

CURS STATISTICĂ - *Unitatea de învățare nr. 6*

ANALIZA STATISTICĂ A DISTRIBUȚIILOR DE FRECVENȚE. INDICATORII TENDINȚEI CENTRALE (2).

Cuprins:

1. Obiectivele unității de învățare.
2. Indicatorii medii de poziție.
 - 2.1 Mediana
 - 2.2. Modul
3. Abordări comparative între principalii indicatori ai tendinței centrale.
4. Răspunsuri și comentarii la testele de autoevaluare.
5. Teme de control.
6. Rezumatul unității de învățare.
7. Bibliografia unității de învățare.

1. Obiectivele unității de învățare

În urma parcurgerii acestei unități de învățare studentul:

- va înțelege când este indicat să se utilizeze mediana și modul
- își va însuși metodologia de calcul a celor doi indicatori (mediana și modul) pentru toate tipurile de serii

2. Indicatorii medii de pozitie

Indicatorii medii de poziție evidențiază tendințele de mijloc sau de concentrare a unităților. Dintre indicatorii medii de poziție, cei mai frecvenți utilizați sunt: mediana și modul.

2.1. Mediana

Mediana face parte din categoria cuantilelor alături de quartile, decile. Cuvântul mediană provine din cuvântul latin “medius” care înseamnă “mijloc”.

Mediana reprezintă acea valoare a unei serii ordonate crescător sau descrescător care împarte seria în două părți egale, așa încât 50% din termenii seriei au valori mai mici decât mediana, iar 50% mai mari decât mediana.

Un avantaj al medianei față de medie este acela că poate fi utilizată în caracterizarea tendinței centrale pentru o serie de date măsurate pe o scară ordinală. Mediana ia în considerație doar poziția termenilor în serie, nu și mărimea acestor valori, deci mediana nu este supusă influenței valorilor foarte mari sau foarte mici care sunt lăsate în afara seriei.

Exemplul 1

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).

Valoarea medianei este invariabilă față de convenția cu care se închid intervalele extreme, spre deosebire de medie care este influențată atât de valori cât și de frecvența lor.

Dacă seria prezintă o repartiție normală atunci mediana poate să înlocuiască valoarea medie deoarece se calculează mai ușor.

Mediana este un indicator utilizat în cercetările medicale, în studiul mortalității, la determinarea duratei medii de viață, la determinarea duratei medii de funcționare a unui produs.

Calculul medianei:

- **pentru o serie simplă (pentru date negrupate)**, întâlnim două situații:

- *seria are un număr impar de termeni* – atunci mediana este egală cu termenul central al seriei ordonate crescător sau descrescător.

Se cunoaște următorul set de valori ale unei caracteristici:

5 7 4 9 12 3 10

Ordonăm seria crescător:

3 4 5 **7** 9 10 12

|

Me

Pentru date ordinale mediana este varianta situată în centrul seriei.

- *seria are un număr par de termeni*, atunci mediana este egală cu media aritmetică simplă a celor 2 termeni centrali ai seriei ordonate crescător sau descrescător.

Fie următorul set de valori:

3 1 5 7 9 4

1 3 4 5 7 9

| |

$$Me = \frac{4 + 5}{2} = 4,5$$

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.

- **pentru o serie de distribuție de frecvențe pe variante** calculul medianei comportă următoarele etape:

Etapa 1: se determină locul medianei în cadrul seriei:

$$L_{Me} = \frac{1}{2} \left(\sum_{i=1}^k n_i + 1 \right)$$

Etapa 2: se cumulează crescător frecvențele absolute și se determină acea frecvență cumulată crescător care este imediat mai mare sau egală cu locul medianei (L_{Me}). Varianta care corespunde frecvenței absolute cumulate ce îndeplinește condiția de mai sus este mediana.

Exemplul 2

80 de apartamente dintr-un bloc au fost sistematizate după numărul de camere rezultând următoarea distribuție de frecvențe:

Nr. Camere (x_i)	Nr. Apartamente (n_i)	n_i cumulat crescător
1	13	13
2	25	38
3	28	66
4	14	80

Total	80	
<p>Calculați mediana.</p> $L_{Me} = \frac{1}{2} \left(\sum_{i=1}^k n_i + 1 \right) = \frac{1}{2} \cdot 81 = 40,5 < 66 \Rightarrow$ <p>$\Rightarrow Me = 3$ camere \Rightarrow 50% dintre apartamente au mai puțin de 3 camere, iar 50% mai mult de 3 camere.</p>		

- **pentru o serie de frecvențe pe intervale de variație**, mediana se poate determina numai în ipoteza în care valorile sunt distribuite uniform în cadrul intervalului de grupare.

Etape:

- se determină locul mediane în cadrul seriei:

$$L_{Me} = \frac{1}{2} \left(\sum_{i=1}^k n_i + 1 \right)$$

- se cumulează crescător frecvențele absolute și se determină acea frecvență cumulată crescător care este imediat mai mare sau egală cu L_{me} . Intervalul care corespunde frecvenței absolute cumulate ce îndeplinește condiția de mai sus este intervalul median.

- se calculează mediana cu relația:

$$Me = x_0 + h \cdot \frac{L_{Me} - n_{pMe}}{n_{Me}}$$

x_0 = limita inferioară a intervalului median;

h = mărimea intervalului median;

n_{pMe} = suma frecvențelor absolute până la intervalul median;

n_{Me} = frecvența absolută a intervalului median.

Exemplul 3

Repartiția sucursalelor unei bănci comerciale în funcție de volumul depozitelor bancare atrase într-o lună este:

Volum depozite bancare (mii euro) (x_i)	Nr. Bănci (n_i)	n_i cumulat crescător
--	---------------------	-------------------------

20 – 40	12	12
40 – 60	14	26
60 – 80	20	46
80 – 100	18	64
100 - 120	16	80
Total	80	-

$$L_{Me} = \frac{1}{2} \left(\sum_{i=1}^5 n_i + 1 \right) = \frac{81}{2} = 40,5 < 46 \Rightarrow Me \in [60, 80]$$

$$Me = 60 + 20 \cdot \frac{40,5 - 26}{20} = 74,5 \text{ mii euro}$$

Deci 50% dintre sucursale au atras depozite în valoare de 74,5 mii euro, iar 50% peste 74,5 mii euro.

Calculul grafic al medianei se poate realiza în două moduri:

- mediana este corespondenta pe abscisă a punctului de intersecție al ogivei crescătoare cu ogiva descrescătoare;
- se trasează doar ogiva crescătoare, iar de pe axa OY din punctul corespunzător locului medianei se duce o paralelă cu axa OX ce intersectează ogiva crescătoare într-un punct. Corespondenta pe abscisă a acestui punct este mediana.

Dezavantajele medianei:

- mediana este mai puțin stabilă decât media;
- nu poate fi supusă cu așa ușurință calculelor algebrice;
- media este preferată în statistica inferențială.

2.2. Modul

Modul (dominanta unei serii) este valoarea cea mai des întâlnită sau căreia îi corespunde cea mai mare frecvență de apariție.

Calculul algebric al modului:

- pentru o serie simplă:



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

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

Caz 3: 35 37 35 40 40

$M_{o_1} = 35$ $M_{o_2} = 40$

Această serie este bimodală.

Există și serii plurimodale.

- pentru o serie de distribuție pe variante, modul este egal cu varianta căreia îi corespunde frecvența absolută sau relativă maximă.

Exemplu 5

Nr. camere (x_i)	Nr. apartamente(n_i)
1	13
2	25
3	28
4	14

$M_o = 3$ camere deoarece variantei 3 îi corespunde frecvența absolută maximă.

- pentru o serie de distribuție de frecvențe pe intervale de variație modul se calculează cu relația:

$$M_o = x_0 + h \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2}$$

x_0 = limita inferioară a intervalului modal;

h = mărimea intervalului modal;

Δ_1 = diferența dintre frecvența intervalului modal și a celui anterior;

Δ_2 = diferența dintre frecvența intervalului modal și a celui următor.

Intervalul modal este intervalul căruia îi corespunde frecvența absolută maximă.

Exemplul 6

Volum depozite bancare (mii euro) (x_i)	Nr. Bănci (n_i)
20 – 40	12
40 – 60	14
60 – 80	20
80 – 100	18
100 - 120	16

Intervalul modal este [60, 80)

$$M_o = 60 + 20 \cdot \frac{20 - 14}{(20 - 14) + (20 - 18)} = 60 + 20 \cdot \frac{6}{6 + 2} = 75 \text{ mii euro}$$

Cele mai multe bănci au atras depozite în valoare de 75 mii euro.

Testul de autoevaluare 1

1. Un auditor bancar a selectat 10 conturi și a înregistrat sumele existente în fiecare dintre aceste conturi. Sumele sunt date în Euro: 150,175, 195, 200,235,240,250,256,275,294

Se cere: să se calculeze mediana și modul.

2. Distribuția salariiștilor unui magazin în funcție de numărul de zile de concediu de odihnă dintr-un an se prezintă astfel:

Zile concediu	14	15	16	17	18	19	20
Nr. salariați	2	6	10	15	8	5	4

Se cere : să se calculeze mediana și modul.

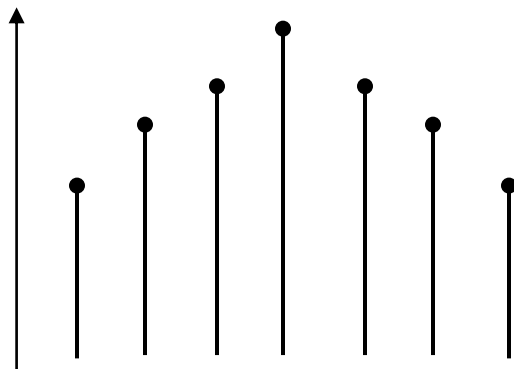
3. Un studiu efectuat asupra unui număr de 50 de cutii de brânză topită la cutie dintr-un magazin a reliefat următoarele informații cu privire la numărul de calorii conținute:

Calorii	75-85	85-95	95-105	105-115	115-125
Nr. cutii cu brânză topită	5	10	15	14	6

Se cere să se calculeze indicatorii tendinței centrale

Calculul grafic al modului:

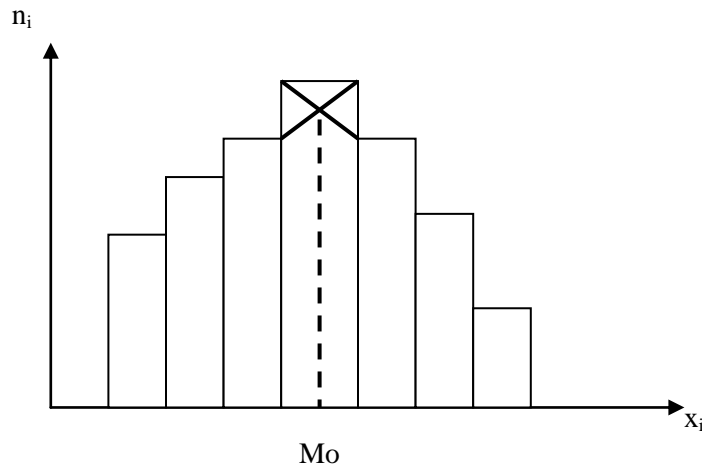
- în cazul seriilor de distribuție pe variante, determinarea grafică a modului se face cu ajutorul diagramei prin bare sau prin bastoane, modul fiind acea valoare de pe abscisă căreia îi corespunde ordonata maximă:



n_i x_i

→

- **în cazul seriilor de distribuție pe intervale de variație**, determinarea grafică a modului se face cu ajutorul histogramei. Se determină punctul de intersecție al segmentului ce unește capătul din stânga al blocului cel mai înalt cu capătul din stânga blocului următor cu segmentul ce unește capătul din dreapta al blocului cel mai înalt cu capătul din dreapta al blocului anterior. Corespondența pe abscisă a acestui punct de intersecție este modul.



Analog cu modul se poate determina, în cazul distribuțiilor în formă de U și valoarea antimodală căreia îi corespunde frecvența minimă.

Modul nu este un indicator al tendinței centrale foarte stabil și poate fi afectat de modul în care au fost construite intervalele de variație. În plus, modul nu se pretează așa ușor la clacule algebrice ca și mediana.

Cu toate aceste dezavantaje, modul este un indicator util în analiza seriilor de dimensiuni mari în care ne interesează valoarea cea mai des întâlnită.

3. Abordarea comparativă a principalilor indicatori ai tendinței centrale

Media este indicatorul cel mai utilizat în analiza tendinței centrale. Calculul mediei se bazează pe ansamblul valorilor caracteristicii x_i , de aceea ea este influențată de valorile extreme.

Mediana este un indicator care, spre deosebire de medie, se calculează în funcție de poziția termenilor în serie și nu este influențată de valorile termenilor.

Modul se determină foarte ușor, dar este indicatorul cel mai sensibil la modul de grupare a datelor. Două grupări diferite ale aceleiași serii vor conduce la două valori modale diferite.

Pentru o serie perfect simetrică cei trei indicatori ai tendinței centrale sunt egali:

$$\bar{x} = Me = Mo$$

Proprietăți (Yule & Kendall)	Modul	Mediana	Media aritmetică
Este definit într-un mod obiectiv	da	da	da
Depinde de numărul de termeni ai seriei	nu	da	da
Este puțin sensibil la mărimea valorilor extreme	da	da	nu
Are o semnificație concretă	da	da	da
Este ușor de calculat	da	da	da și nu
Este puțin sensibil la eșantionare	destul	nu	da
Se pretează la calcule algebrice	nu	nu	da

4. Răspunsuri și comentarii la testele de autoevaluare

1. Mediana

Pentru calculul medianei valorile x_i trebuie ordonate crescător. (sunt, din ipoteză)

Seria are număr par de termeni (10 termeni), deci mediana este media aritmetică a celor doi termeni centrali:

$$Me = \frac{235 + 240}{2} = 237,5 \text{ Euro}$$

Deci, în 50% din conturi sunt mai puțin de 237,5 Euro, iar în 50% din conturi sunt peste 237,5 Euro.

Modul este valoarea cea mai des întâlnită. Fiind serie simplă și neavând date care să se repete, seria nu are mod.

2.

Mediana (valoarea centrală a seriei)

$$L_{Me} : \frac{1}{2} \left(\sum_{i=1}^7 n_i + 1 \right) = \frac{1}{2} \cdot 51 = 25,5 < 33 \Rightarrow Me = 17 \text{ zile}$$

Se cumulează crescător frecvențele absolute (n_i):

x_i	n_i	n_i cumulate crescător
14	2	2
15	6	8
16	10	18
17	15	33
18	8	41
19	5	46
20	4	50

Deci 50% dintre salariați au avut sub 17 zile concediu, iar 50%, peste 17 zile.

Modul (dominanta unei serii)

Modul este, în cazul grupării pe variante, acea valoare a caracteristicii căreia îi corespunde frecvența absolută maximă. Frecvența absolută (n_i) maximă este 15 \Rightarrow $M_o = 17$ zile. Cei mai mulți salariați au avut 17 zile de concediu.

3. Indicatorii tendinței centrale sunt:

a₁) Media – se calculează ca o medie aritmetică ponderată

x_i reprezintă centrul de interval calculat ca medie aritmetică simplă între limita inferioară și limita superioară a fiecărui interval:

$$\bar{x} = \frac{\sum_{i=1}^5 x_i n_i}{\sum_{i=1}^5 n_i} = \frac{80 \cdot 5 + 90 \cdot 10 + 100 \cdot 15 + 110 \cdot 14 + 120 \cdot 6}{50} = \frac{5060}{50} = 101,2 \text{ calorii}$$

a₂) Mediana (valoarea centrală a seriei):

$$L_{Me} : \frac{1}{2} \left(\sum_{i=1}^5 n_i + 1 \right) = \frac{1}{2} \cdot 51 = 25,5 < 30 \Rightarrow Me \in [95, 105]$$

Se cumulează frecvențele absolute și se determină care frecvență absolută este imediat mai mare sau egală cu L_{Me} . Intervalul care corespunde frecvenței absolute cumulate ce îndeplinește

condiția de mai sus este intervalul median.

Me se calculează cu relația:

$$Me = x_0 + h \cdot \frac{L_{Me} - n_{pMe}}{n_{Me}} = 95 + 10 \cdot \frac{25,5 - 15}{15} = 102 \text{ calorii}$$

x_0 = limita inferioară a intervalului median;

h = mărimea intervalului;

n_{pMe} = suma frecvențelor absolute până la intervalul median;

n_{Me} = frecvența absolută a intervalului median.

Deoarece $Me = 102$ calorii \Rightarrow 50% din cutii au sub 102 calorii, iar 50% au peste 102 calorii.

a₃) Modul (dominanta seriei):

$$Mo = x_0 + h \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2} = 95 + 10 \cdot \frac{15 - 10}{(15 - 10) + (15 - 14)} = 103,33 \text{ calorii}$$

x_0 = limita inferioară a intervalului modal;

h = mărimea intervalului modal;

Δ_1 = diferența dintre frecvența absolută a intervalului modal și frecvența absolută a intervalului anterior celui modal;

Δ_2 = diferența dintre frecvența absolută a intervalului modal și frecvența absolută a intervalului următor celui modal.

Intervalul modal este intervalul căruia îi corespunde frecvența absolută maximă. Deci $Mo \in [95, 105]$.

Deoarece $Mo = 103,33$ calorii, rezultă că cele mai multe dintre cutii au 103,33 calorii.

5. Teme de control

1. Se cunosc următoarele date privind numărul de cărți împrumutate în decursul unei luni de abonații unei biblioteci:

Nr. cărți împrumutate	0	1	2	3	4	5	6	7
Nr. abonați	18	39	57	64	42	33	21	4

Calculați mediana acestei serii.

2. Calculați mediana pentru seria statistică următoare:

14 16 12 9 11 18 7 8 9 16 7 9 18

3. Un studiu privind durata de viață în ore a unui produs electrocasnic efectuat pe 100 aparate a condus la următoarele rezultate:

Durata de viață (ore)	Structura numărului de aparate electrocasnice
0 – 1000	8
1000 – 2000	20
2000 – 3000	26
3000 – 4000	22
4000 – 5000	18
5000 - 6000	6
Total	100

Se cere: să se calculeze indicatorii tendinței centrale;

6. Rezumatul Unității de învățare

Indicatorii tendinței centrale reprezintă o categorie deosebit de importantă de indicatori statistici utilizați în analiza variabilelor numerice. Acești indicatori sintetici redau într-o singură măsură ceea ce este tipic, esențial, caracteristic, obiectiv și stabil pentru o serie de date numerice.

Indicatorii tendinței centrale sunt:

- mărimile medii care pot fi calculate atât ca medii simple (pentru date negrupate), cât și ca medii ponderate (pentru date grupate pe variante sau pe intervale)
 - media aritmetică
 - media geometrică
 - media pătratică
 - media armonică
- indicatorii medii de poziție
 - mediana
 - modul

Acești indicatori caracterizează cu atât mai bine tendința centrală cu cât datele pe baza cărora se determină sunt mai omogene.

Cei mai importanți și mai utilizați indicatori ai tendinței centrale sunt: media, mediana, modul.

7. Bibliografia Unității de învățare

1. Chauvat G., Reau J.P., *Statistiques descriptives*, Armand Colin, Paris, 2004

2. Danciu A., Niculescu I., Gruiescu M., *Statistică economică*, Editura Enciclopedică, București, 2009

3. Isaic-Maniu Al., Mitrut C., Voineagu V., *Statistică*, Editura Universitară, București, 2003;
4. Voineagu V., Țițan E., Ghiță S., Boboc C., Todose D. – *Statistică. Baze teoretice și aplicații*, Editura Economică, București, 2007;
5. Wonnacott T.H., Wonnacott R.J., *Statistique, Economica*, Paris, 1995