum f_i} = \frac{94000}{1000} = 94$$

Determinarea intervalului mediu de variaţie:

$$\bar{x} \pm \bar{d} \Rightarrow \begin{cases} \bar{x} + \bar{d} = 640 + 94 = 734 \\ \bar{x} - \bar{d} = 640 - 94 = 546 \end{cases}$$

Putem aprecia că, în medie, salariile acestor muncitori se plasează pe intervalul (546, 734) lei. De asemenea, pornind de la acest interval, putem determina un interval mediu al fondului de salarii, astfel încât conducerea acestei societăţi să ştie la ce nivel al cheltuielilor cu salariile să se aştepte.

- **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:

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

Dispersia, ca şi media, calculată pe baza seriilor de repartiţie după intervale, este mai puţin exactă decât în cazul folosirii datelor individuale negrupate, deoarece se calculează pe baza centrelor intervalelor, în baza ipotezei că frecvenţele sunt repartizate uniform în cadrul fiecărui interval. În practică, însă, această ipoteză este verificată foarte rar, motiv pentru care valoarea dispersiei în această situaţie este afectată de erori.

Proprietăţile dispersiei

- *Dispersia este egală cu diferenţa dintre media pătratelor şi pătratul mediei:*

$$\sigma^2 = \overline{x^2} - \bar{x}^2 .$$

Demonstraţie:

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

$$= \frac{\sum x_i^2 f_i}{\sum f_i} - 2\bar{x}\frac{\sum x_i f_i}{\sum f_i} + \frac{\bar{x}^2 \sum f_i}{\sum f_i} == \overline{x^2} - 2\bar{x}^2 + \bar{x}^2 = \overline{x^2} - \bar{x}^2$$

- *Dispersia unei caracteristici X, pentru care $x_1 = x_2 = ... = x_n$ este $\sigma^2(x) = 0$, deoarece $\bar{x} = x_i$;*
- *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:*

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

Demonstraţie:

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

$$= \overline{x^2} - 2a\bar{x} + a^2 - \bar{x}^2 + 2a\bar{x} - a^2 = \overline{x^2} - \bar{x}^2 = \sigma^2 .$$

- *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* $\boldsymbol{k^2}$ *ori, astfel:*

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

Demonstraţie:

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

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

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

Această nouă relaţie de calcul a dispersiei pare mai complicată, dar, la fel ca în cazul mediei aritmetice (cu *a* – centrul intervalului cu frecvenţa cea mai mare şi *k* – mărimea intervalului de grupare), are loc o reducere a timpului şi calculelor necesare obţinerii dispersiei.

- *Dacă dintr-o serie X (x_1, x_2, ... x_n) construim seria X^* prin micşorarea de k ori a frecvenţelor, atunci dispersia seriei X^* va fi egală cu cea a seriei X:*

  $\sigma^2_{X^*} = \sigma^2$.

  Demonstraţie:

$$\sigma^2_{X^*} = \frac{\sum (x_i - \bar{x})^2 \frac{f_i}{k}}{\sum \frac{f_i}{k}} = \frac{\frac{1}{k}\sum (x_i - \bar{x})^2 f_i}{\frac{1}{k}\sum f_i} = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} = \sigma^2 .$$

- *În cazul în care colectivitatea generală X (x_1, x_2, ... x_r, x_{r+1} ... x_n) este împărţită în două clase omogene de mărime f_a $\left(f_a = \sum_{i=1}^{r} f_i \right)$ şi f_b $\left(f_b = \sum_{i=r+1}^{n} f_i \right)$, dispersia generală este determinată în funcţie de media dispersiilor grupelor σ_a^2 şi σ_b^2 şi de dispersia mediilor parţiale $\bar{x}_a$ şi $\bar{x}_b$ de la media generală $\bar{x}$.* Prin urmare, se determină mediile parţiale $\bar{x}_a$ *şi* $\bar{x}_b$, media generală $\bar{x} = \frac{f_a \bar{x}_a + f_b \bar{x}_b}{f_a + f_b}$ şi dispersiile parţiale σ_a^2 *şi* σ_b^2.

  Dispersia generală va fi:

$$\sigma^2 = \frac{f_a \sigma_a^2 + f_b \sigma_b^2}{f_a + f_b} + \frac{f_a (\bar{x}_a - \bar{x})^2 + f_b (\bar{x}_b - \bar{x})^2}{f_a + f_b} .$$

  Demonstraţie:

  Dispersiile parţiale sunt date de relaţiile următoare:

$$\sigma_a^2 = \frac{\sum_{i=1}^{r} (x_i - \bar{x}_a)^2 f_i}{f_a} \text{ şi } \sigma_b^2 = \frac{\sum_{i=r+1}^{n} (x_i - \bar{x}_b)^2 f_i}{f_b} .$$

  Aplicând proprietatea 3, relaţiile de mai sus, pentru $a = \bar{x}$, devin:

$$\sigma_a^2 = \frac{\sum_{i=1}^{r} (x_i - \bar{x})^2 f_i}{f_a} - (\bar{x}_a - \bar{x})^2 \text{ şi } \sigma_b^2 = \frac{\sum_{i=r+1}^{n} (x_i - \bar{x})^2 f_i}{f_b} - (\bar{x}_b - \bar{x})^2 .$$

  Dispersia generală este:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2 f_i}{n} = \frac{\sum_{i=1}^{r} (x_i - \bar{x})^2 f_i + \sum_{i=r+1}^{n} (x_i - \bar{x})^2 f_i}{n} =$$

$$= \frac{\frac{\sum_{i=1}^{r} (x_i - \bar{x})^2 f_i}{f_a} \cdot f_a + \frac{\sum_{i=r+1}^{n} (x_i - \bar{x})^2 f_i}{f_b} \cdot f_b}{n} =$$

$$= \frac{\sigma_a^2 \cdot f_a + (\bar{x}_a - \bar{x})^2 \cdot f_a + \sigma_b^2 \cdot f_b + (\bar{x}_b - \bar{x})^2 \cdot f_b}{f_a + f_b} =$$

$$= \frac{\sigma_a^2 \cdot f_a + \sigma_b^2 \cdot f_b}{f_a + f_b} + \frac{(\bar{x}_a - \bar{x})^2 \cdot f_a + (\bar{x}_b - \bar{x})^2 \cdot f_b}{f_a + f_b} .$$

- *Pentru colectivităţi de volum redus, dispersia se determină după relaţia:*

$$\sigma^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}.$$

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, $\overline{x} = p$. Dispersia va fi:

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

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

$$\sigma^2 = pq.$$

Exemplul 2.19.

Considerând datele de la exemplul 2.2, să se determine dispersia utilizând atât relaţia de calcul obişnuit, cât şi relaţia de calcul simplificat. Datele necesare calculelor sunt:

Tabelul 2.12.

x_i	f_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2 \cdot f_i$	$\frac{x_i - a}{k}$	$\left(\frac{x_i - a}{k}\right)^2$	$\left(\frac{x_i - a}{k}\right)^2 \cdot f_i$
400	50	-240	2880000	-2	4	200
500	150	-140	2940000	-1	1	150
600	350	-40	560000	0	0	0
700	300	60	1080000	1	1	300
800	100	160	2560000	2	4	400
900	50	260	3380000	3	9	450
-	1000	-	13400000	-	-	1500

Pentru calculul obişnuit, avem:

$$\sigma^2 = \frac{\sum (x_i - \overline{x})^2 f_i}{\sum f_i} = \frac{13400000}{1000} = 13400$$

Pe baza calculului simplificat, avem:

$$\sigma^2 = \frac{\sum \left(\frac{x_i - a}{k}\right)^2 f_i}{\sum f_i} \cdot k^2 - (\overline{x} - a)^2 = \frac{1500}{1000} \cdot 10000 - (640 - 600)^2 =$$

$$= 15000 - 1600 = 13400.$$

Se observă acelaşi nivel al dispersiei pentru ambele procedee (calcul obişnuit şi calcul simplificat).

Dacă analizăm salariul muncitorilor din această unitate prin prisma nivelului de trai şi considerăm că un salariu sub 550 lei este necorespunzător din acest punct de vedere, iar unul peste 550 lei corespunzător, putem regrupa datele din exemplul 2.2. ca în tabelul 2.4. În acest caz, dispersia se determină ţinând cont de caracteristicile variabilei alternative a lui Bernoulli. Astfel, dispersia va fi:

$$\sigma^2 = pq = 0,2 \cdot 0,8 = 0,16 .$$

- **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*:

- $\sigma = \sqrt{\sigma^2} = \sqrt{\dfrac{\sum (x_i - \overline{x})^2}{n}}$ - pentru serii simple;
- $\sigma = \sqrt{\sigma^2} = \sqrt{\dfrac{\sum (x_i - \overline{x})^2 f_i}{\sum f_i}}$ - 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 σ este mic.

La fel ca şi abaterea medie liniară, abaterea standard poate fi folosită pentru determinarea intervalului mediu de variaţie:

$$\overline{x} \pm \sigma \Rightarrow \begin{cases} \overline{x} + \sigma \\ \overline{x} - \sigma \end{cases}$$

În analiza variaţiei fenomenelor economico-sociale, pentru aceeaşi serie de date abaterea standard este mai mare decât abaterea medie liniară ($\sigma > \overline{d}$), rezultând un interval mediu de variaţie mai mare pentru abaterea standard, motiv pentru care este preferat acest indicator.

Dezavantajul abaterii standard constă în faptul că se exprimă în aceeaşi unitate de măsură ca şi variantele caracteristicii. Ea nu permite compararea variaţiei a două colectivităţi în care caracteristica se exprimă în unităţi de măsură diferite. De asemenea, oferă o imagine deformată asupra mărimii variaţiei atunci când se compară două colectivităţi de acelaşi fel în care diferă ordinul de mărime al caracteristicii studiate.

Abaterea standard a variabilei alternative

Abaterea standard pentru variabila alternativă este:

$$\sigma = \sqrt{\sigma^2} = \sqrt{pq} .$$

Frecvenţele relative *p* şi *q* sunt mărimi complementare şi, ca atare, atunci când *p* creşte *q* scade cu aceeaşi valoare cu care *p* a crescut. Datorită acestui lucru, dispersia şi abaterea standard capătă o serie de însuşiri importante şi deosebit de utile pentru folosirea acestor indicatori în analiza statistică, mai ales în practica sondajelor sociologice şi în studiul şi controlul calităţii produselor.

Pornind de la constatarea anterioară (*p* şi *q* sunt mărimi complementare), precum şi de la faptul că atât dispersia, cât şi abaterea standard se calculează doar cu ajutorul acestor frecvenţe, se pot foarte uşor reprezenta grafic ambii indicatori ai variaţiei pentru variabila alternativă (figura 2.6).

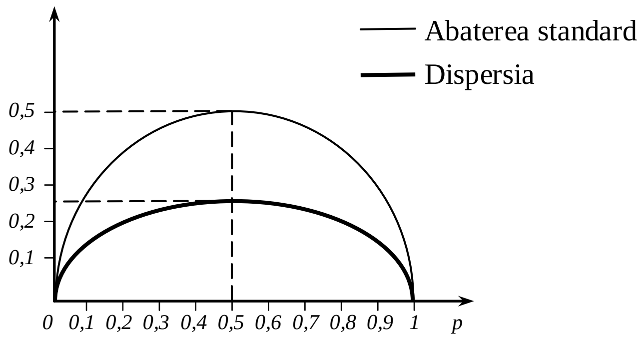


Figura 2.6. Dispersia şi abaterea standard a caracteristicii alternative.

Exemplul 2.20.

Pentru datele de la exemplul 2.2. abaterea standard se determină foarte simplu:

$$\sigma = \sqrt{\sigma^2} = \sqrt{13400} = 115{,}76.$$

Intervalul mediu calculat pe baza abaterii standard este:

$$\bar{x} \pm \sigma \Rightarrow \begin{cases} \bar{x} + \sigma = 640 + 115{,}76 = 755{,}76 \\ \bar{x} - \sigma = 640 - 115{,}76 = 524{,}24 \end{cases}$$

Se observă că intervalul obţinut prin utilizarea abaterii standard (524,24; 755,76) este mai larg decât intervalul rezultat în urma utilizării abaterii medii liniare (546, 734).

Pentru cazul variabilei alternative considerate la exemplul anterior, abaterea standard va fi:

$$\sigma = \sqrt{\sigma^2} = \sqrt{pq} = \sqrt{0{,}16} = 0{,}4\,.$$

- **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:

$$Cv = \frac{\sigma}{\bar{x}} \cdot 100.$$

Coeficientul de variaţie arată câte unităţi din abaterea standard revin la 100 unităţi de medie. Coeficientul de variaţie ia valori între 0 - 100%. Dacă $Cv = 0$, înseamnă că avem de-a face cu o lipsă de variaţie, toate valorile caracteristicii fiind egale între ele şi, respectiv, egale cu media. Dacă $Cv \to 0$ înseamnă că variaţia caracteristicii este mică, colectivitatea cercetată este omogenă, media este reprezentativă, iar gruparea este bine executată. În general, se admite că seria prezintă un grad de omogenitate ridicat dacă $Cv < 35\%$, iar dacă $Cv > 70\text{-}75\%$, se afirmă că variaţia este foarte mare, media nu este semnificativă şi ascunde o structură eterogenă a colectivităţii care necesită repetarea operaţiei de grupare cu respectarea strictă a principiilor teoretice care condiţionează reuşita operaţiei de omogenizare a grupelor. De cele mai multe ori, în asemenea cazuri este necesară împărţirea seriei iniţiale în serii componente pentru a spori gradul de omogenitate a datelor.

Acest indicator urmăreşte, în principal, următoarele:
- verificarea reprezentativităţii mediei variabilei analizate;
- compararea omogenităţii seriilor de date. Astfel, ierarhia coeficienţilor de variaţie ai seriilor de date defineşte ordinea acestora după gradul de omogenitate.

Exemplul 2.21.

Pentru datele de la exemplul 2.2. coeficientul de variaţie se determină imediat:

$$Cv = \frac{\sigma}{\overline{x}} \cdot 100 = \frac{115{,}76}{640} \cdot 100 = 18{,}09\% .$$

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

Utilizarea coeficientului de variaţie se face, însă, cu multă precauţie, întrucât valoarea lui este determinată nu numai de nivelul abaterii standard, ci şi de nivelul mediei. Prezentăm, în acest sens, graficul a două distribuţii X_1 şi X_2 (figura 2.7.), în care a doua distribuţie este obţinută din prima prin translarea termenilor acesteia, astfel încât media celei de-a doua să fie $\overline{x}_2 = 2\overline{x}_1$. Datorită translării, abaterile standard rămân egale: $\sigma_1 = \sigma_2$.

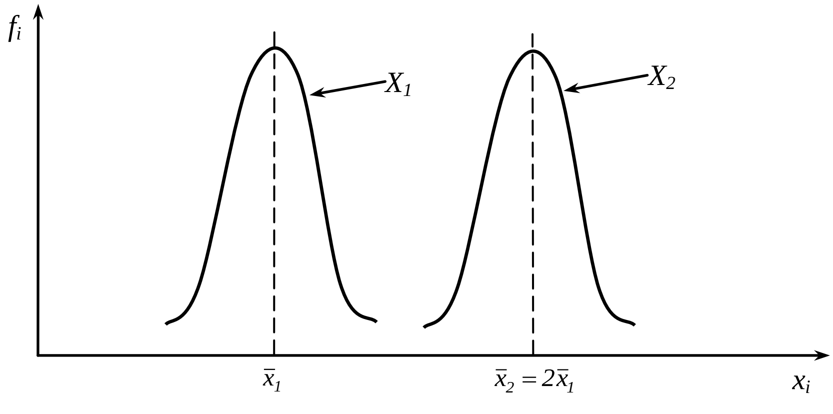


Figura 2.7. Variabile cu medii diferite, dar împrăştiere identică.

Coeficienţii de variaţie, pentru cele două serii, sunt:

$$Cv_1 = \frac{\sigma_1}{\overline{x}_1} \cdot 100,$$

$$Cv_2 = \frac{\sigma_2}{\overline{x}_2} \cdot 100 = \frac{\sigma_1}{2\overline{x}_1} \cdot 100 = \frac{Cv_1}{2} .$$

Deci, la grade identice de dispersare ($\sigma_1 = \sigma_2$) corespund valori diferite ale coeficienţilor de variaţie, primul coeficient fiind de două ori mai mare decât al doilea, deoarece media pentru a doua serie este de două ori mai mare decât prima medie.

2.2.3. 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.

Indicatorii de variaţie intercuantilică se calculează în mărimi absolute şi în mărimi relative. Indicatorii intercuantilici cel mai frecvent utilizaţi sunt: abaterea intercuartilică, coeficientul de variaţie intercuartilică, abaterea interdecilică, coeficientul de variaţie interdecilică.

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

În seriile simetrice abaterea dintre cuartila inferioară şi mediană este egală cu abaterea dintre cuartila superioară şi mediană, iar în interiorul lor se găsesc 50% din numărul cazurilor înregistrate. Ţinând seama de ordinea de creştere a valorilor celor trei cuartile pentru o serie perfect simetrică, putem scrie:

$$Me - x_{Q_1} = x_{Q_3} - Me .$$

În acest caz, media aritmetică a celor două cuartile extreme este egală cu valoarea cuartilei a doua, adică cu mediana seriei:

$$Q = \frac{x_{Q_1} + x_{Q_3}}{2} = x_{Q_2} = Me .$$

Dacă ultimele două relaţii nu sunt verificate, adică $Me - x_{Q_1} \neq x_{Q_3} - Me$ şi $Q \neq Me$, înseamnă că seria prezintă un anumit grad de variaţie intercuartilică, grad care poate şi trebuie să fie măsurată statistic.

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:

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

Dacă seria prezintă un grad mai mare de asimetrie este necesar să se calculeze şi variaţia interdecilică.

- ***Abaterea interdecilică (Q_d)***

Abaterea interdecilică se bazează pe aceleaşi considerente întâlnite anterior, adică într-o serie perfect simetrică distanţele dintre decilele extreme şi mediană sunt egale:

$$Me - x_{D_1} = x_{D_9} - Me ,$$

şi este egală cu media aritmetică a abaterilor decilelor extreme faţă de cuartila centrală a seriei, astfel:

$$Q_d = \frac{(Me - x_{D_1}) + (x_{D_9} - Me)}{2} = \frac{x_{D_9} - x_{D_1}}{2}.$$

Nici acest indicator nu permite comparaţia statistică a mai multor serii, motiv pentru care a fost introdus ***coeficientul de variaţie interdecilică***. Acesta se calculează ca raport între abaterea interdecilică şi valoarea mediană, astfel:

$$V_Q = \frac{Q_d}{Me} \cdot 100 = \frac{x_{D_9} - x_{D_1}}{2Me} \cdot 100.$$

De regulă, calculul variaţiei interdecilice se face pentru serii statistice cu un număr foarte mare de grupe şi cu tendinţă evidentă de asimetrie.

Avantajul major al acestor indicatori îl reprezintă faptul că nu sunt sensibili la existenţa valorilor aberante.

Exemplul 2.22.

Pornind de la exemplul 2.2.şi ţinând cont şi de rezultatele de la exemplele 2.13. ($x_{Q_1} = 564{,}28$ lei; $x_{Q_2} = 635{,}71$ lei; $x_{Q_3} = 716{,}67$ lei) şi 2.14. ($x_{D_1} = 483{,}33$ lei; $x_{D_9} = 800$ lei), să se determine abaterile intercuantilice.

Rezolvare

Abaterea intercuartilică este:

$$Q_c = \frac{x_{Q_3} - x_{Q_1}}{2} = \frac{716{,}67 - 564{,}28}{2} = 76{,}19 \text{ lei.}$$

Coeficientul de variaţie intercuartilică este:

$$V_Q = \frac{Q_c}{Me} \cdot 100 = \frac{76{,}19}{635{,}71} \cdot 100 = 11{,}98\%.$$

Abaterea interdecilică este:

$$Q_d = \frac{x_{D_9} - x_{D_1}}{2} = \frac{800 - 483{,}33}{2} = 158{,}33 \text{ milioane lei.}$$

Coeficientul de variaţie interdecilică este:

$$V_D = \frac{Q_d}{Me} \cdot 100 = \frac{158{,}33}{635{,}71} \cdot 100 = 24{,}91\%.$$

Cum valorile coeficienţilor de variaţie intercuantilică (V_Q şi V_D) sunt mici (ambele mai mici de 35%) putem trage concluzia că seria prezintă un grad de omogenitate ridicat, iar media este reprezentativă.

2.2.4. Momentele

Vom arăta în continuare că indicatorii media şi dispersia nu reprezintă altceva decât cazuri particulare de valori ce semnifică momentele unei serii statistice.

Numim ***moment de ordinul t*** în raport cu o valoare cunoscută ***a***, parametrul:

$$m_t(a) = \frac{\sum (x_i - a)^t f_i}{\sum f_i}.$$

În funcţie de valorile pe care le ia ***a*** putem avea următoarele tipuri de momente:

- *momente iniţiale (m_t^0)* – în acest caz $a=0$, iar relaţia de calcul a momentelor este:

  $$m_t^0 = \frac{\sum x_i^t f_i}{\sum f_i};$$

- *momente centrate (μ_t)* – în această situaţie $a = \bar{x}$ şi se determină pe baza relaţiei:

  $$\mu_t = \frac{\sum (x_i - \bar{x})^t f_i}{\sum f_i};$$

- *momente obişnuite sau ordinare (m_t)* – sunt cele în care $a \neq 0$ şi $a \neq \bar{x}$, iar relaţia de calcul este cea descrisă iniţial.

Între momentele centrate şi cele obişnuite pot fi stabilite o serie de relaţii. Pentru determinarea acestor relaţii vom face următoarele notaţii:

$$\left.\begin{array}{l} x_i - a = p \\ x_i - \bar{x} = r \\ \bar{x} - a = s \end{array}\right\} \Rightarrow (x_i - a) = (x_i - \bar{x}) + (\bar{x} - a); \;\; p = r + s.$$

Momentul obişnuit de ordinul t va fi:

$$m_t = \frac{\sum p^t f_i}{\sum f_i} = \frac{\sum (r+s)^t f_i}{\sum f_i} = \frac{\sum r^t f_i}{\sum f_i} + \frac{C_t^1 s \sum r^{t-1} f_i}{\sum f_i} + \frac{C_t^2 s^2 \sum r^{t-2} f_i}{\sum f_i} + \ldots + s^t =$$

$$m_t = \mu_t + C_t^1 s \mu_{t-1} + C_t^2 s^2 \mu_{t-2} + \ldots + s^t.$$

În mod asemănător, momentul centrat de ordinul t va fi:

$$\mu_t = \frac{\sum r^t f_i}{\sum f_i} = \frac{\sum (p-s)^t f_i}{\sum f_i} = \frac{\sum p^t f_i}{\sum f_i} - \frac{C_t^1 s \sum p^{t-1} f_i}{\sum f_i} +$$

$$+ \frac{C_t^2 s^2 \sum p^{t-2} f_i}{\sum f_i} + \ldots + (-1)^t s^t =$$

$$\mu_t = m_t - C_t^1 s m_{t-1} + C_t^2 s^2 m_{t-2} + \ldots + (-1)^t s^t.$$

Momentele iniţiale pentru diferite valori ale lui t sunt:

- $t=0 \rightarrow m_0^0 = \frac{\sum x_i^0 f_i}{\sum f_i} = \frac{\sum f_i}{\sum f_i} = 1;$
- $t=1 \rightarrow m_1^0 = \frac{\sum x_i^1 f_i}{\sum f_i} = \bar{x};$
- $t=2 \rightarrow m_2^0 = \frac{\sum x_i^2 f_i}{\sum f_i} = \sigma^2 + \bar{x}^2.$

Momentele obişnuite pentru diferite valori ale lui t sunt:

- $t=0 \rightarrow m_0 = \frac{\sum (x_i - a)^0 f_i}{\sum f_i} = \frac{\sum p^0 f_i}{\sum f_i} = \frac{\sum f_i}{\sum f_i} = 1;$
- $t=1 \rightarrow m_1 = \frac{\sum (x_i - a)^1 f_i}{\sum f_i} = \frac{\sum x_i f_i}{\sum f_i} - \frac{a \sum f_i}{\sum f_i} = \bar{x} - a;$
- $t=2 \rightarrow m_2 = \frac{\sum (x_i - a)^2 f_i}{\sum f_i} = \frac{\sum (r+s)^2 f_i}{\sum f_i} = \mu_2 + 2s\mu_1 + s^2 \Rightarrow m_2 = \sigma^2 + s^2;$

- $t=3 \rightarrow m_3 = \dfrac{\sum(x_i - a)^3 f_i}{\sum f_i} = \dfrac{\sum(r+s)^3 f_i}{\sum f_i} = \mu_3 + 3s\mu_2 + 3s^2\mu_1 + s^3 \Rightarrow$

$$m_3 = \mu_3 + 3s\mu_2 + s^3.$$

Pentru diferite valori ale lui *t* momentele centrate sunt:

- $t=0 \rightarrow \mu_0 = \dfrac{\sum(x_i - \bar{x})^0 f_i}{\sum f_i} = \dfrac{\sum r^0 f_i}{\sum f_i} = \dfrac{\sum f_i}{\sum f_i} = 1;$
- $t=1 \rightarrow \mu_1 = \dfrac{\sum(x_i - \bar{x})^1 f_i}{\sum f_i} = \dfrac{\sum(p-s)^1 f_i}{\sum f_i} = 0$, deoarece suma abaterilor de la medie este nulă (a se vedea, în acest sens, proprietăţile mediei aritmetice);
- $t=2 \rightarrow \mu_2 = \dfrac{\sum(x_i - \bar{x})^2 f_i}{\sum f_i} = \sigma^2$ (din definiţia dispersiei);
- $t=3 \rightarrow \mu_3 = \dfrac{\sum(x_i - \bar{x})^3 f_i}{\sum f_i} = \dfrac{\sum(p-s)^3 f_i}{\sum f_i} = m_3 - 3sm_2 + 3s^2 m_1 - s^3 \Rightarrow$

$$\mu_3 = m_3 - 3sm_2 + 2s^3.$$

Momentele prezentate sunt folosite foarte frecvent în statistică, atât ca bază de calcul în determinarea unor indicatori specifici seriilor de distribuţie (excesul), cât şi ca modalitate practică de simplificare a calculului unuia dintre indicatorii sintetici utilizaţi în analiza statistică (coeficientul de corelaţie liniară). La rândul lor, momentele pot fi determinate cu ajutorul procedeului de calcul simplificat. Acest procedeu nu diferă, în principiu, de cel prezentat la calculul mediei aritmetice ponderate şi dispersiei, astfel:

- $m_t = \dfrac{\sum\left(\dfrac{x_i - a}{k}\right)^t f_i}{\sum f_i} \cdot k^t.$

Cu ajutorul acestor rezultate vor fi calculate foarte uşor momentele centrate.

Exemplul 2.23.

Să se determine momentele iniţiale, obişnuite şi centrate pentru diferite valori ale lui *t*, corespunzătoare distribuţiei prezentate în exemplul 2.2.

Rezolvare

a) Momentele iniţiale sunt:

- $m_0^0 = \dfrac{\sum x_i^0 f_i}{\sum f_i} = \dfrac{\sum f_i}{\sum f_i} = 1;$
- $m_1^0 = \dfrac{\sum x_i^1 f_i}{\sum f_i} = \bar{x} = 640;$
- $m_2^0 = \dfrac{\sum x_i^2 f_i}{\sum f_i} = \sigma^2 + \bar{x} = 13400 + 640 = 14040.$

b) Vom utiliza calculul simplificat pentru obţinerea momentelor obişnuite. Pentru aceasta vom construi tabelul 2.13.

Tabelul 2.13.

x_i	f_i	$\frac{x_i-a}{k}$	$\left(\frac{x_i-a}{k}\right)\cdot f_i$	$\left(\frac{x_i-a}{k}\right)^2\cdot f_i$	$\left(\frac{x_i-a}{k}\right)^3\cdot f_i$
400	50	-2	-100	200	-400
500	150	-1	-150	150	-150
600	350	0	0	0	0
700	300	1	300	300	300
800	100	2	200	400	800
900	50	3	150	450	1350
-	1000	-	400	1500	1900

a = 600; k = 100

Momentele obişnuite sunt:

- $m_0=\frac{\sum(x_i-a)^0 f_i}{\sum f_i}=\frac{\sum p^0 f_i}{\sum f_i}=\frac{\sum f_i}{\sum f_i}=1$;
- $m_1=\frac{\sum\left(\frac{x_i-a}{k}\right)f_i}{\sum f_i}\cdot k=\frac{400}{1000}\cdot 100=40$;
- $m_2=\frac{\sum\left(\frac{x_i-a}{k}\right)^2 f_i}{\sum f_i}\cdot k^2=\frac{1500}{1000}\cdot 10000=15000$;
- $m_3=\frac{\sum\left(\frac{x_i-a}{k}\right)^3 f_i}{\sum f_i}\cdot k^3=\frac{1900}{1000}\cdot 1000000=1900000.$

c) Momentele centrate sunt:

- $\mu_0=\frac{\sum(x_i-\bar{x})^0 f_i}{\sum f_i}=\frac{\sum r^0 f_i}{\sum f_i}=\frac{\sum f_i}{\sum f_i}=1$;
- $\mu_1=\frac{\sum(x_i-\bar{x})^1 f_i}{\sum f_i}=0$;
- $\mu_2=\frac{\sum(x_i-\bar{x})^2 f_i}{\sum f_i}=\sigma^2=13400$;
- $\mu_3=m_3-3sm_2+2s^3=1900000-3\cdot 40\cdot 15000+2\cdot 40^3=228000.$

2.2.5. 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 (tabelul 2.14.):

Tabelul 2.14.

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
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	f_{i1}	f_{i2}	...	f_{ij}	...	f_{im}	f_i	$\bar{y}_i$	σ_i^2
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$	$\vdots$
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:

$$\sum_{j=1}^{m} f_{ij} = f_{i1} + f_{i2} + ... + f_{ij} + ... + f_{im} = f_i .$$

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

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

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:

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

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

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

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:

$$\overline{\sigma}^2 = \frac{\sum_{i=1}^{n}\sigma_i^2 f_i}{\sum_{i=1}^{n} f_i}.$$

- **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:

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

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:

$$\bar{y}_0 = \frac{\sum_{j=1}^{m} y_j f_j}{\sum_{j=1}^{m} f_j} = \frac{\sum_{i=1}^{n} \bar{y}_i f_i}{\sum_{i=1}^{n} f_i}.$$

În exemplul considerat la dispersia de grupă, dacă presupunem că variaţia vechimii în muncă nu ar avea nici o influenţă asupra salariului, atunci mediile de grupă ar trebui să fie egale. Într-o astfel de situaţie, media generală ar fi egală cu mediile de grupă, iar dispersia dintre grupe ar fi nulă. Însă, vechimea în muncă este unul din factorii importanţi de influenţă asupra salariului, iar această influenţă este cuantificată de către indicatorul dispersia dintre grupe. Nivelul acestui indicator este cu atât mai mare cu cât influenţa vechimii în muncă este mai consistentă.

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** $\left(\sigma_0^2\right)$ – 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:**

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

Regula de adunare a dispersiilor mai este utilă şi pentru a calcula o dispersie atunci când se cunosc celelalte două dispersii.

Exemplul 2.24.

Distribuţia muncitorilor unei societăţi comerciale din oraşul Craiova după salariu (lei) şi vechime (ani) în luna decembrie 2006 este redată în tabelul 2.15.

Tabelul 2.15.

X Vechime \ Salariul Y		→ 450	450 – 550	550 – 650	650 – 750	750 – 850	850 →	Total f_x
		400	500	600	700	800	900	
→ 15	10	20	30	30	20	-	-	100
15-25	20	30	80	100	90	30	-	330
25-35	30	-	40	140	120	50	30	380
35 →	40	-	-	80	70	20	20	190
Total f_y		50	150	350	300	100	50	1000

Date convenţionale

Să se determine indicatorii variaţiei pentru această serie.

Rezolvare

a) Dispersia de grupă

Pentru determinarea dispersiilor de grupă vom avea nevoie de mediile de grupă:

$$\bar{y}_1 = \frac{\sum_{j=1}^{m} y_j f_{1j}}{\sum_{j=1}^{m} f_{1j}} = \frac{400\cdot 20 + 500\cdot 30 + 600\cdot 30 + 700\cdot 20}{100} = \frac{55000}{100} = 550$$

$$\bar{y}_2 = \frac{\sum_{j=1}^{m} y_j f_{2j}}{f_2} = \frac{400\cdot 30 + 500\cdot 80 + 600\cdot 100 + 700\cdot 90 + 800\cdot 30}{330} = \frac{199000}{330} = 603{,}03$$

$$\bar{y}_3 = \frac{\sum_{j=1}^{m} y_j f_{3j}}{f_3} = \frac{500\cdot 40 + 600\cdot 140 + 700\cdot 120 + 800\cdot 50 + 900\cdot 30}{380} = \frac{255000}{380} = 671{,}05$$

$$\bar{y}_4 = \frac{\sum_{j=1}^{m} y_j f_{4j}}{f_4} = \frac{600\cdot 80 + 700\cdot 70 + 800\cdot 20 + 900\cdot 20}{190} = \frac{131000}{190} = 689{,}47$$

Dispersiile de grupă vor fi:

$$\sigma_1^2 = \frac{\sum_{j=1}^{m} (y_j - \bar{y}_1)^2 f_{1j}}{f_1} = \frac{(400-550)^2 20 + (500-550)^2 30 + (600-550)^2 30 +}{12}$$

$$\frac{+(700-550)^2 20}{100} = \frac{1050000}{100} = 10500$$

$\sigma_2^2 = 12415{,}06$; $\sigma_3^2 = 11530{,}47$; $\sigma_4^2 = 9362{,}88$

Dispersie mai mică apare pentru grupa 4, respectiv grupa de vechime de peste 35 ani. Urmează grupele 1, 3 şi 2 (în grupa 2 dispersia fiind maximă). Deci, pentru ultimele două grupe factorii întâmplători au influenţat puternic nivelul salariului.

b) Media dispersiilor de grupă

$$\bar{\sigma}^2=\frac{\sum_{i=1}^{n}\sigma_i^2 f_i}{\sum_{i=1}^{n} f_i}=\frac{10500{,}00\cdot 100+12415{,}06\cdot 330+11530{,}47\cdot 380+9362{,}88\cdot 190}{1000}\Rightarrow$$

$$\bar{\sigma}^2=11307{,}5$$

c) Dispersia dintre grupe

Pentru determinarea acestui indicator avem nevoie de media generală:

$$\bar{y}_0=\frac{\sum_{j=1}^{m} y_j f_j}{\sum_{j=1}^{m} f_j}=\frac{400\cdot 50+500\cdot 150+600\cdot 350+700\cdot 300+800\cdot 100+900\cdot 50}{100}=640$$

$$\delta^2=\frac{\sum_{i=1}^{n}(\bar{y}_i-\bar{y}_0)^2 f_i}{\sum_{i=1}^{n} f_i}=\frac{(550-640)^2 100+(603{,}03-640)^2 330+}{1000}$$

$$\frac{+(671{,}05-640)^2 380+(689{,}47-640)^2 190}{1000}=\frac{2092503{,}99}{1000}=2092{,}5$$

Comparând nivelul dispersiei dintre grupe cu nivelul mediei dispersiilor de grupă constatăm faptul că factorii întâmplători, la nivelul întregii colectivităţi, au exercitat o influenţă mai puternică decât factorii consideraţi esenţiali (vechimea în muncă).

d) Dispersia generală

$$\sigma_0^2=\frac{\sum_{j=1}^{m}(y_j-\bar{y}_0)^2 f_j}{\sum_{j=1}^{m} f_j}=\frac{(400-640)^2 50+(500-640)^2 150+(600-640)^2 350+}{1000}$$

$$\frac{+(700-640)^2 300+(800-640)^2 100+(900-640)^2 50}{1000}$$

$$\sigma_0^2=\frac{13400000}{1000}=13400.$$

Regula adunării dispersiilor este verificată:

$$\sigma_0^2=\bar{\sigma}^2+\delta^2 \Rightarrow 13400 = 11307{,}50 + 2092{,}50.$$

2.3. 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.

2.3.1. 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.

Pentru cazul în care variaţia este simetrică faţă de valoarea centrală a caracteristicii, compensarea abaterilor se face nu numai pe ansamblul ei, ci şi în interiorul seriei, ca urmare a faptului că frecvenţele de apariţie ale acestor abateri sunt egale de ambele părţi ale valorii centrale. Dacă frecvenţele de apariţie ale variantelor nu urmează această regularitate înseamnă că seria prezintă o tendinţă de asimetrie fie spre valorile mai mari, fie spre valorile mai mici ale caracteristicii.

O serie perfect simetrică va corespunde acelei forme de variaţie statistică în care şi influenţa factorilor întâmplători urmează o anumită regularitate, astfel încât are loc o repartiţie uniformă în ambele sensuri.

Pentru determinarea tipului de asimetrie se poate recurge la metode elementare, precum: metoda grafică şi momentul centrat de ordinul 3.

- ***Metoda grafică*** – la interpretarea gradului de asimetrie se porneşte de la poziţia şi valoarea pe care le au cei trei indicatori ai tendinţei centrale: media, mediana şi modul. Astfel, în funcţie de raportul dintre aceşti indicatori, putem avea una din următoarele situaţii:
  - $\bar{x} = Me = Mo$ - serie simetrică (figura 2.8.a);
  - $\bar{x} < Me < Mo$ - serie cu asimetrie spre stânga (negativă) – figura 2.8.b;
  - $\bar{x} > Me > Mo$ - serie cu asimetrie spre dreapta (pozitivă) – figura 2.8.c.

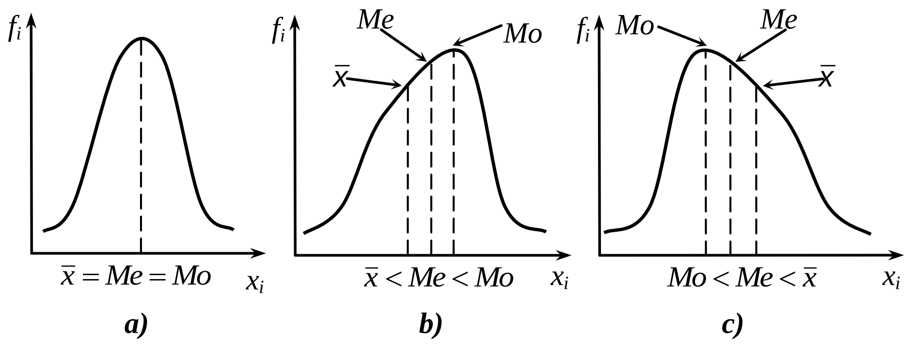


Figura 2.8. Tipuri de serii de repartiţie: a) simetrică; b) cu asimetrie spre stânga (negativă); c) cu asimetrie spre dreapta (pozitivă) .

- ***Momentul centrat de ordinul 3:***

$$\mu_3 = \frac{\sum (x_i - \bar{x})^3 f_i}{\sum f_i}.$$

Interpretarea acestui indicator porneşte de la observaţia că momentele centrate de ordin impar ale seriilor de distribuţie perfect simetrice sunt egale cu zero (deci şi $\mu_3=0$). Pentru seriile în care predomină termenii cu abateri negative faţă de medie ($x_i - \bar{x} < 0$), vom avea $\mu_3<0$, iar pentru seriile în care predomină termenii cu abateri pozitive faţă de medie ($x_i - \bar{x} > 0$), vom avea $\mu_3>0$. Ca atare, în funcţie de valoarea lui μ_3 vom avea:

- serie simetrică – pentru $\mu_3 = 0$;
- serie cu asimetrie spre stânga (negativă) – pentru $\mu_3 < 0$;
- serie cu asimetrie spre dreapta (pozitivă) – pentru $\mu_3 > 0$.

Pentru măsurarea statistică a asimetriei se folosesc coeficientul de asimetrie al lui Pearson şi coeficientul lui Fisher.

- ***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).

În cazul seriilor uşor asimetrice bazate pe un număr mare de cazuri observate, când se verifică relaţia $Mo \approx \bar{x} - 3(\bar{x} - Me)$, se poate folosi un alt coeficient de asimetrie, calculat după relaţia:

$$Cas^* \approx \frac{3(\bar{x} - Me)}{\sigma}.$$

Acest coeficient ia valori în intervalul (–3,3) şi va arăta un grad mai mare de simetrie cu cât se va apropia mai mult de 0.

Pentru caracterizarea asimetriei, Pearson a mai propus şi un *al doilea coeficient de asimetrie* bazat pe momentele centrate de ordinul 2 şi 3:

$$\beta_1 = \frac{(\mu_3)^2}{(\mu_2)^3} = \frac{1}{\mu_2} \cdot \left(\frac{\mu_3}{\mu_2}\right)^2.$$

Din această formulă se observă că $\beta_1 \geq 0$. Deci, acest indicator nu poate fi folosit în aprecierea sensului asimetriei. Interpretarea coeficientului este următoarea:

- $\beta_1 = 0$ - serie simetrică;
- $\beta_1 > 0$ - serie cu asimetrie (spre dreapta sau spre stânga).

Imposibilitatea furnizării de informaţii asupra sensului asimetriei a condus la redefinirea acestui coeficient de către Fisher în forma prezentată în continuare.

- ***Coeficientul lui Fisher*** – se determină astfel:

$$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{(\mu_2)^{3/2}}.$$

Deoarece numitorul va fi întotdeauna pozitiv (neinfluenţând semnul indicatorului), interpretarea coeficientului lui Fisher este asemănătoare cu cea a momentului centrat de ordinul 3 (μ_3).

Exemplul 2.25.
Pornind de la exemplul 2.2., să se analizeze asimetria seriei.
Rezolvare

$$Cas = \frac{\overline{x} - Mo}{\sigma} = \frac{640 - 630}{115{,}76} = 0{,}086.$$

Rezultă că avem o asimetrie moderată spre dreapta sau pozitivă.

2.3.2. 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.

Pe cale grafică, boltirea se apreciază comparând curba frecvenţelor unei distribuţii empirice cu modelul corespunzător distribuţiei normale. Curba frecvenţelor poate să apară în una din următoarele trei situaţii[1] (figura 2.9.):

- *curbă mezocurtică* – coincide modelului (curba normală);
- *curbă platicurtică* – prezintă o variaţie puternică a variabilei X în paralel cu o variaţie slabă a frecvenţelor;
- *curbă leptocurtică* - prezintă o variaţie slabă a variabilei X în paralel cu o variaţie puternică a frecvenţelor.

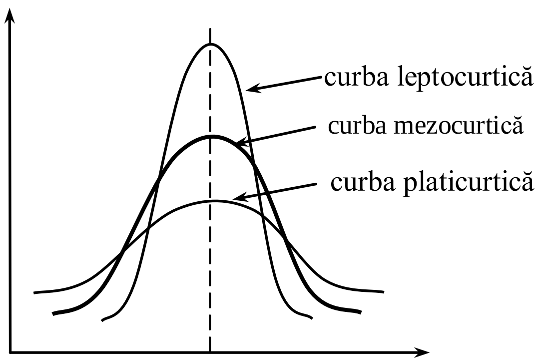


Figura 2.9. Boltirea.

[1] Denumirile folosite în continuare îşi regăsesc rădăcinile etimologice în limba greacă: *kurtos* = cocoşat; *platos* = larg, lat; *leptos* = îngust, subţire.

Pe calea calculelor algebrice boltirea se determină pe baza unor coeficienţi.

- ***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, iar pentru $\beta_2 < 3$ avem o curbă platicurtică.

- ***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.

Exemplul 2.26.
Considerând datele de la exemplul 2.2., să se analizeze boltirea (aplatizarea) seriei.
Rezolvare
a) Coeficientul de boltire Pearson:
Mai întâi vom determina momentul centrat de ordinul 4:

$$\mu_4 = \frac{\sum (x_i - \bar{x})^4 f_i}{\sum f_i} = \frac{522320000000}{1000} = 52232000.$$

Coeficientul va fi:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{522320000}{179560000} = 2{,}91 < 3$$ ⇒ avem o curbă platicurtică.

b) Coeficientul de boltire Fisher:
$\gamma_2 = \beta_2 - 3 = -0{,}09 < 0$ ⇒ avem o curbă platicurtică.

2.4. Concentrare / diversificare

Corelat cu analiza dispersării valorilor individuale înregistrate ale unei anumite variabile are loc şi analiza fenomenului de concentrare. Acest fenomen a fost studiat pentru prima dată de statisticianul italian Corrado Gini în 1912 şi viza distribuţia veniturilor populaţiei.

Prin ***concentrare*** *se înţelege aglomerarea unităţilor unei populaţii statistice sau a valorilor globale în jurul unei anumite valori a caracteristicii de grupare.*

Practic, această definiţie ne prezintă concentrarea ca pe o noţiune conexă celei de dispersare. În prezent, printre numeroasele aplicaţii ale concentrării se regăsesc:

- măsurarea concentrării întreprinderilor în scopul stabilirii taliei lor în funcţie de numărul angajaţilor, de valoarea producţiei, de cifra de afaceri;

- măsurarea concentrării sarcinilor de serviciu în scopul organizării eficiente a activităţii în funcţie de numărul orelor de lucru necesare rezolvării lor;
- evidenţierea inegalităţilor dintre repartiţiile de structură, după o variabilă dată, a indivizilor şi a veniturilor unei societăţi, în scopul caracterizării nivelului de trai al populaţiei;
- caracterizarea structurii pieţelor, situaţie în care studiul concentrării se completează cu măsurarea diversificării.

Analiza concentrării necesită studierea comparată a structurii unităţilor dintr-o populaţie statistică şi a structurii valorii globale pe aceleaşi variante / intervale de variaţie a caracteristicii de grupare. În felul acesta s-ar putea evidenţia atât inegalităţile dintre distribuţiile de structură comparate cât şi concentrarea valorii globale pe un număr redus de unităţi din populaţia statistică observată: cu cât sunt mai mari diferenţele dintre cele două distribuţii de structură cu atât mai mari sunt diferenţele dintre grupe, concentrarea tinzând să crească, şi invers, cu cât disparităţile de distribuţie sunt mai mici, cu atât concentrarea este mai slabă, tinzându-se spre o distribuţie egalitară.

Studierea concentrării este aplicabilă numai variabilelor continue cu valori pozitive. Se poate extinde şi în domeniul seriilor calitative atributive cu scopul stabilirii gradului de concentrare pe tipuri calitative. În general, concentrarea este aplicabilă oricărui fenomen care posedă caracteristici ce pot fi însumate.

Ca atare, analiza seriilor de distribuţie cu ajutorul concentrării se face în condiţiile îndeplinirii a două cerinţe: să aibă sens însumarea variabilei de distribuţie şi să fie posibilă împărţirea valorii globale a variabilei între unităţile colectivităţii. Aceste două cerinţe sunt îndeplinite de distribuţii precum distribuţia populaţiei pe clase de venituri, distribuţia întreprinderilor după cifra de afaceri – cazuri în care valorile globale cumulate ar evidenţia diferenţele existente în repartiţia veniturilor colectivităţii analizate. Însă, în cazul distribuţiei pe vârste a indivizilor unei colectivităţi, spre exemplu, nu ar fi respectate, deoarece atât însumarea, cât şi împărţirea vârstei indivizilor ar fi operaţii fără sens pentru colectivitate.

2.4.1. Indicatorii concentrării

Caracterizarea statistică a fenomenului concentrării se poate realiza atât prin procedee grafice, cât şi prin calcule numerice.

- **Curba de concentrare**

Procedeul grafic de caracterizare a concentrării a fost elaborat de Corrado Gini şi de americanul Lorentz şi se bazează pe construirea curbei de concentrare (curba Lorentz-Gini), determinându-se pe baza ei gradul de concentrare (indicele de concentrare Gini).

Curba de concentrare este construită într-un sistem de axe rectangulare, pe baza frecvenţelor relative cumulate. Se parcurg următoarele etape:

1) se determină frecvenţele relative cumulate corespunzătoare efectivelor f_i după relaţia:

$$F_k(f)=\sum_{i=1}^{k} p_i=\sum_{i=1}^{k}\left(\frac{f_i}{\sum f_i}\right)$$

şi se fixează pe axa absciselor (unde $p_i=\frac{f_i}{\sum f_i}$ reprezintă frecvenţele relative);

2) pe ordonată se fixează procentele cumulate ale valorilor $x_i f_i$ calculate după relaţia:

$$F_k(xf) = \sum_{i=1}^{k} \left(\frac{x_i f_i}{\sum x_i f_i} \right);$$

3) se construieşte pătratul ABCD (pătratul lui Gini) şi curba de concentrare prin unirea punctelor de coordonate ($F_k(f)$; $F_k(xf)$) ca în figura 2.10.

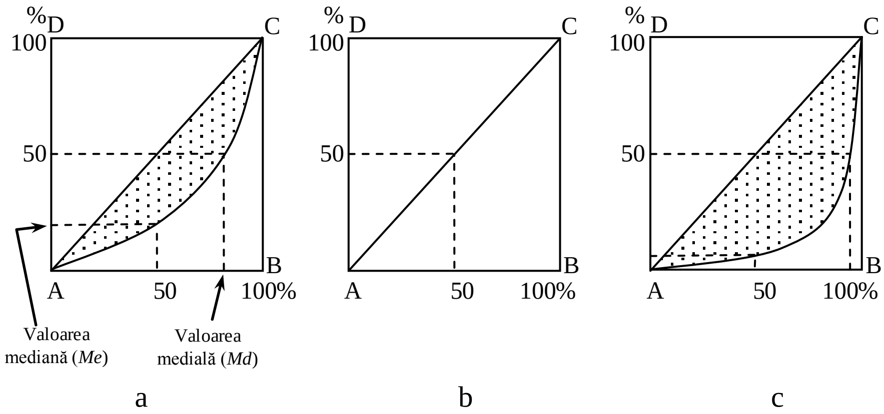


Figura 2.10. Curba de concentrare cu grade diferite: a) concentrare slabă; b) lipsa concentrării; c) concentrare puternică.

Ca mijloc de apreciere a gradului de concentrare, curba Gini se bazează pe faptul că prin reprezentarea grafică a concordanţei ponderilor cumulate ale efectivelor unei colectivităţi ($F_k(f)$) cu ponderile cumulate ale valorilor globale ale unei caracteristici de distribuţie ($F_k(xf)$) se arată cât din valoarea globală a caracteristicii se concentrează în primele două grupe, în primele trei grupe ş.a.m.d.

Gradul de concentrare se poate aprecia în funcţie de mărimea suprafeţei de concentrare. Astfel, cu cât abaterea curbei de concentrare este mai mare faţă de diagonala pătratului ABCD, cu atât este mai mare suprafaţa de concentrare şi, ca atare, diferenţele dintre grupe sunt mai mari şi concentrarea este mai puternică.

Când valorile celor două variabile sunt egale ($F_k(f)$=($F_k(xf)$), curba de concentrare se suprapune pe diagonala pătratului, respectiv este cazul unei echirepartiţii (figura 2.10. b).

Când întreaga valoare globală este concentrată la o singură unitate a colectivităţii, curba coincide cu laturile pătratului. În acest caz, concentrarea este maximă, adică o singură unitate din colectivitate deţine întreaga valoare globală a caracteristicii.

Curba de concentrare are numeroase aplicaţii în domeniul economico-social, şi anume:

- mijloc de apreciere a gradului de concentrare a unei distribuţii;
- metodă de aproximare a valorilor centrale ale unei distribuţii (*Me* şi *Md*), a indicelui de concentrare Gini;
- metodă de depistare a tipurilor calitative dintr-o distribuţie;
- mijloc de comparare calitativă a gradului de concentrare etc.

Exemplul 2.27.

Considerând datele de la exemplul 2.2., să se aprecieze gradul concentrării muncitorilor în funcţie de salariul lunar cu ajutorul curbei de concentrare.

Rezolvare

Pentru determinarea curbei de concentrare sunt necesare calculele din tabelul 2.16.

Tabelul 2.16.

x_i	f_i	$p_i = \frac{f_i}{\sum f_i}$	$F_k(f) = \sum_{i=1}^{k} p_i$	$x_i f_i$	$\frac{x_i f_i}{\sum x_i f_i}$	$F_k(xf) = \sum_{i=1}^{k} \left(\frac{x_i f_i}{\sum x_i f_i} \right)$
400	50	0,050	0,050	20000	0,031	0,031
500	150	0,150	0,200	75000	0,117	0,148
600	350	0,350	0,550	210000	0,328	0,477
700	300	0,300	0,850	210000	0,328	0,805
800	100	0,100	0,950	80000	0,125	0,930
900	50	0,050	1	45000	0,070	1
Total	1000	1	-	640000	1	-

Se construieşte pătratul lui Gini şi curba de concentrare prin unirea punctelor de coordonate ($F_k(f)$; $F_k(xf)$) ca în figura 2.11.

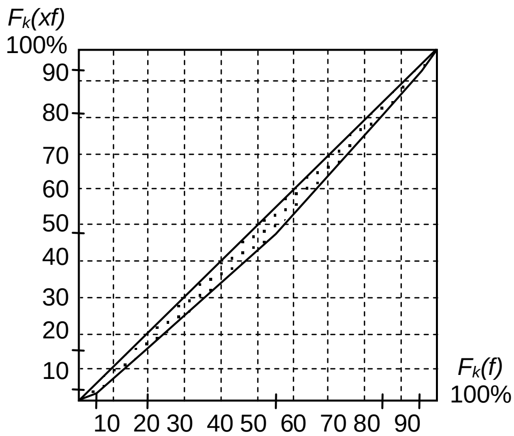


Figura 2.11. Curba de concentrare Gini.

Se observă o curbă de concentrare foarte apropiată de diagonala pătratului, de unde rezultă că avem o concentrare slabă. De altfel, corespondenţa $F_k(f)$ - $F_k(xf)$ ne arată disparităţi foarte mici între cele două repartiţii de structură: 3,1% din masa salarială este deţinută de 5% din mincitori, 14,8% de 20%, 47,7% de 55% ş.a.m.d.

- **Indicele concentrării Gini**

Indicele de concentrare este un indicator sintetic al concentrării unei distribuţii, care a fost formulat pentru prima dată de către C. Gini, cu ocazia studierii distribuţiei salariilor şi veniturilor populaţiei. Indicele de concentrare a luat diferite forme, în funcţie de procedeul folosit, dar este cunoscut, în continuare, sub denumirea de indicele lui Gini. Relaţia de calcul a indicelui este următoarea:

$$I_G = \frac{Suprafata\, de\, concentrare}{Aria\, triunghiului\, ABC},$$

unde suprafaţa de concentrare este cuprinsă între diagonala pătratului (AC) şi curba de concentrare.

Valoarea indicelui de concentrare ia întotdeauna valori cuprinse în intervalul **[0, 1]**, reflectând o variaţie de la o concentrare nulă la o concentrare maximă. Este transferabil în timp şi spaţiu, permiţând efectuarea de comparaţii. Datorită eficacităţii sale, indicele de concentrare Gini este unul dintre cei mai utilizaţi.

Din graficul de concentrare se poate observa că acesta poate fi utilizat şi pentru aproximarea valorilor centrale.

Determinarea indicelui Gini poate fi făcută utilizând diverse metode, printre care:

- *metoda grafică* – presupune construirea curbei de concentrare pe hârtie milimetrică. Evaluarea suprafeţei de concentrare se face prin numărarea pătratelor întregi cuprinse în suprafaţa de concentrare şi raportarea la jumătate din numărul total de pătrăţele ale pătratului ABCD;
- *metoda trapezelor* – are ca punct de plecare curba de concentrare. Evaluarea suprafeţei de concentrare se face considerând că suprafaţa de concentrare este egală cu suprafaţa triunghiului ABC minus suma suprafeţelor trapezelor (figura 2.12). Se poate observa că există atâtea trapeze câte intervale de variaţie sunt.

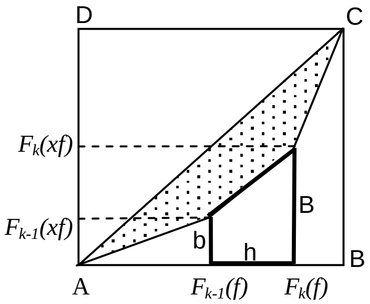


Figura 2.12. Metoda trapezelor.

Suprafaţa triunghiului ABC este egală cu jumătate din suprafaţa pătratului de concentrare (ABCD) care are valoarea 1.

Suprafaţa trapezelor se determină pornind de la formula ariei unui trapez $\left(S = \frac{(b+B)h}{2} \right)$ adaptată notaţiilor unei distribuţii statistice, folosite în figura 2.12.

Suprafaţa de concentrare (Sc) va fi calculată după relaţia următoare:

$$Sc = \frac{1}{2} - \sum \frac{(b+B)h}{2} = \frac{1}{2} \cdot \left[1 - \sum \left(F_{k-1}(xf) + F_k(xf) \right) p_k \right],$$

unde: $F_k(xf)$ – ponderea valorilor globale cumulate până la nivelul i al variabilei X;
p_k – ponderea efectivelor în totalul colectivităţii.

În acest moment, indicele de concentrare Gini se determină foarte uşor împărţind valoarea suprafaţei de concentrare la aria triunghiului ABC.

- *metoda triunghiurilor* – indicele de concentrare se determină pe baza relaţiei următoare:

$$I_G = \sum_{k=1}^{n-1} \left(F_k(f) \cdot F_{k+1}(xf) - F_{k+1}(f) F_k(xf) \right).$$

- **Abaterea medială-mediană**

Reprezintă un alt procedeu numeric de calcul a concentrării şi se determină cu ajutorul relaţiei următoare:

$\Delta M = Md - Me$.

Semnificaţia acestui indicator este următoarea: cu cât valoarea abaterii este mai mare, cu atât concentrarea este mai puternică şi, invers, cu cât valoarea abaterii este mai mică, cu atât concentrarea este mai slabă. Dacă abaterea este zero (adică $Md = Me$) nu există concentrare, distribuţia reprezentând o echirepartiţie.

Spre exemplu, vom considera distribuţia unei echipe de muncitori după caracteristica salariu. Mediana va fi acea valoare x_i care împarte colectivitatea în două părţi egale. Primii 50% dintre muncitori au salarii mai mici decât ceilalţi 50% şi, ca atare, valoarea globală a salariului primilor este mai mică decât valoarea globală a salariilor ultimilor. Din acest motiv, valoarea care împarte salariile globale în două părţi egale (mediala) va fi mai mare decât mediana (pentru a echilibra acea diferenţă rezultată din împărţirea făcută de mediană). Dacă, însă, distribuţia ar fi uniformă, am avea o medială egală cu mediana (situaţia echirepartiţiei).

- **Coeficientul de concentrare**

Determinarea coeficientului de concentrare constă în compararea, sub formă de raport, a abaterii medială-mediană (ΔM) cu amplitudinea absolută a variaţiei caracteristicii de grupare ($A_a = X_{max} - X_{min}$), după relaţia:

$$\Delta M_{\%} = \frac{\Delta M}{A_a} \cdot 100.$$

Coeficientul ia valori în intervalul [0,100]. Pentru valori mici (tinzând către zero) avem o concentrare slabă, iar pentru valori mari (tinzând către 100) avem o concentrare puternică, adică există mari diferenţe între valorile globale pe clase de variaţie.

Spre deosebire de abaterea medială-mediană, coeficientul de concentrare permite, datorită exprimării relative, compararea gradului de concentrare pentru diferite distribuţii statistice indiferent de unitatea de măsură folosită pentru exprimarea variabilelor de grupare.

Deşi uşor de calculat, ambii indicatori (abaterea medială-mediană şi coeficientul de concentrare) prezintă dezavantajul unor mărimi aproximative, datorită faptului că mărimile comparate (*Md şi Me*) nu exprimă toţi termenii seriei, ci doar valorile ce ocupă o poziţie centrală într-o distribuţie.

Pentru aprecierea ***concentrării în seriile calitative atributive*** este necesară cunoaşterea structurii populaţiei statistice investigate, cu ajutorul ponderilor sau greutăţilor specifice (p_i) şi calculul şi interpretarea unor indicatori, dintre care cei mai importanţi sunt:

- raportul de concentrare;
- energia informaţională Onicescu;
- diferenţa Hirschman;
- coeficientul de concentrare Gini;
- coeficientul de concentrare Strück;
- lungimea vectorului de structură.

2.4.2. Indicatorii diversificării

Diversificarea este procesul invers concentrării, considerându-se, spre exemplu, că o întreprindere care fabrică mai mult de un produs este diversificată. Determinarea gradului de diversificare este făcut în mod asemănător cu cel al concentrării, utilizând pentru aceasta o serie de indicatori. Cei mai importanţi indicatori utilizaţi pentru determinarea diversificării sunt:

- **raportul de diversificare (*d*)** – se calculează ca valoare complementară la unitate a raportului de specializare (S_i). Cei doi indicatori se determină pe baza relaţiilor următoare:

  $$S_i = \frac{f_i}{N},\; d = 1 - \sum S_i = 1 - \sum_{i=1}^{n} \frac{f_i}{N},$$

  unde: f_i – efectivul principal al unei întreprinderi;
  N – totalul efectivului.

  Valoarea raportului de diversificare este cuprinsă în intervalul $\left[0; \frac{n-1}{n}\right]$. Valoarea minimă (0) corespunde cazului în care toţi angajaţii unei întreprinderi lucrează într-o singură activitate, iar valoarea maximă apare în cazul unei împărţiri echivalente în cele *k* activităţi;

- **indicele de diversificare** – se determină pe baza relaţiei următoare:

  $$I_D = 1 - \sum_{i=1}^{n} \left(\frac{f_i}{N} \right)^2 = 1 - \sum_{i=1}^{n} S_i^2 .$$