

STATISTICA DESCRIPTIVĂ II

DATE CANTITATIVE

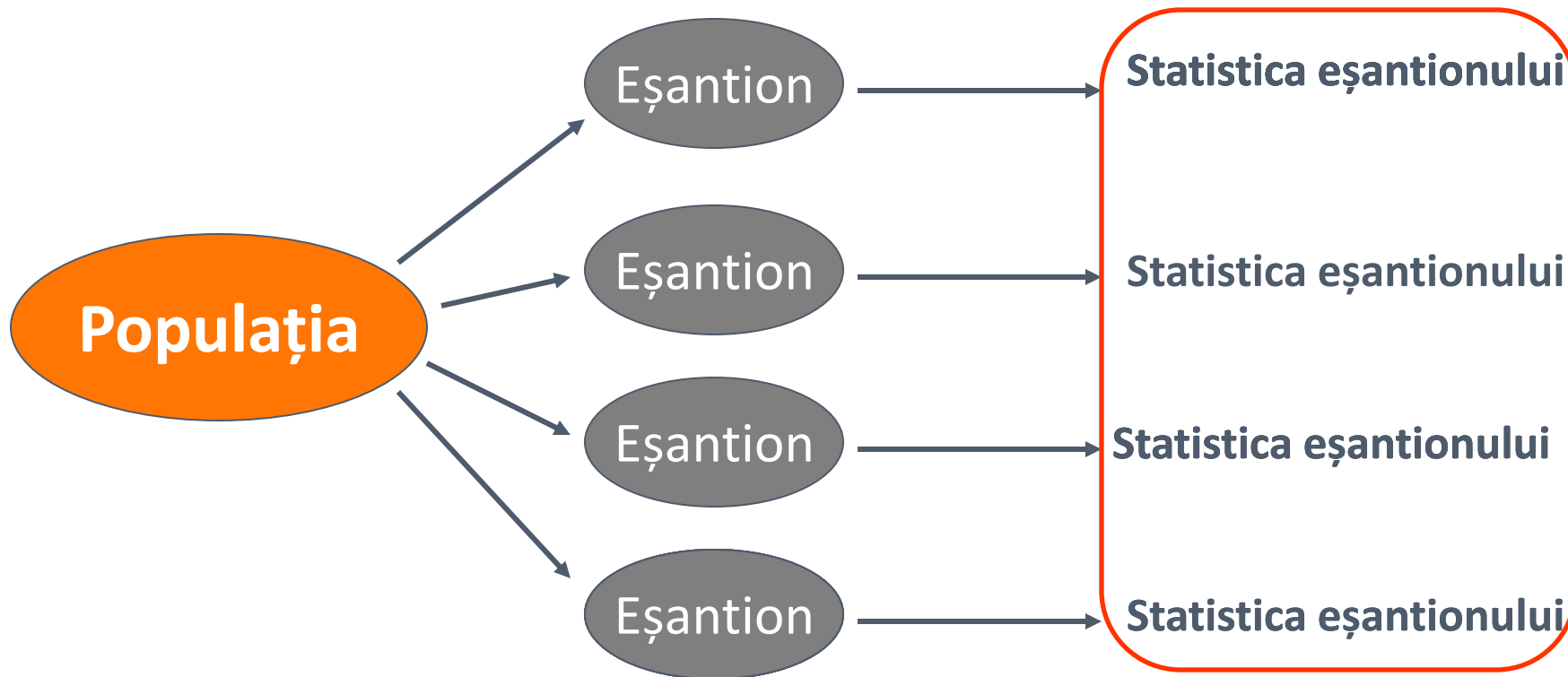
PARAMETRII DE CENTRALITATE

» Serii unidimensionale - Parametrii de centralitate:

- > Media aritmetică
- > Mediana
- > Modulul
- > Valoarea centrală
- > Media geometrică
- > Media ponderată
- > ...

» Aplicații practice

Cuprins



**Distribuția
eșantionului**

≠

**Distribuția
de eșantionare**



N = volumul populației

Copii de 1 an din RO

AB: $x_{AB,1}, x_{AB,2}, \dots, x_{AB,1000}$

...

GJ: $x_{GJ,1}, x_{GJ,2}, \dots, x_{GJ,1000}$

...

VN: $x_{VN,1}, x_{VN,2}, \dots, x_{VN,1000}$

\bar{x}_{AB}
...
 \bar{x}_{GJ}
...
 \bar{x}_{VN}

$$\mu = \frac{x_{AB} + \dots + x_{GJ} + \dots + x_{VN}}{N}$$

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

Distribuția de eșantionare

media (\bar{x}) $\approx \mu$

$n \uparrow \Rightarrow \downarrow s < \sigma$

Măsuri de centralitate

- ◆ Media
- ◆ Mediana
- ◆ Modulul
- ◆ Valoarea centrală
- ◆ Media geometrică, armonică

Măsuri de împrăștiere

- ◆ Amplitudine
- ◆ Variația
- ◆ Deviația standard
- ◆ Coeficientul de variație
- ◆ Eroarea standard

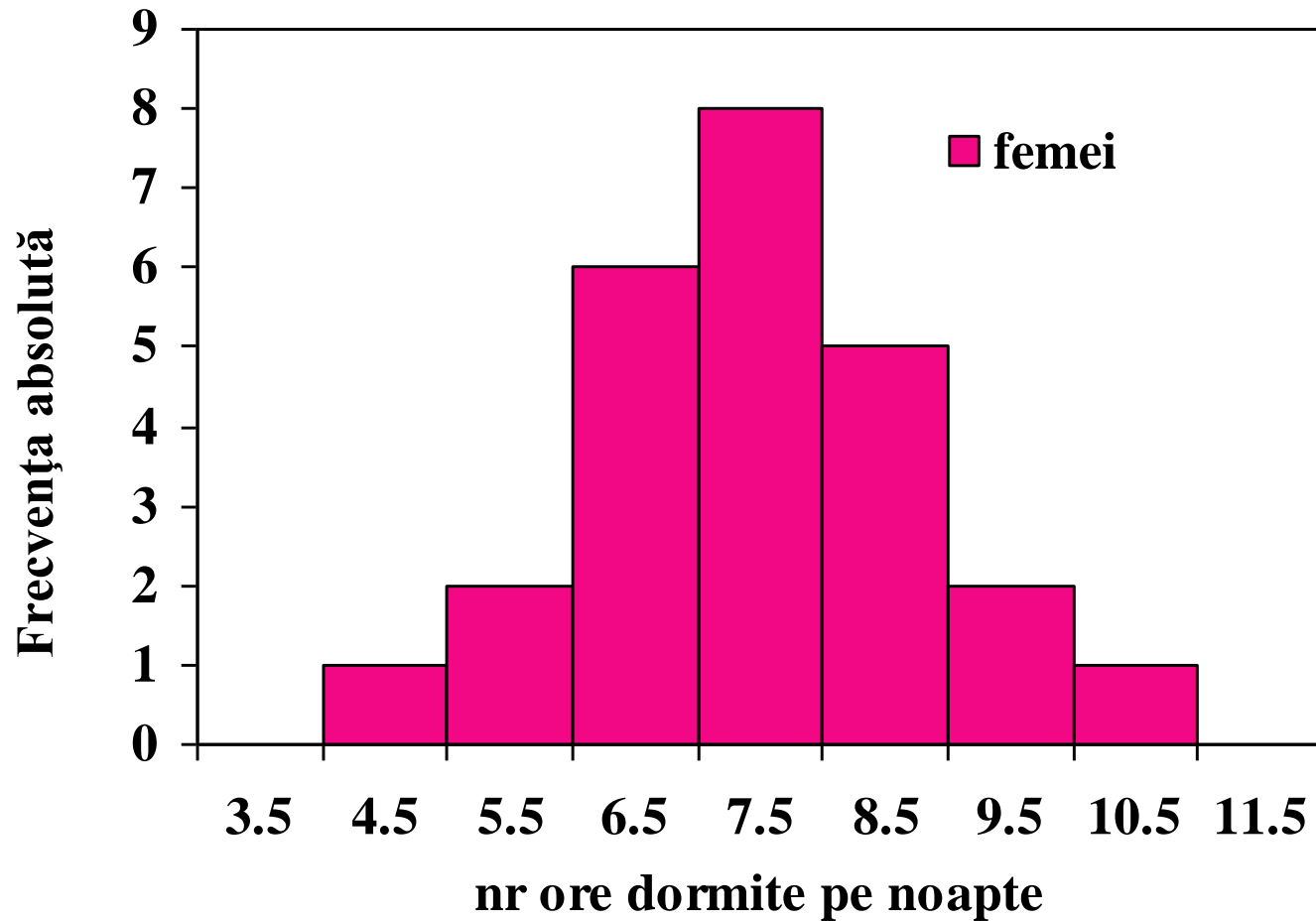
Măsuri de simetriei

- ◆ Asimetria
- ◆ Boltirea

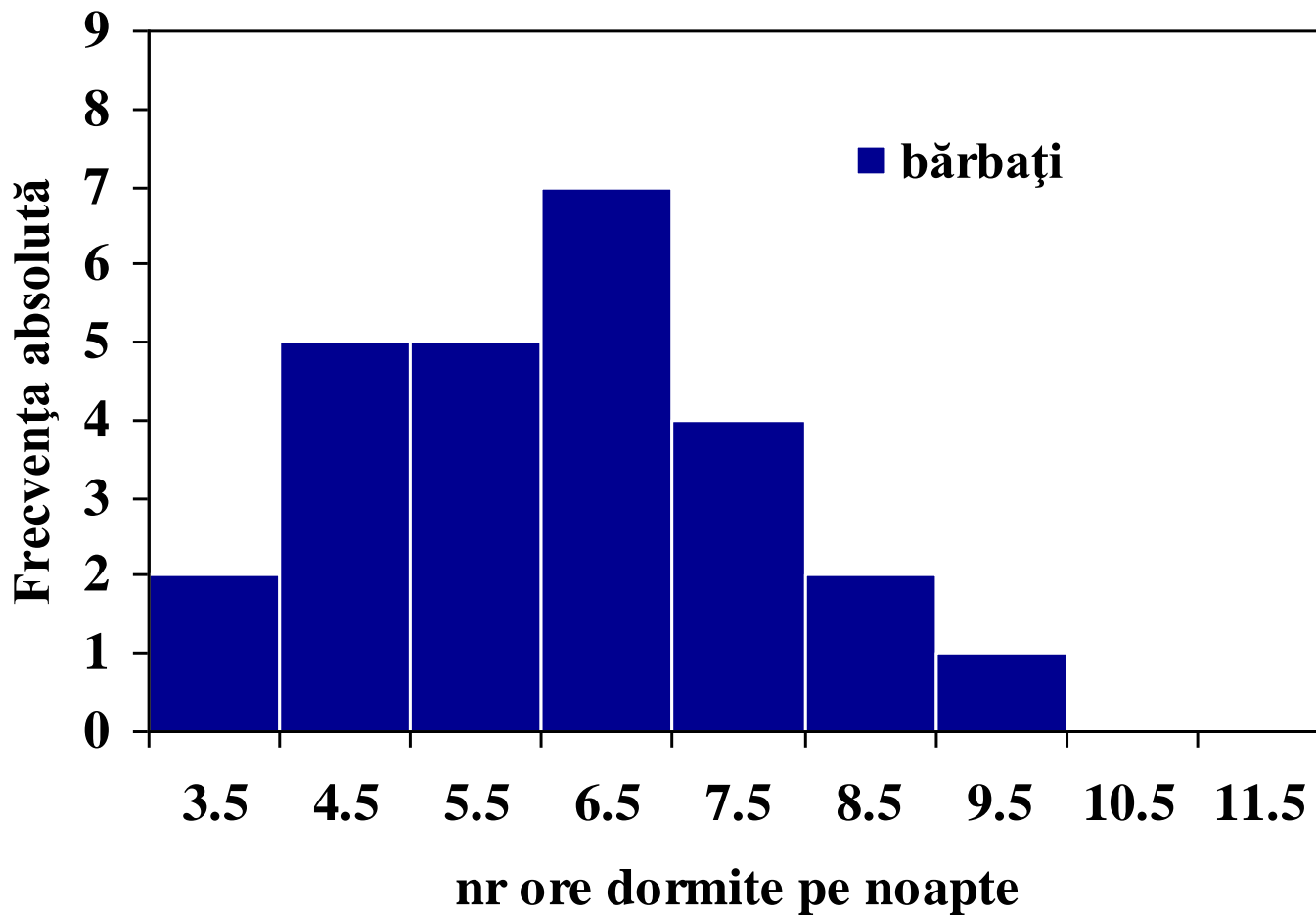
Măsuri de localizare

- ◆ Cvartiel (decile; percentile)

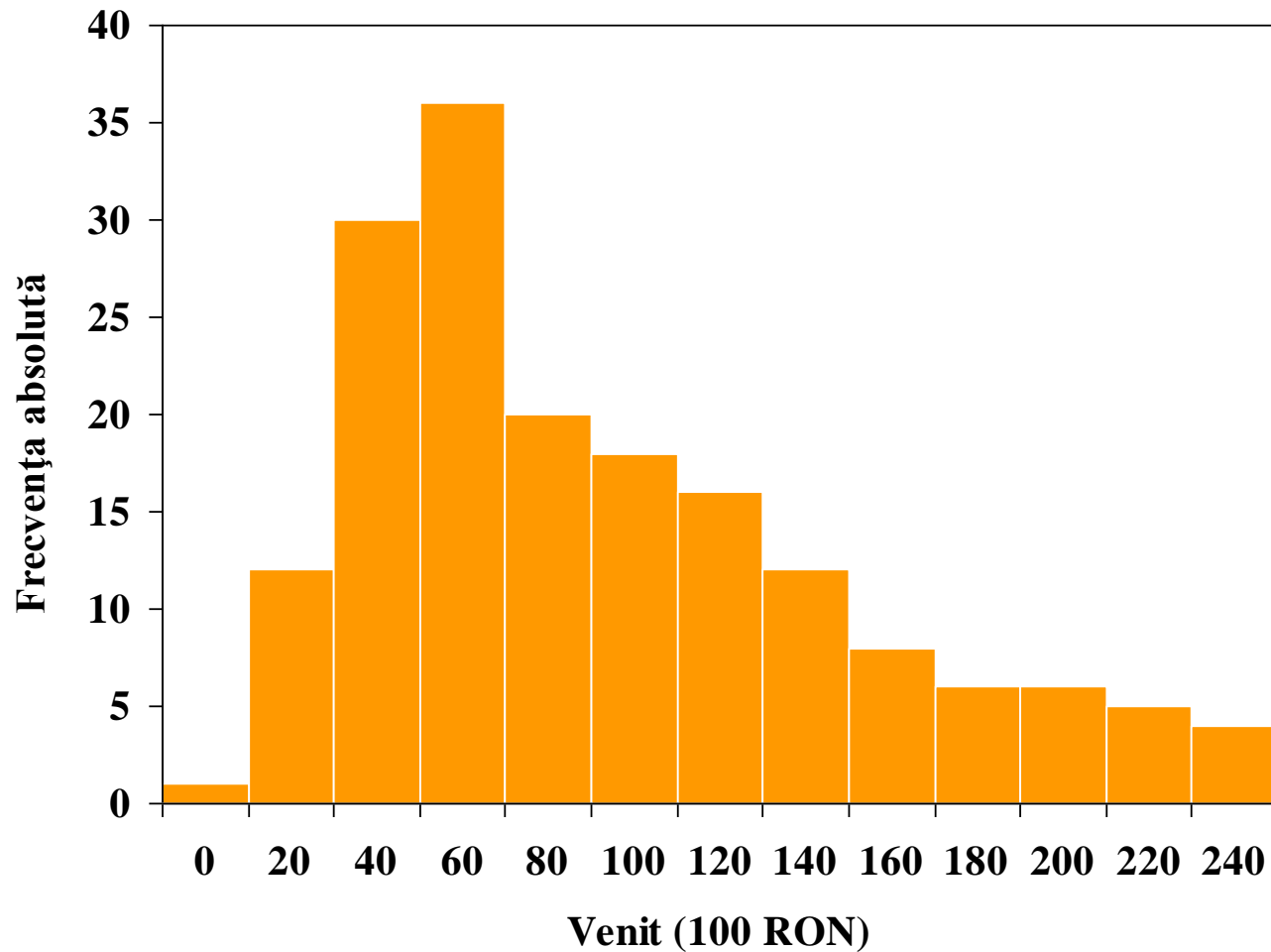
Parametrii statistici descriptivi



Măsuri de centralitate

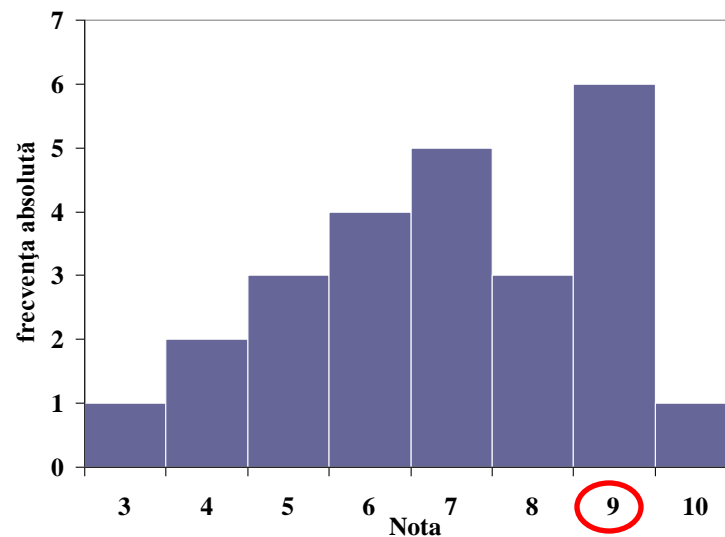


Măsuri de centralitate



Măsuri de centralitate

- » Denumit și **valoarea modală**: este o valoare care are cea mai mare frecvență din serie
- » Nu există formulă matematică de calcul
- » Corespunde punctului cel mai înalt pe distribuția grafică de frecvențe.
 - > *Care este valoare modală pentru cele trei reprezentări grafice anterioare?*
- » Nu este influențat de valorile extreme



Măsuri de centralitate: Modulul

» Serie unimodală:

> Vârsta pacienților internați în Clinica Pediatrie I cu sindrom diareic în perioada 1.11-8.11.2008

2	1	2	1	1
---	---	---	---	---

» Serie bimodală:

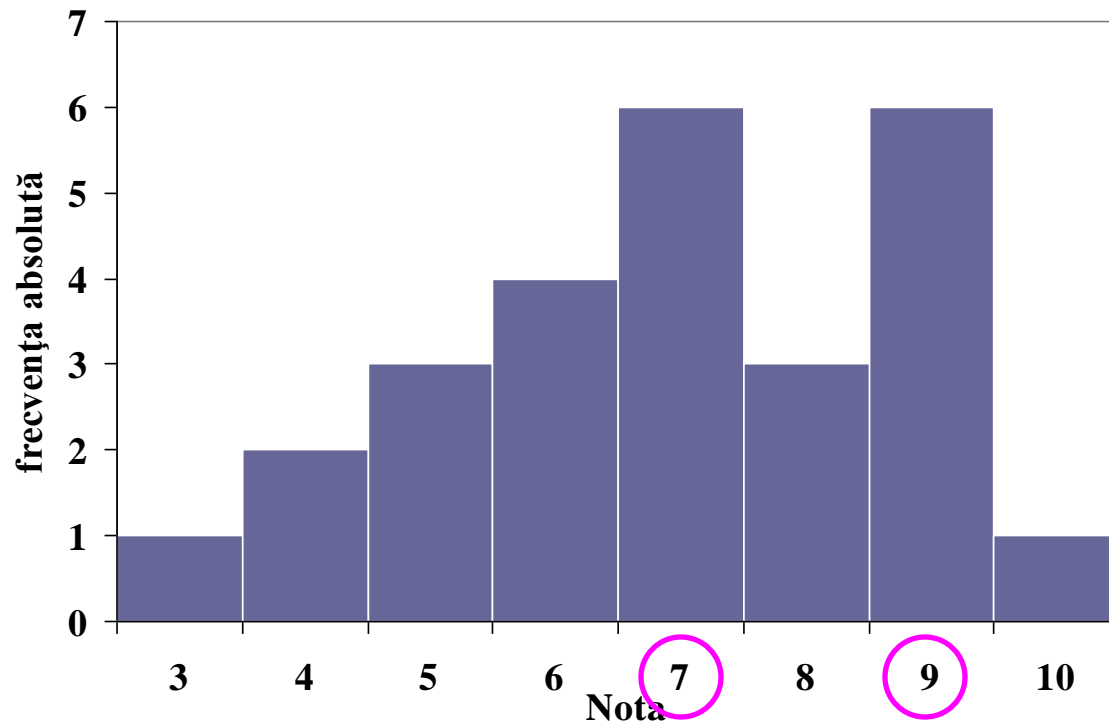
2	1	2	1	1
2	2	1	3	3

» Serie multimodală:

2	1	2	1	1
2	3	3	3	4

Măsuri de centralitate: Modulul

» Distribuție bi-modală



Măsuri de centralitate: Modulul

- » Valoarea care împarte distribuția în jumătate
- » Pași în calculul medianei:
 - > Se ordonează datele seriei în ordine crescătoare.
 - > Se localizează poziția medianei în acest șir și se determină valoarea ei.
 - > Valoarea este egală cu valoarea percentilei 50

- » Dacă volumul n al seriei este impar, atunci mediana este dată prin formula:

$$Me = X_{\frac{n+1}{2}}$$

- » Dacă n este par, atunci mediana este dată prin formula:

$$Me = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$$

Măsuri de centralitate: Mediana

1. Mediana nu este afectată de valorile extreme ale seriei de date.

	A	B	C	D	E	F	G	H	I	J	K	L
1	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}		Mediana
2	3	5	5	6	7	8	8	9	9	10		7.5

	A	B	C	D	E	F	G	H	I	J	K	L
1	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}		Mediana
2	3	5	5	6	7	8	8	9	9	100		7.5

2. Valoarea obținută pentru mediană poate fi nereprezentativă pentru distribuția datelor seriei dacă valorile individuale nu se grupează înspre valoarea centrală (mediana).

	A	B	C	D	E	F	G	H	I	J	K	L
1	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}		Mediana
2	3	5	500	6	7	8	800	900	9	100		8.5

3. Mediana este o măsură de tendință centrală care minimizează suma valorilor absolute ale abaterilor de la o valoare X de pe dreapta numerelor reale

Măsuri de centralitate: Mediana

3, 9, 9, 5, 8, 6, 10, 7, 8, 5

	A	B	C	D	E	F	G	H	I	J
1	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀
2	3	9	9	5	8	6	10	7	8	5

→ Ordonarea crescător

	A	B	C	D	E	F	G	H	I	J
1	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀
2	3	5	5	6	7	8	8	9	9	10

$$Me = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$$

- » Volumul eșantionului: $n = 10$ (număr par)
- » $Me = (X_5 + X_6) / 2 = (7 + 8) / 2 = 7,5$
- » **Excel:** =median(A2:J2) = 7,5

Serie
multimodală

Măsuri de centralitate: Mediana

→ Ordonarea crescător

	A	B	C	D	E	F	G	H	I	J	K
1	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉		Mediana
2	5	6	7	8	8	9	9	10	10		8

$$Me = X_{\frac{n+1}{2}}$$

- » Volumul eșantionului: $n = 9$ (număr impar) ○
- » $Me = X_{(9+1)/2} = X_5 = 8 = 8$
- » **Excel:** =median(A2:I2) = 8

**Serie
multimodală**


Măsuri de centralitate: Mediana

- » Suma tuturor datelor seriei împărțită la numărul de date din serie
- » Modificarea unei singure date din serie nu afectează valoarea modală sau mediana dar va afecta media aritmetică

- » Populație (media populației în problemele de statistică e cunoscută):

$$\mu = \frac{\sum_{i=1}^N X_i}{N}$$

- » Eșantion (se calculează):


$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

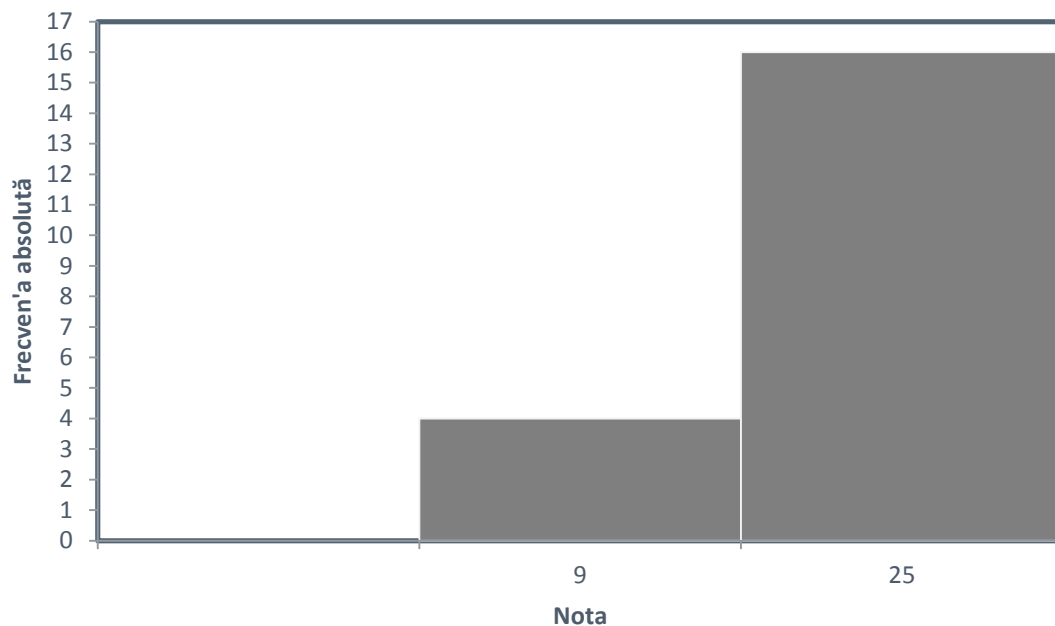
Simbol diferit pentru populație respectiv eșantion dar formulă de calcul similară

Măsuri de centralitate: Media aritmetică

» Media aritmetică: $= (5+6+7+8+8+9+9+10+10)/9 = 8$

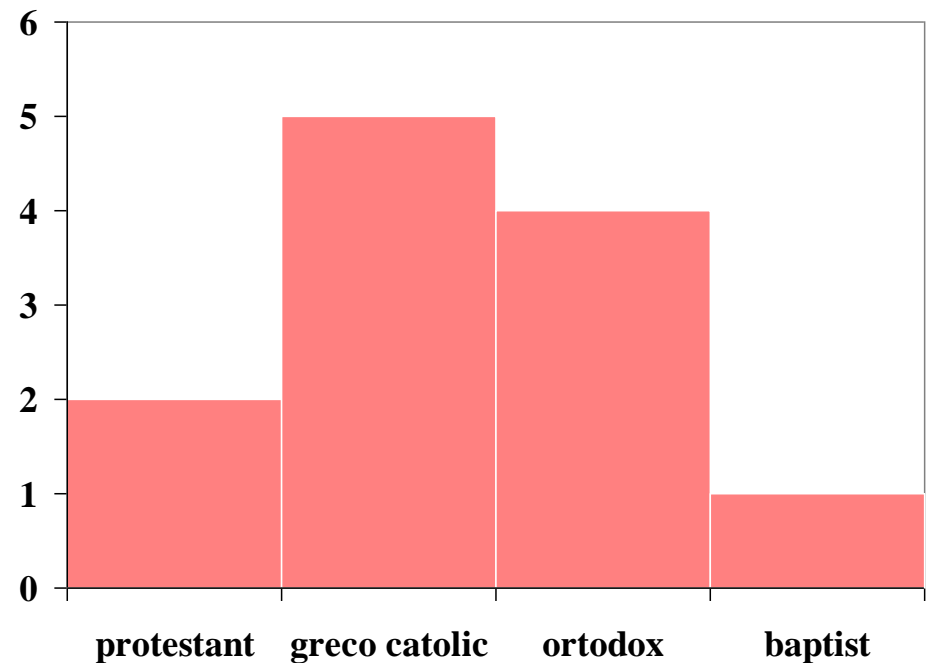
	A	B	C	D	E	F	G	H	I
1	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
2	5	6	7	8	8	9	9	10	10

» **Excel:** `=AVERAGE(A2:I2)`



Măsuri de centralitate: Media aritmetică

- » Este parametrul cel mai preferat ca măsură de centralitate atât ca și parametru de descriere a datelor cât și ca estimator
- » **Dar**, are semnificație **DOAR DACĂ** variabila de interes este *cantitativă*.



Măsuri de centralitate: Media aritmetică

Proprietăți:

	A	B	C	D	E	F	G	H	I
1	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
2	5	6	7	8	8	9	9	10	10

1. Orice valoare a seriei este luată în considerare în calculul mediei.

$$\text{Media aritmetică} = (5+6+7+8+8+9+9+10+10)/9 = 8$$

2. Valorile extreme pot influența media aritmetică distrugându-i reprezentativitatea.

	A	B	C	D	E	F	G	H	I	J	K
1	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉		Media aritmetică
2	5	6	7	8	8	9	9	10	100		18

3. Media aritmetică se situează printre valorile seriei de date.

$$\text{Minimum} = 5$$

$$\text{Maximum} = 10$$

$$\text{Media aritmetică} = 8$$

	A	B	C	D	E	F	G	H	I
1	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
2	5	6	7	8	8	9	9	10	10

Măsuri de centralitate: Media aritmetică

Proprietăți:

	A	B	C	D	E	F	G	H	I
1	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
2	5	6	7	8	8	9	9	10	10

4. Suma diferențelor dintre valorile individuale din serie și medie este zero:

$$(5-8) + (6-8) + (7-8) + (8-8) + (8-8) + (9-8) + (9-8) + (10-8) + (10-8) =$$

$$-3 -2 -1 + 0 + 0 + 1 + 1 + 2 + 2 = -6+6 = 0$$

$$\sum_{i=1}^n (X_i - \bar{X}) = 0$$

5. Schimbarea originii scalei de măsură a variabilei X din care provine seria de date are influență asupra mediei. Fie $X''=X+C$ (unde C este o constantă)

	A	B	C	D	E	F	G	H	I	J	K
1	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9		Media aritmetică
2	5	6	7	8	8	9	9	10	100		18
3											
4	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9		Media aritmetică
5	95	96	97	98	98	99	99	100	190		108


C = 90

Măsuri de centralitate: Media aritmetică

Proprietăți:

6. Transformarea scalei de măsură a variabilei X , de asemenea, influențează media aritmetică. Dacă se ia $X'' = h \cdot X$, h fiind o constantă reală.

	A	B	C	D	E	F	G	H	I	J	K
1	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9		Media aritmetică
2	5	6	7	8	8	9	9	10	100		18
3											
4	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9		Media aritmetică
5	13	15	18	20	20	23	23	25	250		45


$$C = 2,5$$

7. Suma pătratelor abaterilor valorilor seriei de la media aritmetică este minimul sumei pătratelor abaterilor valorilor seriei de la o valoare X a dreptei reale

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \min_{X \in \mathbb{R}} \sum_{i=1}^n (X_i - X)^2$$

Măsuri de centralitate: Media aritmetică

Formula: $= (X_{\max} + X_{\min}) / 2$

Estimator rar folosit în analiza statistică deoarece:

- Eficiență redusă deoarece ia în considerare doar valorile

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Xmax	Xmin	Valoare centrală
8	7	8	9	9	10	6	8	9	7	10	6	8

- Putere redusă deoarece valoarea este modificată semnificativ de orice outlier

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Xmax	Xmin	Valoare centrală
8	7	8	9	9	10	6	8	9	70	70	6	38

- Lipă de reprezentativitate: ia aceeași valoare pentru orice serie care are valor minime si maxime identice

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Xmax	Xmin	Valoare centrală
8	7	8	9	9	10	4	8	9	7	10	4	7

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Xmax	Xmin	Valoare centrală
6	6	5	7	4	10	4	8	9	4	10	4	7

Valoarea centrală

- » Fiecare valoare X_i este înmulțită cu o pondere W_i pozitivă, care indică importanța valorii respective în raport cu celelalte valori:

$$m_x = \frac{\sum_{i=1}^n W_i X_i}{\sum_{i=1}^n W_i}$$

- » Dacă ponderile W_i sunt alese egale și pozitive atunci se obține media aritmetică obișnuită

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Credite	6	5	2	4	4	5	2	2			
2	Nr. Crt.	Nume, prenume (initiale)	Anatomie și Histologie	Fiziologie	Educație pentru sănătate	Farmacologie	Nursing	Comportament & Psihologie & Sociologie	Comunicare	Opțional	Media aritmetica	Media ponderata	Media generală
3	1	BR	7	7	7	8	8	5	7	8	7.13	7.00	7.06
4	2	BA	8	8	8	9	9	7	8	6	7.88	7.97	7.92
5	3	BS	9	9	7	7	10	5	7	9	7.88	7.93	7.90
6	4	BR	10	6	8	8	8	6	8	9	7.88	7.80	7.84
7	5	BA	8	7	9	9	9	9	8	8	8.38	8.33	8.35
8	6	CD	9	8	6	9	6	7	8	9	7.75	7.83	7.79
9	7	CE	5	5	7	10	7	10	9	9	7.75	7.43	7.59
10	8	CR	6	10	5	10	5	7	6	6	6.88	7.17	7.02
11	9	CS	5	5	6	5	7	4	8	9	6.13	5.63	5.88
12	10	CT	6	10	10	6	9	9	6	10	8.25	8.10	8.18

Măsuri de centralitate: Media aritmetică ponderată

$$\sqrt[n]{X_1 \times X_2 \times \dots \times X_n}$$

- » Utilă în descrierea creșterii proporționale
- » Înălțimea exprimată în cm a unui copil în creștere măsurată o dată la 3 luni a avut următoarele valori: 100, 102, 116 și 128
- » Media geometrică = $\sqrt[4]{100 \times 102 \times 116 \times 128}$
- » Media geometrică = 112 \Rightarrow există o creștere anuală de 12%
- » Media aritmetică = $(100+102+116+128)/4 = 111 \Rightarrow$ există o creștere anuală medie de 11%

Măsuri de centralitate: Media geometrică

	+++	---
MODUL	<ul style="list-style-type: none"> • ușor de calculat • utilă pentru datele nominale 	<ul style="list-style-type: none"> • slabă stabilitatea de eșantionare
MEDIANANA	nu e afectată de valorile extreme	<ul style="list-style-type: none"> • Într-o oarecare măsură slabă stabilitate de eșantionare
MEDIA	<ul style="list-style-type: none"> • stabilitate de eșantionare • în legătură cu varianța 	<ul style="list-style-type: none"> • Nu este utilă pentru datele discrete • E afectată de distribuția asimetrică a datelor

De reținut! Măsuri de centralitate