

STATISTICA DESCRIPTIVĂ III

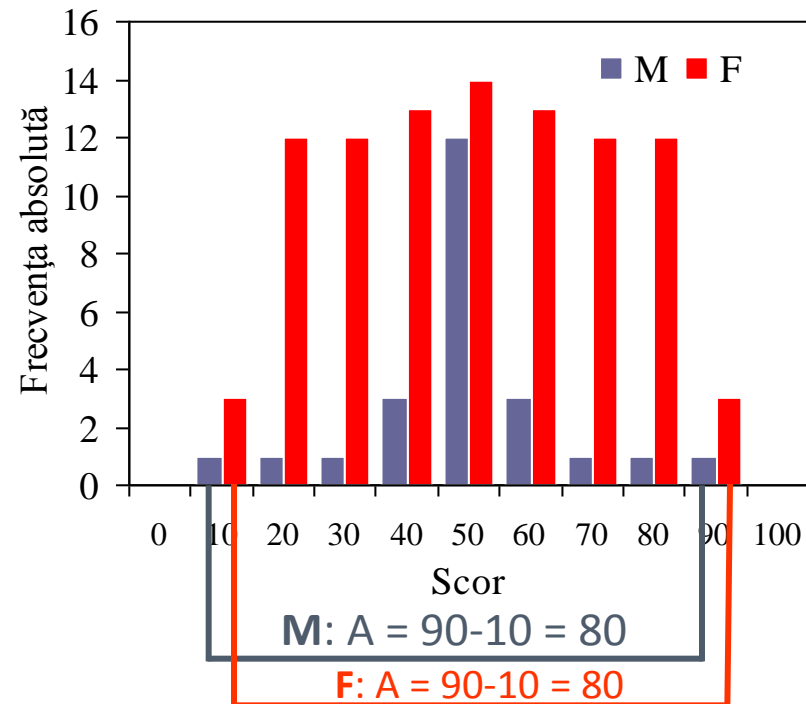
PARAMETRII DE DISPERSIE/ÎMPRĂȘTIERE

- » Calcularea parametrilor de dispersie/împrăștiere
 - > Amplitudinea
 - > Intervalul dintre cvartiala 1 și cvartila 3
 - > Deviația standard de la medie
- » Calcularea și interpretarea coeficientului de variație

Obiective educaționale

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

- » 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

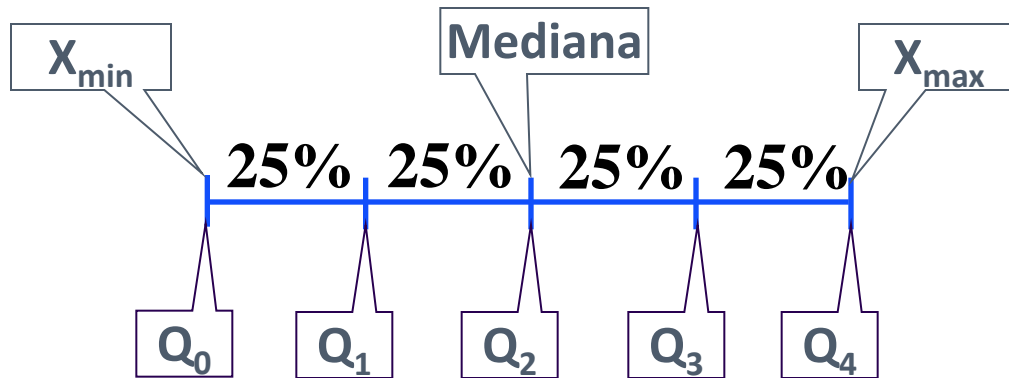
$$A = X_{\max} - X_{\min} = 10 - 3 = 7$$

Aplicabilitate: evaluarea fluctuațiilor temperaturii zilnice

Amplitudinea

$$IC = Q_3 - Q_1$$

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



Măsură a dispersiei pentru 50% din datele de mijloc.

Intervalul dintre cvartilele 1 și 3

» Variabila: note la examenul practic

X_1	9
X_2	6
X_3	4
X_4	9
X_5	4
X_6	8
X_7	8
X_8	9
X_9	7
X_{10}	4
X_{11}	10
X_{12}	10

Ordonare →

	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

$Q_3 - Q_1 = 9 - 5.5 = 3.5$

$A = 10 - 4 = 6$

$$Me = [X_{12/2} + X_{(12/2+1)}] / 2 = (X_6 + X_7) / 2 = (8 + 8) / 2 = 8$$

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 cvartilele 1 și 3

» 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

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

$$A = 10 - 4 = 6$$

Me: ½ din studenți au avut nota la examenul practic < 8 și ½ au avut nota > 8

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

Intervalul dintre cvartilele 1 și 3

- » Măsoară media distanței dintre fiecare observație față de valoarea centrală: media aritmetică sau mediana
- » Dă o pondere egală fiecărei observații
- » Sensibilitatea mai mare decât amplitudinea sau intervalul dintre cvartila 3 și 1
- » Se utilizează în calcularea variației și respectiv a deviației standard

Abateră medie

- » Variația eșantionului = media corectată (n-1) a sumei pătratelor abaterilor de la medie

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

- » Deviația standard = abaterea standard = ecartul tip
- » Are aceeași unitate de măsură ca și media și datele seriei

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

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

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

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

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

$$s = 2,31$$

Variația și deviația standard

» 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

Coeficientul de variație

- » Măsură a variabilității relative utilizată pentru:
 - > Măsurarea modificărilor care au apărut în populație în timp
 - > Compararea variabilității a două populații când unitățile de măsură sunt diferite (mg/dL vs mmol/L – colesterol)
 - > Frecvent exprimat procental

	Greutate (kg)	Înălțime (cm)
Media aritmetică	72,6	168
Deviația standard	13,6	10,2

- » Care din variabilele de mai sus are împrăștierea mai mare?
 - > Nu se poate răspunde la întrebare
- » Care din variabilele de mai sus are împrăștierea relativă la medie mai mare ?

Greutate: $CV = 13,6/72,6 * 100 = 19\%$

Înălțime: $CV = 10,2/168 * 100 = 6,1\%$

Coeficientul de variație

» Aplicabilitate: dispersia în două seturi de date

> A: 12, 13, 16, 18, 18, 20

> B: 120, 130, 160, 180, 180, 200

	Grup A	Grup B
Media	16	162
Deviația standard	3	29
CV (%)	18	18

» Aplicabilitate: dispersia în două seturi de date

> A: 12, 13, 16, 18, 18, 20

> B: 2, 3, 160, 18, 200, 300

	Grup A	Grup B
Media	16	114
Deviația standard	3	114
CV (%)	18	100

» Care din seria de mai jos are deviația standard cea mai mică

A. 10, 11, 13, 13, 13, 15, 16

B. 10, 10, 10, 10, 10, 10, 12

C. 10, 11, 12, 13, 14, 15, 16

D. 20, 22, 25, 210, 212, 215, 2100

E. 5, 50, 51, 52, 500, 510, 520

	A	B	C	D	E
1	A	B	C	D	E
2	10	10	10	20	5
3	11	10	11	22	50
4	13	10	12	25	51
5	13	10	13	210	52
6	13	10	14	212	500
7	15	10	15	215	510
8	16	12	16	2100	520
9					
10	2	1	2	755	252