PRELEGEREA IV
STATISTICĂ MATEMATICĂ

I. Indicatori de măsură a împrăştierii

Distribuţia unei variabile nu poate fi descrisă complet numai prin cunoaşterea mediei, ci este necesar să avem informaţii şi despre gradul der împrăştiere în jurul mediei. În general, se poate afirma că, o împrăştiere mai mare va fi reflectată de existenţa mai multor valori frecvente.

Reamintim indicatorii de măsură a împrăştierii valorilor într-o distribuţie: **amplitudine**, **quantila**, **momentul de ordin r**, **momentul centrat de ordin r**, **dispersia**, **abaterea medie**, **abaterea standard**, **coeficientul de variaţie**, **coeficientul de asimeterie**, **coeficientul de boltire** şi **coeficientul de exces**.

4.2.1. Amplitudinea

Amplitudinea (sau **întinderea domeniului** seriei statistice) este cel mai simplu indicator de măsură a împrăştierii.

Amplitudinea este egală cu diferenţa dintre cea mai mare şi cea mai mică valoare observată: $\mathbf{A = x_{max} - x_{min}}$.

A nu se confunda amplitudinea cu numărul de valori posibile: $\mathbf{x_{max} - x_{min} + 1}$.

Aplicaţie. Pentru seria experimentală:
42, 44, 10, 57, 64, 120, 33 se obţine:
A = 120 – 10 = 110.

Observaţii. 1. Amplitudinea este dependentă de volumul seriei statistice avută în vedere. Cu cît este mai mare, cu atît valoarea calculată a amplitudinii este mai credibilă. Deci, modificarea dimensiunii eşantioanelor studiate poate afecta valoarea acestui indicator.

2. Dacă valorile **extreme** (**aberante**) sînt într-un număr foarte mic, se pot înlătura practic din operaţiunile de grupare, dar se recomandă ca acest operaţie să se facă cu discernămînt, deoarece rareori există limite reale superioare sau inferioare pe care valorile unei v.a. să nu le poată depăşi.

3. Amplitudinea, alături de amplitudinea interquartilică, este o măsură a **variabilităţii**, mai ales, cînd se realizează comparaţii între varaibilităţile a două distribuţii cu un număr egal de scoruri. De exemplu:

- eşantionul **1**: **11**, **16**, **18**, **23**, **29**, **31**, **37**;

- eşantionul **2**: **18**, **19**, **21**, **23**, **24**, **26**, **29**. Medianele lor fiind egale: $\mathbf{Me_1 = 23}$, iar $\mathbf{Me_2 = 23}$ se determină amplitudinile lor: $\mathbf{A_1 = 37 - 11 = 26}$, iar $\mathbf{A_2 = 29 - 18 = 11}$. Rezultă că primul eşantion este mai eterogen decît al doilea.

4. Amplitudinea nu ţine cont de forma repartiţiei, aceeaşi valoare se poate obţine atît în situaţia unei curbe cu frecvenţe simetrice, cît şi în aceea a unei curbe de frecvenţe în formă de **J** sau **L**.

4.2. Quantila

Quantila (percentila**,** cvantila**) de ordin** α permite aprecierea modului de împrăştiere a valorilor celor mai frecvente pe întregul segment al datelor observate şi fixarea unei observaţii în raport cu mediana distribuţiei. Cu alte cuvinte, quantila este valoarea din seria experimentală care depăşeşte o proporţie **α%** de observaţii şi este depăşită de (**1-α**)**%** observaţii.

În tabelul următor se evidenţiază cele mai frecvente sisteme de quantile.

Nr. crt.	**Număr de părţi egale**	**Număr de quantile**	**Denumire**
1	**3**	**2**	**Trecilă**
2	**4**	**3**	**Quartilă**
3	**5**	**4**	**Quintilă**
4	**6**	**5**	**Sextilă**
5	**7**	**6**	**Septilă**
6	**8**	**7**	**Octilă**
7	**9**	**8**	**Nonilă**
8	**10**	**9**	**Decilă**
9	**100**	**99**	**Centilă**

Tabelul 4.2.2.1. Sisteme de quantile

Doar pentru o repartiţie uniformă, quartilele sînt egal depărtate (ca număr de categorii) între ele.

De exemplu, **quartilele** împart observaţiile în patru părţi egale:

$$Q_1 = Q_{1/4},\ Q_2 = Q_{1/2},\ Q_3 = Q_{3/4}.$$

Se utilizează termenii:

- **quartila întîi** (sau **quartila de 25 %**);
- **quartila a doua** (sau **quartila de 50%**);
- **quartila a treia** (sau **quartila de 75%**).

Se observă că, quantila de ordin ½ este egală cu mediana, adică:

$$Me(X) = Q_{1/2}.$$

Practic:

- pentru **datele discrete**, etapele de aflare a unei quntile sînt:

1) se ordonează şirul valorilor observate $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, unde **n** este numărul total de observaţii;

2) se află **indicele** (**rangul**) valorii care împarte şirul de valori observate în proporţiile date: $k = [\alpha n] + 1$, iar **[]** reprezintă funcţia **partea întreagă a unui număr**.;

3) quantila de ordin α este atunci valoarea observată de rang **k**, $Q_\alpha = x_{(k)}$.

În general, **α = (numărul de observaţii mai mici decît Q_α)/(numărul total de observaţii)**, şi se exprimă în procente.

- pentru **datele continue**, se procedează astfel:

1) se calculează indicele **k** al observaţiei care determină quantila, $k = [\alpha n] + 1$;

2) se determină intervalul căruia îi aparţine observaţia $x_{(k)}$;

3) se utilizează formula: $Q_\alpha = l_i + l\dfrac{k - f_{i-1}\downarrow}{f_i}$, unde

l_i – este limita inferioară exactă a intervalului,

l – este lungimea intervalului,

f_i – este frecvenţa (necumulată) a intervalului care conţine quantila,

$f_{i-1}\downarrow$ - este frecvenţa cumulată pentru intervalul anterior intervalului care conţine quantila.

Utilizarea unui sistem de quantile cu mai multe elemente (**decile**, **centile**) duce la o înţelegere mai fină a modului în care sînt împrăştiate valorile observate în domeniul respectiv.

Observaţii. 1) La caracterizarea gradului de împrăştiere a valorilor observate se pot utiliza următoarele mărimi statistice derivate din sistemele de quantile (**intervalul interquantilic**, **abaterea quartilică**).

a) intervalul interquartilic, **IQR** (**Interquartile Range**) = $\mathbf{Q_3 - Q_1}$, unde $\mathbf{Q_1}$ şi $\mathbf{Q_3}$ sînt quartilele **1** şi **3** ale distribuţiei. În acest interval se află **50%** dintre valorile distribuţiei. Cu cît intervalul quartilic este mai mare, cu atît valorile sînt mai împrăştiate.

Valorile extreme sînt eliminate din calculul acestui indicator. Două serii de date cu acelaşi interval **IQR** pot să difere semnificativ ca distribuţie a valorilor.

b) intervalul interdecilic, **IQR** = $\mathbf{Q_9 - Q_1}$, unde $\mathbf{Q_1}$ şi $\mathbf{Q_9}$ sînt decilele **1** şi **9** ale distribuţiei. În acest interval se află **90%** dintre valorile distribuţiei.

c) intervalul intercentilic, **IQR** = $\mathbf{Q_{99} - Q_1}$, unde $\mathbf{Q_1}$ şi $\mathbf{Q_{99}}$ sînt centilele **1** şi **99** ale distribuţiei.

d) abaterea quartilică (sau **amplitudinea semiinterquartilică**), **AQ** = $\mathbf{(Q_3 - Q_1)/2}$.

Acest indicator permite determinarea unui interval centrat pe mediană (adică, quartila a doua), interval în care sînt incluse aproximativ **50%** dintre observaţii. Limitele acestui interval sînt **Me – AQ** şi **Me + AQ** şi este simetric faţă de mediana **Me**. Această afirmaţie este valabilă doar în cazul repartiţiilor simetrice (cum este repartiţia normală) şi erorile sînt majore pentru repartiţiile cu asimetrie pronunţată. Lungimea intervalului definit în acest mod este egală cu cea a intervalului interquartilic. Pentru o distribuţie normală se poate considera că amplitudinea este de aproximativ **7.5** abateri quartilice.

e) criteriul 1.5IQR, utilizat pentru depistarea valorilor aberante (extreme). Orice observaţie situată la o distanţă mai mare decît **1.5IQR** sub prima quartilă sau peste a treia quartilă, poate fi identificată ca o **valoare aberantă**.

Utilizînd statisticile $\mathbf{x_{min}}$ $\mathbf{Q_1}$ **Me** $\mathbf{Q_3}$ $\mathbf{x_{max}}$ calculate din valorile observate, în ordinea dată, se obţine o imagine sintetică a amplitudinii datelor (prin includerea valorilor extreme), a tendinţei centrale (prin mediană) şi a împrăştierii datelor (prin quartile).

4.2.3. Momentul de ordin r

Momentul de ordin r al v.a. **X**, cu notaţia $\mathbf{M_r}$, se defineşte ca valoarea medie a v.a. $\mathbf{X^r}$.

$$\mathbf{M_r(X) = M(X^r)}.$$

Pentru **cazul continuu**, avem: $\mathbf{M_r(X) = M(X^r)} = \int_{-\infty}^{+\infty} x^r f(x)dx$.

Dacă **r = 1**, momentul de ordin **r** este egal cu valoarea medie, iar dacă **r = 2**, momentul de ordin **r** este egal cu pătratul mediei pătratice.

În plus, în momentul de ordin **r** se amplifică ponderea valorilor extreme, valorile fiind ridicate la puterea **r**. Diferenţele de semn dispar doar dacă **r** este par.

Observaţie. Momentul de ordin **r** în raport cu un număr **a** se defineşte doar pentru cazul discret: $a^{M_r(X)} = \sum_i (x_i - a)^r / n$.

4.2.4. Momentul centrat de ordin r

Momentul centrat de ordin r al lui **X**, cu notaţia $\boldsymbol{\mu_r}$, se defineşte ca momentul de ordin **r** al abaterii v.a. **X** de la media ei:

$$\boldsymbol{\mu_r} = \mathbf{M((X - M(X))^r)}.$$

Calculul momentului centrat de ordin r al v.a. **X** se face în următoarea ordine:

- se determină valoarea medie;
- se află abaterea de la medie;
- se ridică abaterea medie la puterea **r**;
- se află valoarea medie pentru variabila obţinută în etapa anterioară.

În momentul centrat de ordin **r** este amplificată ponderea valorilor mai depărtate de medie.

Observaţii. 1. Momentul centrat de ordinul întîi este nul: $\mu_1 = 0$.

2. Vom vedea în paragraful următor că, momentul centrat de ordin **2** este egal cu dispersia: $\mu_2 = D^2(X)$.

3. Se demonstrează că momentul centrat de ordin **r** este independent la transformări liniarte de genul, $z_i = x_i - x_0$, unde x_0 este originea seriei experimentale: $\mu_2(X) = \mu_2(Z)$.

4.2.5. Dispersia

Dispersia (sau **varianţa**, **momentul centrat de ordinul al doilea**) unei v.a **X**, cu notaţia σ^2 sau $D^2(X)$, se defineşte ca momentul centrat de ordin **2**, adică media pătratului abaterii **X - M(X)**.

În aprecierea împrăştierii datelor observate pentru o anumită caracteristică, dispersia rămîne indicatorul cel mai folosit, indicînd cît de mult se abat valorile măsurate faţă de medie. Cu cît este mai mare această valoare, cu atît mai mult se împrăştie valorile în jurul valorii centrale:

- pentru **date discrete** avem:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 .$$

Prin ridicarea abaterilor de la medie la pătrat se realizează ponderea suplimentară a valorilor extreme (mai depărtate de medie). De exemplu, o valoare depărtată de medie cu **10** contribuie la calculul dispersiei cu **100**, în timp ce o valoare mai apropiată de medie, avînd o abatere de **2** de la medie, contribuie doar cu **4**. Se observă că "**greutatea**" contribuţiei la dispersie este amplificată de valorile mai depărtate de medie.

Observaţie. Oricare ar fi o altă valoare de referinţă a abaterilor, **W**, are loc inegalitatea:

$$\sum_{i=1}^{N}(x_i - W)^2 \geq \sum_{i=1}^{N}(x_i - \bar{x})^2$$

, adică suma pătratelor abaterilor de la medie este minimă şi că media este cea mai bună aproximare, în sensul celor mai mici pătrate a seriei de valori respectivă.

- pentru **date continue** se utilizează: $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} n_i(c_i - \bar{x})^2$. Deoarece $\sum_{i=1}^{N}(x_i - \bar{x})^2 = \sum_{i=1}^{N} x_i^2 - N\bar{x}^2$, formulele de calcul pentru dispersie se pot scrie, mai simplu, în modul următor:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \bar{x}^2 ,$$

$\sigma^2 = \frac{1}{N}(\sum_{i=1}^{N} x_i^2 - \frac{1}{N}(\sum_{i=1}^{N} x_i)^2)$ şi respectiv

$\sigma^2 = \frac{1}{N}(\sum_{i=1}^{N} n_i c_i^2 - \frac{1}{N}(\sum_{i=1}^{N} n_i c_i)^2)$, unde $\mathbf{n_i}$ este frecvenţa intervalului **i**, $\mathbf{c_i}$ este centrul (sau valoarea medie) intervalului **i**, iar **N** este volumul seriei experimentale, **N** = $\sum_{i=1}^{k} fa_i$, iar **k** este numărul claselor de grupare.

În situaţia utilizării dispersiei de sondaj (calculată dintr-un eşantion mic), ca estimaţie a dispersiei întregii populaţii statistice,
formulele de mai sus sînt corectate, în sensul că împărţirea se face prin **n-1** în loc de **N**:

$$\mathbf{s^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \text{ sau respectiv}$$

$$\mathbf{s^2} = \frac{1}{n-1}\sum_{i=1}^{k} n_i(c_i - \overline{x})^2 \text{ şi, corespunzător:}$$

$$\mathbf{s^2} = \frac{1}{n-1}(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2) \text{ şi}$$

$$\mathbf{s^2} = \frac{1}{n-1}(\sum_{i=1}^{k} n_i c_i^2 - \frac{1}{n}(\sum_{i=1}^{k} n_i c_i)^2).$$

În cazul eşantioanelor mari (**n > 30**) diferenţele dintre rezultate sînt neglijabile.

Proprietăţi.

a) Dispersia este o cantitate pozitivă.

b) Dispersia unei constante **c** este egală cu zero: $\mathbf{D^2(c) = 0}$.

c) Dispersia produsului dintre o constantă **c** şi o va.. **x** este egală cu: $\mathbf{D^2(cX) = c^2D^2(X)}$, iar $\mathbf{D^2(\frac{1}{c}X) = \frac{1}{c^2}D^2(X)}$,

d) Dispersia unei sume finite de v.a. independente este:

$\mathbf{D^2(X_1 \pm X_2 \pm \ldots \pm X_k) = D^2(X_1) + D^2(X_2) + \ldots + D^2(X_k)}$.

e) Dispersia sumei dintre o constantă **c** şi o v.a. **X** este egală cu:
$\mathbf{D^2(c \pm X) = D^2(X)}$.

f) Se demonstrează că dispersia este independentă la transformări liniare de genul, $\mathbf{z_i = x_i - x_0}$, şi $\overline{z} = \overline{x} - \overline{x_0}$, unde $\mathbf{x_0}$ este originea seriei experimentale:

$$\mathbf{D^2(X) = D^2(Z)}.$$

g) Nu acelaşi lucru se întîmplă, atunci cînd pe lîngă origine se schimbă şi unitatea de măsură: dacă $\mathbf{z_i} = \frac{x_i - x_0}{c}$, atunci:

$$\mathbf{D^2(X) = c^2D^2(Z)}.$$

Aplicaţii. 1. $\mathbf{D^2(5) = 0}$.
2. În experienţa aruncării unui zar ideal, se ştie că $\boldsymbol{\mu} = \mathbf{3.5}$.
Dispersia va fi egală cu:

$$\mathbf{D^2(X) = (1 - 3.5)^2 \frac{1}{6} + (2 - 3.5)^2 \frac{1}{6} + (3 - 3.5)^2 \frac{1}{6} + (4 - 3.5)^2 \frac{1}{6} + (5 - 3.5)^2 \frac{1}{6} + (6 - 3.5)^2 \frac{1}{6} = 2.92}.$$

3. Dacă în aruncarea cu două zaruri se notează cu **X** numărul obţinut la primul zar , iar cu **Y** numărul obţinut la doilea zar, atunci suma celor două numere este o nouă variabilă aleatoare discretă, **Z**. V.a. **X** şi **Y** fiind independente, atunci:

$$\mathbf{D^2 (Z) = D^2 (X) + D^2 (Y) = 2.92 + 2.92 = 5.84}.$$

4. Propietăţile menţionate mai sus permit reducerea volumului de calcul sau facilitarea calculelor prin micşorarea ordinului de mărime a valorilor observate. Este mai simplu să calculăm dispersia pentru valorile $\mathbf{X_1}$**: 2**, **3**, **4**, **1** decît pentru $\mathbf{X_2}$**: 200**, **300**, **400**, **100**, ştiind că:

$$\mathbf{D^2(X_2) = 100^2 D^2(X_1) = 10000 D^2(X_1)}.$$

Dispersia este o măsură a împrăştierii valorilor cu probabilităţi suficient de mari (încît realizarea lor să fie plauzibilă). În plus, proprietăţile menţionate mai sus aduc precizări legate de împrăştierea valorilor mai multor v.a., de exemplu, diferenţa a două v.a este tot atît de împrăştiată ca şi suma lor.

Observaţii. **1.** Notaţii. Utilozate în statistică în mod curent:

Nume entitate	**Populaţie**	**Eşantion**
Numărul unităţilor	**N**	
Valoarea medie	$\overline{X}$ sau μ	
Dispersie	$\mathbf{D^2(X)}$ sau σ^2	
Amplitudine	**A**	**a**
Coeficient de variaţie	**CV**	**cv**

2. Interpretarea practică a dispersiei este îngreunată pentru că nu se exprimă în aceleaşi unităţi de măsură ca datele iniţiale. Dacă dispersia se calculează pentru lungimi exprimate în **cm**, atunci dispersia se exprimă în $\mathbf{cm^2}$. Acest neajuns este înlăturat de indicatorul statistic, abaterea standard, utilizat în locul dispersiei în mod frecvent.

3. Pentru **n** suficient de mare, valorile mediei şi ale dispersiei trebuie să fie apropiate. Discrepanţele dintre ele se datorează, în principal, împărţiri pe clase a unei selecţii de volum mic.

4.2.6. Abaterea medie

Abaterea medie (**abaterea medie aritmetică** sau **abaterea medie absolută**) se defineşte prin expresia:

$$\mathbf{AM} = \frac{1}{N}\sum_{i=1}^{N} | x_i - \overline{x} | .$$

Însă, în locul ei, în ultima vreme, se foloseşte abaterea standard.

4.2.7. Abaterea standard

Abaterea standard(sau **abaterea medie pătratică**), cu notaţia σ, se defineşte ca fiind rădăcină pătrată din dispersia populaţiei studiată:

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

Abaterea medie pătratică se exprimă în aceleaşi unităţi de măsură ca şi valorile experimentale, iar pentru un eşantion cu dispersia $\mathbf{s^2}$, în mod similar, abaterea standard este: $\mathbf{s} = \sqrt{s^2}$.

Observaţii. 1. Abaterea medie pătratică este o cantitate particulară de abatere de la medie faţă de abaterea medie pătratică de la o valoare arbitrară.

2. Aproximativ **2/3** dinte valorile normale sau tipice ale unei distribuţii se găsesc în intervalul cuprins de o parte şi de alta a mediei cu o abatere standard: ($\mathbf{M - \sigma}$, $\mathbf{M + \sigma}$). De aici, provine atributul **standard** a abaterii, deoarece indiferent de forma distribuţiei, aproximativ **2/3** dinte valori se vor găsi în acest interval.

Aplicaţie. Abaterea standard a v.a. **X** din experimentul aruncării unui zar ideal este:

$\boldsymbol{\sigma} = \sqrt{2.92} = \mathbf{1.71}$.

4.2.7.1 Calculul practic al mediei şi abaterii standard

Pentru calculul practic al mediei şi abaterii standard se consideră următoarea:

Aplicaţie. La controlul final al unor produse textile s-a analizat o caracteristică obţinîndu-se următoarele date de sondaj:

168	**176**	**170**
148	**173**	**192**
169	**158**	**162**
171	**170**	**184**
174	**178**	**166**
152	**142**	**162**
184	**156**	**155**
166	**158**	**158**
155	**162**	**168**
151	**155**	**148**

Se cere: a) Să se împartă datele în **k = 5** clase de amplitudine egală (se adoptă intervale de tipul **[;)**);

b) Să se calculeze media;

c) Să se calculeze abaterea medie pătratică.

Rezolvare: a) Valorile extreme sînt: $\mathbf{x_{min}} = \mathbf{142}$ şi x_{max} = **192**. Deci, amplitudinea este: **A = 192 – 142 = 50**.

Lungimea unui interval (sau noua unitate) se obţine, astfel: **l = A/k = 10**.

Împărţirea datelor de sondaj în **5** intervale egale se prezintă după cum urmează:

Nr. clase	$[x_i;x_{I+1})$	c_i	n_i	**Val. simpl.** (z_i)	$n_i.z_i$	$n_i.z_i^2$
1	**[142;152)**	**147**	**4**	**-2**	**-8**	**16**
2	**[152;162)**	**157**	**8**	**-1**	**-8**	**8**
3	**[162;172)**	**167**	**11**	**0**	**0**	**0**
4	**[172;182)**	**177**	**4**	**1**	**4**	**4**
5	**[182;192)**	**187**	**3**	**2**	**6**	**12**
Total			**30**		**-6**	**40**

b) şi **c)** Valorile simplificate se obţin prin transformarea liniară:

$z_i = \dfrac{c_i - 167}{10}$, unde $\mathbf{c_i}$ este centrul unei clase, iar noua origine este $\mathbf{x_0}$ = **167**.

În rezolvarea problemei vom utiliza următoarele formule pentru:

- volumul seriei experimentale: **n** = $\sum_{i=1}^{k} n_i$;

- media adevărată: $\overline{x} = x_0 + c.\overline{z}$, unde **c = l**;

- dispersia reală: **σ^2(x) = c^2. σ^2(z)**;

- abaterea medie pătratică reală: **σ(x) = c.** $\sqrt{\sigma^2(z)}$.

În ordine, se obţin:

n = 30, $\overline{z}$ **= -6/30 = -0.2**,

σ^2(z) = 40/30 – (-0.2)^2 = 1.33 – 0.04 = 1.29,

$\overline{x}$ **= 167 + 10x (-0.2) = 165**,

σ^2(x) = 10^2x 1.29 = 129,

σ(x) = 10x $\sqrt{1.29}$ **= 10x1.14 = 11.4**.

4.2.7.1.1. Corecţia lui Sheppard

Erorile sistematice introduse în calculul dispersiei, în general, sînt semnificative, pe cînd nu acelaşi lucru se poate spune despre medie, ele fiind neglijabile. De aceea, cînd nu sînt suficiente intervale (clase) la dispoziţie, practic, mai puţin de **10**, se aplică corecţia Sheppard asupra valorii calculate a dispersiei:

$$\boldsymbol{\sigma^2}\textbf{(corectată)} = \boldsymbol{\sigma^2}\textbf{(calculată)} - \frac{c^2}{12}, \text{ cu } \mathbf{c} > \frac{1}{4}\sigma .$$

Aplicaţie. Corecţia lui Sheppard se aplică valorii calculată a dispersiei în exemplul din paragraful anterior, deoarece **10>** $\frac{1}{4}\sigma$. Deci, valoarea corectată a dispersiei este:

σ^2(corectată) = 129 – 100/12 = 129 – 8.33 = 120.67.

Prin urmare, valoarea corectată a abaterii standard va fi:

σ = $\sqrt{120.67}$ **= 10.98**.

4.2.8. Coeficientul de variaţie

Coeficientul de variaţie nu depinde de unităţile de măsură folosite în obţinerea datelor experimentale (este adimensional) şi se află fie cu formula:

CV = σ/M(X), fie procentual:

CV = 100σ/M(X)%.

Aplicaţie. Pentru problema rezolvată în paragraful **4.2.7.1.1**, se obţine:

CV = 100x 10.98/165% = 6.65%.

Coeficientul de variaţie se utilizează în operaţii de comparare a două sau mai multe serii statistice.

4.2.9. Coeficientul de asimetrie

Acest indicator măsoară, pe de o parte, gradul de simetrie al unei distribuţii., iar, pe de alta, permite compararea a două distribuţii, atunci cînd valorile obţinute pentru $\boldsymbol{\mu_3}$ sînt comparabile cu valorile variabilei studiate.

Coeficientul de asimetrie, cu notaţia $\boldsymbol{\beta_1}$, se află astfel:

$$\beta_1 = \frac{\mu_3^2}{(\sigma^2)^3}$$, unde $\boldsymbol{\sigma^2}$ este dispersia v.a. **X**.

În formula coeficientului de asimetrie intervine momentul centrat de ordin **3**, care pondereazā suplimentar valorile mai depărtate de medie şi care conservă semnul.

Valorile negative ale momentului centrat de ordin **3**, reflectă **repartiţii cu asimetrie spre stînga**, iar **cele pozitive repartiţii cu asimetrie spre dreapta**. Mai mult, deoarece coeficientul de asimetrie este o valoare pozitivă totdeauna, aprecierea sensului simetriei se obţine cu:

$$\beta_1' = \frac{\mu_3}{(\sigma^2)^{3/2}}$$. Valorile accentuat apropiate de zero a celor doi indicatori statistici, reflectă o repartiţie simetrică.

4.2.10. Alţi coeficienţi de asimetrie pentru distribuţii unimodale

Am văzut că forma unei distribuţii de date poate fi simetrică sau asimetrică în raport cu media ei. Astfel, pentru distribuţii simetrice moda, mediana şi media, cei trei parametri de tendinţă centrală, se confundă, iar pentru distribuţii asimetrice, cele trei valori sînt foarte diferite. Dacă moda, mediana şi media diferă foarte puţin, atunci distribuţia este una uşor asimetrică.

În plus, asimetria unei distribuţii unimodale poate fi stabilită şi cu ajutorul **coeficienţilor lui Pearson**, **Fisher** şi **Yule**.

a) Coeficientul lui Pearson. Ţinînd cont de relaţia de legătură între modă, mediană şi medie, pentru coeficientul lui Pearson se cunosc trei expresii:

1. $\beta_P = \frac{\bar{x} - Mo}{\sigma}$, **2.** $\beta_P = \frac{3(\bar{x} - Me)}{\sigma}$ şi **3.** $\beta_P = \frac{3(Me - Mo)}{2\sigma}$.

Cu cît valoarea acestui coeficient se apropie de zero, cu atît mai mult distribuţia este asimetrică**.**

b) Coeficientul lui Fisher. În expresia de calcul a acestui coeficient intervine momentul centrat de ordin **3** şi cubul abaterii standard: $\beta_F = \frac{\mu^3}{\sigma^3}$. În mod analog, forma distribuţiei, avută în vedere, va fi una pronunţat asimetrică, dacă valoarea lui β_F tinde către zero.

c) Coeficientul lui Yule. Cu interpretare similară celorlaţi doi coeficienţi de mai sus, pentru acest indicator se prezintă două expresii:

1. $\beta_Y = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}$ şi

2. $\beta_Y = \frac{Q_1 + Q_3 - 2Me}{Q_3 - Q_1}$. A doua expresie se obţine din prima, ştiind că mediana este egală cu quartila a doua: $\mathbf{Me = Q_2}$.

4.2.11. Coeficientul de boltire

Coeficientul de boltire măsoară gradul de înălţare în apropierea valorii mod a unei repartiţii unimodale.

Coeficientul de boltire (**coeficientul de aplatizare**- kurtosis), cu notaţia β_2, se determină în modul următor:

$\beta_2 = \frac{\mu_4}{(\sigma^2)^2}$, unde σ^2 este dispersia, iar μ_4 este momentul centrat de ordin **4**.

O valoare mai mare a coeficientului de boltire descrie o supraînălţare în apropierea valorii mod. Mai tîrziu, vom vedea că pentru graficul densităţii de probabilitate a unei variabile aleatoare **X**, o valoare mai mare a coeficientului de boltire descrie faptul că valorile respective au probabilităţi de realizare mai mari decît valorile mai depărtate, nălţare care de cele mai multe ori este raportată la boltirea repartiţiei normale, $\beta_2 = 3$.

4.2.12. Coeficientul de exces

Coeficientul de exces se calculează în funcţie de β_2:

$$E = \beta_2 - 3.$$

Valoarea excesului reflectă:

1. O repartiţie **platicurtică**, dacă $E < 0$ (sau $\beta_2 < 3$),

2. O repartiţie **mezocurtică**, dacă $E = 0$ (sau $\beta_2 = 3$),

3. O repartiţie **leptocurtică**, dacă $E > 0$ (sau $\beta_2 > 3$).

În concluzie, din tipul repartiţiei se pot extrage informaţii despre concentrarea valorilor cu probabilităţi de realizare mai mari, iar valorile coeficienţilor de boltire şi de exces reflectă tocmai forma repartiţiei.