



# **SEMINAR BAZELE STATISTICII**

**SEMINAR 3**



## RECAPITULARE SEMINAR 2

$X$  – variabila

$x_i$  – valorile variabilei  $X$

$n_i$  – frecventa absoluta (numarul de unitati statistice)

$f_i$  – frecventa relativa

$n$  - volumul esantionului

# 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  $i$

# FRECVENȚE RELATIVE CUMULATE CRESCĂTOR ( $F_i \downarrow$ ) SAU DESCRESCĂTOR ( $F_i \uparrow$ )

- exprimă **ponderea unităților statistice** cumulate “până la” sau “peste” nivelul considerat al caracteristicii, adică valori  $\leq x_i$ , respectiv  $\geq x_i$ .

# ***ANALIZA SERIEI FOLOSIND METODE NUMERICE***

- ▶ Presupune calculul indicatorilor statisticii descriptive, cunoscuți și sub denumirea de caracteristici numerice ale unei distribuții.

## ***Indicatori ai tendinței centrale (mărimi medii)***

*Definire:*

- mediile sunt acele valori în jurul cărora se repartizează unitățile unei populații.
- cele mai importante mărimi medii sunt media aritmetică, modul și mediana.

- Media aritmetică este valoarea pe care am observa-o dacă unitățile statistice ar înregistra aceleași valori ale variabilei (dacă nu ar exista variații ale valorilor înregistrate de unitățile statistice).

MEDIA ARITMETICĂ (  $\bar{x}$  )



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);**

- Media unei distribuții este o valoare internă:

$$x_{\min} \leq \bar{x} \leq x_{\max}.$$

- Media este o mărime normală: suma abaterilor valorilor individuale ale unei variabile  $X$  de la media lor este egală cu zero.

*PROPRIETĂȚI ALE MEDIEI ARITMETICE*

Se consideră vânzările unor firme prezentate astfel: 10, 8, 9, 7, 7 (mil. lei).

$$\bar{X} = \frac{\sum_i x_i}{n} = \frac{10+8+9+7+7}{5} = \mathbf{8.2 \text{ mil lei}}$$

*Interpretare: Vanzarile medii au fost de 8.2 mil lei.*

- ▶ 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)

Se consideră vânzările unor firme prezentate astfel: 10, 8, 9, 7, 7 (mil. lei).

$$n_{i \max} = 2 \Rightarrow \text{pentru } x_i = 7 \Rightarrow Mo = 7$$

*Interpretare: Cele mai multe firme au inregistrat vanzari de 7 mil lei.*

La un magazin de pantofi s-au vândut într-o oră pantofi având următoarele mărimi:

Caz 1: 35 37 39 40 42

Această serie nu are mod.

Caz 2: 35 37 35 40 42

$Mo = 35$  deoarece este valoarea cea mai des întâlnită

Caz 3: 35 37 35 40 40

$Mo_1 = 35$        $Mo_2 = 40$

Această serie este bimodală.

Există și serii plurimodale.

- 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.

Se consideră vânzările unor firme prezentate astfel: 10, 8, 9, 7, 7 (mil. lei).

5 7 7 8 9 10

Se ordoneaza crescator seria: 7, 7, 8, 9, 10  $\Rightarrow$  Me= 8

*Interpretare: Jumatate (50%) din numarul total de firme au realizat vanzari de cel mult 8 milioane lei, iar cealalta jumatate (50%) au realizat vanzari de cel putin 8 milioane lei.*



Pentru 10 șobolani care încearcă să iasă dintr-un labirint se cunosc următorii timpi de parcurgere: 9 șobolani au parcurs labirintul în mai puțin de 15 minute, în timp ce un șobolan a reușit să parcurgă labirintul după 24 de ore. Pentru a calcula timpul mediu în care un șobolan parcurge labirintul valoarea reprezentativă este mediana și nu media (care ar fi afectată de acea durată mare de peste 24 ore).

Exemplu:

Se cunoaște următorul set de valori ale unei caracteristici: 5 7 4 9 12 3 10  
3 4 5 7 9 10 12

### *Interpretare*

*Jumatate (50%) din setul de valori ale unei caracteristici sunt pana la 7, iar cealalta jumătate (50%) din setul de valori ale unei caracteristici sunt peste la 7.*

Exemplu:

Fie următorul set de valori: 3 1 5 7 9 4  
1 3 4 5 7 9

### *Interpretare*

*Jumatate (50%) din setul de valori ale unei caracteristici sunt pana la 4.5, iar cealalta jumătate (50%) din setul de valori ale unei caracteristici sunt peste la 4.5.*

Arată forma unei distribuții:

1. atunci când cele 3 medii sunt egale, distribuția este simetrică.
2. atunci când media este mărimea medie cea mai mare, distribuția este asimetrică la dreapta (asimetrie pozitivă).
3. atunci când media este mărimea medie cea mai mică, distribuția este asimetrică la stânga (asimetrie negativă).

*RELAȚII ÎNTRE CELE TREI MĂRIMI MEDII*

## Serie simpla (cu frecvente egale):

a) **Repartitia notelor studentilor:** 3, 4, 5, 6, 8, 10

Intrebare: de care ori apare fiecare valoare  $x_i$  a variabilei X?

b) **Repartitia notelor studentilor:** 6, 6, 7, 7, 9, 9

Intrebare: de care ori apare fiecare valoare  $x_i$  a variabilei X?

# APLICATII - 1

## Serie cu frecvente diferite:

**Repartitia notelor studentilor:** 5, 6, 6, 7, 7, 7, 8, 8, 8, 8

- Definiti variabila.
- Definiti unitatea statistica.
- Care este volumul esantionului?

$x_i$	$n_i$	$f_i$
5	1	1/10
6	2	2/10
7	3	3/10
8	4	4/10
Total	$n = 10$	1

# APLICATII - 2

Se înregistrează cheltuielile anuale (mii lei) realizate de 3 familii pentru produsul A:

Familia 1: 1, 3, 7, 6, 9.

Familia 2: 1, 3, 7, 3, 5, 3, 7.

Familia 3: 1, 3, 1, 7, 7, 3.

Să se precizeze frecvențele absolute în cazul celor 3 serii.

## APLICATII - 3

Distribuția unui eșantion de angajați ai firmei X în funcție de numărul de mijloace carti impumutate de pe Bookster:

$x_i$	1	2	3	4
$n_i$	5	75	27	3

► Se cere să se afle:

- a) să se precizeze frecvențele absolute simple;
- b) să se calculeze frecvențele absolute cumulate crescător și descrescător;
- c) să se calculeze frecvențele relative simple;
- d) să se calculeze frecvențele relative cumulate crescător și descrescător;
- e) să se interpreteze toate frecvențele pentru  $i = 3$  ( $x_3$ );

## APLICATII - 4

1	2	3	4	5	6	7
$x_i$	$n_i$ frecvențele absolute simple	$N_i \downarrow$ - crescător frecvențele absolute cumulate crescător	$N_i \uparrow$ - decreșcător frecvențele absolute cumulate decreșcător	$f_i$ frecvențele relative simple	$F_i \downarrow$ (%) frecvențele relative cumulate crescător	$F_i \uparrow$ (%) frecvențele relative cumulate decreșcător
1	5	5	110 (5+75+27+3)	4.5% (5/110)*100	4.5% (5/110)*100	100% ((5+75+27+3)/110)*100
2	75	80 (5+75)	105 (75+27+3)	68.2% (75/110)*100	72.7% ((5+75)/110)*100	95.5% ((75+27+3)/110)*100
3	27	107 (5+75+27)	30 (27+3)	24.5% (27/110)*100	97.3% ((5+75+27)/110)*100	27.3% ((27+3)/110)*100
4	3	110 (5+75+27+3)	3	2.7% (3/110)*100	100% ((5+75+27+3)/110)*100	2.7% (3/110)*100
TOTAL	110	-	-	100%	-	-



Distribuția numărului de copii pentru familiile dintr-un bloc se prezintă astfel:

Nr. copii( $x_i$ )	0	1	2	3	4
Nr. familii ( $n_i$ )	10	15	30	3	2

Se cere să se afle:

- a) numărul de familii care au cel mult 3 copii;
- b) numărul de familii care au cel puțin 2 copii;
- c) ponderea familiilor care au cel puțin 2 copii;
- d) ponderea familiilor care au cel mult 1 copil;
- e) ponderea familiilor care au 2 copii.

- ▶ Se consideră seria: 2, 5, 6, 7, 8, 1000. Să se precizeze mărimea medie cea mai reprezentativă. Explicați.
- ▶ Sa se calculeze mediana seriei:

$x_i$	1	2	3	4
$n_i$	5	75	27	3

## APLICATII - 6

