

STATISTICA DESCRIPTIVĂ

SUMARIZAREA NUMETICĂ

Despre ...

Variabile calitative:

- Estimatori punctuali: Raportul, Proporția, Rata

Variabile cantitative:

- Mărusi de centralitate
- Măsuri de dispersie/împrăștiere
- Măsuri de simetrie
- Măsuri de localizare

VARIABILE CALITATIVE

Estimatorul punctual

- Estimarea parametrilor vs. testarea ipotezelor:
 - Estimare: proporția unei populații ce verifică un anumit criteriu
 - Testarea ipotezelor: identificarea estimatorului diferit
- Estimator punctual = o singură valoare care caracterizează populația
- Estimatori punctuali pe date calitative:
 - Raportul
 - Proporția
 - Rata

Raportul

- Numere raționale pozitive a și b , $b \neq 0$
- Simbolică
 - $a:b$
 - a/b
- Numitorul nu include în mod obligatoriu subiecții numărătorului
- Într-o grupă de 12 studenți avem 4 fumători.
 - Raportul nefumători/fumători = $8/4 = 2/1 = 2 \Leftrightarrow$ adică la 3 nefumători există un fumător
 - Raportul fumători/nefumători = $4/8 = 0,5$

Proporția

- O proporție este un raport în care numărătorul face parte din numitor. Astfel o proporție are forma generală:

$$a/(a+b)$$

- Ia valori între:
 - 0 și 1
 - 0 și 100 dacă se exprimă procentual
- Toți indivizii de la numărător sunt incluși la numitor
- Prevalența este o proporție
- Într-o grupă de 12 studenți avem 4 fumători.
 - » Proporția fumătorilor = $4/12$
 - » Proporția nefumătorilor = $8/12$

Proporția

- La serviciul de urgențe ale unui spital județean s-au prezentat pentru consultație 1200 pacienți. Dintre aceștia 420 au fost internați (200 femei și 220 bărbați):
 - » Proporția pacienților internați = $420/1200*100 = 35\%$
 - » Proporția pacienților de sex feminin internați = $200/420*100 = 48\%$
 - » Proporția pacienților de sex masculin internați = $220/420*100 = 52\%$

Prevalența

- » = proporția de indivizi dintr-o populație care au boala la un moment dat
- » estimează probabilitatea ca un individ să aibă boala la un moment dat
- » Formula:
$$\text{Prevalența} = (\text{numărul de cazuri de boală}) / (\text{total populație})$$

Rata

- O rată calculată reflectă riscul de a surveni în timp un anumit eveniment.
- Ia valori de la 0 la infinit
- Număr de indivizi raportat la unitatea de timp (oră / zi / săptămână / lună / an etc.)
 - Rate de morbiditate
 - Prevalența
 - Incidența
 - Rata de atac
 - Rate de mortalitate
 - Rate de natalitate
- Într-un oraș cu populație de 100000 locuitori s-au înregistrat în anul 1999 200 născuți vii.
- Rata de natalitate = $200/100000 * 1000 = 2$ nou născuți la mia de locuitori
- Rata de fertilitate = (nr. nașteri)/(nr. femei cu vârsta între 15-45 ani) * 1000
- Rata de morbiditate = frecvența cazurilor de îmbolnăvire pentru o populație specificată într-o perioadă de timp
- Rata de mortalitate = (nr. decese)/(populație)*1000

Incidența

- numărul de cazuri noi de boală care apar într-o populație la risc pentru boală într-un interval de timp
- *Incidența cumulată*: proporția de indivizi dintr-o populație fixă care dezvoltă boala respectivă, într-un interval de timp (populația la risc)
= (numărul de cazuri noi într-o perioadă de timp) / (populația totală la risc în aceeași perioadă de timp)
- *Incidența densitate* = măsura teoretică a numărului de cazuri noi care apar pe unitate de populație-timp, de exemplu persoane-ani la risc
= (numărul de cazuri noi într-o perioadă de timp) / (total persoane-timp de observație)

» Boli acute: incidența > prevalența
» Boli cronice: incidența < prevalența

Incidența

- Incidența cumulată: proporția unei populații fixe care se îmbolnăvește într-o perioadă de timp definită. Nu este o rata ci este o proporție dar adesea se denumește ca și rată (ex. rata de atac)
 - Rata de natalitate: 8 copii la 1000 locuitori (8‰)
 - Rata de atac = $(\text{numărul de persoane bolnave}) / (\text{numărul de persoane supuse riscului})$
 - Rata de mortalitate:
 - = $(\text{decese dintr-o perioadă}) / (\text{populația la mijlocul perioadei})$
 - 2.000 de decese la 100.000 de locuitori

Riscul

- Riscul = probabilitatea ca un subiect să dezvolte o patologie de interes într-o perioadă de timp dată
 - Riscul relativ (RR) = raportul dintre rata de incidență la cei expuși și rata de incidență la cei neexpuși
 - $RR = 1 \Leftrightarrow$ nu există risc
 - $RR < 1 \Leftrightarrow$ factor de protecție
 - Riscul atribuabil = diferența dintre rata de incidență la cei expuși și rata de incidență la cei neexpuși

Tabel de contingență 2X2

	Boala+	Boala-	Total
Expus	a	b	a+b
Neexpus	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

Subiecți expuși

Subiecți neexpuși

Subiecți cu boala

Subiecți indemni

n = volumul eșantionului

- $RR = [a/(a+b)]/[c/(c+d)]$
- $RS \text{ (rata șansei - OR)} = (a*d)/(b*c)$

	Cancer+	Cancer-	Total
Fumător	83	72	155
Nefumător	3	14	17
Total	86	86	172

- Prevalența = $86/172*100 = 50\%$
- De câte ori cancerul este mai frecvent la fumători față de nefumători?
 - Prevalența la fumători (P1) = $83/155*100 = 54\%$
 - Prevalența la nefumători (P2) = $3/17*100 = 18\%$
 - $P1/P2 = 54/18 = 3$

Estimatori punctuali: Tabela de contingență

	Boală+	Boală -	Total
Expuns	AP	FP	= AP+FP
Neexpus	FN	AN	= FN+AN
Total	= AP+FN	=FP+AN	= n

Estimatori ai testului diagnostic

Denumire parametru	Formula
Rata falșilor pozitivi	$=FP/(FP+AN)$
Rata falșilor negativi	$=FN/(FN+AP)$
Sensibilitatea	$=AP/(AP+FN)$
Specificitatea	$=AN/(AN+FP)$
Acuratețea	$=(AP+AN)/n$
Valoarea predictivă pozitivă	$=AP/(AP+FP)$
Valoarea predictivă negativă	$=AN/(AN+FN)$
Riscul relativ	$=AP(FP+AN)/FN(AP+FP)$
Rata șansei	$=(AP \cdot AN)/(FN \cdot FP)$
Riscul atribuabil	$=AP/(AP+FP) - FN/(FN+AN)$

Teste diagnostice

- Sensibilitatea = persoane bolnave diagnosticate pozitiv/număr total de persoane
- Specificitatea = persoane indemne diagnosticate negativ/număr de persoane indemne
- Acuratețea = suma dintre persoanele adevărat pozitive și adevărat negative raportat la totalul eșantionului

Testul de sarcină

	Sarcină+	Sarcină-	Total
Test+	80	70	150
Test-	6	14	20
Total	86	84	170

- Prevalența = ?
- Sensibilitatea = ?
- Specificitatea = ?
- Acuratețea = ?
- Rata falșilor pozitivi = ?
- Rata falșilor negativi = ?

PASUL 1

Fie plasați capătul absorbant al testului direct sub jetul de urină, pentru cel puțin 5 secunde, fie colectați urina într-un recipient curat și uscat, după care introduceți capătul absorbant în urină pentru cel puțin 5 secunde.



PASUL 2

Așteptați 5 minute afișarea rezultatului



PASUL 3

CITIREA REZULTATULUI



Apariția unei singure linii roz în zona de control (C) indică un rezultat negativ – **NU SUNTEȚI ÎNSĂRCINATĂ**.

Apariția a două linii roz indică un rezultat pozitiv – **SUNTEȚI ÎNSĂRCINATĂ**.

Dacă în 5 minute nu apare nici o linie în zona de control (C), sau apare doar linia din zona de test (T), **TESTUL NU ESTE VALID**. Repetați testarea cu un nou Test de sarcină Barza.

VARIABILE CANTITATIVE

Parametrii de centralitate

- Estimatori care pun la dispoziție informații cu privire la distribuția datelor
- Estimatori:
 - Modulul
 - Mediana
 - Media (media aritmetică)
 - Valoarea centrală
 - Media geometrică
 - Media ponderată

Centralitate: MODULUL

- Denumit și valoare modelă
 - valoarea cu cea mai mare frecvență în seria statistică
- Nu există formulă de calcul
- Corespunde punctului cel mai înalt al graficului de distribuție

Excel (funcție predefinită):
=MODE(array)

Centralitate: MODULUL

- Seria unimodală:

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

- Vârsta pacienților internați cu diaree în Clinica Pediatrie 1 în perioada 11.01 – 11.08.2008

- Seria bimodală:

2	1	2	1	1
2	2	1	3	3

- Seria multimodală (multimodală):

2	1	2	1	1
2	3	3	3	4

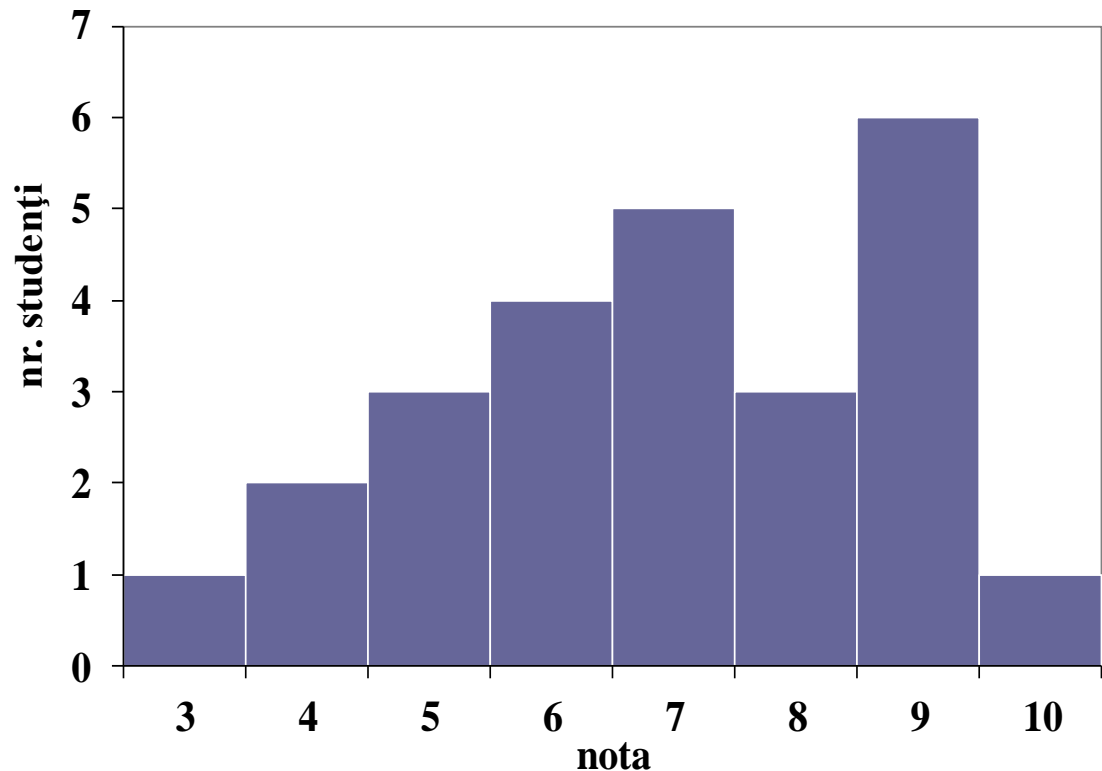
Centralitate: MODULUL

- Nu este influențată de valorile extreme

Note la examenul
practic pentru un
eșantion de 25
studenți sunt:

3, 4, 9, 5, 4, 6, 7, 7, 8,
5, 9, 7, 9, 5, 6, 9, 10,
6, 7, 7, 8, 9, 8, 9, 6

Moulul = 9



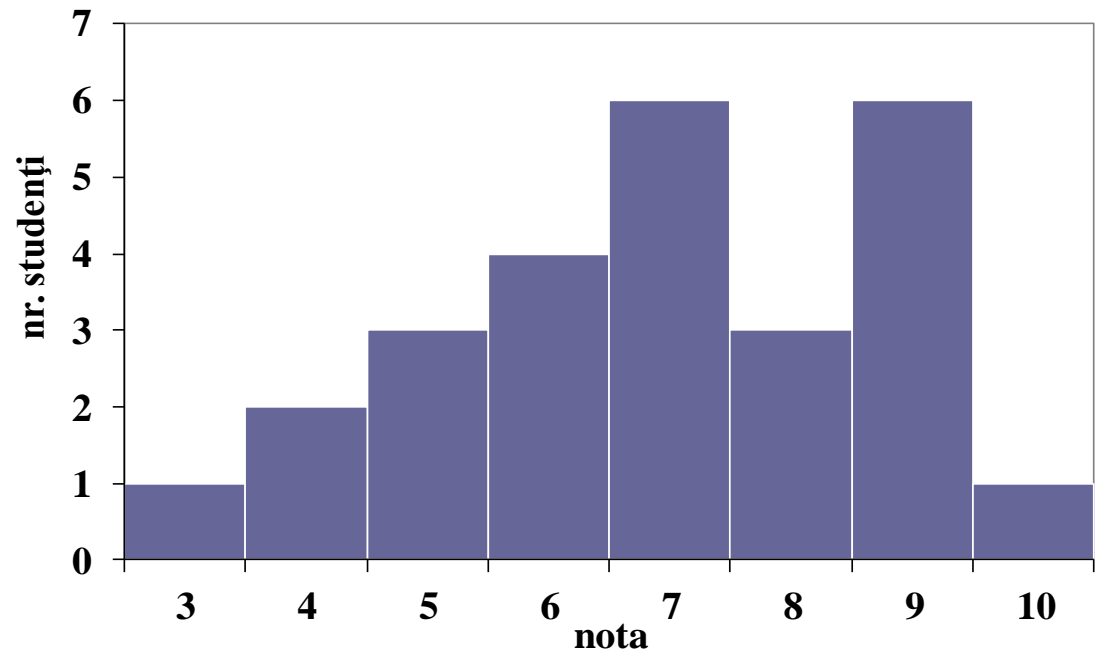
Centralitate: MODULUL

- Seria bimodală

Note la examenul practic pentru un eșantion de 26 studenți sunt:

3, 4, 9, 5, 4, 6, 7, 7, 8, 5,
9, 7, 9, 5, 7, 6, 9, 10, 6,
7, 7, 8, 9, 8, 9, 6

Modulul = 7 & 9



Centralitate: MEDIANA

- 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:

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

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

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

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

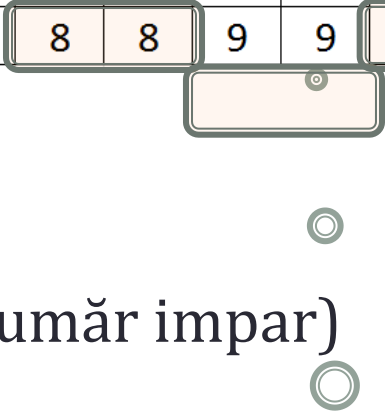
- Volumul eșantionului: $n = 10$ (număr par)
- $Me = (X_5 + X_6) / 2 = (7 + 8) / 2 = 7,5$
- **Excel:** $=\text{median}(A2:J2) = 7,5$

**Serie
multimodală**

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:** $=\text{median}(A2:I2) = 8$

**Serie
multimodală**

Centralitate: MEDIA ARITMETICĂ

- 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):

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

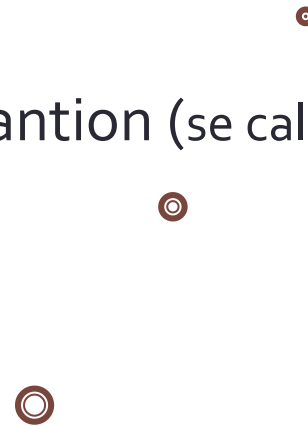
Simbol diferit pentru populație respectiv eșantion dar formulă de calcul similară

Centralitate: MEDIA ARITMETICĂ

- 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):

- Eșantion (se calculează):



Simbol diferit pentru populație respectiv eșantion dar formulă de calcul similară

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.

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.

Minimum = 5

Maximum = 10

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

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


$$C = 90$$

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

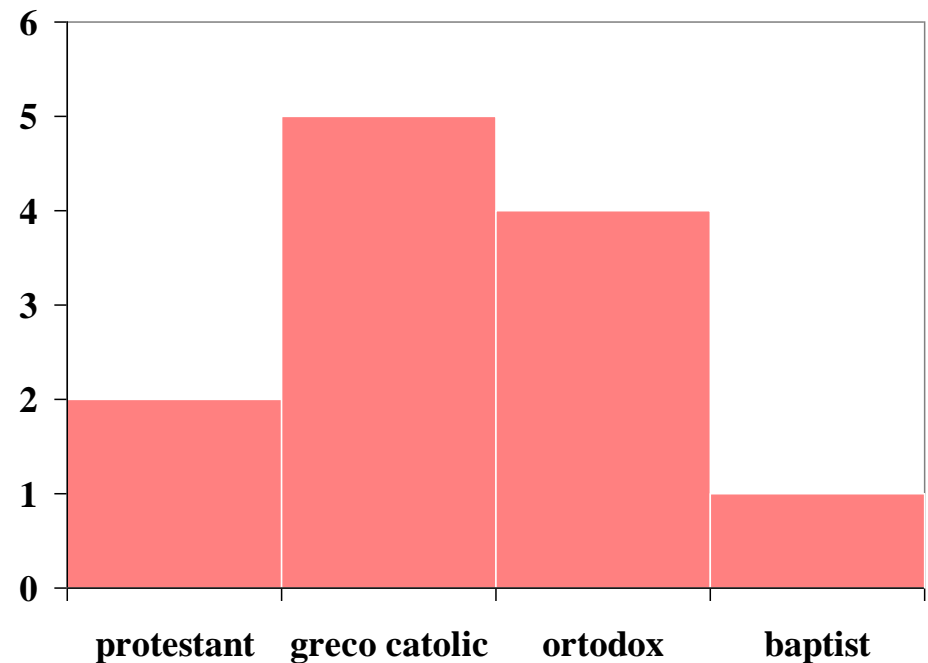

$$C = 2,5$$

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

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

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*.



Centralitate: VALOAREA CENTRALĂ

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 și 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

Centralitate: MEDIA GEOMETRICĂ

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

- 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 =
- 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%

Centralitate: MEDIA PONDERATĂ

- 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

AVANTAJE ȘI DEZAVANTAJE

Estimator	Avantaj	Dezavantaj
Media	Utilizează toate datele Ușor de aplicat (formulă ușoară)	Influențată de outliers Nerepresentativă dacă datele nu au o distribuție simetrică
Mediana	Nu e influențată de outliers Neinfluențată de asimetria datelor	Ignoră majoritatea datelor din serie
Modulul	Aplicabil și variabilelor calitative	Ignoră majoritatea datelor din serie
Media geometrică	Aplicabilă datelor asimetrice spre dreapta	Appropriate if the log transformation produce a symmetrical distribution
Media ponderată	Cuantifică importanța relativă a fiecărei observații	Ponderile trebuie să fie cunoscute sau estimate

DISPERSIE

- Împrăștierea relative la valoarea mediei sau a medianei
- Datele sunt cu atât mai împrăștiate cu cât valorile sunt mai diferite

Estimatori:

1. Amplitudinea
2. Variația (VAR.S)
3. Deviația standard (STDEV.S)
4. Coeficientul de variație
5. Eroarea standard

AMPLITUDINEA

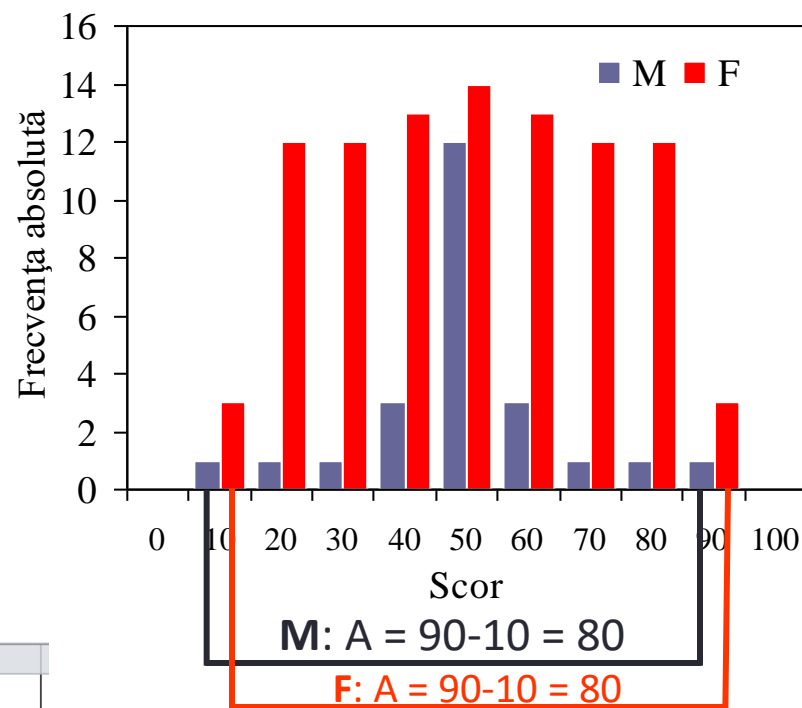
$$A = X_{\max} - X_{\min}$$

- Nu ne spune nimic despre modalitatea în care datele variază în jurul valori centrale
- Valorile extreme afectează semnificativ valoarea amplitudinii
- Excel: RANGE (Descriptive Statistics)

	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

↓

$$A = X_{\max} - X_{\min} = 10 - 3 = 7$$



Aplicabilitate: evaluarea fluctuațiilor temperaturii zilnice

VARIAȚIA ȘI DEVIAȚIA STANDARD

- Variația eșantionului (s^2) = media corectată (n-1) a sumei pătratelor abaterilor de la medie
- Deviația standard (s) = abaterea standard = ecartul tip
- Are aceeași unitate de măsură ca și media și datele seriei

Populație

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \quad s = \sqrt{s^2}$$

$$S^2 = \frac{n}{n-1} s^2 \quad S = \sqrt{S^2}$$

Eșantion

VARIAȚIA ȘI DEVIAȚIA STANDARD

Obs.	Nota	Deviația de la medie	$(X_i - m)^2$
X_1	9	$=9 - 7,3 = 1,67$	$=1,67^2 = 2,78$
X_2	6	-1,33	1,78
X_3	4	-3,33	11,11
X_4	9	1,67	2,78
X_5	4	-3,33	11,11
X_6	8	0,67	0,44
X_7	8	0,67	0,44
X_8	9	1,67	2,78
X_9	7	-0,33	0,11
X_{10}	4	-3,33	11,11
X_{11}	10	2,67	7,11
X_{12}	10	2,67	7,11

Variația eșantionului

Media aritmetică (m)
m = 7,3

$$S^2 = \text{sum}(X_i - m)^2 / (n - 1)$$

$$S^2 = 58,67 / 11 = 5,3$$

Deviația standard a
eșantionului

$$S = 2,31$$

DISPERSIE: DEVIAȚIA STANDRAD

Interval	% din observațiile conținute
$\bar{X} \pm 1 \cdot s$	68
$\bar{X} \pm 2 \cdot s$	95
$\bar{X} \pm 3 \cdot s$	99.7

DISPERSIE: COEFICIENTUL DE VARIATIE

- Măsură relativă a dispersiei datelor
- Formula de calcul:

$$CV = \frac{S}{\bar{X}}$$

- Evaluare a abaterii standard în raport cu valoarea medie
- Are avantajul de a fi un indicator independent de unitățile de măsură
- Se poate exprima și procentual
- Interpretare: populația poate fi considerată

CV < 10%	<u>Omogenă</u>
10% ≤ CV < 20%	<u>Relativ omogenă</u>
20% ≤ CV < 30%	<u>Relativ hetrogenă</u>
> 30%	<u>Heterogenă</u>

DISPERSIE: EROAREA STANDARD

- E utilizată ca estimator de dispersie asociat mediei aritmetice
- Estimator utilizat în calcularea intervalelor de încredere

$$ES = \frac{s}{\sqrt{n}}$$

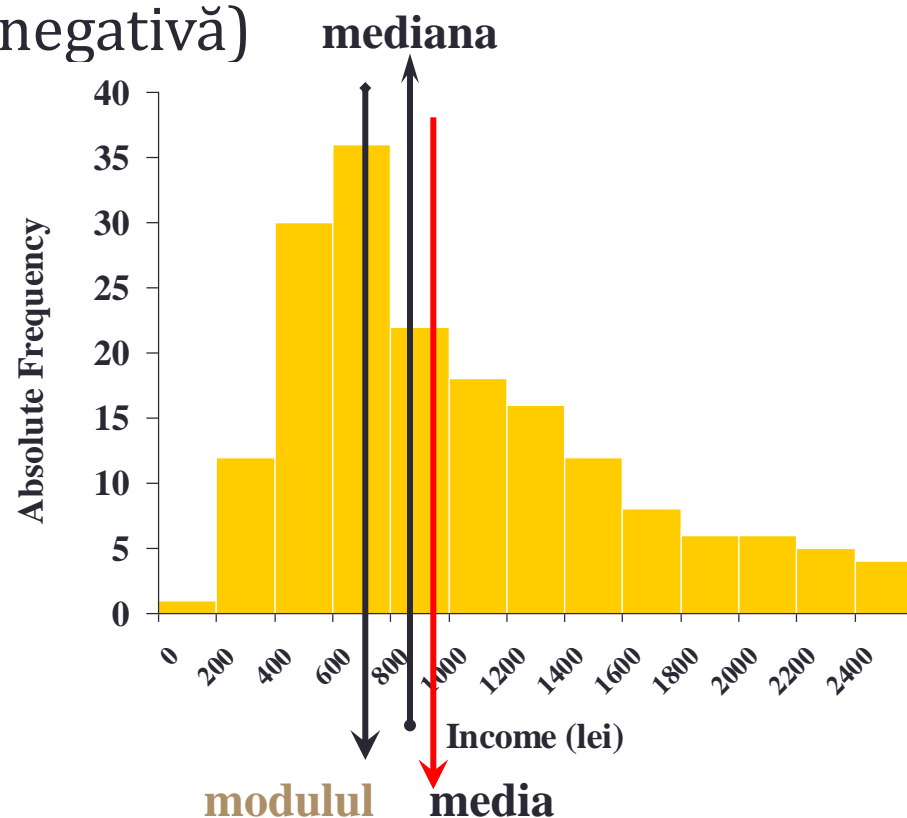
SIMETRIE: ASIMETRIA

- Indică pentru o serie statistică:
 - Deviația de la simetrie
 - Direcția deviației (pozitivă/negativă)

$$M_3 = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n}$$

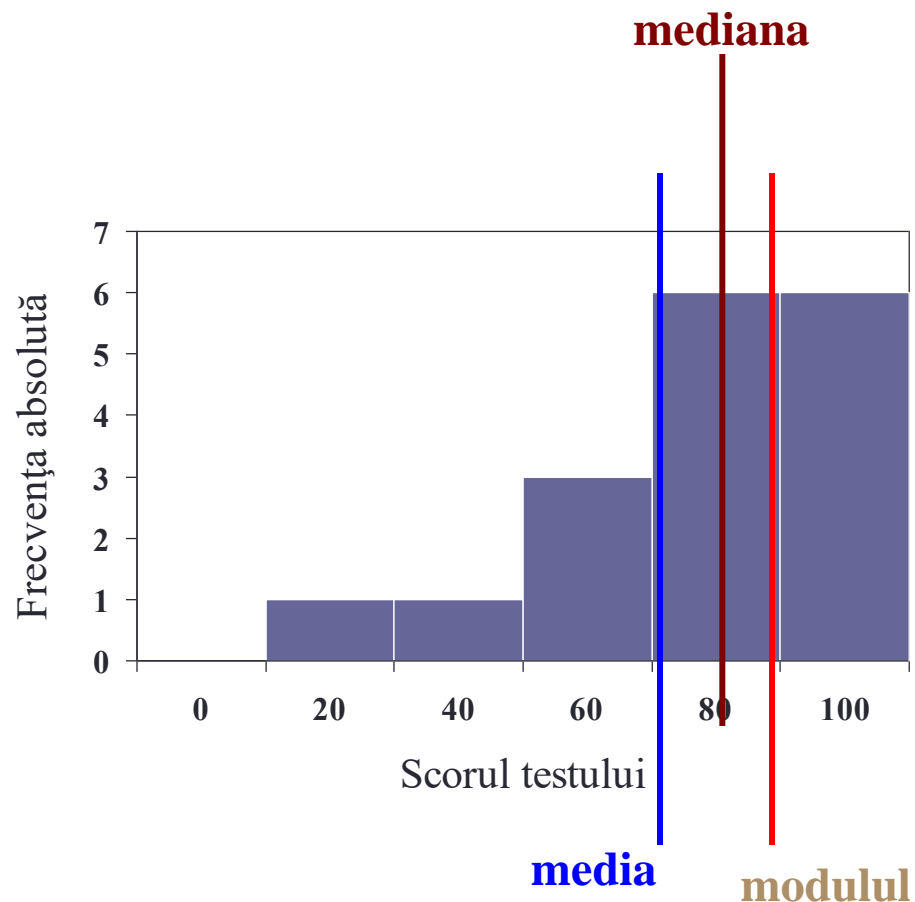
- Asimetrie pozitivă:**

- Modulul** = 7000 Ron
 - Mediana** = 8870 Ron
 - Media** = 9360 Ron
-
- Modulul < Mediana < Media**



SIMETRIE: ASIMETRIA

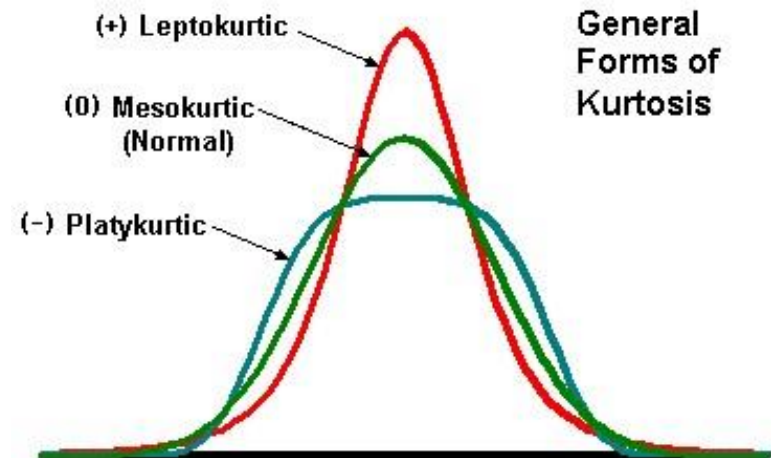
- Asimetrie negativă:
 - Modulul > Mediana > Media
= SKEW(number1, ..., numbern)
- Interpretarea [Bulmer MG, Principles of Statistics, Dover, 1979,] – aplicată populației
 - Asiemtria <-1 sau > +1 → **înalt asimetrică**
 - Asimetria cu valoare între -1 și -1/2 sau +1/2 și +1 → **moderat asimetrică**
 - Asimetria cu valoare între -1/2 și +1/2 → **aproximativ simetrică**



SIMETRIE: BOLTIREA

- = KURT(array)
- Distribuția normală are boltirea în jurul valorii 3.
- Excesul de boltire (ceea ce calculează Excel-ul cu funcția KURT) = Boltire - 3
- Distribuția normală: boltirea $\cong 3$ (excesul de boltire $\cong 0$)
→ **mezocurtic**
- Distribuția cu boltirea < 3 (excesul de boltire < 0) se numește platocurtică
- Distribuția cu boltirea > 3 (excesul de boltire > 0) se numește leptocurtică

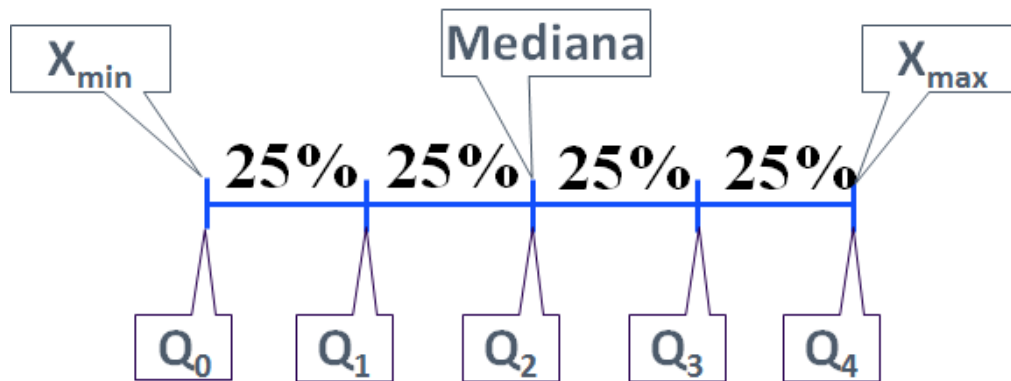
$$\alpha_4 = \frac{\frac{1}{n} \cdot \sum_{i=1}^n (X_i - \bar{X})^4}{S^4} - 3$$



LOCALIZARE: CVARTILE

$$IC = Q_3 - Q_1$$

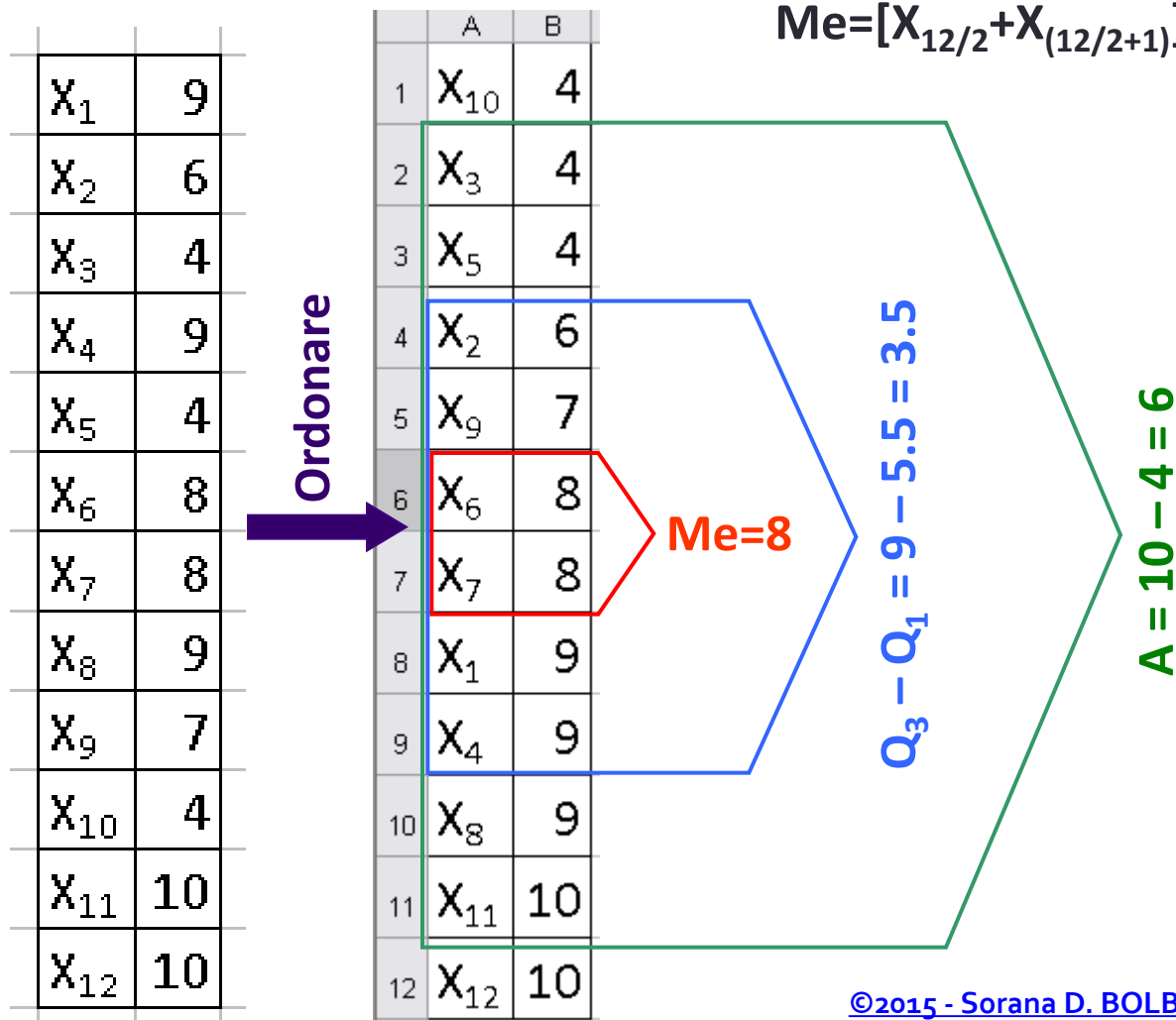
unde Q_3 = cvartila 3 (percentila 75%), Q_1 = cvartila 1 (percentila 25%)



Măsură a dispersiei pentru 50% din datele de mijloc.

INTERVALUL DINTRE Q1 și Q3

- Variabila: note la examenul practic



Formule Excel:

(Mediana) Me:

=MEDIAN(B1:B12)

(Intervalul dintre cvartila 3 și 1) IC:

=QUARTILE(B1:B12,3)-
QUARTILE(B1:B12,1)

(Amplitudinea) A:

=MAX(B1:B12)-
MIN(B1:B12)

INTERVALUL DINTRE Q1 și Q3

- Variabila: note la examenul practic

	A	B
1	X ₁₀	4
2	X ₃	4
3	X ₅	4
4	X ₂	6
5	X ₉	7
6	X ₆	8
7	X ₇	8
8	X ₁	9
9	X ₄	9
10	X ₈	9
11	X ₁₁	10
12	X ₁₂	10

Me=8

$$Q_3 - Q_1 = 9 - 5.5 = 3.5$$

$$A = 10 - 4 = 6$$

Q1: 25% din studenți au note ≤ 5.5

Q3: 75% din studenți au note ≤ 9

IC: 50% din studenți au note care nu diferă una față de alta cu mai mult de 3.5 puncte

A: Diferența dintre nota maximă și nota minimă a fost de 6 puncte

IMPORTANT!

	Nominal	Ordinal	Cantitativ
Modulul	Da	Da (Nu e recomandat)	Da (Nu e recomandat)
Mediana	Nu	Da	Da
Media	Nu	Nu	Da (dacă seria e unimodală și simetrică)

	Amplitudinea	Deviația standard
Nominal	NU	NU
Ordinal	DA (nu e cea mai bună metodă)	NU
Numeric	DA (nu e cea mai bună metodă)	DA (dacă seria e unimodală și simetrică)

IMPORTANT!

- Dacă fiecare dată a seriei se adună sau se scade la o constantă:
 - Media va crește sau va descrește cu valoarea constantei
 - Deviația standard nu se va schimba
- Dacă fiecare dată a seriei se înmulțește sau împarte la o constantă :
 - Media va fi înmulțită sau împărțită la valoarea constantei
 - Deviația standard va fi înmulțită sau împărțită la valoarea constantei

RECALL!

- Unitățile de măsură au influență asupra estimatorilor punctuali
- Estimatorii punctuali se calculează în funcție de tipul variabilei
- Media aritmetică, deviația standard și amplitudinea sunt sensibile la valoril extreme
- Sumarizarea datelor cu ajutorul estimatorilor se face cu pierdere de informație

