



SEMINAR BAZELE STATISTICII

**SEMINAR 7-
RECAPITULARE**

POPULAȚIA STATISTICĂ

- pentru a fi bine definită, o populație statistică trebuie să aibă precizată omogenitatea elementelor sale componente din punct de vedere **calitativ, de timp și de spațiu**.

Exemplu: Studenții cu vârsta de peste 25 de ani de la FEAA Iași, la 01 octombrie 2021

- volumul unei populații se notează cu N

EȘANTIONUL

- ▶ Reprezintă un sub-ansamblu de unități ale unei populații extras din populația statistică după anumite procedee.
- ▶ Volumul eșantionului se notează cu n

*Exemplu: Studenții **de la specializarea EAI**, cu vârsta de peste 20 de ani de la FEAA Iași, la 01 octombrie 2021*

UNITĂȚILE STATISTICE

➤ sunt elementele componente ale unei populații statistice.

Exemplu:

- Un student de peste 30 ani

VARIABILA STATISTICĂ

- ▶ reprezintă însușirea, trăsătura esențială purtată de unitățile statistice ale unei populații.
- ▶ se notează cu X .
- ▶ Valorile (variantele) variabilei se notează cu x_i , cu $i=1, m$.

TIPURI DE VARIABILE STATISTICE

a) după modul de exprimare

1. Variabile numerice (cantitative):

- ❖ variabile ale cărei valori se exprimă prin cifre, numere.

Exemplu: vârsta, înălțimea, salariul, greutatea, etc.

- ❖ variabile discrete: variabile ale cărei valori se exprimă prin numere întregi (fără subdiviziuni zecimale).

Exemplu: numărul de persoane, numărul de copii/familie, numărul de țări, numărul de camere, etc

- ❖ variabile continue: sunt variabile care pot lua o infinitate de valori într-un interval.

Exemplu: PIB/loc., salariul, venitul, înălțimea, etc.

TIPURI DE VARIABILE STATISTICE

2. Variabile nenumerice (calitative):

- ❖ sunt variabile ale cărei valori se exprimă prin cuvinte.

Exemplu: nivelul de studii, sexul persoanei, etc.

- ❖ variabile nominale: sunt variabile pentru care între valorile acestora nu există o ierarhie.

Exemplu: sexul persoanei: 1- masculin, 2 – feminin

Caz particular: variabile alternative (dichotomice, dummy) sunt variabilele care au 2 valori cărora li se acordă codurile 0 (Nu) și 1 (Da).

Exemplu: Produsul poate fi 1 – bun sau 0 - rău

TIPURI DE VARIABILE STATISTICE

❖ *variabile ordinale: există o ierarhie între categoriile acesteia.*

Exemplu: nivelul de studii, preferința pentru un produs cu valorile

Foarte rău Rău Nici rău, nici bun Bun Foarte bun

1 2 3 4 5

SCALE DE MĂSURARE

❖ Definire:

Scala este un continuum de cifre sau de simboluri, plasate ierarhic, de la inferior la superior.

❖ Tipuri de scale

1. Variabile calitative

α. **Scala nominală** presupune acordarea de numere (coduri) fiecărei categorii a unei populații (doi indivizi care aparțin unor categorii distincte au valori diferite).

✓ este caracteristică variabilelor nominale

Exemplu: Sexul persoanei – 1 M, 2 F

SCALE DE MĂSURARE

b. Scala ordinală presupune ca atribuirea de coduri numerice pentru fiecare categorie să se realizeze în ordine descrescătoare, după importanța lor.

✓ este caracteristică variabilelor ordinale.

Exemplu: Preferința pentru un produs

SCALE DE MĂSURARE

2. Variabile cantitative

a. Scala interval are ca proprietăți identitatea, ordinea și faptul că intervalul între numere are un sens.

❖ se poate, astfel, compara diferența dintre două valori ale aceleiași variabile.

Exemplu:

- *măsurarea temperaturii în sistemul Celsius și în sistemul Fahrenheit: diferența dintre două temperaturi are un sens.*

SCALE DE MĂSURARE

b. Scala raport

- ❖ are aceleași proprietăți ca scala interval și, în plus, posedă un zero absolut (considerat punct de referință).
- ❖ diferența și raportul dintre două valori au un sens.

Exemplu: PIB., salariul, vârsta, etc.

PREZENTAREA SERIEI: SERII SIMPLE ȘI SERII CU FRECVENȚE DIFERITE

O serie statistică este reprezentată de termenii (x_i, n_i) sau (x_i, f_i) , cu $i=1, m$.

seria simplă $X: (x_i)$, cu $i=1, m$, când $n_1 = n_2 = \dots = n_i$.

seria cu frecvențe diferite $X: \begin{pmatrix} x_i \\ n_i \end{pmatrix}$, când $n_i \neq n_j$.

$X: \begin{pmatrix} x_i \\ f_i \end{pmatrix}$, cu $f_i = n_i/n$.

FRECVENȚE ABSOLUTE CUMULATE CRESCĂTOR ($N_i \downarrow$) SAU DESCRESCĂTOR ($N_i \uparrow$)

- exprimă **numărul de unități statistice** cumulate “până la” (\downarrow) sau “peste” (\uparrow) nivelul considerat al caracteristicii, adică valori $\leq x_i$, respectiv $\geq x_i$.

$$N_i \downarrow = N_{i-1} \downarrow + n_i = \sum_{h=1}^i n_h$$

$$N_i \uparrow = N_{i+1} \uparrow + n_i = \sum_{h=i}^m n_h$$

Unde:

n_i - efectivul corespunzator modalității i

n - efectivul total

N_i - efectivul cumulat până la” sau “peste” modalitatea i

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)

Mod de calcul în cazul seriilor simple și seriilor cu frecvențe diferite (variabilă discretă).

► Media simplă: $\bar{x} = \frac{\sum_i x_i}{n}$

► Media ponderată: $\bar{x} = \frac{\sum_i x_i \cdot n_i}{\sum_i n_i}$ sau $\bar{x} = \sum_i x_i \cdot f_i$

Observații:

- de cele mai multe ori, valoarea mediei nu coincide cu niciuna din valorile individuale din care s-a calculat (s-a extras esențialul din grup);
- **media aritmetică este sensibilă la prezența valorilor extreme (outliers);**

- ▶ este valoarea variabilei cea mai frecvent observată într-o distribuție, adică valoarea x_i care corespunde frecvenței maxime ($n_{i\max}$).

Observații:

- ▶ modul poate fi aflat doar în cazul seriilor **cu frecvențe diferite**.
- ▶ o distribuție poate avea una, două sau mai multe valori modale (serii unimodale, bimodale sau plurimodale).

Interpretare: Cele mai multe unități înregistrează valoarea modală.

MODUL (MO)

- este acea valoare a variabilei unei serii ordonate, crescător sau descrescător, până la care și peste care sunt distribuite în număr egal unitățile colectivității:

jumătate din unități au valori mai mici decât mediana, iar jumătate au valori mai mari decât mediana.

- corespunde locului unității mediane calculate astfel:

$$U^{Me} = \frac{n + 1}{2}$$

MEDIANA (ME)

AFLAREA MEDIANEI SE FACE DIFERIT ÎN FUNCȚIE DE TIPUL SERIEI:

1. Serii simple:

- număr impar de termeni - atunci mediana este egală cu termenul central al seriei ordonate crescător sau descrescător
 - număr par de termeni - atunci mediana este egală cu media aritmetică simplă a celor 2 termeni centrali ai seriei ordonate crescător sau descrescător.
- Pentru un șir de date ordinale format din număr par de termeni, mediana este egală cu una din cele două variante din centrul seriei dacă aceste variante sunt egale, iar dacă variantele nu sunt egale mediana ia 2 valori deoarece nu se poate face media lor

2. Serii cu frecvențe diferite

- se calculează unitatea mediană (U^{Me}).
- se calculează $N_i \downarrow$
- se află prima valoare $N_i \downarrow \geq U^{Me}$
- valoarea x_i corespunzătoare acesteia este Me.

Observație:

- mediana nu este influențată de valorile extreme.

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

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

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 univariante 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.

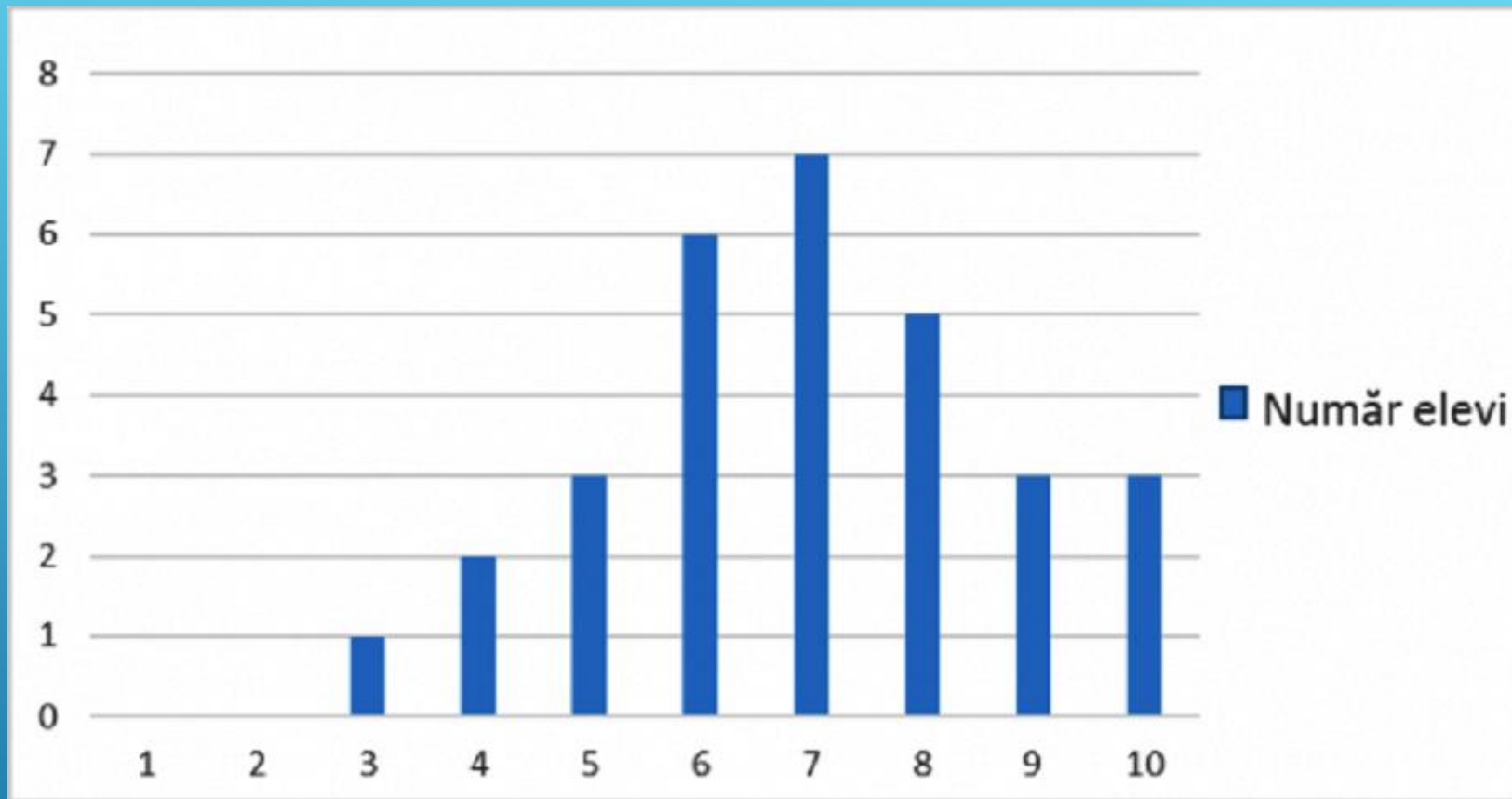
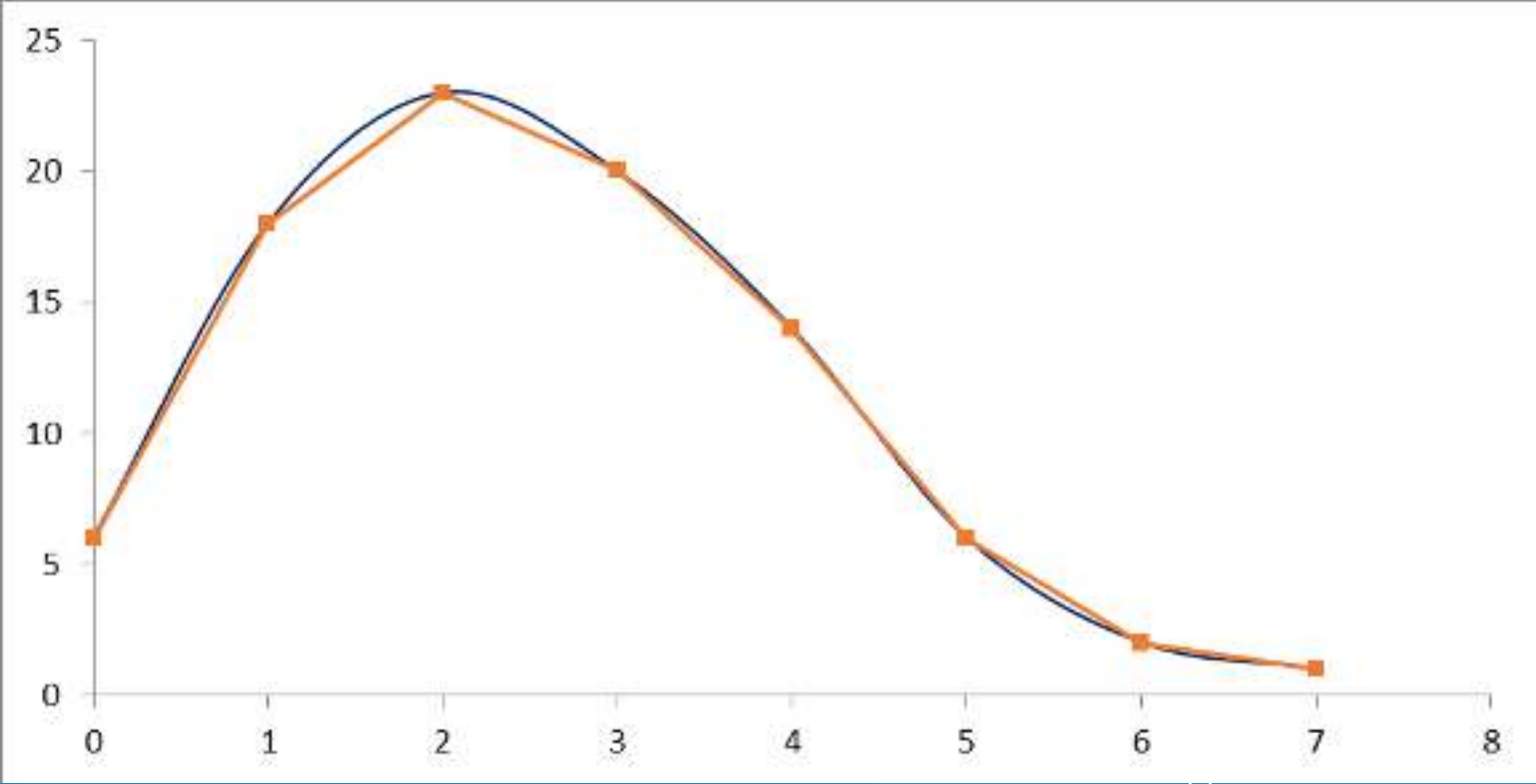
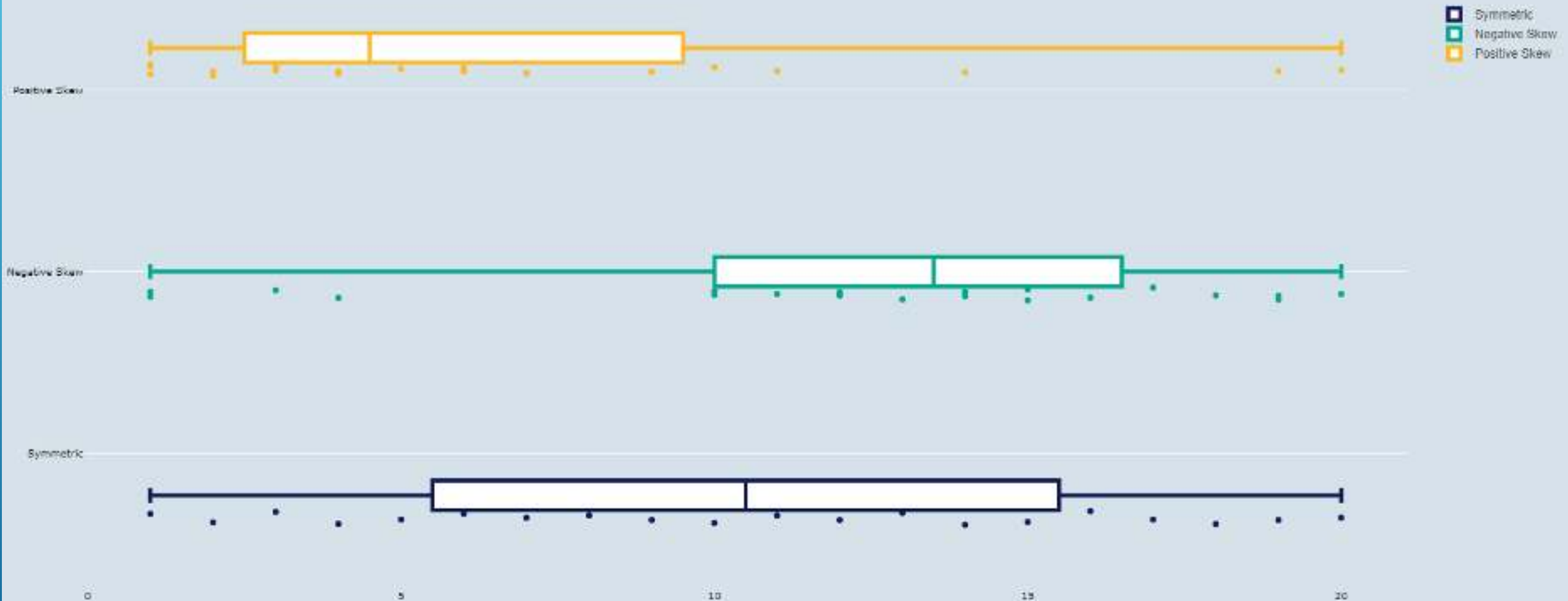


Figura 1. Distribuția elevilor de la un liceu după nota obținută la un examen, în martie 2016

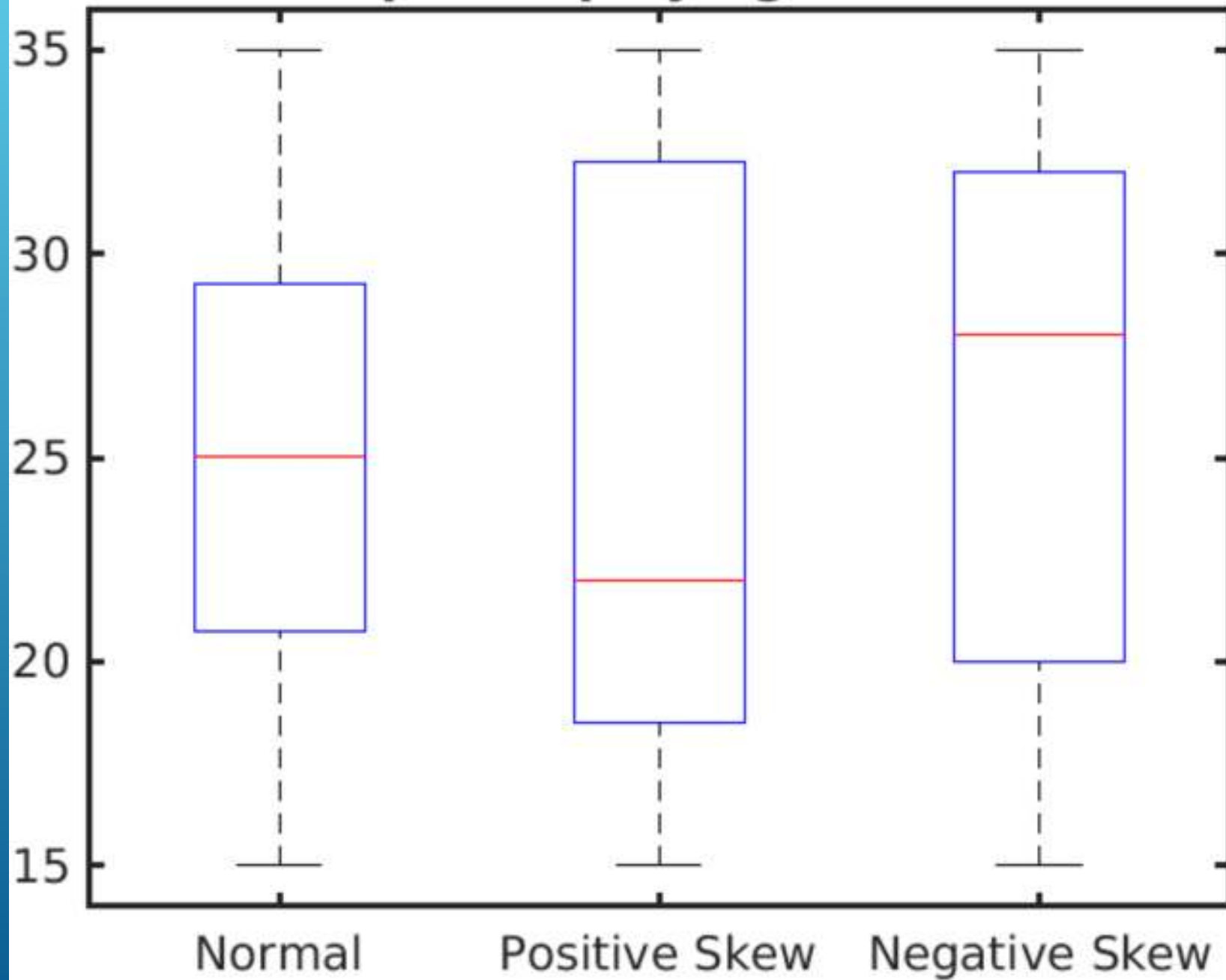
x_i	n_i
0	6
1	18
2	23
3	20
4	14
5	6
6	2
7	1
Total	90

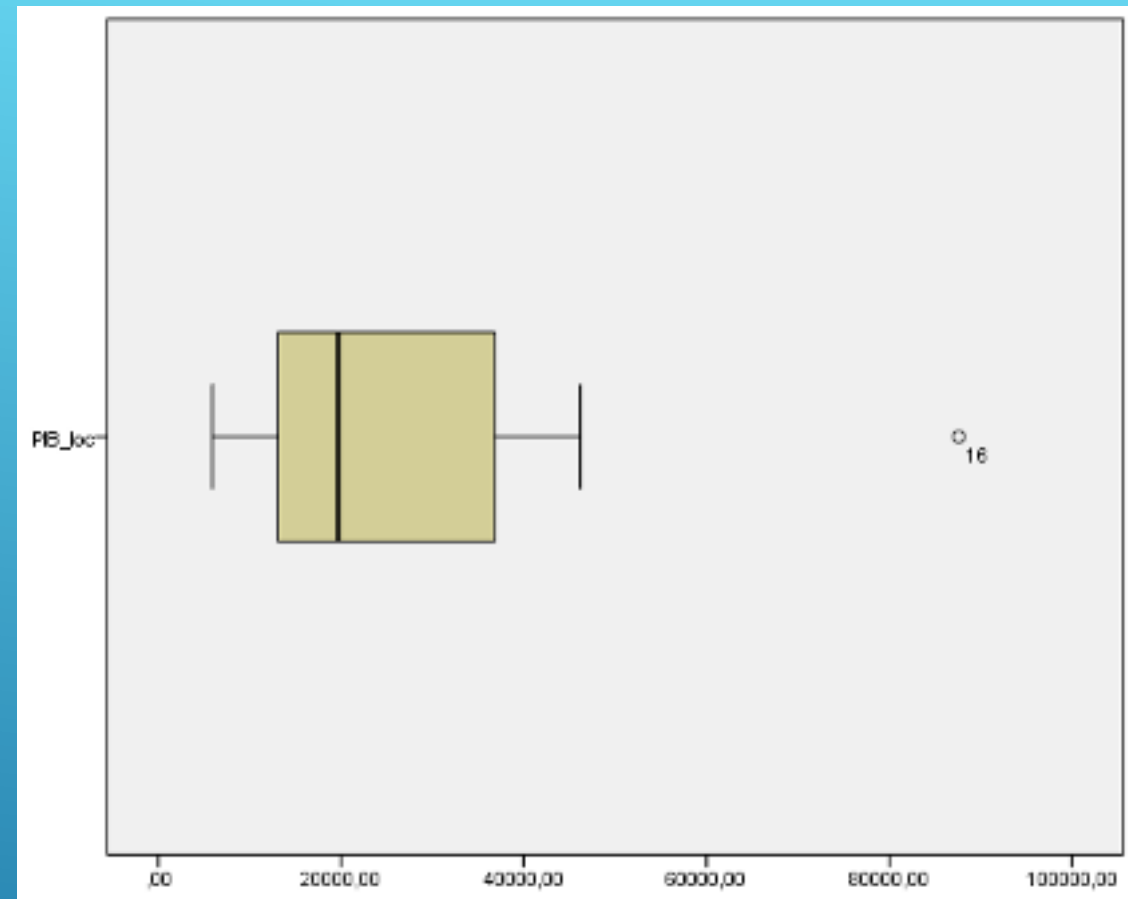
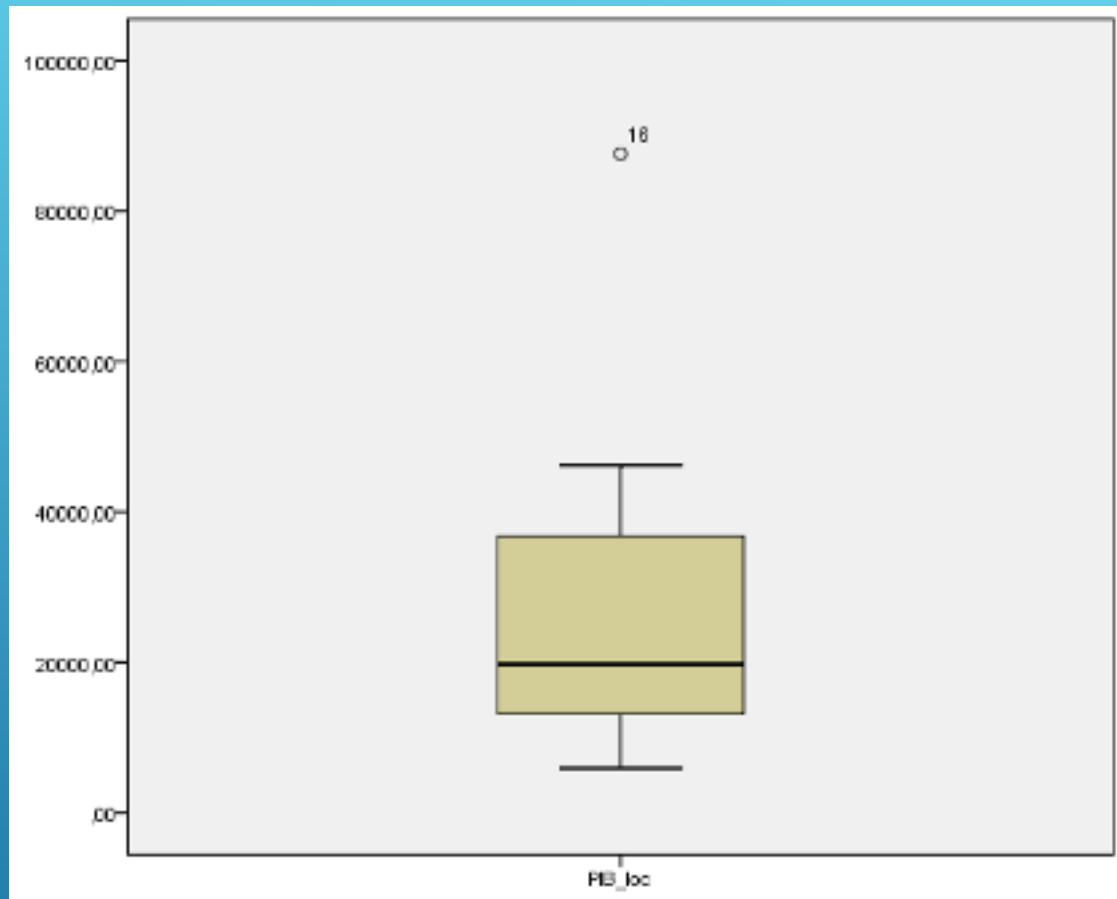


Box Plots & Skew

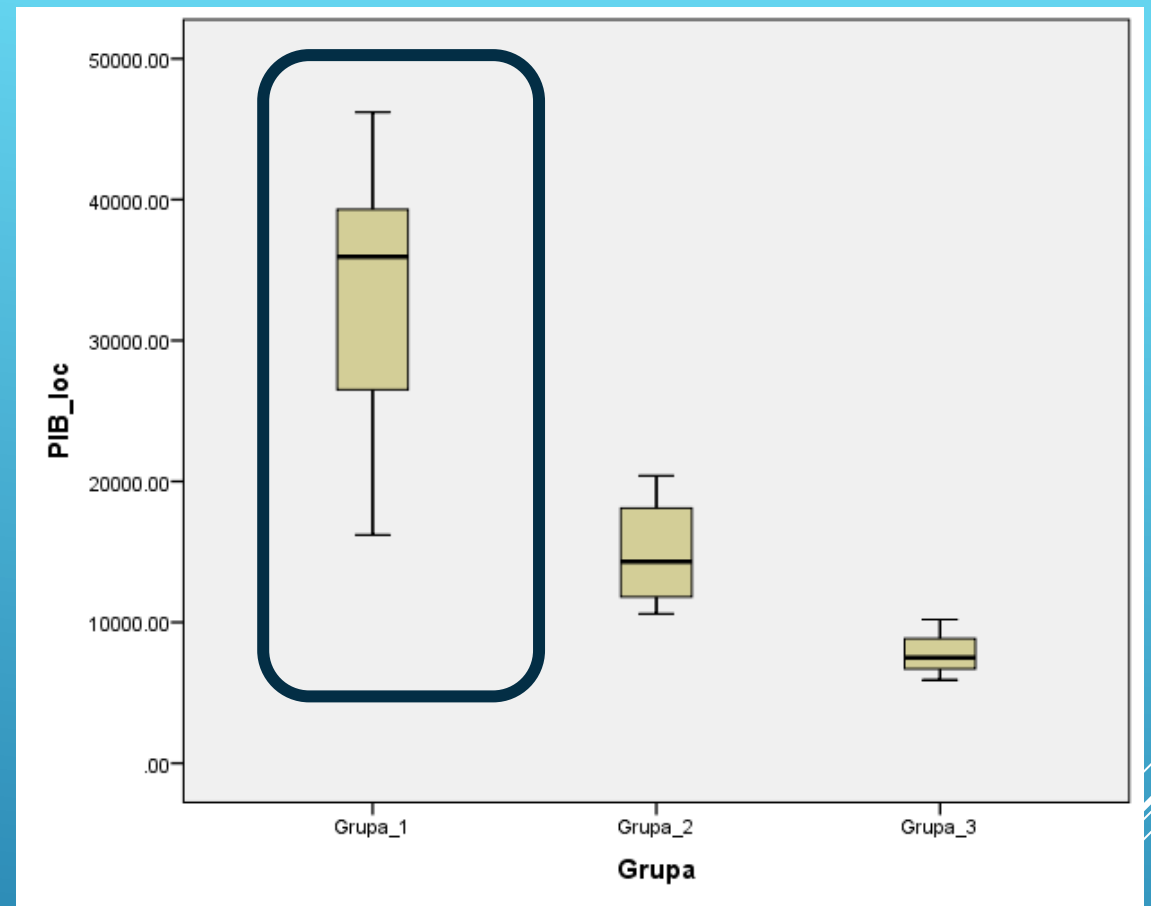
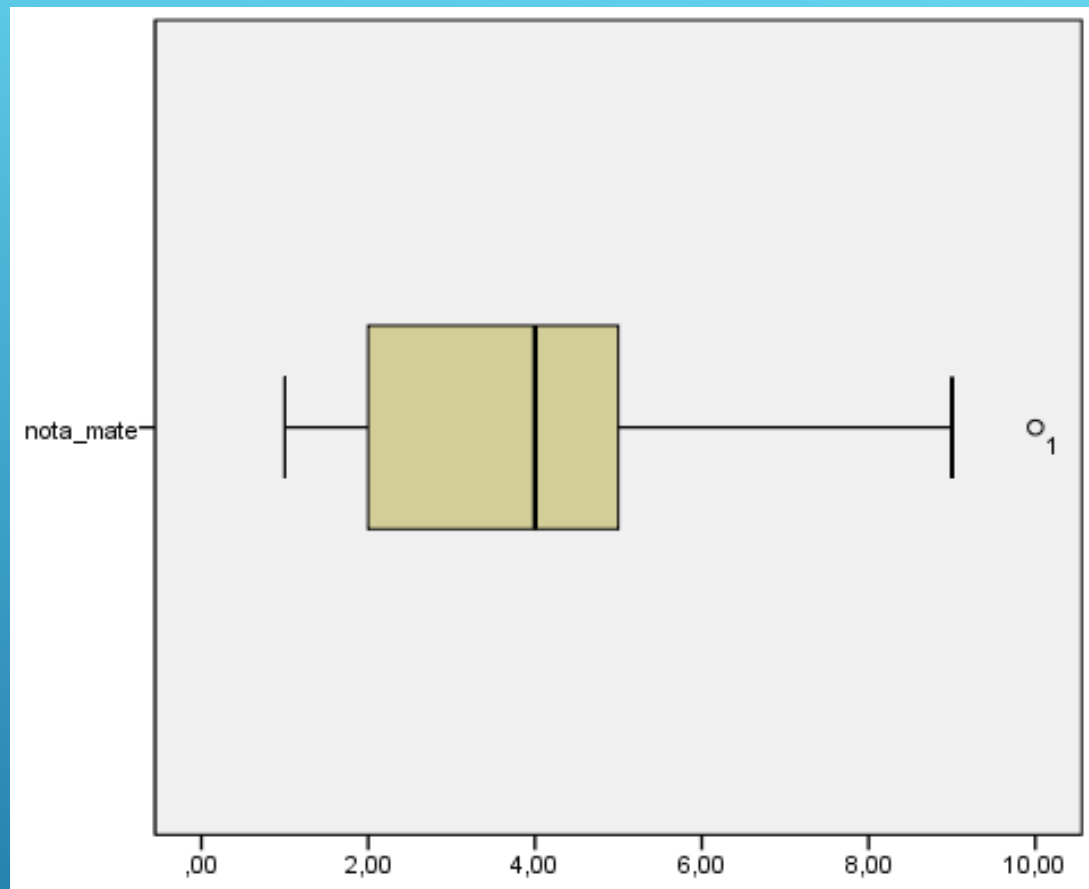


Example Displaying Skewness

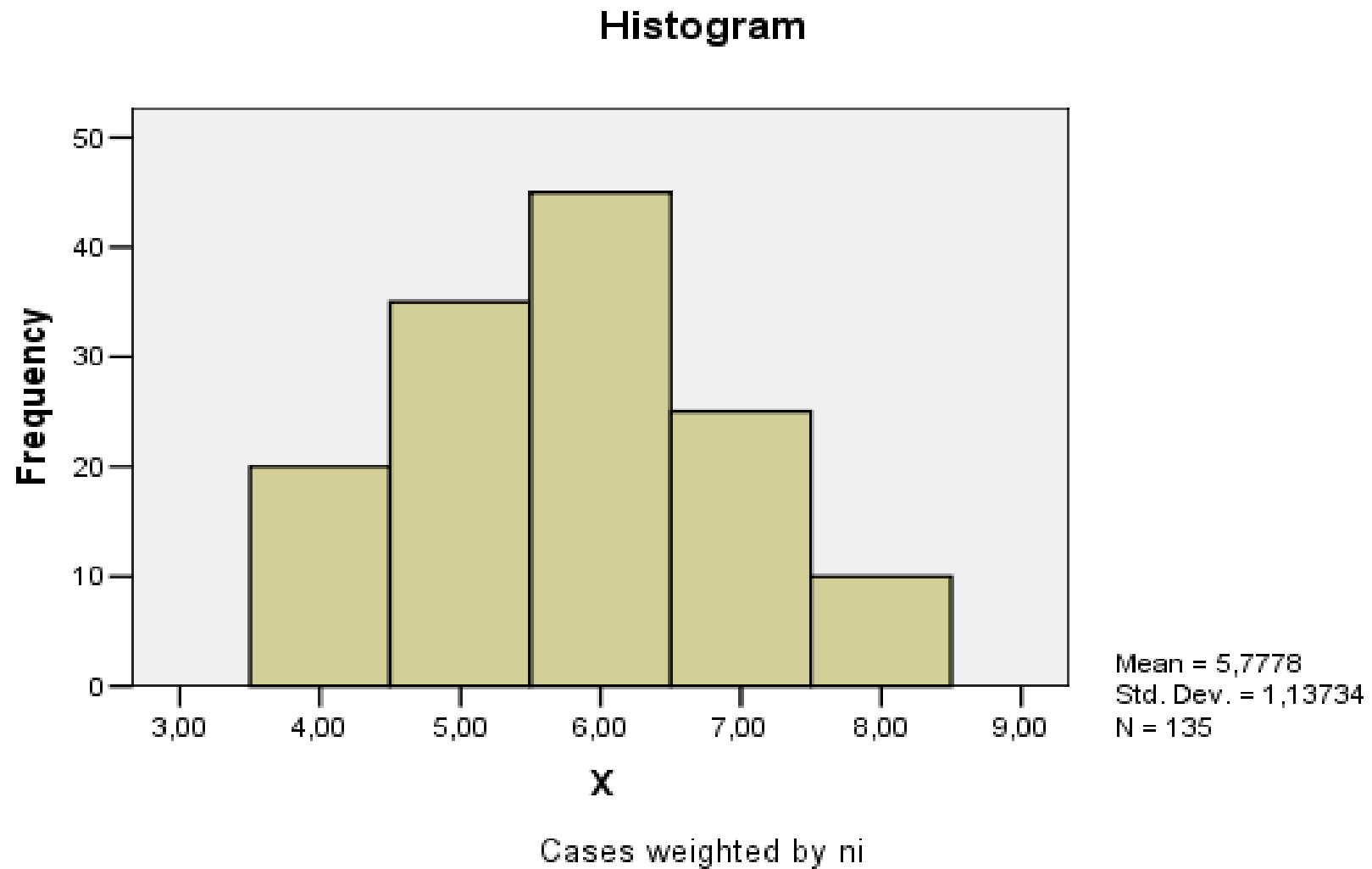




Asimetrie la dreapta



Asimetrie la stanga



a. Histograma

Analiza unei serii univariate

2.1. Variabilă cantitativă

- A. Variabilă discretă
- B. Variabilă continuă

2.2. Variabilă calitativă

I. Tipuri de variabile

- A. Variabile nominale
 - B. Variabile ordinale
- 
- Several white lines of varying lengths and orientations are positioned in the bottom right corner of the slide, creating a modern, abstract design element.

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**.

Analiza unei serii bidimensionale

1. Prezentarea seriei

► O serie bidimensională prezintă variația unităților unui eșantion după două variabile de grupare în mod simultan:

- variabilele X_i cu valorile $x_i, i = \overline{1, m}$ și Y_j cu valorile $y_j, j = \overline{1, p}$

Efectivele (unitățile) eșantionului care poartă simultan valoarea x_i și valoarea y_j sunt n_{ij} .

Distribuția bivariată este definită de:

$$(x_i, y_j, n_{ij}), i = \overline{1, m}, j = \overline{1, p}$$

$x_i \backslash y_j$	$j=1$	$j=2$...	$j=p$	Total
$i=1$	n_{11}	n_{12}	n_{1p}	$n_{1.}$
$i=2$	n_{21}	n_{22}	n_{2p}	$n_{2.}$
:					:
:	n_{ij}	$n_{i.}$
:					:
$i=m$	n_{m1}	n_{m2}	...	n_{mp}	$n_{m.}$
Total	$n_{.1}$	$n_{.2}$...	$n_{.j}$	$n_{.p}$	$n_{..}$

- 🚦 Variații pe coloană – variații după i , întrucât nivelul j este constant, este același
- 🚦 Variații pe linii – variații după j , întrucât nivelul i este același
- 🚦 Variabila X are **m niveluri de variație**
- 🚦 Variabila Y are **p niveluri de variație**

2. Tipuri de variabile

- o variabilă numerică și o variabilă nenumerică;
- ambele variabile numerice;
- ambele variabile nenumerice.

3. Distribuția după o variabilă cantitativă și o variabilă calitativă

În cadrul unei distribuții bidimensionale se disting:

a). *Două distribuții marginale*

► Distribuția marginală în X : $X : (x_i, n_{i\bullet}), i = 1, \dots, m$

$$n_{i\bullet} = \sum_{j=1}^p n_{ij}$$

- Distribuția marginală în Y :

$$Y : (y_j, n_{\bullet j}), \quad j = 1, \dots, p$$

$$n_{\bullet j} = \sum_{i=1}^m n_{ij}$$

Observație:

- Convenim să notăm variabila numerică cu X și variabila nenumerică cu Y . Prin urmare, există **m** valori ale variabilei numerice și **p** valori ale variabilei nenumerice.

b) Distribuții condiționate (m+p distribuții)

- ▶ Distribuția condiționată a variabilei X în funcție de valorile variabilei Y
 - este definită pentru fiecare valoare y_j

$(X / Y = y_j) : (x_i, n_{ij}), i = 1, \dots, m \quad \text{si} \quad j \text{ valoare fixă}$

- Distribuția condiționată a variabilei Y în funcție de valorile variabilei X
 - este definită pentru fiecare valoare \mathbf{x}_i

$$(Y / X = x_i) : (y_j, n_{ij}), j = 1, \dots, p \quad \text{și} \quad i \text{ valoare fixă}$$

4. Frecvențe absolute

- ▶ *Frecvențe absolute marginale* - sunt efectivele grupurilor create de X sau Y

$$n_{i.} \text{ și } n_{.j}$$

- ▶ *Frecvențe absolute parțiale*: n_{ij} .

5. Frecvențe relative

► *Frecvențe relative marginale*

►
$$f_{i\bullet} = \frac{n_{i\bullet}}{n_{\bullet\bullet}}; \quad f_{\bullet j} = \frac{n_{\bullet j}}{n_{\bullet\bullet}}$$

► *Frecvențe relative parțiale:*

$$f_{ij} = \frac{n_{ij}}{n_{\bullet\bullet}}$$

► Frecvențe relative condiționate

$$f_{i/j} = \frac{n_{ij}}{n_{\bullet j}} \quad j \text{ valoare fixa}, i = 1, \dots, m$$

$$f_{j/i} = \frac{n_{ij}}{n_{i\bullet}} \quad i \text{ valoare fixa}, j = 1, \dots, p$$

6. Medii condiționate (pe grupe)

Dacă X este variabila numerică, atunci media variabilei X pe grupe este:

$$\bar{x}_j = \frac{\sum_{i=1}^m x_i \cdot n_{ij}}{n_{\bullet j}}, \text{ cu } n_{\bullet j} = \sum_{i=1}^m n_{ij}, j = \overline{1, p}$$

7. Media pe total:

$$\bar{x} = \frac{\sum_{j=1}^p \bar{x}_j \cdot n_{\bullet j}}{\sum_{j=1}^p n_{\bullet j}}.$$

8. Variația general

- masoara, la nivelul populatiei, diferentele dintre indivizi, negrupati

$$s_X^2 = s_{\bar{x}_j}^2 + \bar{s}^2$$

a) Variante între grupe (varianțe intergrupe)

- masoara, la nivel general, diferentele dintre indivizii din grupe diferite

$$s_{\bar{x}_j}^2 = \frac{\sum_{j=1}^p (\bar{x}_j - \bar{x})^2 \cdot n_{\bullet j}}{\sum_{j=1}^p n_{\bullet j}}$$

b) Media varianțelor de grupă (varianța intra-grupe)

- măsura, la nivel general, diferențele dintre indivizii de același fel (din același grup)

$$\bar{s}^2 = \frac{\sum_j s_j^2 \cdot n_{\bullet j}}{\sum_j n_{\bullet j}}$$

unde s_j^2 sunt varianțe condiționate (varianțe de grupă). Acestea măsoară variația în cadrul unei grupe (intragrupă).

Se calculează astfel:

$$s_j^2 = \frac{\sum_{i=1}^m (x_i - \bar{x}_j)^2 \cdot n_{ij}}{n_{\bullet j}}$$

pentru

$$Y = y_j$$

Măsurarea gradului de influență a factorului de grupare și a factorilor aleatori

□ Plecând de la regula de adunare a varianțelor se pot calcula:

a) Coeficientul influenței factorului de grupare

$$k_1 = \frac{s_{\bar{x}_j}^2}{s_X^2} \cdot 100$$

b) Coeficientul influenței factorilor întâmplători

$$k_2 = \frac{\bar{s}^2}{s_X^2} \cdot 100$$

