



Autor: Conf. Dr. Bondor Cosmina-Ioana

Sumarizarea și prezentarea datelor



ALWAYS



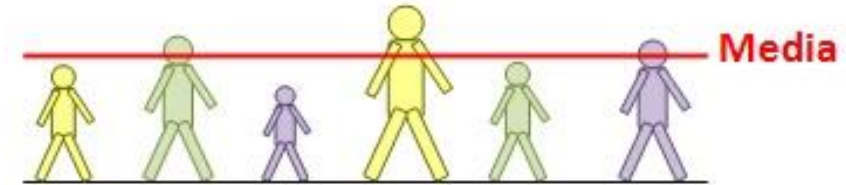
SEEK



KNOWLEDGE

Variabile numerice (cantitative)

Măsurarea tendinței centrale



Media aritmetică \bar{X} = media observațiilor.

Mod de calcul: se adună observațiile pentru a obține suma și apoi se împarte la numărul de observații.

Formula pentru medie:

$$\bar{X} = \frac{1}{n} \sum X$$

Σ înseamnă adunare, X reprezintă observațiile individuale, n este numărul de observații.

Media este o măsură ce indică mijlocul
(centralitatea)?

- Ex. 6 cadre medicale câștigă pe lună 2400, 2500, 2900, 2900, 3000, 3100 Euro.

Media:

$$\bar{X} = \frac{2400+2500+2900+2900+3000+3100}{6} = 2800 \text{ Euro}$$

Media este o măsură ce indică mijlocul
(centralitatea)? Nu tot timpul

Folosim media dacă datele sunt numerice fără cazuri aberante

Utilizăm valoarea mediană dacă datele sunt numerice cu valori extreme sau datele sunt ordinale

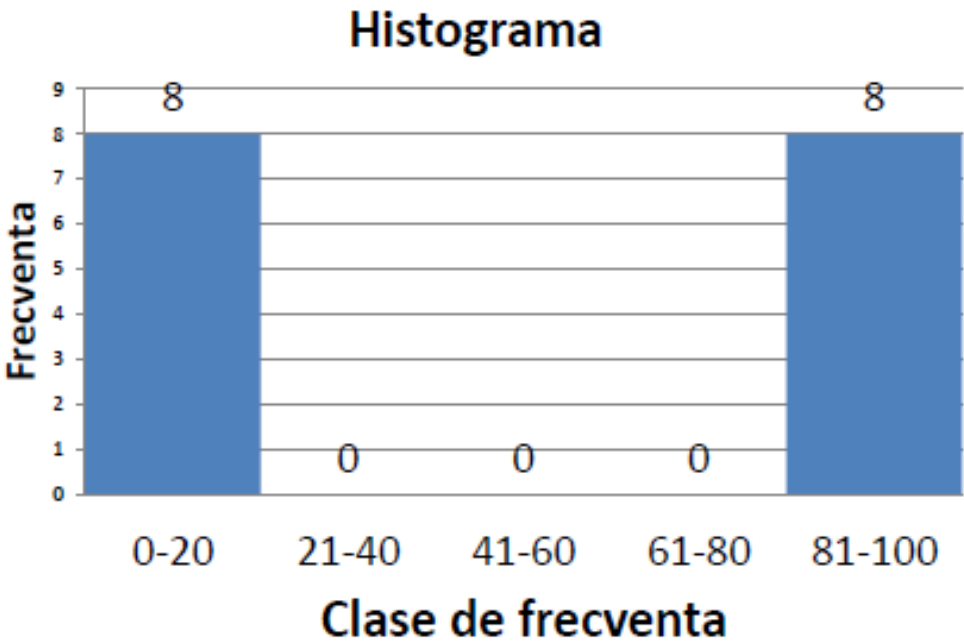


Seria 1
1
1
2
3
5
6
6
7
93
94
94
95
97
98
98
100

Tabel de frecvență

Seria 1 – Clase de frecvență	Frecvența absolută	Frecvența relativă %
0-20	8	50
21-40	0	0
41-60	0	0
61-80	0	0
81-100	8	50
Total	16	100

grafic Histograma



Distribuția definiție

Arată cât de des apare o valoare (frecvența) sau un grup de valori (clase de frecvență)

- reprezentare tabelară
- reprezentare grafică

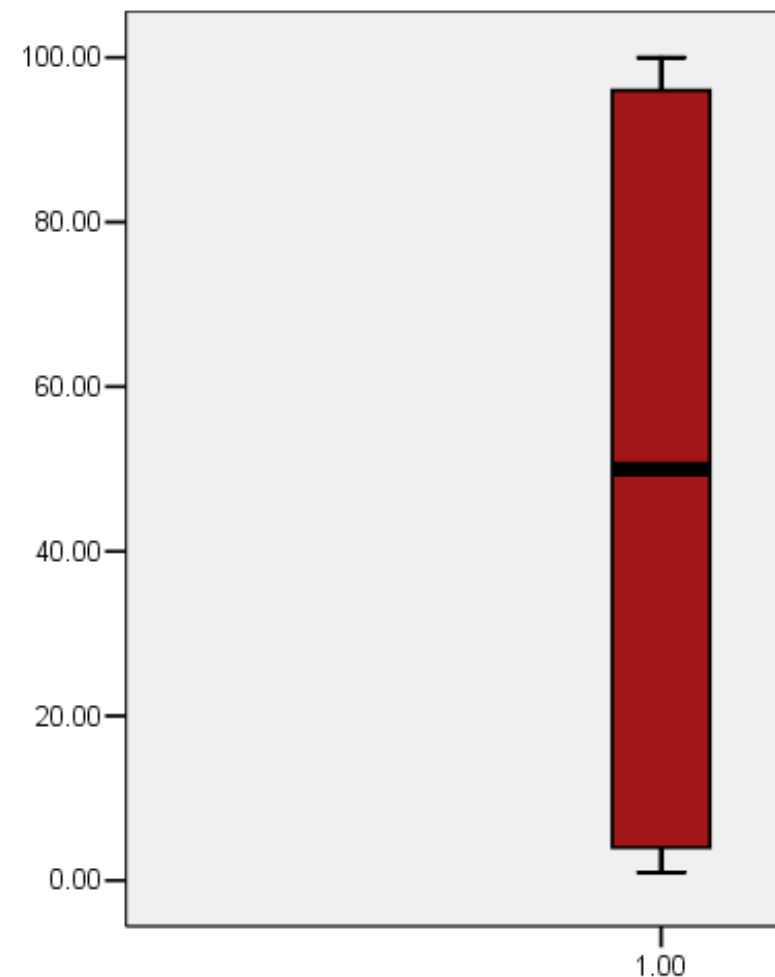
Seria 1
1
1
2
3
5
6
6
7
93
94
94
95
97
98
98
100

Mediana $= (7+93)/2 = 50$

Percentila 25 $= (3+5)/2 = 4$

Percentila 75 $= (95+97)/2 = 96$

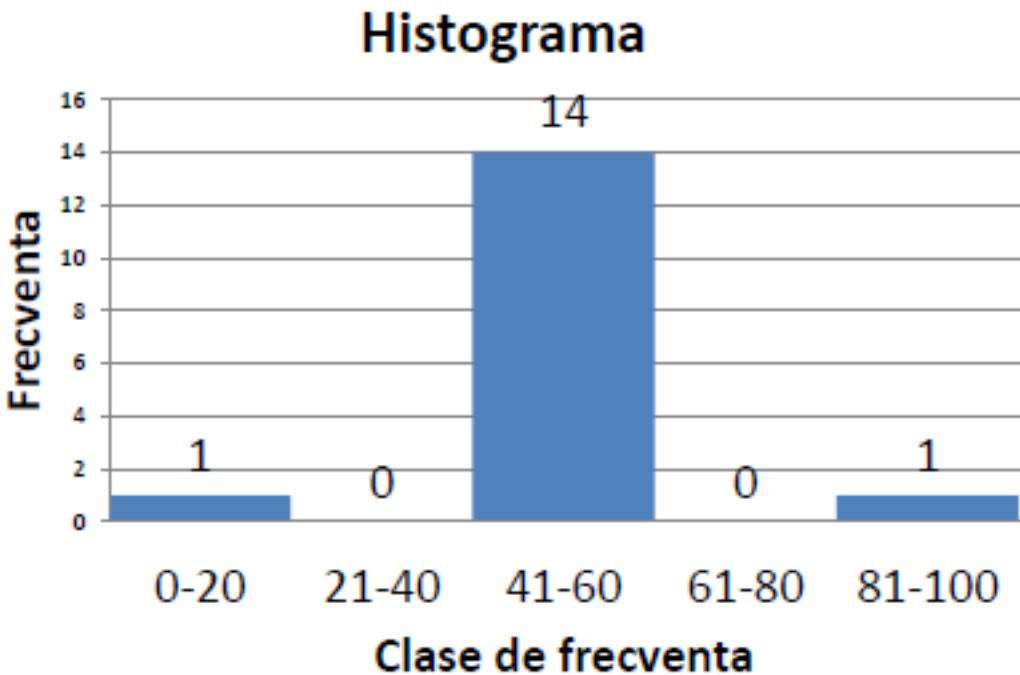
grafic cutie cu mustăți



Seria 2
1
44
45
46
48
48
49
50
50
51
52
52
54
55
55
100

Distribuția

Seria 2 – Clase de frecvență	Frecvența absolută	Frecvența relativă %
0-20	1	6,25
21-40	0	0
41-60	14	87,50
61-80	0	0
81-100	1	6,25



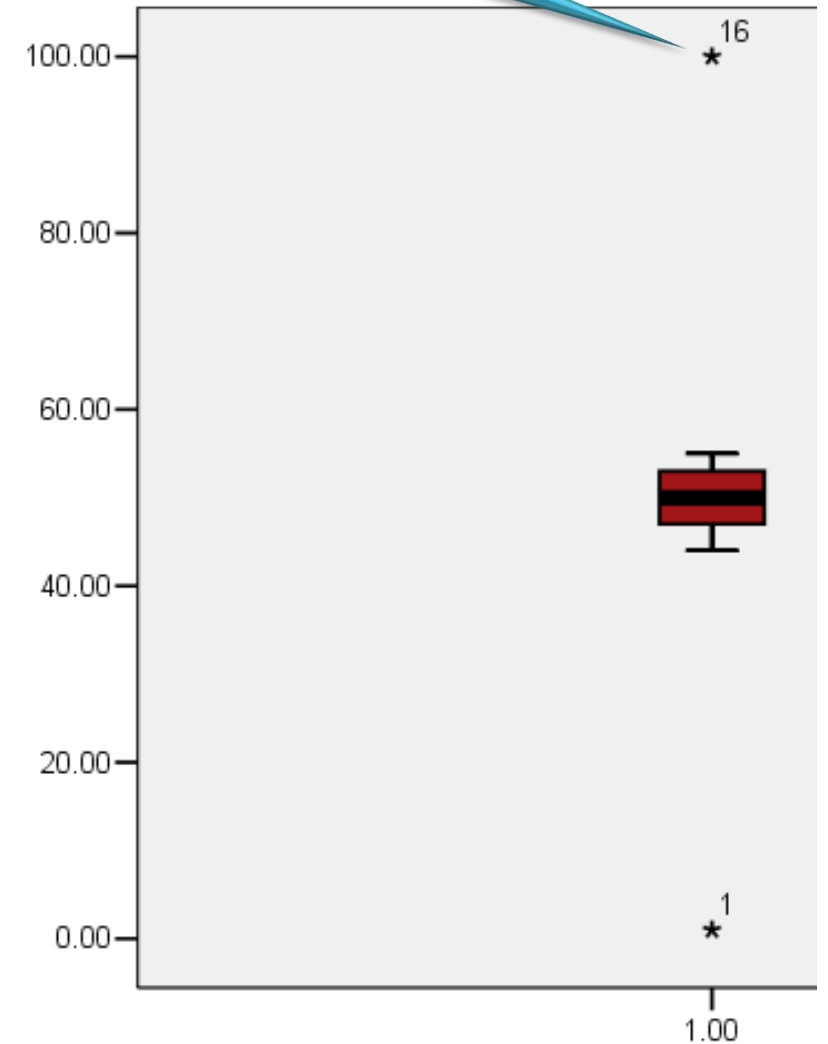
Seria 2
1
44
45
46
48
48
49
50
50
51
52
52
54
55
55
100

Mediana $= (50 + 50) / 2 = 50$

Percentila 25 $= (46 + 48) / 2 = 47$

Percentila 75 $= (52 + 54) / 2 = 53$

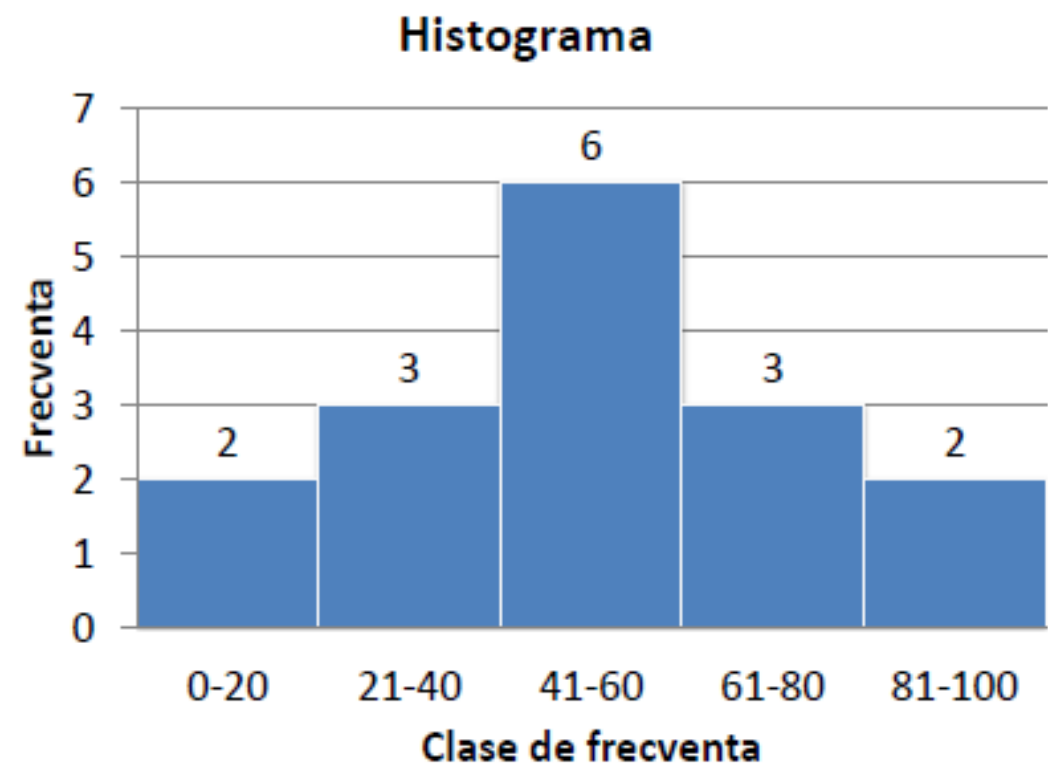
Caz extrem



Seria 3
1
11
24
29
36
41
45
49
51
55
59
64
71
76
88
100

Distribuția

Seria 3 – Clase de frecvență	Frecvența absolută	Frecvența relativă %
0-20	2	12,50
21-40	3	18,75
41-60	6	37,50
61-80	3	18,75
81-100	2	12,50

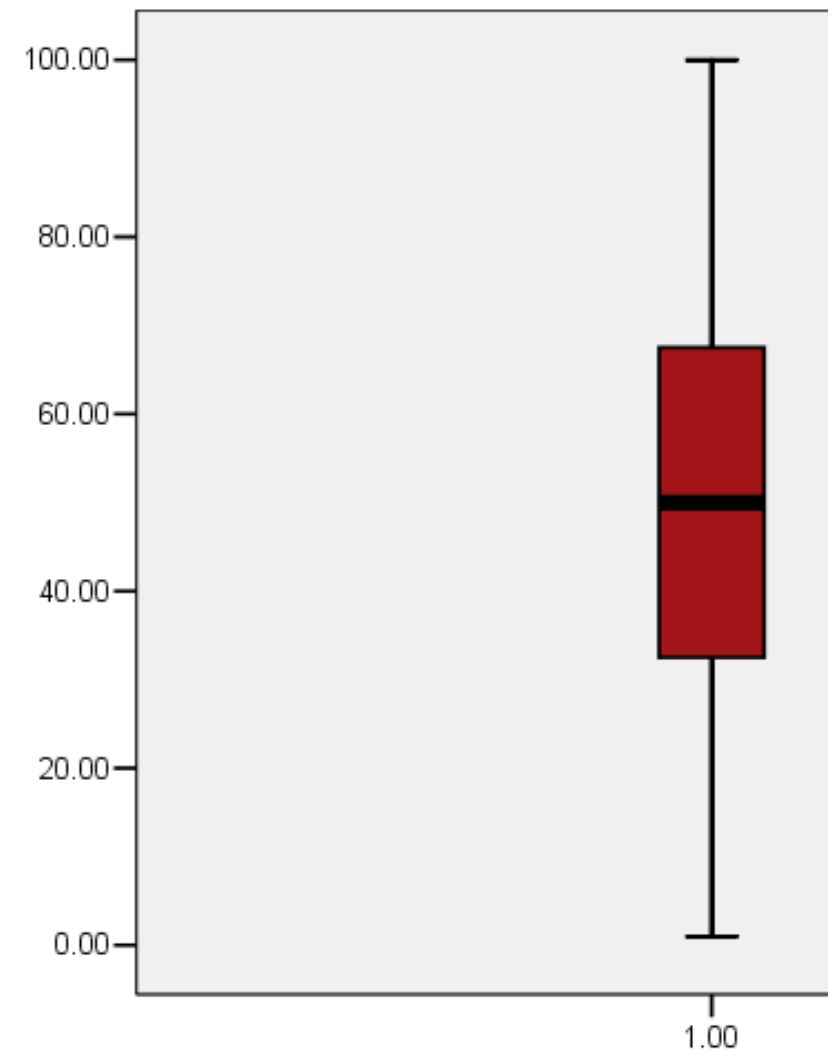


Seria 3
1
11
24
29
36
41
45
49
51
55
59
64
71
76
88
100

Mediana $= (49 + 51) / 2 = 50$

Percentila 25 $= (29 + 36) / 2 = 32,5$

Percentila 75 $= (64 + 71) / 2 = 67,5$



Măsuri de dispersie

Deviația standard – Media deviației de la medie

Formula pentru **populație** (toate observațiile):

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

Σ înseamnă adunare, X reprezintă observațiile individuale, μ este media aritmetică a întregii populații, N este numărul de observații.

Formula pentru un subset al populației (eșantion)

Deviația standard pentru **eșantion** (cu corecție)

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N-1}} = \sqrt{\frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + \dots + (x_N - \bar{X})^2}{N-1}}$$

unde

N – numărul total de observații

\bar{X} - media aritmetică

x_1, \dots, x_N - observațiile

Exemplu

Nr. de dinți pentru 6 pacienți: 20, 24, 26, 25, 25, 30

Media aritmetică $\bar{X} = (20+24+26+25+25+30)/6 = 25$

Deviația standard:

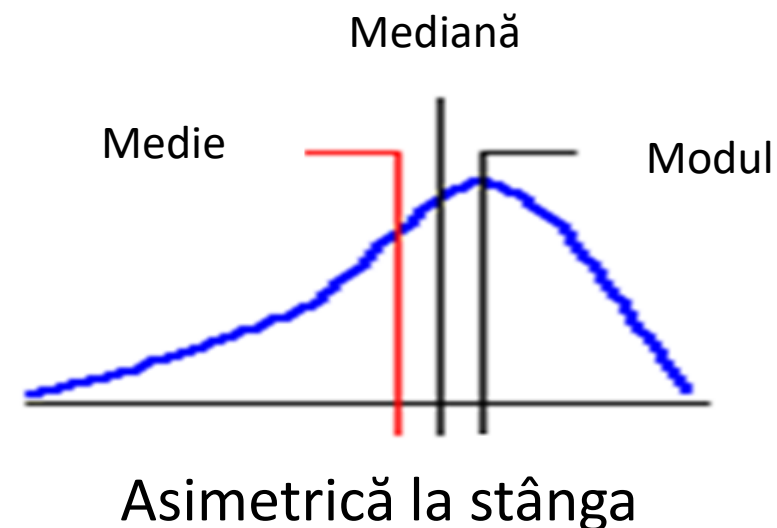
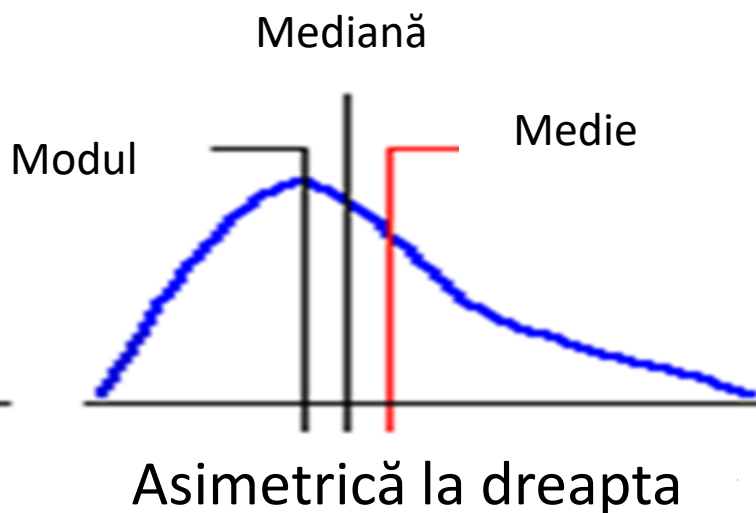
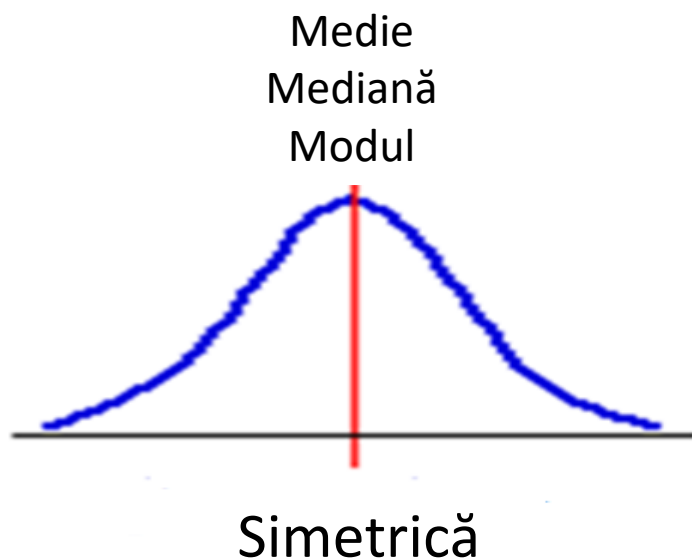
$$S = \sqrt{\frac{(20-25)^2 + (24-25)^2 + (26-25)^2 + (25-25)^2 + (25-25)^2 + (30-25)^2}{6-1}} =$$
$$= \sqrt{\frac{(-5)^2 + (-1)^2 + 1^2 + 0^2 + 0^2 + 5^2}{5}} = \sqrt{\frac{25+1+1+25}{5}} = \sqrt{\frac{52}{5}} = 3,22$$

Deviația
pozitivă de
la medie

Media deviației de la medie :-P

Măsuri de centralitate - aplicabilitate

- Cum să interpretăm media, mediana și modulul:



Asimetria

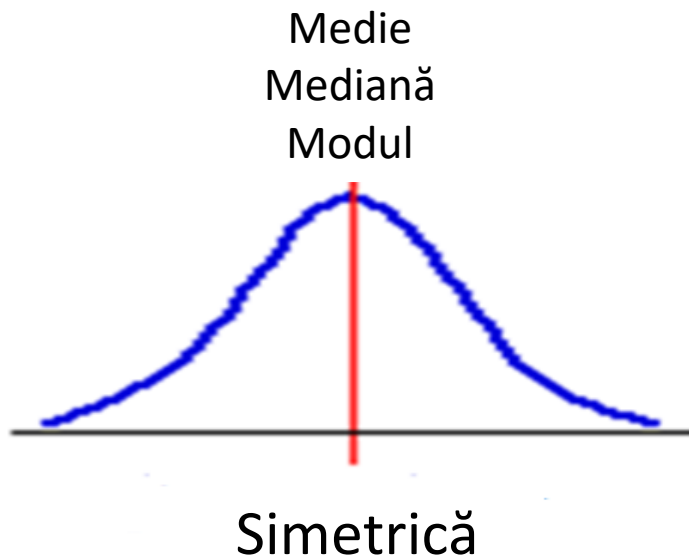
Asimetria – Asimetria datelor

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

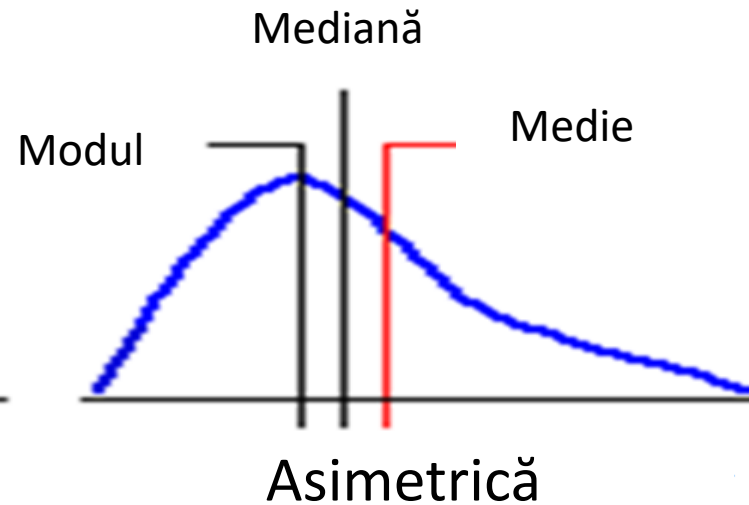
*Unde s – deviația standard, n nr. de indivizi,
 X_1, \dots, X_n – valorile seriei, \bar{X} media aritmetică*

Asimetria

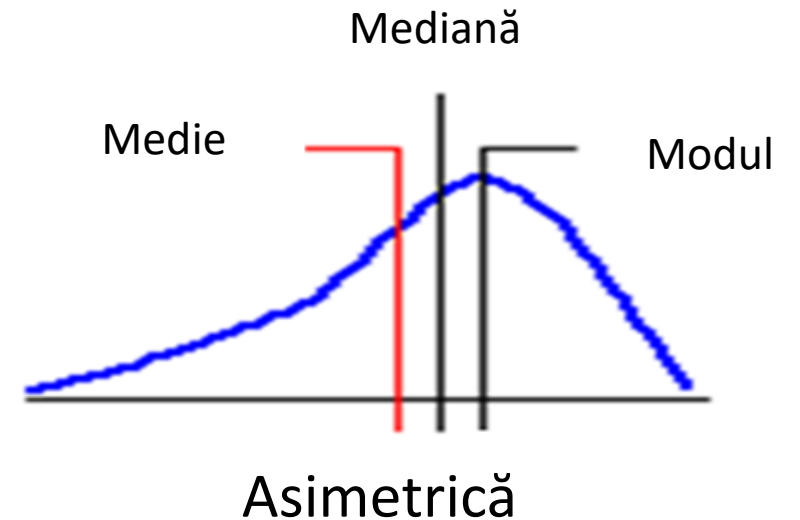
Asimetria este o măsură care descrie distribuția unei variabile cantitative continue



$\text{Simetria} \approx 0$
Distribuție simetrică



$\text{Simetria} > 1$
asimetrică



$\text{Simetria} < -1$
asimetrică

Boltirea

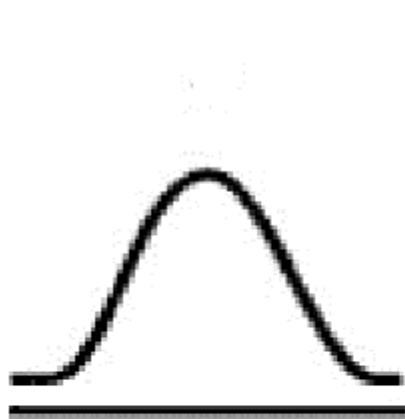
Boltirea – Platitudinea

$$K_E \approx \frac{1}{n} \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{s^4} - 3$$

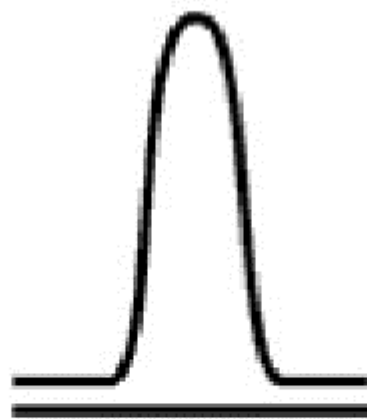
*Unde s – deviația standard, n nr. de indivizi,
 X_1, \dots, X_n – valorile seriei, \bar{X} media aritmetică*

Boltirea (kurtosis)

Boltirea este o măsură care descrie distribuția unei variabile cantitative continue



Mesocurtică
Boltirea ≈ 0



Leptocurtică
Boltirea > 1



Platicurtică
Boltirea < -1



Descrierea datelor numerice (Continue, Discrete)



Măsuri ale tendinței centrale

- Media aritmetică
- Mediana
- Modul

Măsuri de dispersie (împrăștiere)

- Varianța
- Deviația Standard

Alte măsuri

- Asimetria
- Boltirea
- Cvartile
- Percentile

Grafice

- Histograma
- Box-plots

Muțumesc!