

DATE NUMERICE

# POPULAȚIE

DATE ALFANUMERICE



INDIVID

DATE  
ORDINALE

DATE  
NOMINALE

EȘANTION

# Cursul I – Indicatori statistici

---

- Minim, maxim
  - Media
  - Deviația standard
  - Mediana
  - Cuartile
  - Centile, decile
-

# Tabel de date

FO	NUME	PRENUME	VIRSTA	SEX	MEDIU	DIAGCOD	HAV	ALCOOL
1	Voinea	Radu	55	b	u	CHAUD	da	da
2	Diaconescu	Ion	54	b	r	CHA	nu	da
3	Netcu	Tudor	45	b	r	CHA	nu	da
4	Dobrescu	Ana	56	f	r	CHA	nu	da
5	Cosmelita	Ion	56	b	r	CHA	nu	da
6	Iordanescu	Elena	65	f	u	CHALB	da	nu
7	Craciunescu	Misu	47	b	r	CHA	nu	da
8	Nedelcu	Maria	55	f	u	CHAUD	nu	da
9	Radulescu	Vasile	55	b	r	CHA	nu	da
10	Armeanca	Dumitru	55	b	u	CHA	da	da
11	Dumitrana	Alexandru	48	b	u	CHA	da	da
12	Ignat	Maria	50	f	u	CHA	nu	da
13	Serbanescu	Stefan	62	b	u	CHA	da	nu
14	Fronie	Sofica	28	f	u	CHA	da	nu
15	Dobrinescu	Florin	50	b	r	CHA	da	da
16	Ciuta	Maria	57	f	r	CHA	nu	da
17	Busuioc	Nicolita	65	f	r	CHA	da	nu

De cele mai multe ori, datele pe care le înregistrăm despre pacienți se prezintă în tabele ca cel de mai sus

---

# Indicatorii pentru serii de date

---

- Dacă am înregistrat valorile unui parametru la mai mulți indivizi, le considerăm o

**Serie de valori (de date) = Serie statistică**

- Hb: 12,5; 13,5; 15,3; 16,4; 11,7,.....etc
  - Vârsta: 36; 54; 73; 46; 31; 46; .....etc
  - Vom scrie astfel:
    - $X: x_1, x_2, x_3, \dots, x_n$
    - $Y: y_1, y_2, y_3, \dots, y_m$
-

# Minimul și maximul

---

□ X: 58, 74, 70, 71, 56, 68, 70, 82, 62, 62 (vârste)

□ Minim=56, Maxim=82

Y: 58, 74, 70, 71, 56, 68, 70, 82, 62, 62, 59, 46, 57, 71  
(aceleași vârste și încă 4)

□ Minim=46, Maxim=82

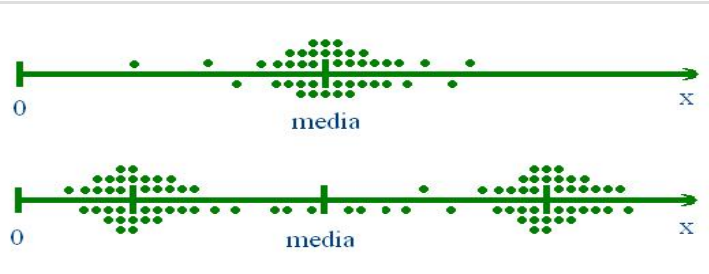
□ Amplitudinea absolută

$$A = \text{Max} - \text{Min} = 36$$

# Media aritmetică

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} = m$$

- Media unei serii statistice este raportul dintre suma valorilor seriei și numărul lor.
- Media este indicatorul care arată tendința centrală a seriei, și **de obicei** arată unde tind datele să se aglomereze.
- Media mai multor valori egale este egală cu fiecare din valori
- Deși se obține din valori concrete, măsurate în practică, media este un număr abstract



# Calculul mediei, influența unei valori extreme

---

$$\bar{X} = \frac{58 + 74 + 70 + 71 + 56 + 68 + 70 + 82 + 62 + 62}{10} = \frac{673}{10} = 67,3$$

$$\bar{Y} = \frac{58 + 74 + 70 + 71 + 56 + 68 + 70 + 82 + 62 + 62 + 59 + 46 + 57 + 71}{14} = \frac{906}{14} = 64,7$$

- **Seria Y este alcătuită din valorile lui X la care se adaugă patru noi valori. Existența valorii 46, extrem de mică, modifică media sensibil**

[Excel](#)

---

# Formula mediei ponderate

- $F_1, F_2, \dots, F_m$  sunt frecvențele de apariție în serie ale valorilor  $x_1, x_2, \dots, x_m$ .

$$\bar{X} = \frac{x_1 \cdot F_1 + x_2 \cdot F_2 + \dots + x_n \cdot F_m}{F_1 + F_2 + \dots + F_m}$$

Nr	Copii	Nr familii	Nr copii
1	0	102	0
2	1	406	406
3	2	388	776
4	3	143	429
5	4	54	216
6	5	16	80
7	6	5	30
8	7	1	7
		1115	1944
		1.74	



# Calculul dispersiei

(Engl: variance)

---

$$D = \frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + \dots + (x_n - \bar{X})^2}{n - 1}$$

- Este bine de reținut că:
    1. La medii aproximativ egale, este mai împrăștiată seria cu dispersia mai mare.
    2. La dispersii aproximativ egale, este mai împrăștiată seria cu media mai mică.
-

# Abaterrea standard

(Engl: standard deviation)

---

## □ Dispersia are dezavantaje:

1. Se exprimă cu unitățile de măsură ale variabilei, ridicate la pătrat
2. Are în general valori foarte mari comparativ cu media.

## □ De aceea se mai folosește un alt indicator, numit abatere standard care este radicalul dispersiei. Se notează cu $\sigma$ și are deci formula:

$$\sigma = \sqrt{D}$$

## □ Abaterrea standard = Abaterrea medie pătratică

---

# Exemplu de calcul

---

- Să presupunem că am măsurat zilnic tensiunea arterială la doi pacienți timp de 10 zile, obținând pentru fiecare următoarele valori pentru tensiunea arterială maximă:
    - 170, 180, 160, 180, 190, 190, 180, 190, 170, 190, pentru primul pacient și
    - 160, 170, 190, 160, 190, 190, 200, 180, 180, 180, pentru al doilea pacient.
-

# Aranjarea calculelor

$\bar{X} = 180$  170, 180, 160, 180, 190, 190, 180, 190, 170, 190

$$D_y = 1600 / 9 = 177,7.$$

$\bar{Y} = 180$  160, 170, 190, 160, 190, 190, 200, 180, 180, 180

$$\sigma_x = \sqrt{D_x} = \sqrt{111,1} = 10,5$$

$x_i - \bar{X}$ : -10, 0, -20, 0, 10, 10, 0, 10, -10, 10.

$y_i - \bar{Y}$ : -20, -10, 10, -20, 10, 10, 20, 0, 0, 0.

$(x_i - \bar{X})^2$ : 100, 0, 400, 0, 100, 100, 0, 100, 100, 100.

$$\sigma_y = \sqrt{D_y} = \sqrt{177,7} = 13,3$$

$(y_i - \bar{Y})^2$ : 400, 100, 100, 400, 100, 100, 400, 0, 0, 0.

$$D_x = \frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + \dots + (x_{10} - \bar{X})^2}{10 - 1} = \frac{400 + 6 \cdot 100}{9} = \frac{1000}{9} = 111,1$$

- Este mai împrăștiată seria Y, care are o abatere standard de 13,3, față de 10,5 la seria X, în condițiile în care ambele serii au aceeași medie.

# Coeficientul de variație

---

- Este raportul dintre deviația standard și medie, atunci când media este diferită de 0, și se măsoară în procente:

$$C.V. = \frac{\sigma}{\bar{X}}$$

- Pentru seriile de mai sus, coeficientul de variație este mai mare pentru cea mai împrăștiată, adică pentru cea cu deviația standard mai mare:

1.  $C.V._x = 10,5 / 180 = 0,058 = 5,8 \%$ .

2.  $C.V._y = 13,3 / 180 = 0,073 = 7,3\%$ .

# Indicatori de asimetrie

---

- **Mediana** (Engl: median) este o valoare, pentru parametrul măsurat (greutate, înălțime etc.), situată la mijlocul seriei de valori, dacă indivizii sunt ordonați crescător sau descrescător.
  - Jumătate din valori sunt mai mari decât mediana și jumătate mai mici.
  - **Cuartila Q1** este valoarea parametrului măsurat pentru care un sfert din indivizi au valori mai mici și trei sferturi mai mari decât această valoare.
  - Analog, **cuartila Q3** este valoarea față de care trei sferturi din indivizi au valori mai mici și un sfert - mai mari.
-

# Mediana – exemplu

Tensiunea arterială maximă la un bolnav în 10 zile

---

*150, 160, 160, 170, 160, 170, 150, 160, 170, 160*

□ Ordonând valorile crescător, obținem:

*150, 150, 160, 160, 160, 160, 160, 170, 170, 170*

□ În acest caz, mediana este între a cincea și a șasea valoare din șirul ordonat, adică 160

➤ Dacă aceste două valori de mijloc diferă, considerăm mediana ca fiind media lor aritmetică.

➤ Dacă numărul de măsurători este impar atunci mediana este chiar valoarea din mijloc.

---

# Alți indicatori statistici

---

- **Decile.** Pe eșantioane mai mari de multe sute de indivizi. Sunt 9 decile, fiecare corespunzând unui procent de 10%, 20%,.....,90% din eșantion, asemănător cuartilelor. Decila a 5-a este mediana.
- **Centile.** Folosite, în studii pe mii de cazuri, de obicei de un interes mai larg, național, internațional, și sunt corespunzătoare procentelor de 1%, 2%,...,99% din lot. Centila a 50-a este mediana.
- **Modul.** Clasa sau valoarea corespunzătoare frecvenței maxime dintr-o serie statistică se numește mod.

Excel

---