



SEMINAR BAZELE STATISTICII

SEMINAR 5

Analiza unei serii statistice univariate. Cazul unei variabile cantitative discrete

1. Mărimi medii:

- media aritmetică, modul, mediana.
- Quartile
- Decile

2. Indicatori ai dispersiei

- varianta (s^2),
- abaterea standard (s)
- coeficientul de variație (v)
- amplitudinea intervalului interquartilic
- abaterea medie liniară

3. Indicatori ai formei

- Asimetrie (Skewness)
- Boltire (Kurtosis)

RECAPITULARE SEMINAR 4

Quartile

$$Q_1 : U^{Q_1} = \frac{n+1}{4}$$

Decile

$$U^{D_1} = \frac{n+1}{10}$$

Abaterea medie liniară

$$\bar{d} = \frac{\sum_i |x_i - \bar{x}|}{n} \quad \bar{d} = \frac{\sum_i |x_i - \bar{x}| \cdot n_i}{\sum_i n_i}$$

Varianța

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2 \cdot n_i}{\sum_i n_i}$$

Abaterea standard (s) (deviația standard)

$$s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n}}$$

Coeficientul de variație (v)

$$v = \frac{s}{\bar{x}} \cdot 100$$

Amplitudinea intervalului interquartilic

$$I_Q = Q_3 - Q_1$$

Indicatori ai forme

Asimetria (Skewness):

$$\gamma_1 = \frac{\mu_3}{s^3}$$

$\gamma > 0$ – asimetrie la dreapta

$\gamma = 0$ – distributie simetrica

$\gamma < 0$ – asimetrie la stanga

Boltirea (Kurtosis)

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{\mu_4}{s^4} - 3$$

$\gamma_2 > 0$ – distributie leptocurtica

$\gamma_2 = 0$ – distributie mezocurtica (normala)

$\gamma_2 < 0$ – distributie platicurtica

Analiza unei serii univariate după o variabilă cantitativă discretă. Reprezentare grafică.

a. *Poligonul frecvențelor:*

- construirea acestuia presupune găsirea locului geometric al punctelor A_i de coordonate (x_i, n_i) sau (x_i, f_i) și unirea acestora prin segmente de dreaptă.
- aproximează forma unei distribuții.

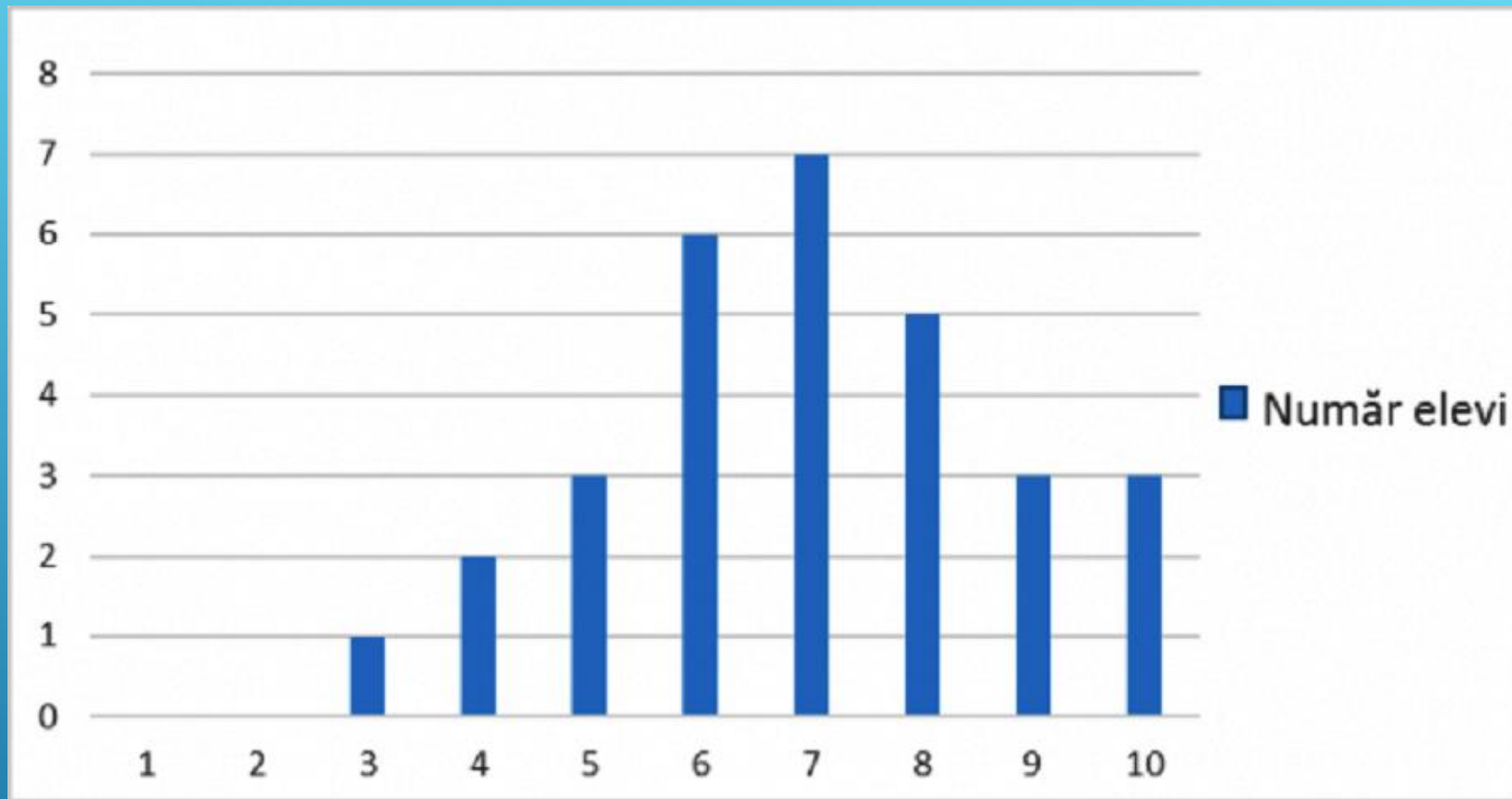
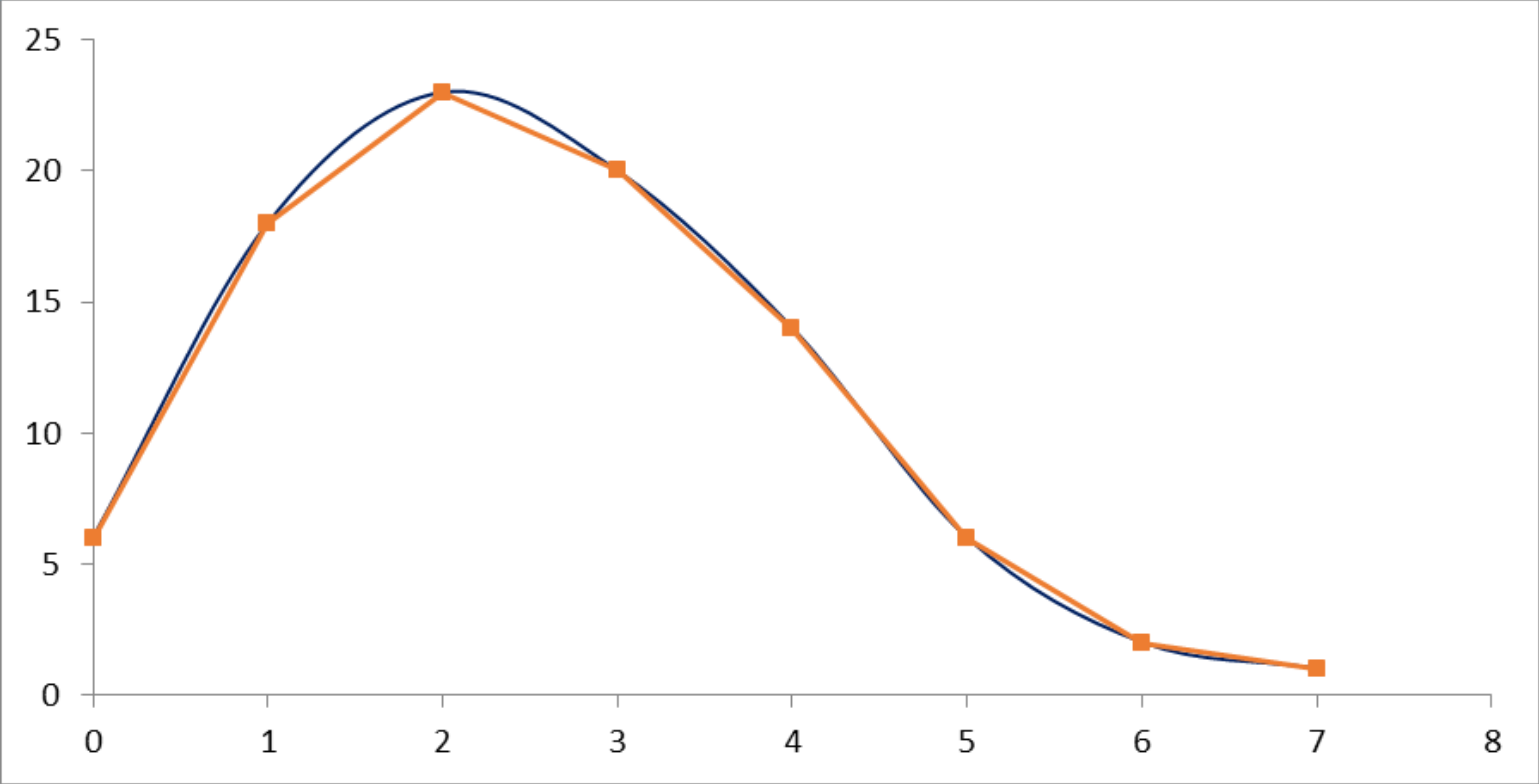
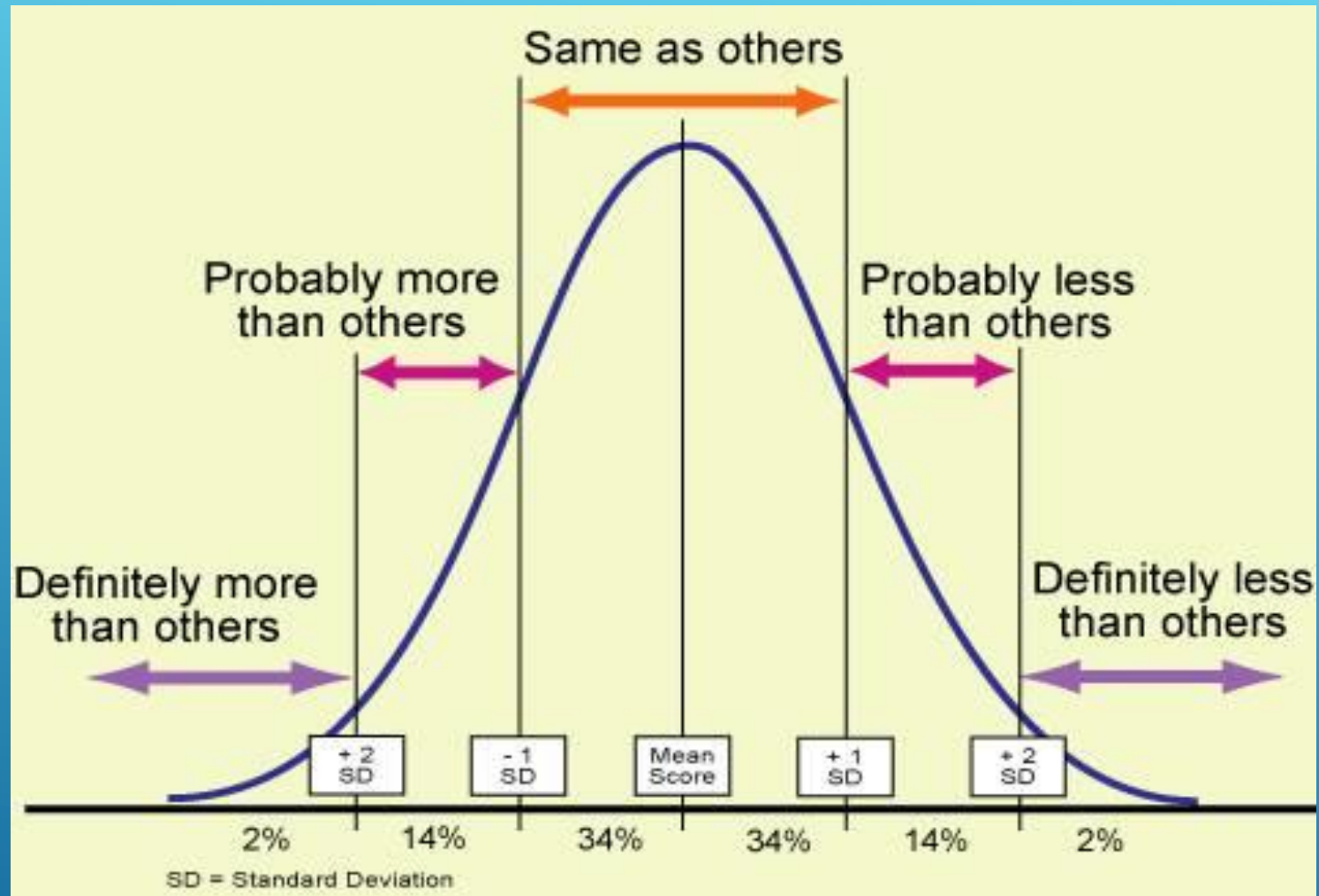


Figura 1. Distribuția elevilor de la un liceu după nota obținută la un examen, în martie 2016

x_i	n_i
0	6
1	18
2	23
3	20
4	14
5	6
6	2
7	1
Total	90





c. Diagrama “box-plot” sau “box-and-whiskers” (“cutie și mustăți”)

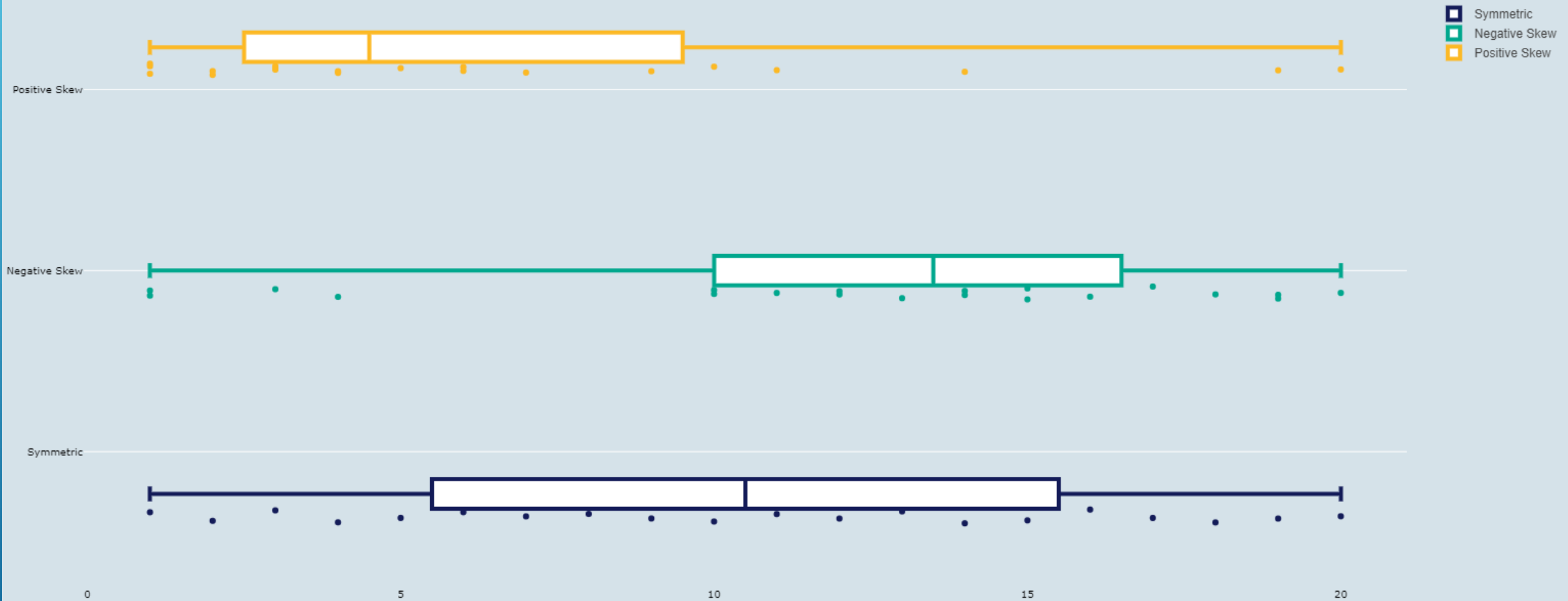
Diagrama bloxpot este un instrument deosebit de util pentru studiul distribuției variabilelor continue, și mai ales pentru detectarea rapidă a outlierilor unei serii de date (valori extreme sau aberante).

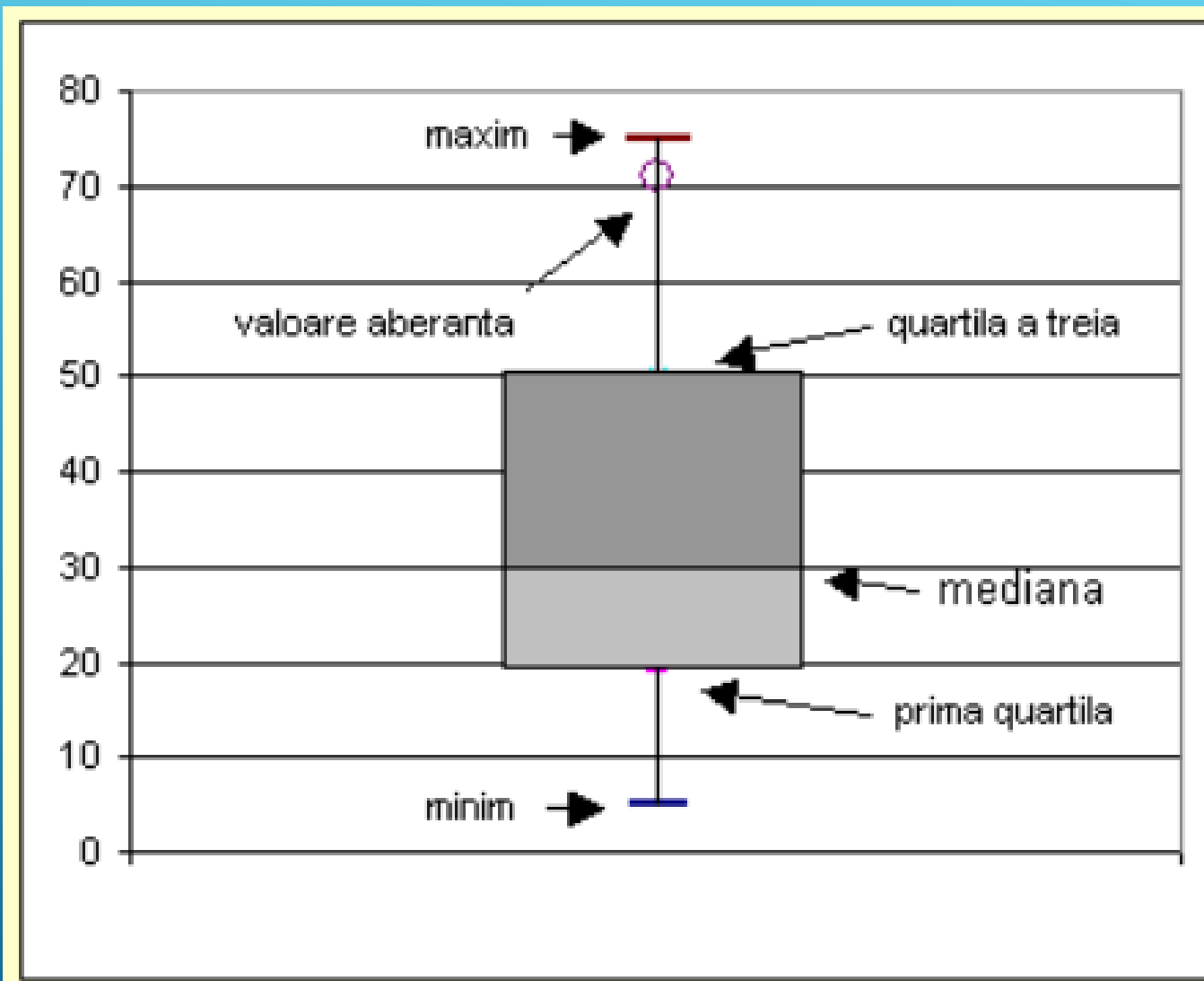
Forma diagramei (x_{min} , D_1 , Q_1 , Q_2 , Q_3 , x_{max} , D_9);

Avantaje:

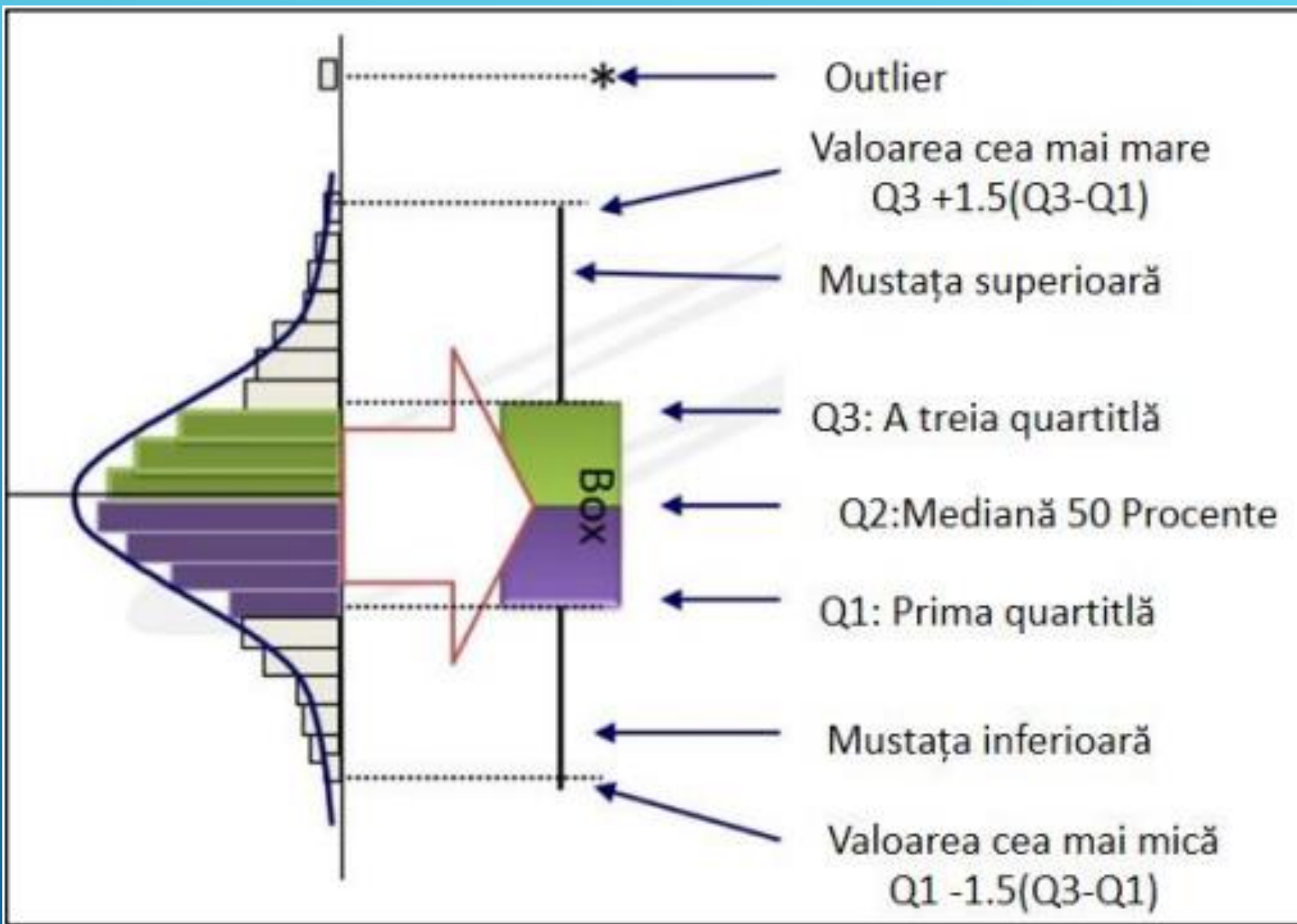
- permite identificarea punctelor extreme;
- permite aprecierea nivelului mediu (Me), dispersiei și asimetriei unei distribuții;
- facilitează compararea mai multor distribuții (prin reprezentarea simultană a diagramelor).

Box Plots & Skew

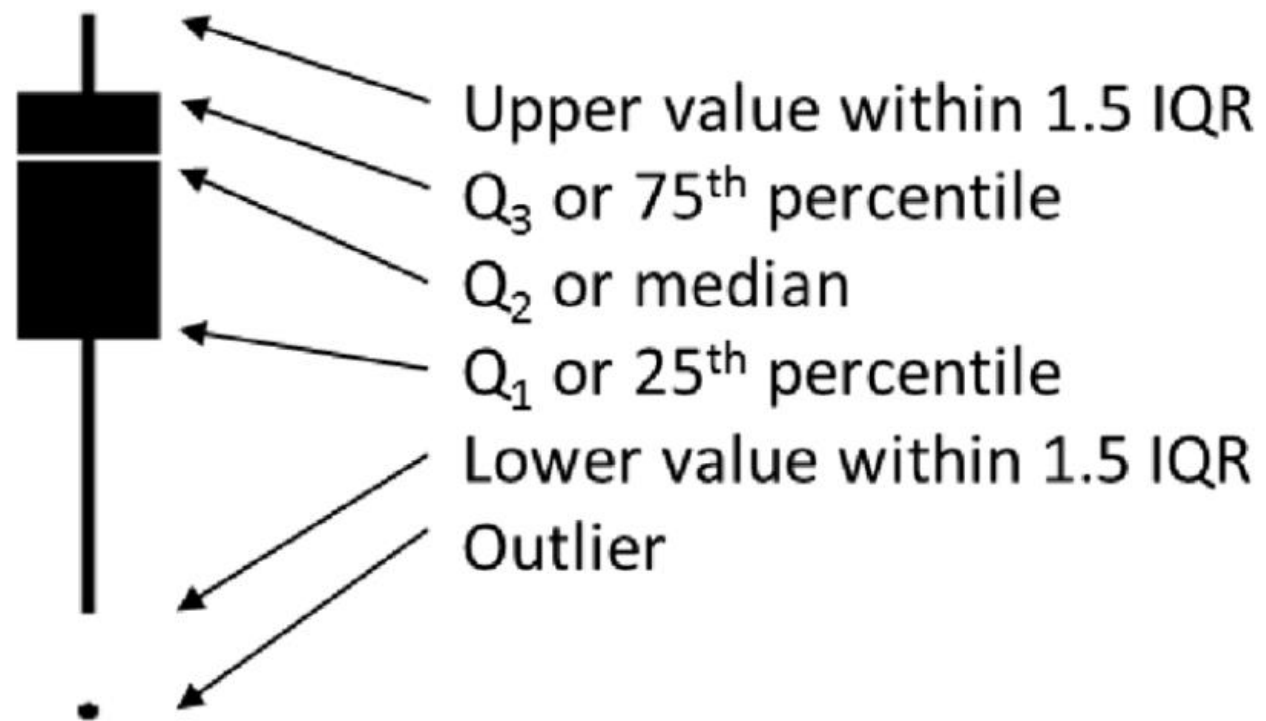




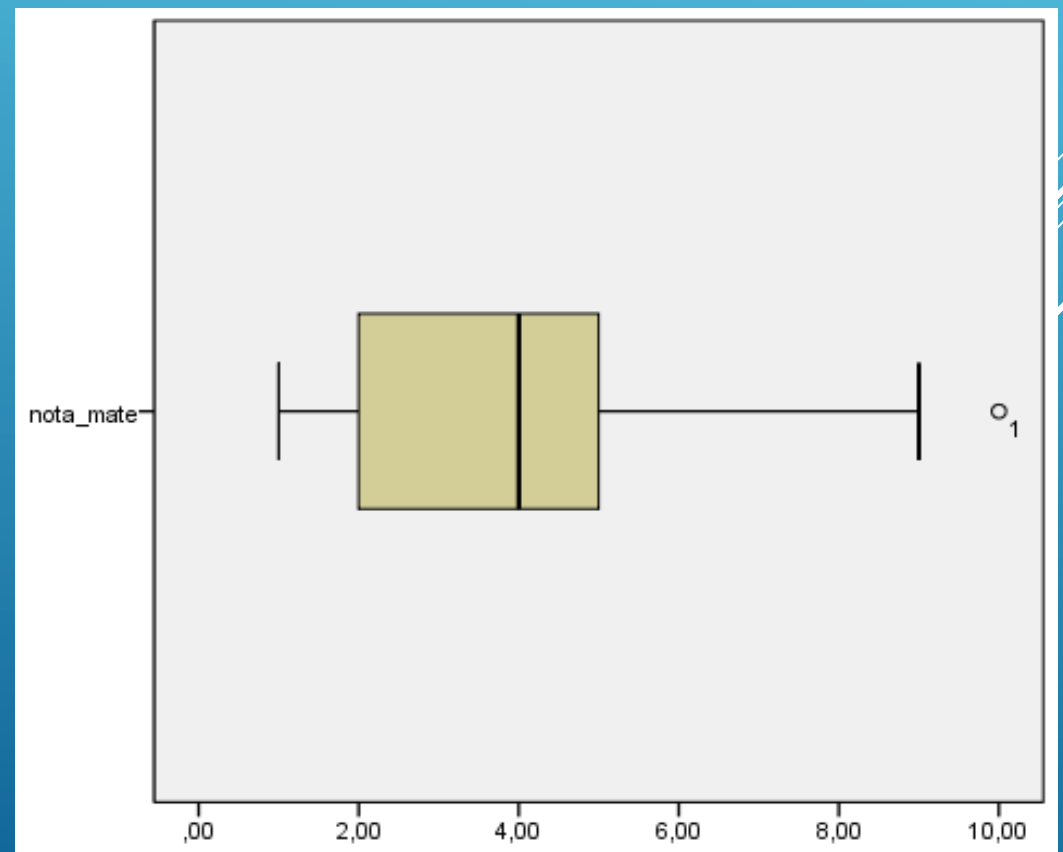
1. dacă distribuția este asimetrică la dreapta (predomină scorurile mari), cutia va fi apropiată de “mustașa” inferioară, iar mediana va fi și ea situată în partea inferioară a cutiei (aproape de prima quartilă)



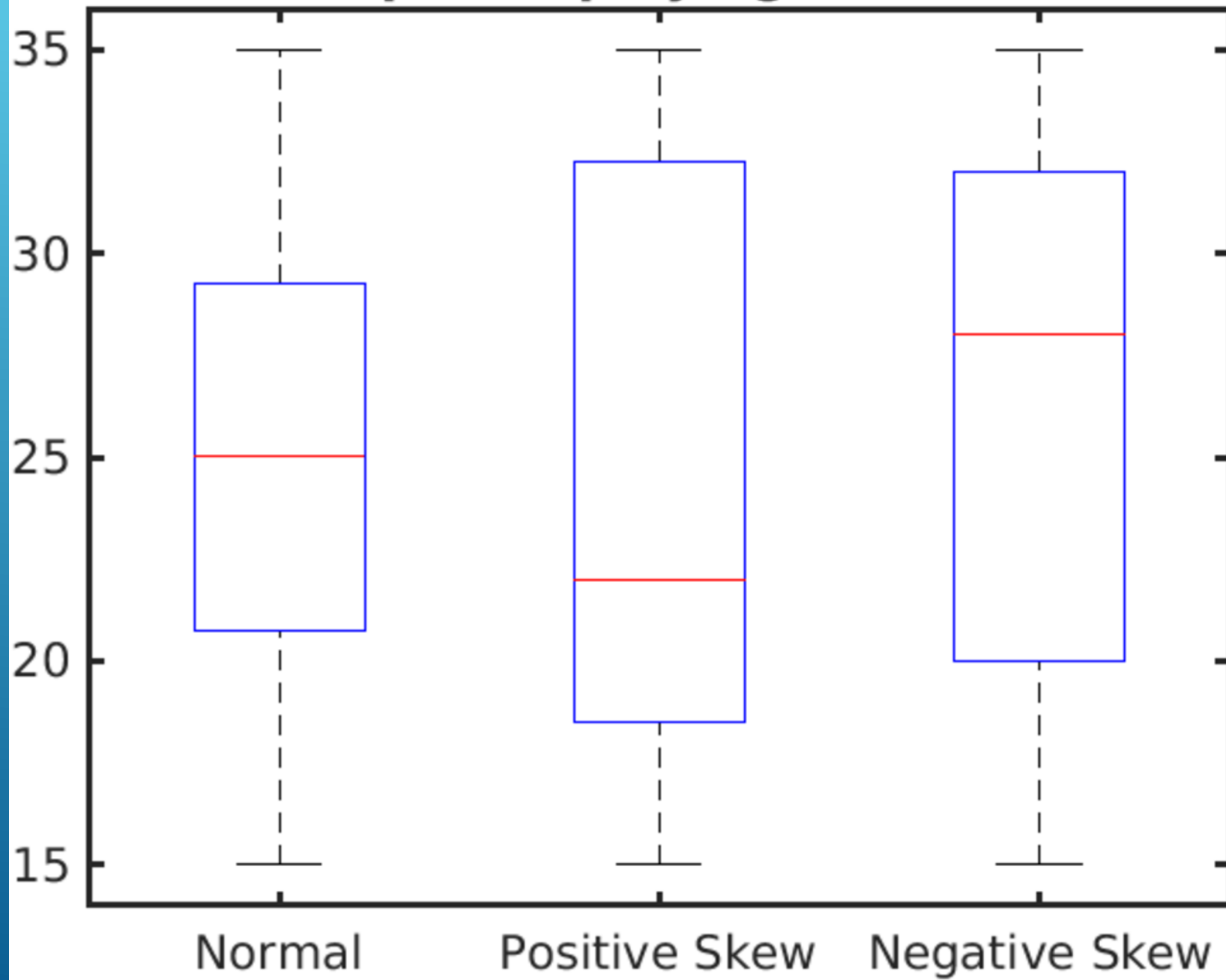
2. dacă distribuția este (aproximativ) simetrică, diagrama va fi de asemenea simetrică, cu alte cuvinte **cutia se va situa la mijlocul distanței dintre “mustăți”, iar mediana la mijlocul cutiei**

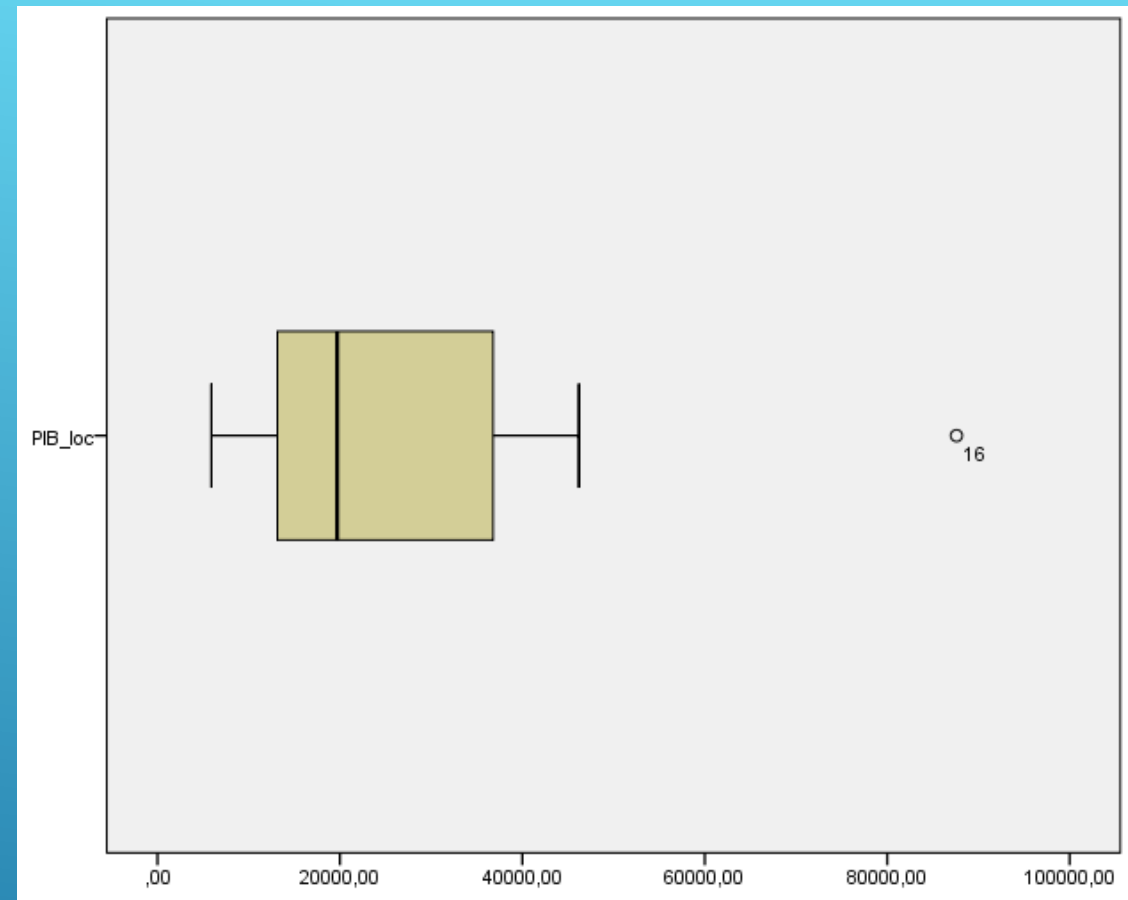
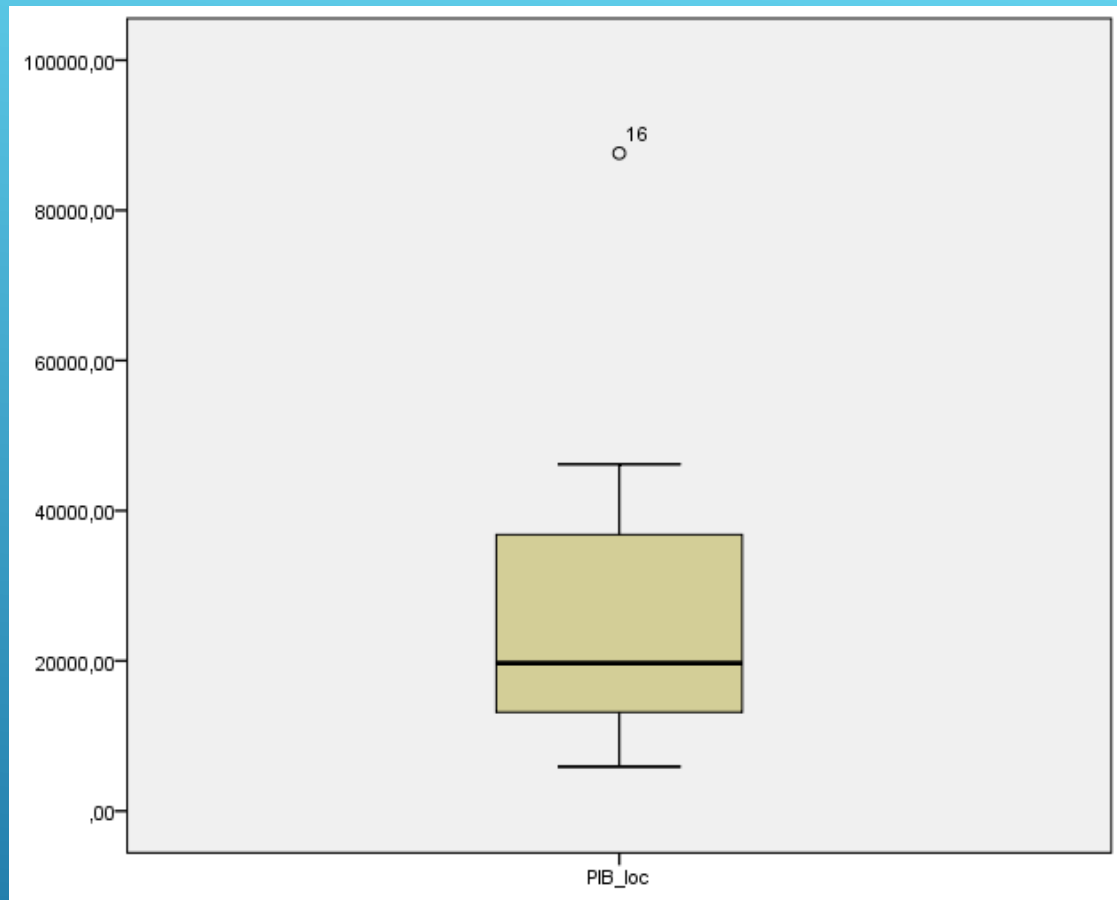


3. dacă distribuția este asimetrică la stânga (predomină scorurile mici), cutia este apropiată de mustața superioară, iar mediana de latura superioară a cutiei (a treia quartilă).

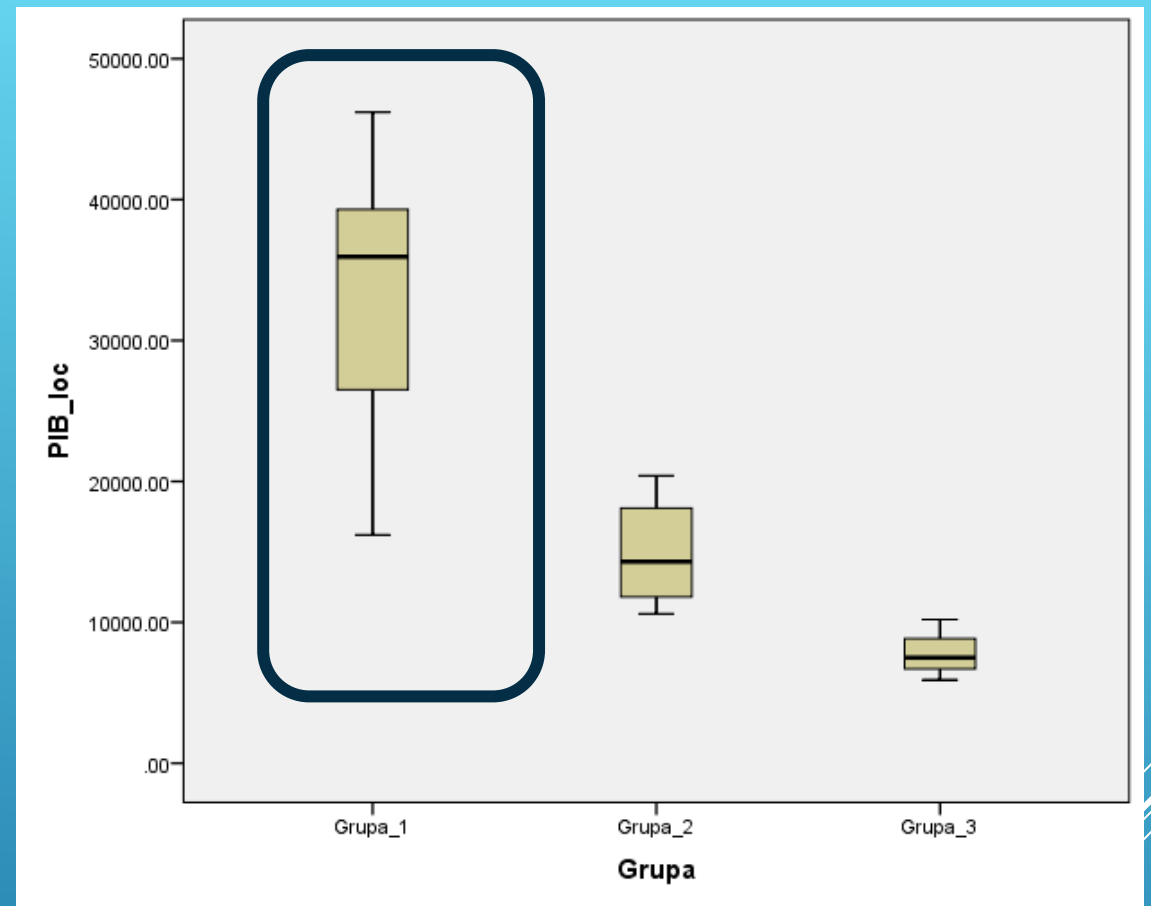
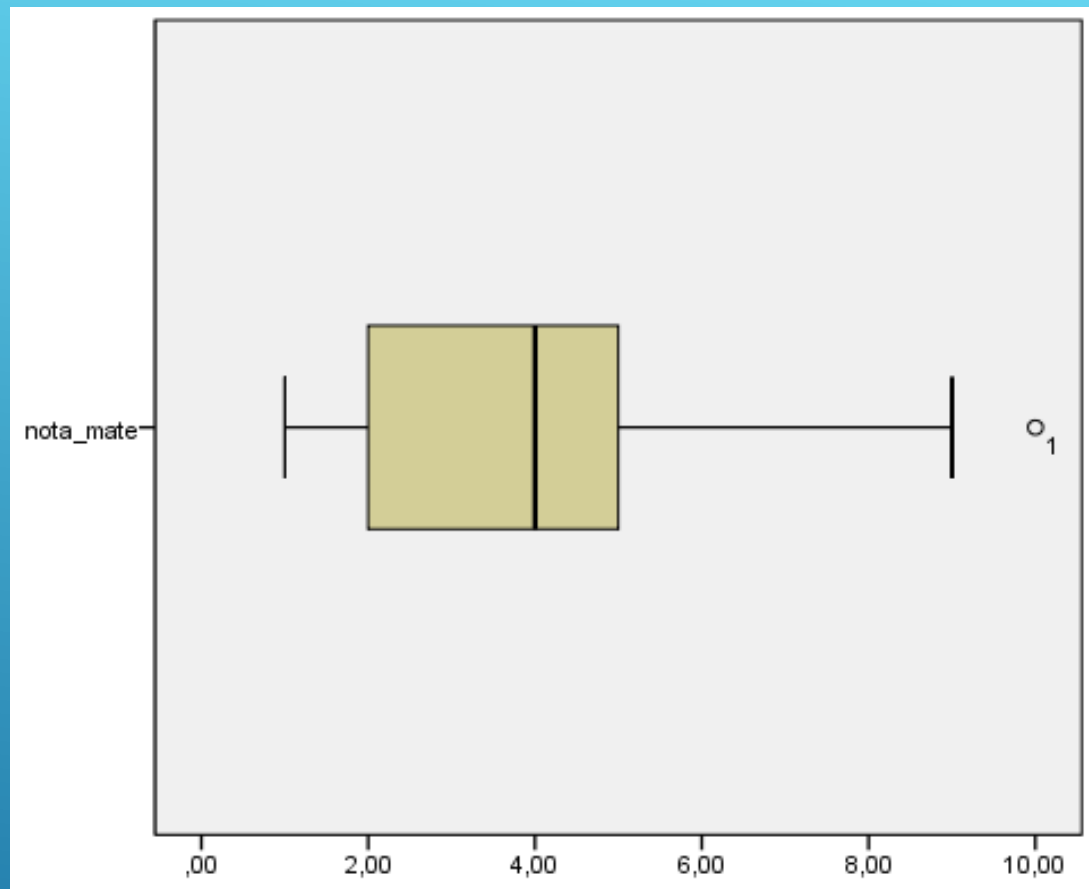


Example Displaying Skewness





Asimetrie la dreapta



Asimetrie la stanga

Analiza unei serii univariate după o variabilă cantitativă continuă:

1. Prezentarea seriei statistice

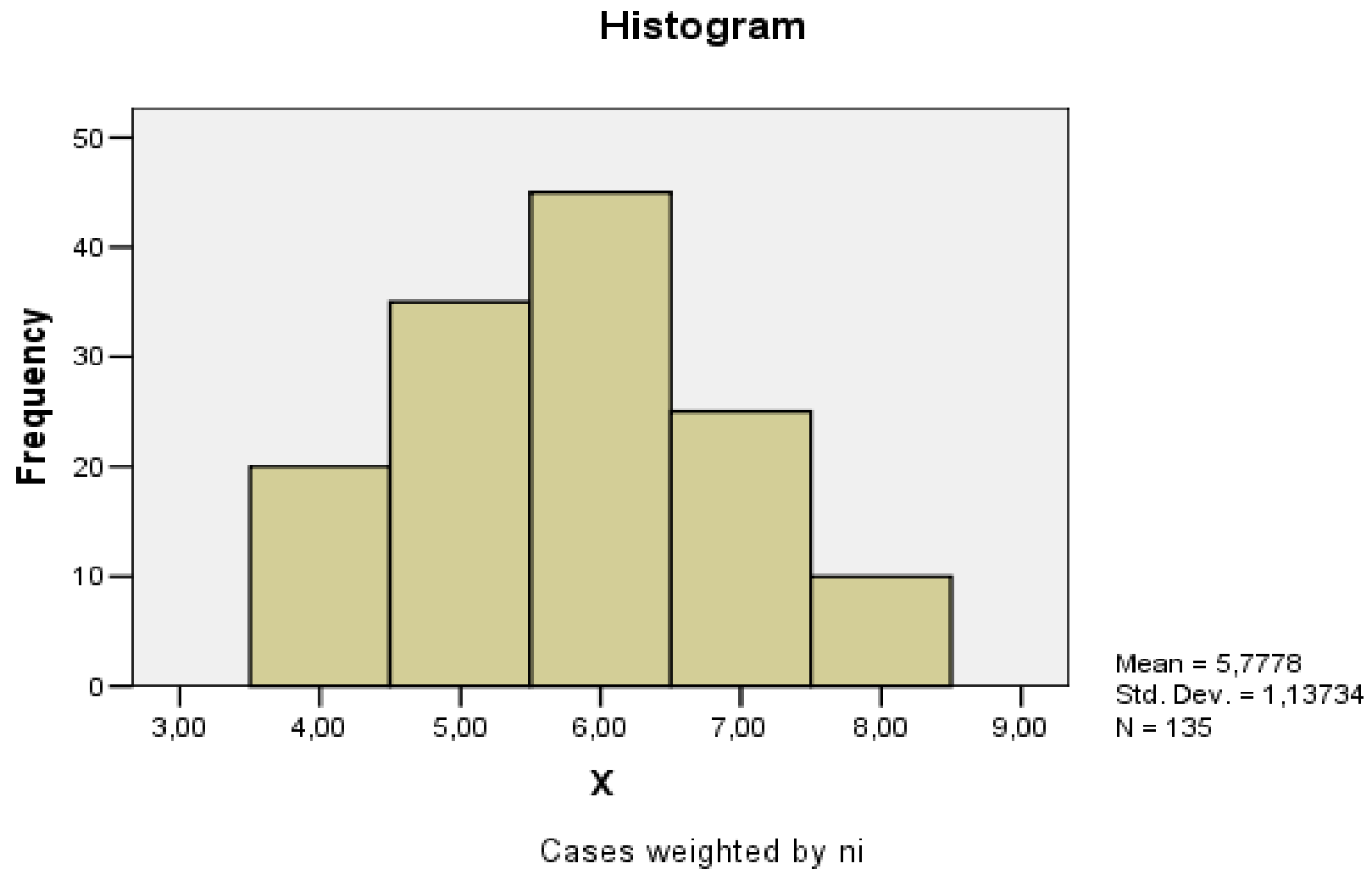
- gruparea unităților statistice este realizată pe *intervale de variație*.

Observație:

- Gruparea pe intervale de variație duce la pierderea unei părți a informației inițiale.

2. Prelucrarea seriei statistice folosind metode grafice

- a. Histograma
- b. Poligonul frecvențelor
- c. Curba frecvențelor



a. Histograma

Analiza unei serii univariate

1. Variabilă cantitativă

- A. Variabilă discretă
- B. Variabilă continuă

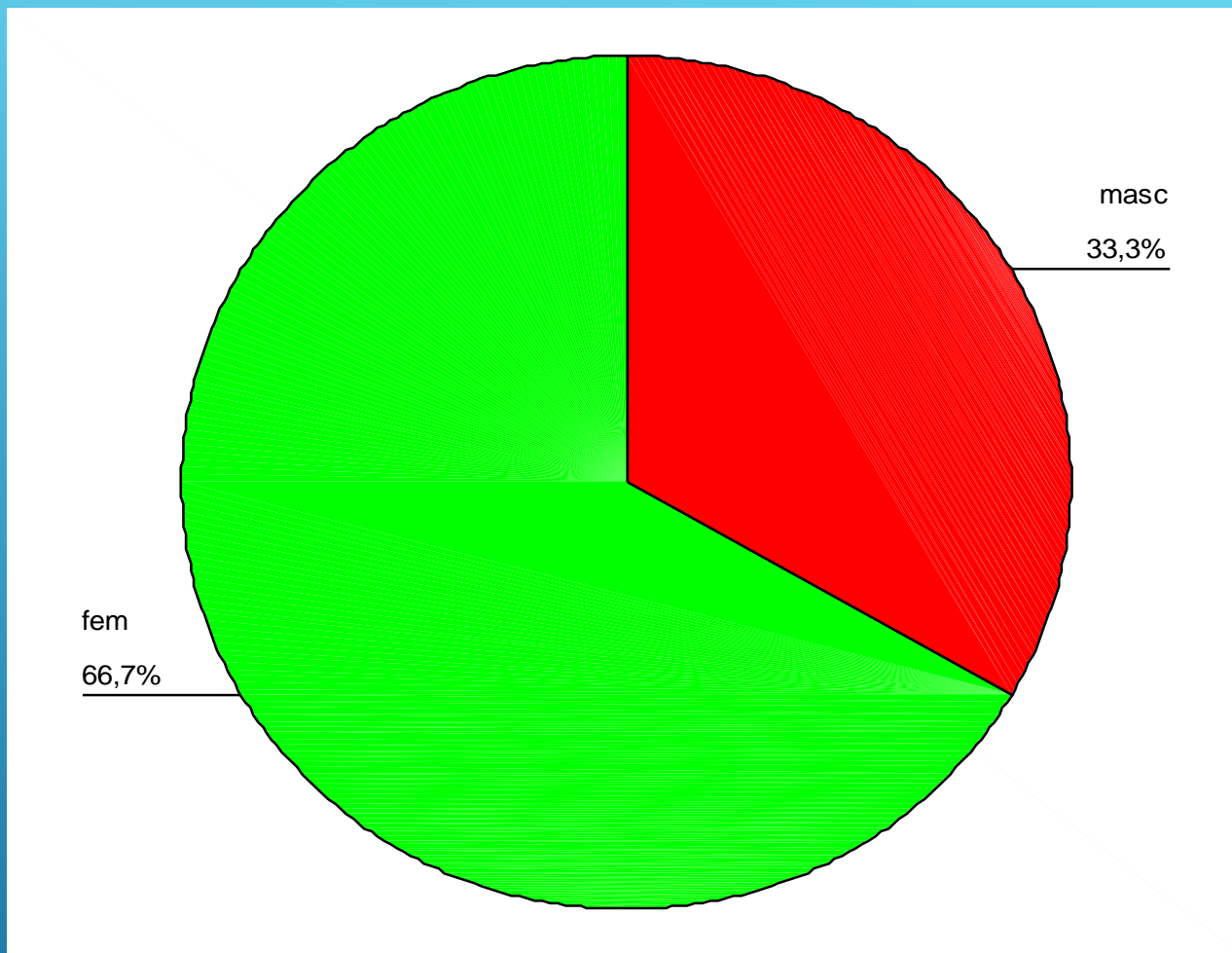
2. Variabilă calitativă

I. Tipuri de variabile

- A. Variabile nominale
- B. Variabile ordinale

A) Variabile nominale

- Pentru a reprezenta structura pe categorii la nivelul unui eșantion se calculează **frecvențe relative**;
- Reprezentarea frecvențelor unui eșantion se realizează folosind diagramele: dreptunghiul și cercul de structură (*Bar Chart* sau *Pie Chart*).
- Indicatori specifici: **frecvențe relative, modul**.



Exemplu

Aplicatii

1. În urma prelucrării datelor privind notele obținute de un grup de studenți obținut, s-au obținut următoarele rezultate:

Se cere să se aprecieze:

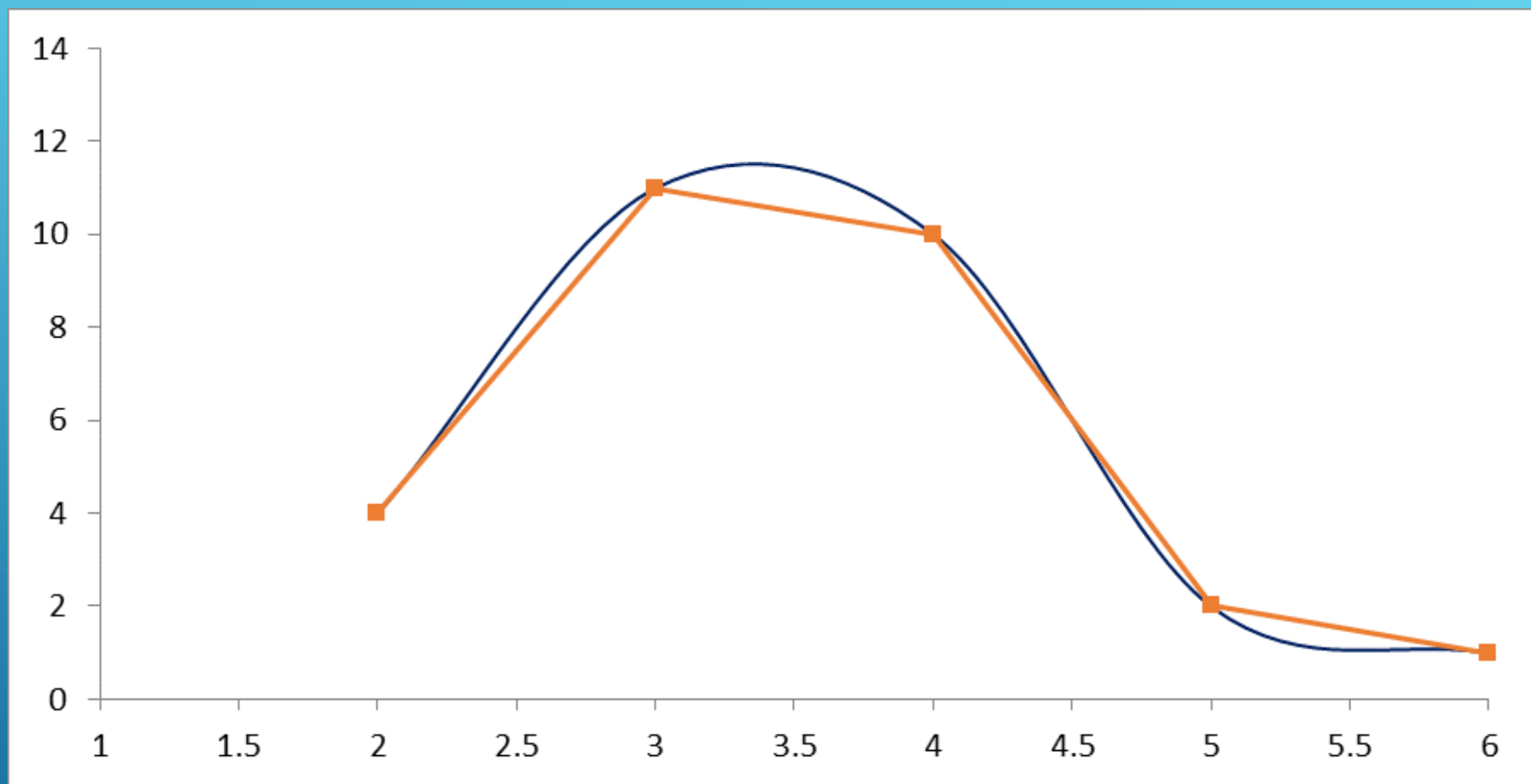
- a) reprezentativitatea mediei;
- b) să se aprecieze forma distribuției;
- c) să se interpreteze valorile quartilelor;

Statistics		
note		
N	Valid	71
	Missing	0
Mean		7.34
Median		7.00
Mode		8
Std. Deviation		1.253
Variance		1.570
Skewness		.268
Kurtosis		-.523
Percentiles	25	6.00
	50	7.00
	75	8.00

2. Se consideră distribuția familiilor după numărul de membri prezentată astfel:

Să se reprezinte poligonul și curba frecvențelor folosind frecvențele absolute.

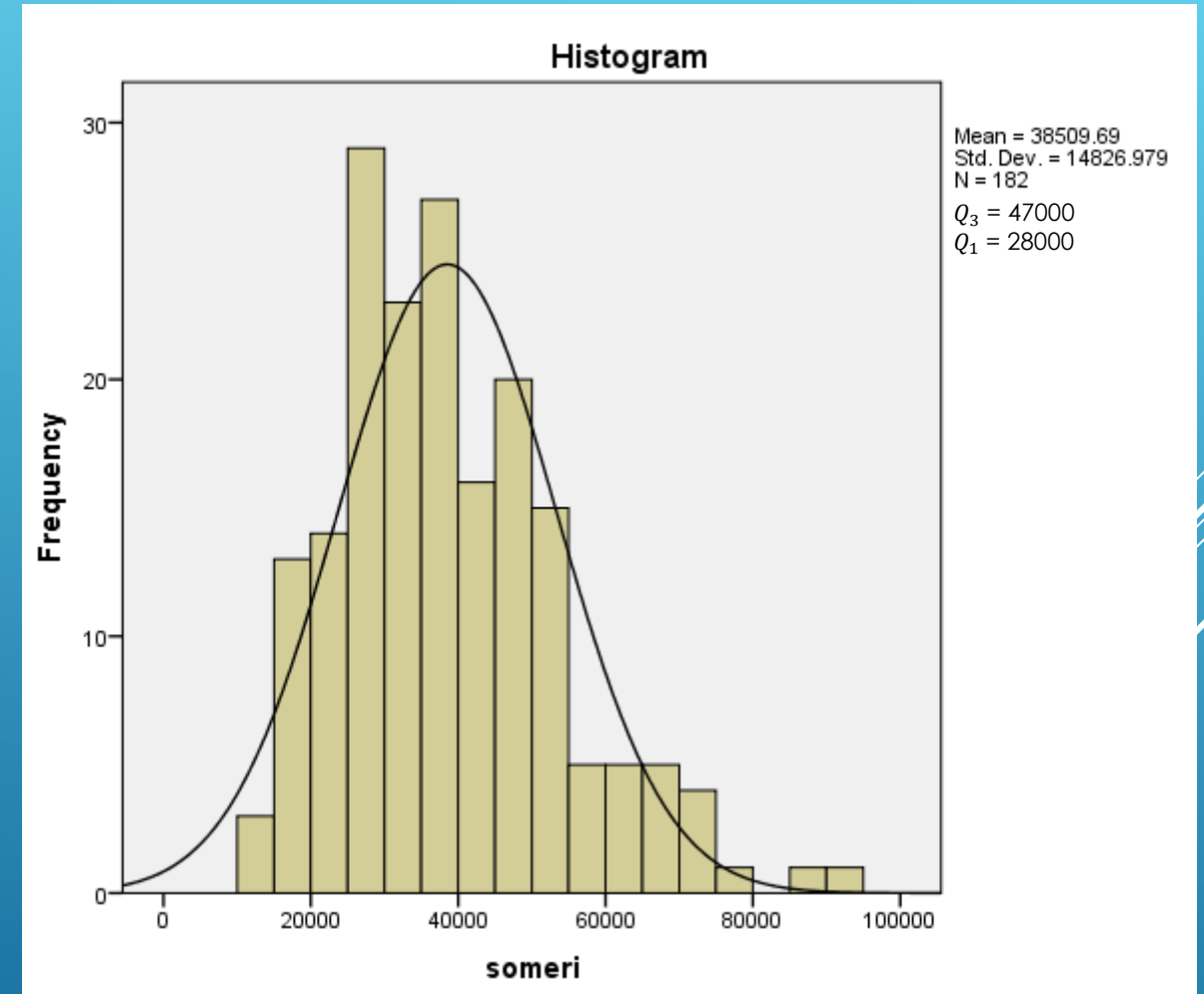
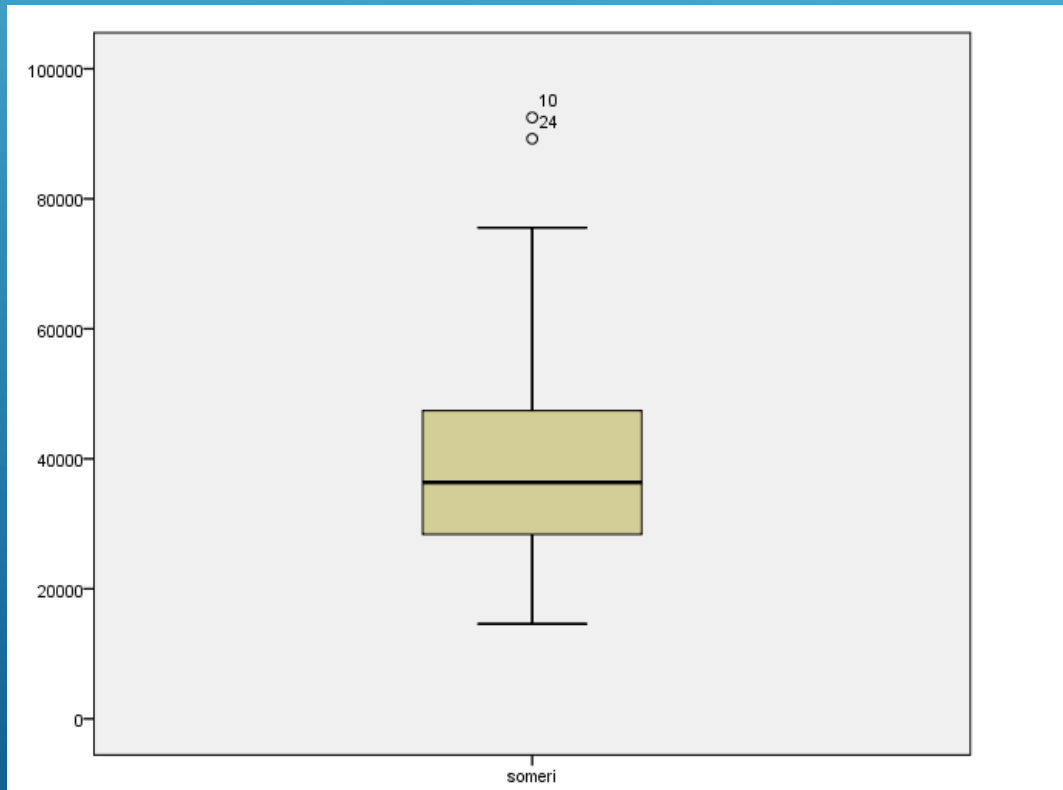
Număr membri (x_i)	Număr familii (n_i)
2	4
3	11
4	10
5	2
6	1
TOTAL	n=28



3. Pentru distribuția numărului de someri din Romania la nivelul anului 2013, s-au înregistrat următoarele rezultate:

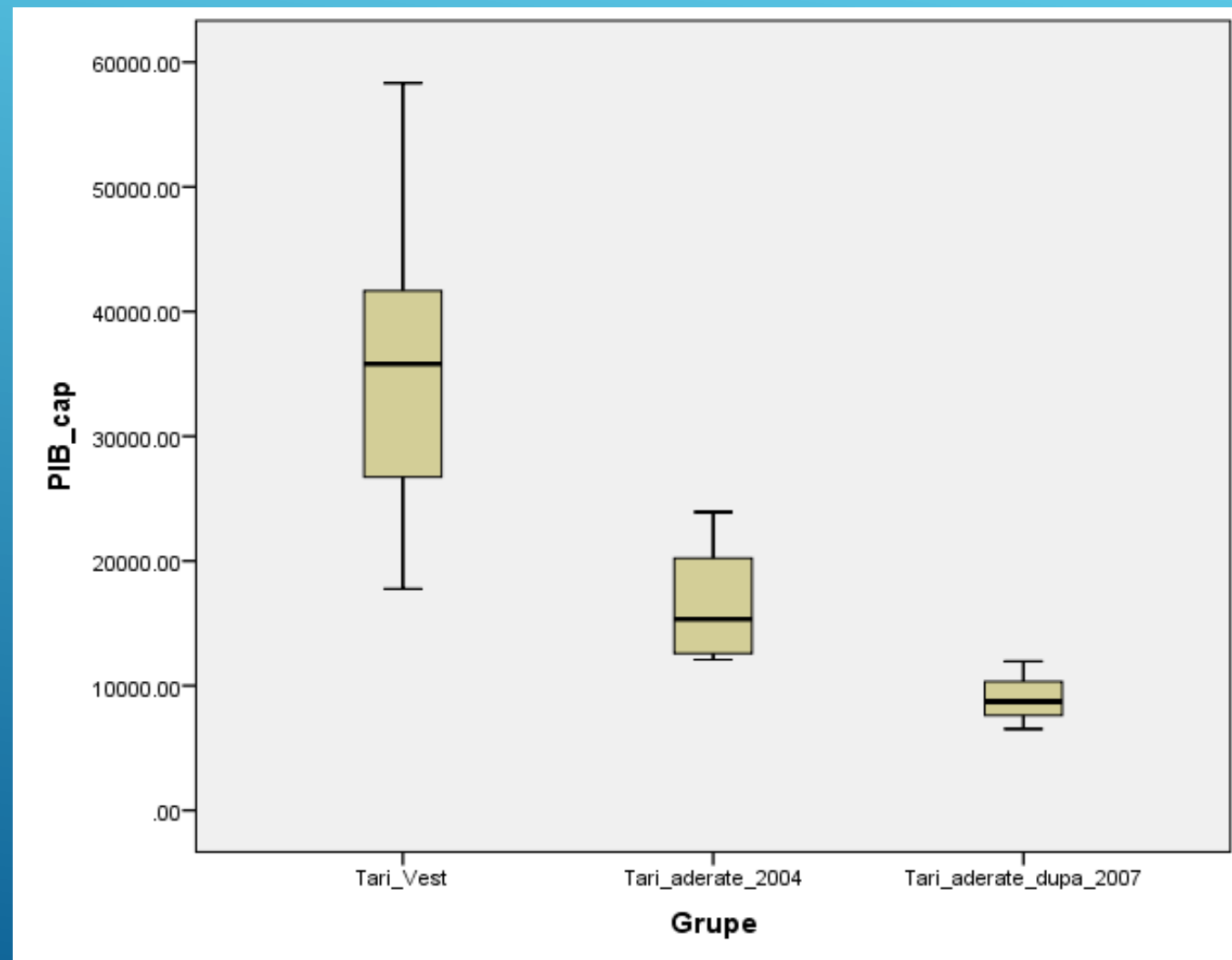
Se cere:

- a) să se precizeze forma distribuției.
- b) să se interpreteze valoarea quartilei trei.
- c) să se interpreteze amplitudinea intervalului interquartilic.



4. Se consideră distribuția PIB/locuitor (euro) în țările Uniunii Europene, în anul 2018, în funcție de momentul aderării (Grupa_1 - formată din țările care erau înainte de extinderea spre Europa Centrală și de Est (fără Luxembourg); Grupa_2 - formată din țările care au aderat în anul 2004; Grupa_3 - formată din țările care au aderat în anii 2007 și 2013) . Reprezentarea grafică a acestor distribuții este prezentată în figura de mai jos:

Se cere să se caracterizeze cele trei grupe de țări (comparând nivelurile medii și dispersia).

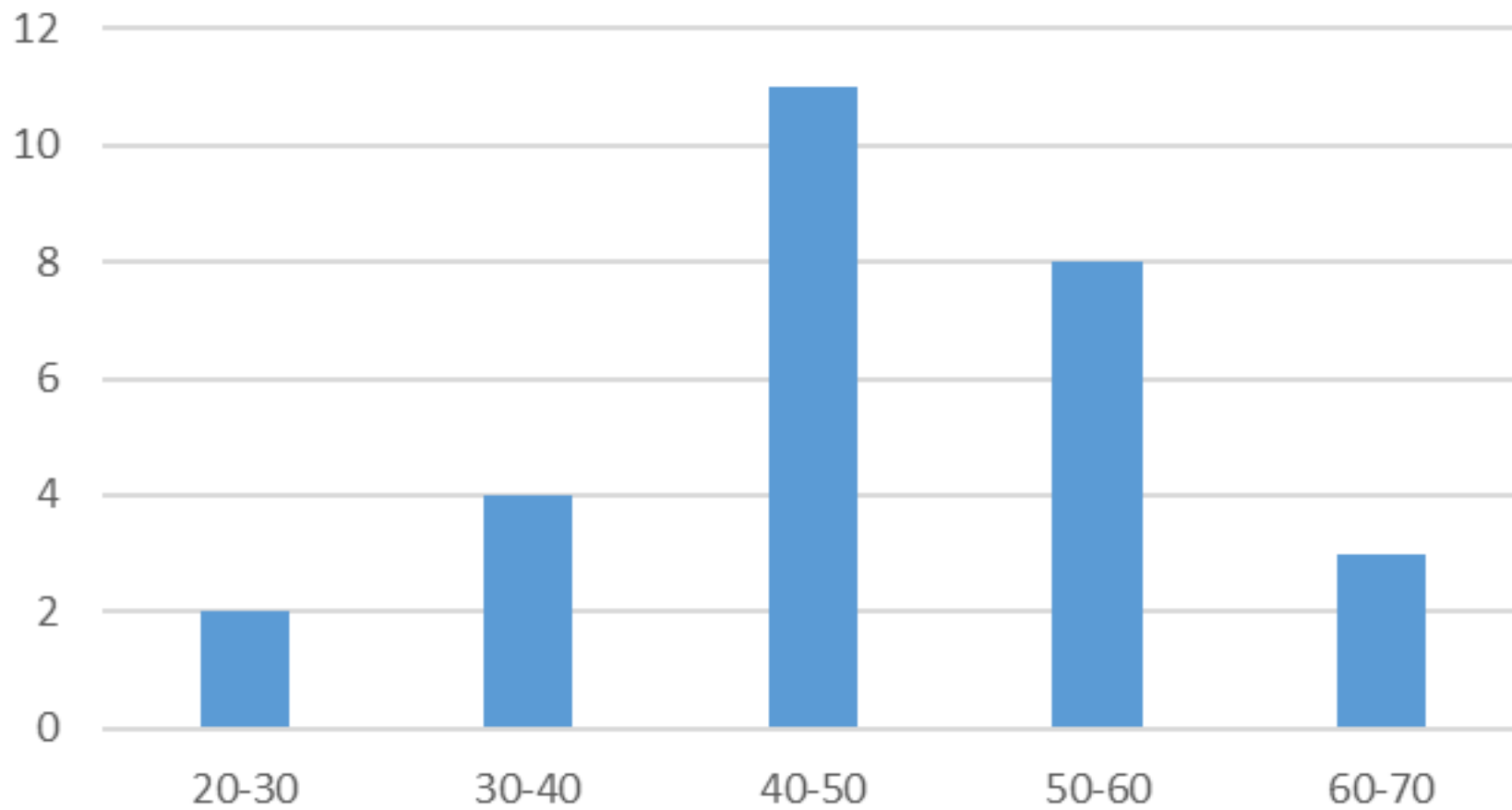


5. Se consideră distribuția numărului de minute alocate de fiecare student zilnic pentru pregătirea seminarului bazele statisticii, prezentată în tabelul de mai jos:

Număr minute (x_{i-1}, x_{i+1})	Număr studenti (n_i)
20-30	2
30-40	4
40-50	11
50-60	8
60-70	3
TOTAL	28

Să se calculeze numărul mediu de minute.

Minutele alocate de studenti pentru pregatirea seminarului bazele statisticii

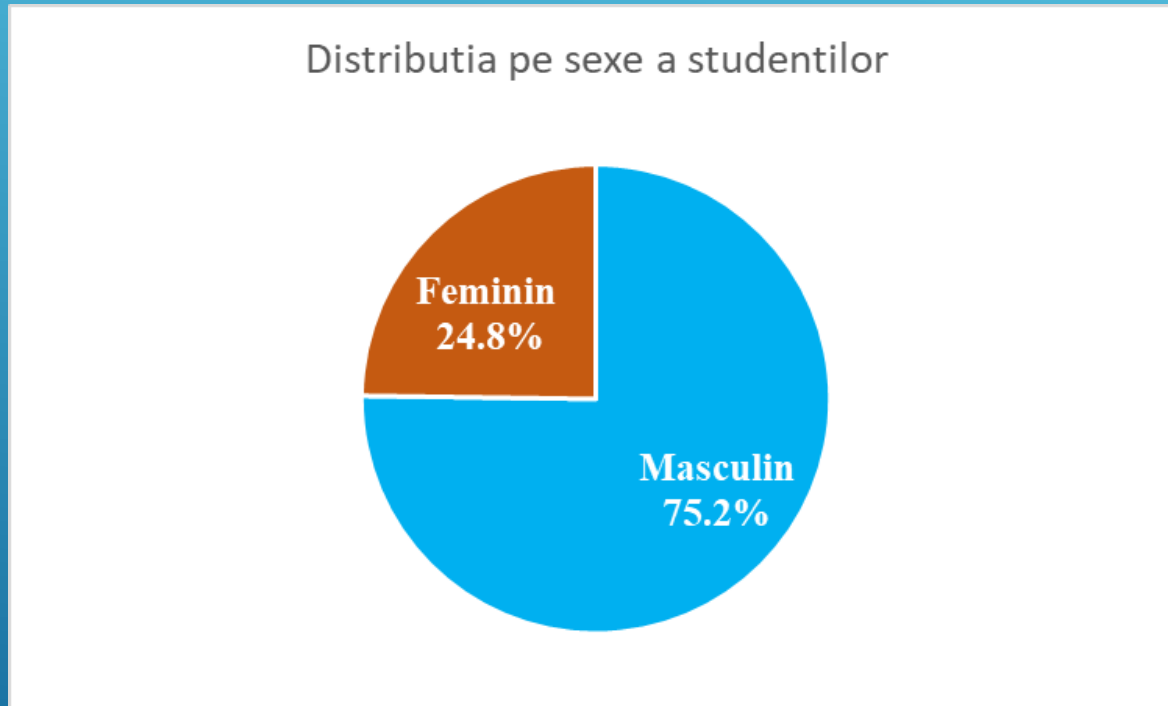


6. Pentru distribuția persoanelor după gen (masculin și feminin) prezentată în tabelul de mai jos, să se afle valoarea modului.

Genul persoanei (x_i)	Număr persoane (n_i)	Pondere persoane (f_i)
Masculin	25	22.7%
Feminin	85	77.3%
TOTAL	$n=110$	100%

7. Reprezentarea grafică a unei serii de studenți pe sexe este realizată în figura de mai jos:

Să se interpreteze valoarea modului.



8. Pentru distribuția persoanelor după nivelul de studii absolvite, prezentată în tabelul de mai jos, să se afle valoarea mediană și quartila 3.

Nivel studii (x_i)	Număr persoane (n_i)
Studii liceale	23
Studii universitare	16
Studii post-universitare	17
Doctorat	3
TOTAL	$n = 59$

10. Pentru un eșantion de salariați, se înregistrează salariul lunar (mii lei) la 1 ianuarie 2017 și se obțin următoarele rezultate:

Se cere:

a) Să se aprecieze reprezentativitatea mediei.

b) Să se aprecieze forma distribuției
(asimetria și boltirea).

Column1	
Mean	2.76
Standard Error	0.301552
Median	2
Mode	2
Standard Deviation	1.507758
Sample Variance	2.273333
Kurtosis	-0.77295
Skewness	0.523621
Range	5
Minimum	1
Maximum	6
Sum	69
Count	25

1. Graficul a carui reprezentare se utilizeaza: valoarea minima, quartile 1, mediana, quartile 3 si valoarea maxima

maxima se numeste:

- a) Poligonul frecventelor
- b) Poligonul frecventelor cumulate
- c) Diagrama Box-Plot

2. Diagrama Box-Plot este:

- a) O distributie de frecventa
- b) Acelasi lucru cu poligonul frecventelor
- c) O reprezentare grafica pe baza indicatoriilor medii de pozitie

