



Autor: Conf. Dr. Bondor Cosmina-Ioana

# Sumarizarea și prezentarea datelor



ALWAYS



SEEK



KNOWLEDGE

Variabile numerice (cantitative)

C	G	H	I
Varsta mamei	Numar saptamani de sarcina	Numar nasteri mama	Numar sarcini
22	40	2	2
25	41	0	3
32	39	0	2
28	41	0	0
25	34	2	3
26	41	0	4
31	41	0	0
35	36	3	3
26	41	1	1
24	39	0	0
25	41	1	1
27	41	0	8
29	41	1	1
30	40	1	1
26	41	2	2
21	39	1	4
29	41	1	1
33	28	0	1
41	29	0	1
28	40	0	0

# Descrierea datelor numerice (Continue, Discrete)



## Măsuri ale tendinței centrale

- Media aritmetică
- Mediana
- Modul

## Măsuri de dispersie (împrăștiere)

- Varianța
- Deviația Standard
- Coeficientul de variație
- Eroarea Standard

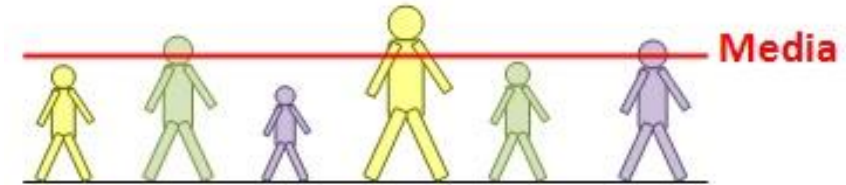
## Alte măsuri

- Asimetria
- Boltirea
- Cvartile
- Percentile

## Grafice

- Histograma
- Box-plots
- Medie/Error plot

# 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:

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

*$\Sigma$  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:

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

# Venitul personalului medical

2600€ 2600 €



Media  
aritmetică=  
2800 €

3000 €



3000 €



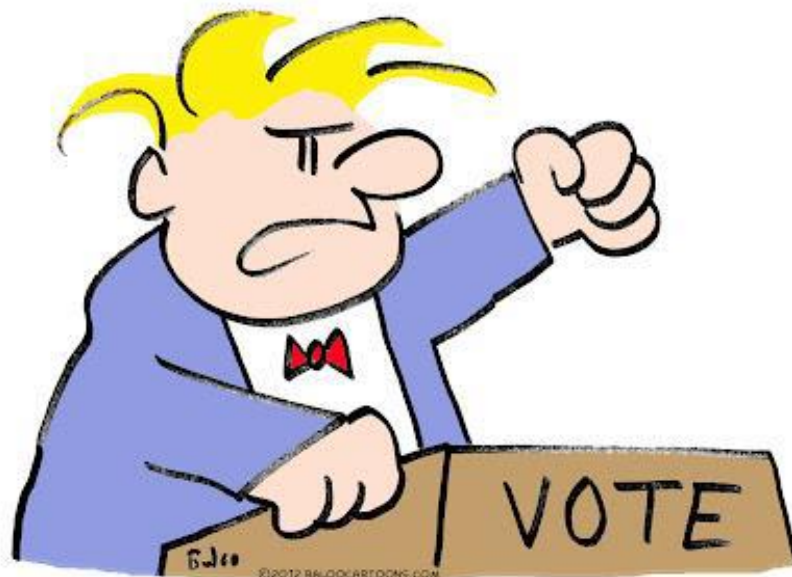
2800 €



2800 €

Media veniturilor cadrelor medicale 2800 €

# Se schimbă politicile sau apare o criză



"Desperate times call for  
desperate politicians!"

După 12 luni politicienii anunță că **s-au dublat**  
salariile cadrelor medicale de la 2800€ la 5600€

S-au **dublat** veniturile personalului medical după 12 luni

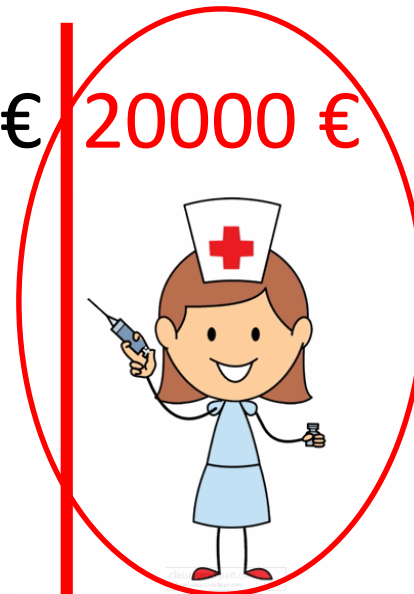
2600€ 2600 €



2800 € 2800 €

A fost 2800 €

3000 € ~~20000 €~~



Media aritmetică= 5633 €



Concluzie: cazurile aberante (salar 20000€) pot influența media

Mediana?

## Venitul personalului medical

2600€ 2600 €



Mediana =  
2800 €

3000 €



3000 €



2800 €



2800 €

## Veniturile personalului medical după 12 luni

2600€ 2600 €



Mediana =  
2800 €



2800 €



2800 €

3000 €



20000 €



Mediana = 2800 €

A fost 2800 €

Concluzie: cazurile aberante NU pot influența mediana

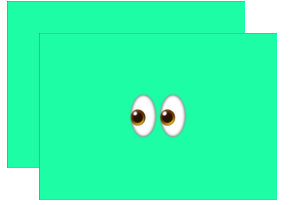
Media este o măsură ce indică mijlocul  
(centralitatea)? Nu tot timpul

Folosim media dacă datele sunt numerice fără cazuri aberante

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



# Întrebări



<https://app.wooclap.com/BFKRI03?from=event-page>

Grafice

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

amplitudinea = maxim-minim = 99



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

Împărțim amplitudinea = 99  
în **4-10** clase de frecventa

Clase de frecvență = intervale egale

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

Împărțim amplitudinea = 99 în 4-10 clase de frecventa

Clase de frecvență = **intervale** egale

Seria 1 –			
Clase de			
frecvență			
0-20			
21-40			
41-60			
61-80			
81-100			

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

8

8

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

8 valori între 0 și 20

8 valori între 81 și 100

Seria 1

1

1

2

3

5

6

6

7

93

94

94

95

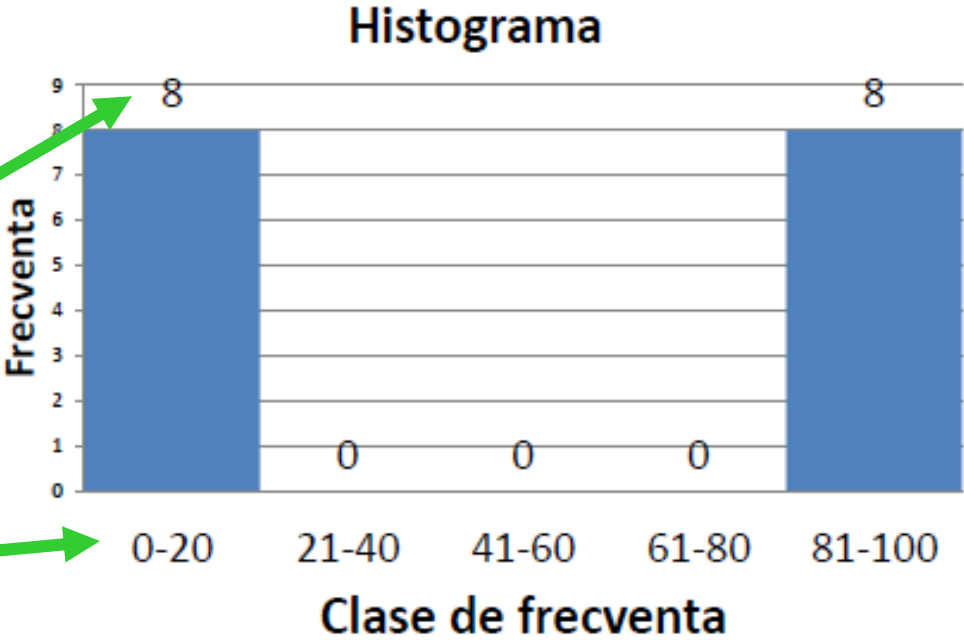
97

98

98

100

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



# 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

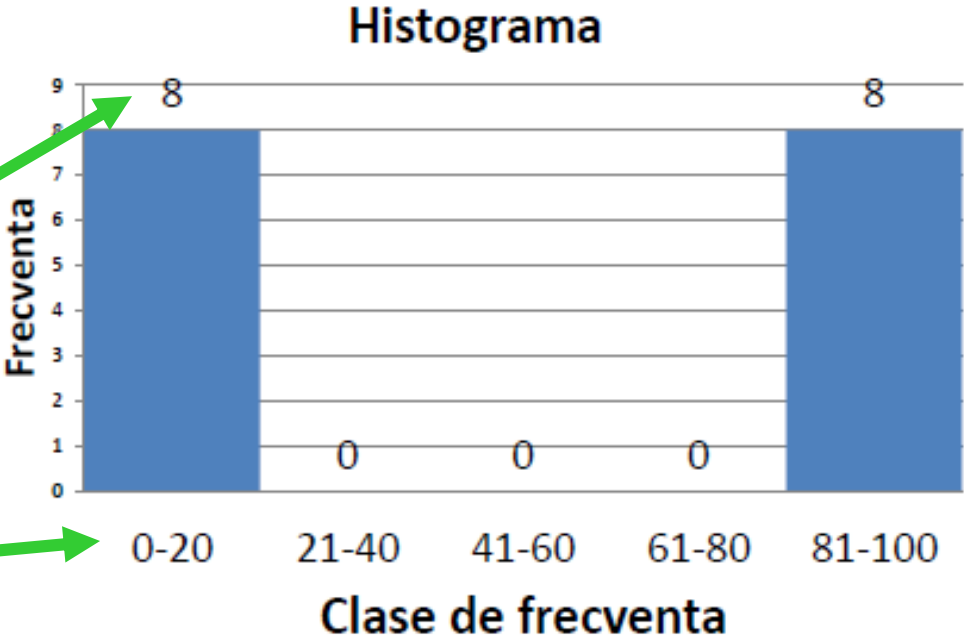
98

98

100

# Distribuția

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



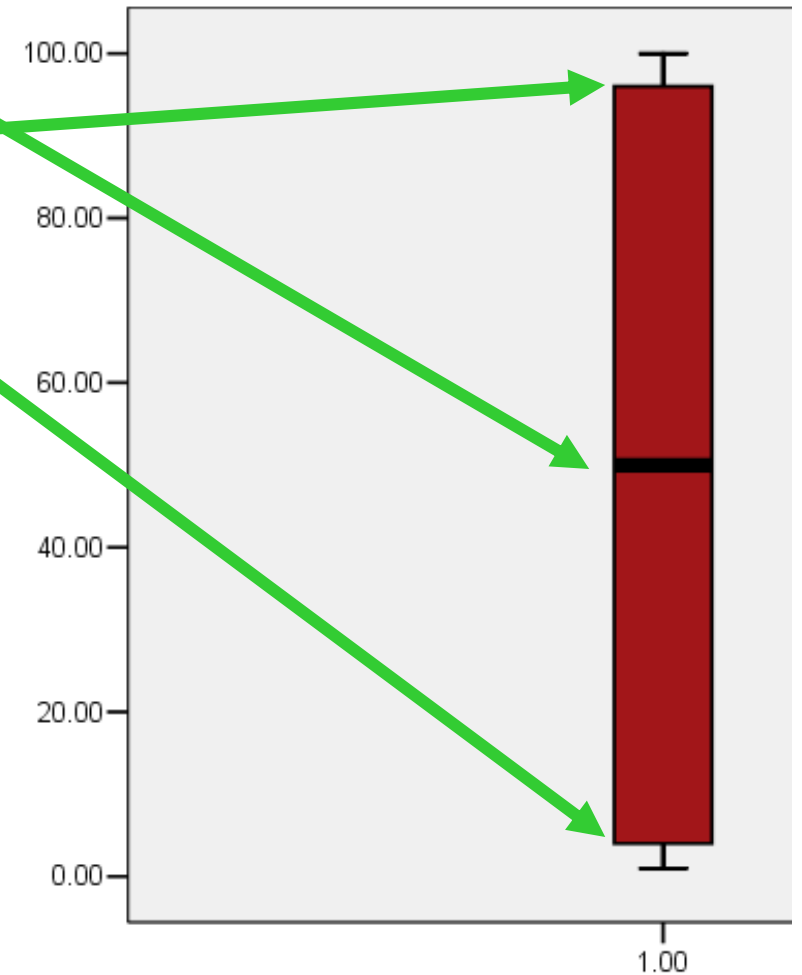
Seria 1
1
1
2
3
5
6
6
7
93
94
94
95
97
98
98
100

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

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

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

Aceeși serie, alt tip de grafic – cutie cu mustăți



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

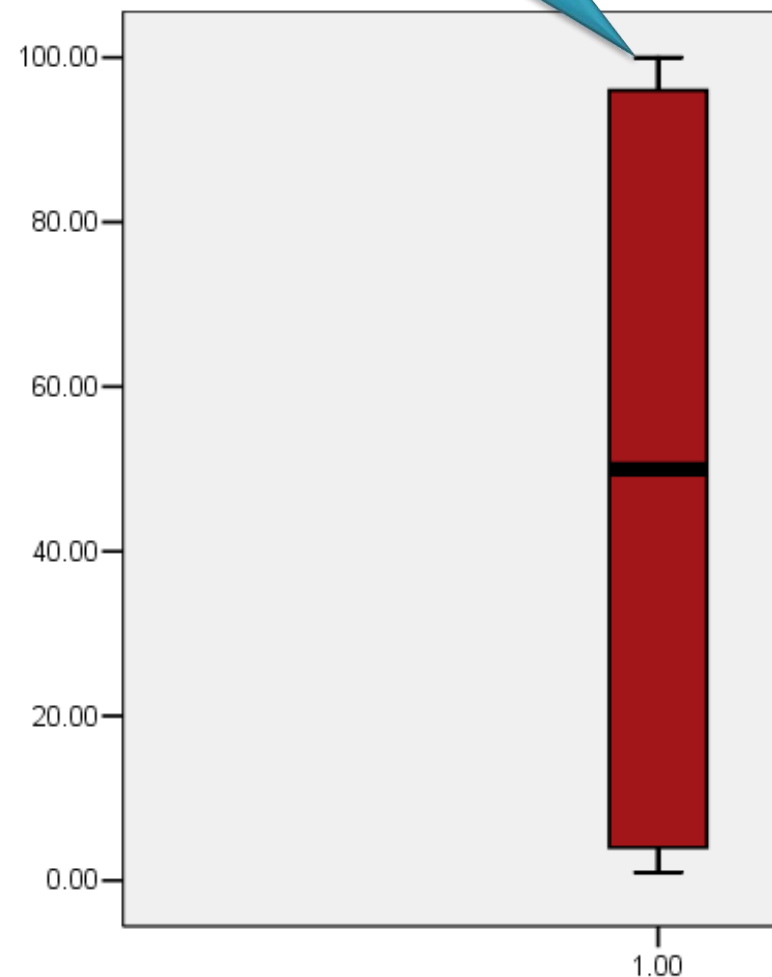
Mediana  $= (7+93)/2 = 50$

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

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

Aceeași serie, alt tip de grafic

Între minim și percentila 25 este o distanță mică ceea ce înseamnă că în acest interval avem multe date





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

Seria 2

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

## Distribuția

14

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

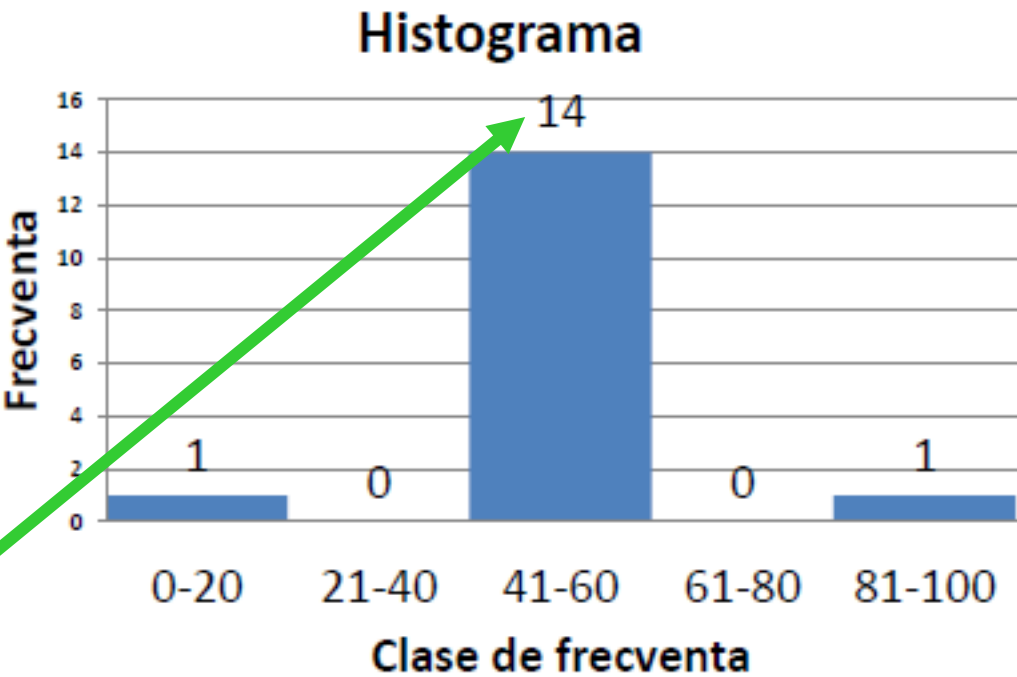
55

55

100

# 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



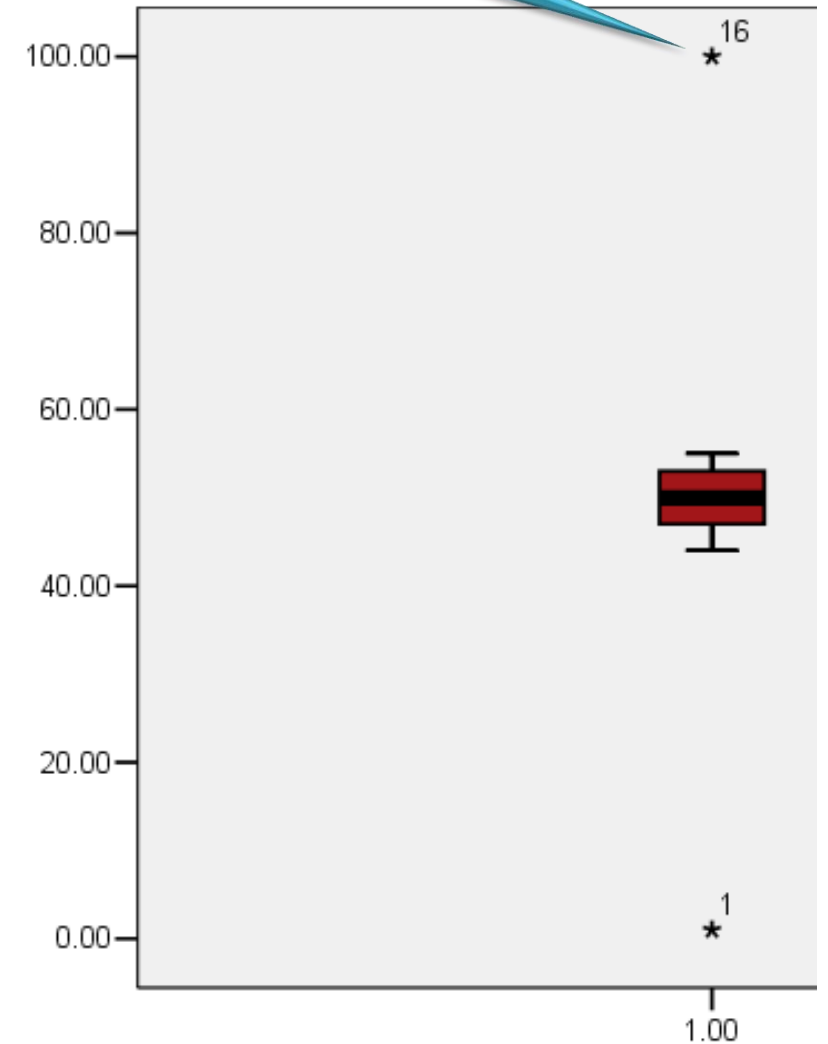
Seria 2
1
44
45
46
48
48
49
50
50
51
52
52
54
55
55
100

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

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

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

Caz extrem



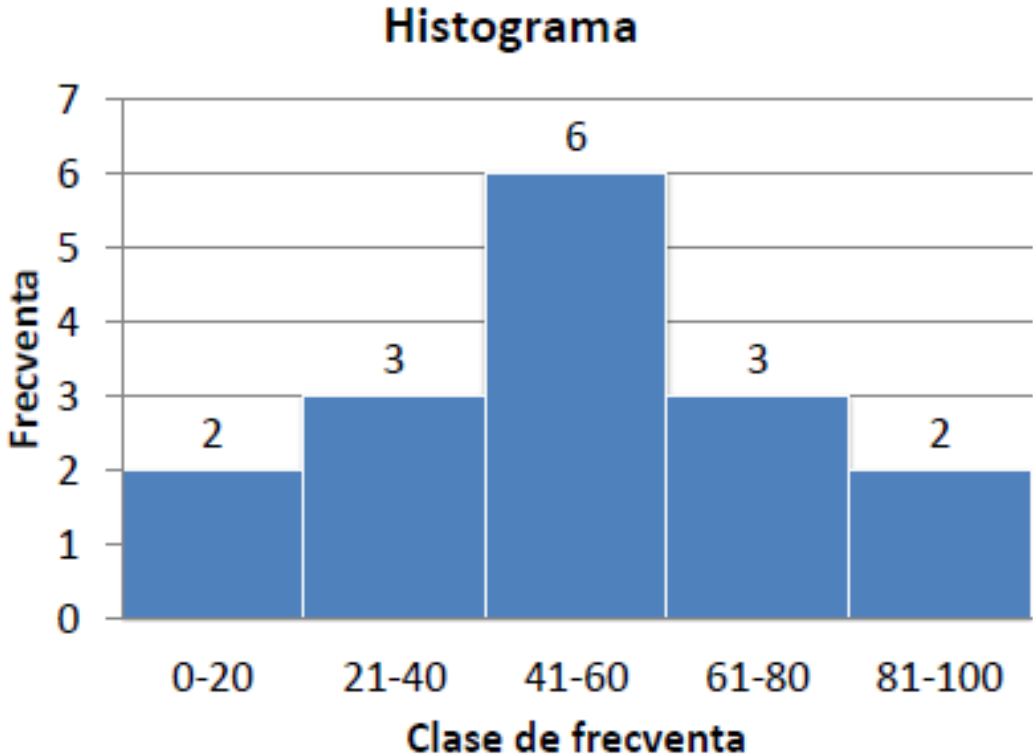
Seria 3
1
11
24
29
36
41
45
49
51
55
59
64
71
76
88
100

Seria 3

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

# Distribuția

Seria 2 – 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

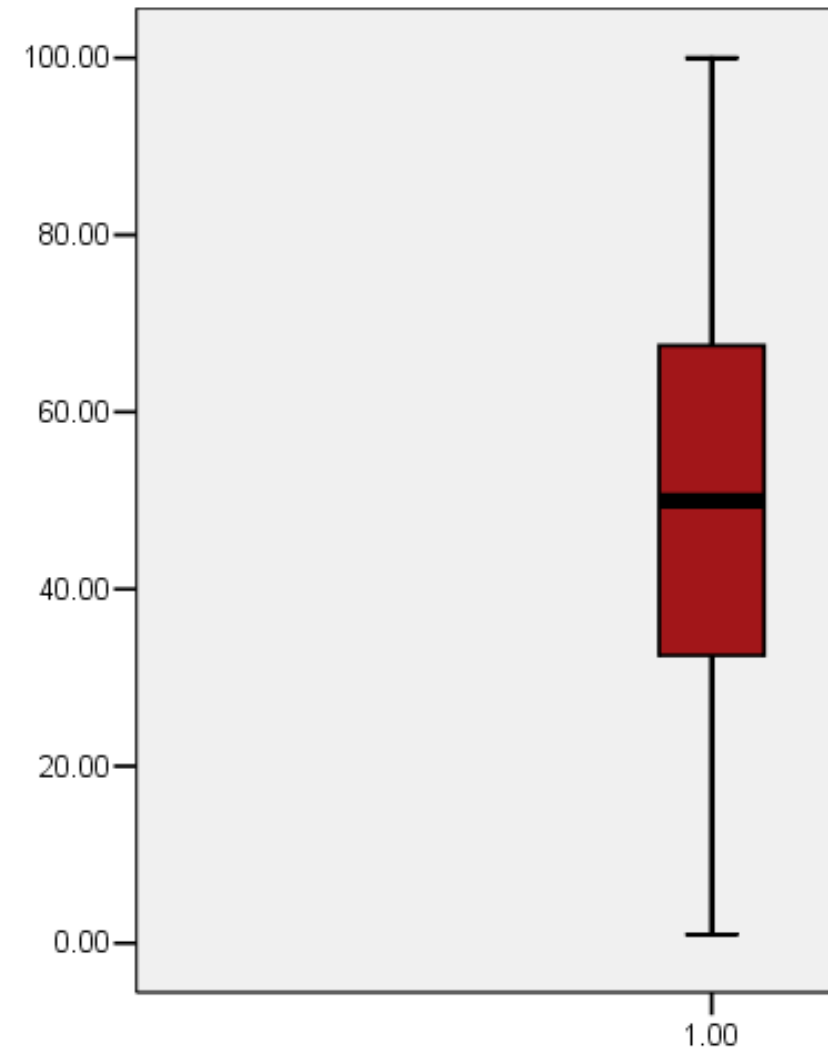


Seria 3
1
11
24
29
36
41
45
49
51
55
59
64
71
76
88
100

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

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

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



# Măsuri de dispersie

Seria 1	Seria 2	Seria 3
1	1	1
1	44	8
2	45	11
3	46	14
5	48	28
6	48	30
6	49	37
7	50	48
93	50	52
94	51	62
94	52	70
95	52	72
97	54	84
98	55	91
98	55	92
100	100	100
800	800	800

De ce avem nevoie de măsurile de dispersie când vrem să descriem datele?

Numărul de observații = 16

**Media** pt. fiecare set = 50

**Mediana** pt. fiecare set = 50



# Dispersie

- Nucile in cozonac (uniform dispersate sau aglomerate la baza?)
- Gemul pe paine (este raspandit pe toata felia?)
- Norii pe cer (ce grosime au si unde e aceasta maxima?)
- Conducta sparta (ce raspandire are apa?)
- Ce raspandire are populatia peste 65 de ani in Romania?
- Ce raspandire are obezitatea ( $IMC > 30$ ) in Romania?
- Ce raspandire are poluarea cu praf in Cluj-Napoca? Care sunt zonele cele mai putin poluate?

# Măsuri de dispersie

Deviația standard – Media deviației de la medie

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

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

*$\Sigma$  înseamnă adunare,  $X$  reprezintă observațiile individuale,  $\mu$  este media aritmetică a întregii populații,  $N$  este numărul de observații.*

# Formula pentru un subset al populației (eșantion)

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

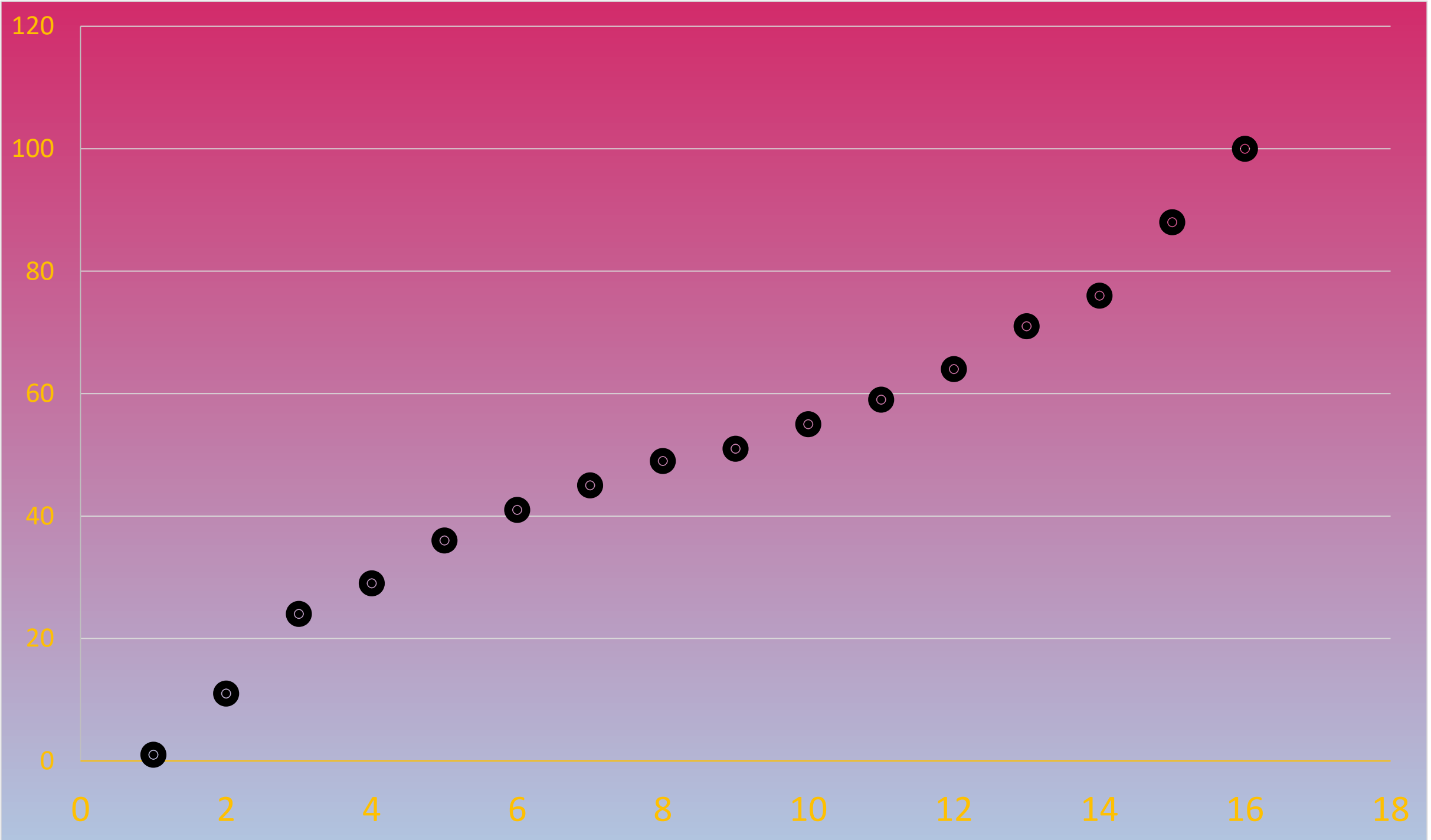
$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N-1}} = \sqrt{\frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + \dots + (x_N - \bar{X})^2}{N-1}}$$

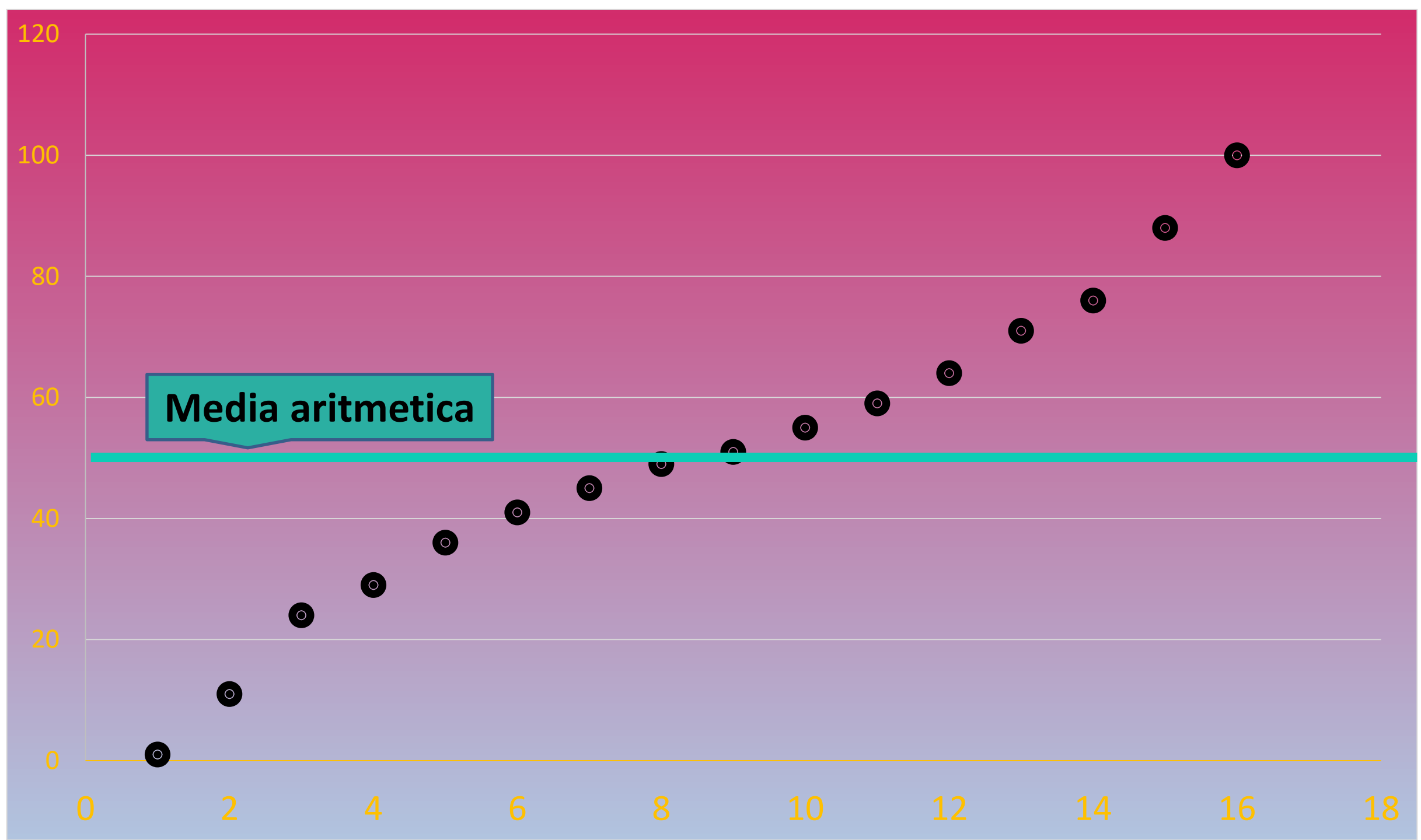
*unde*

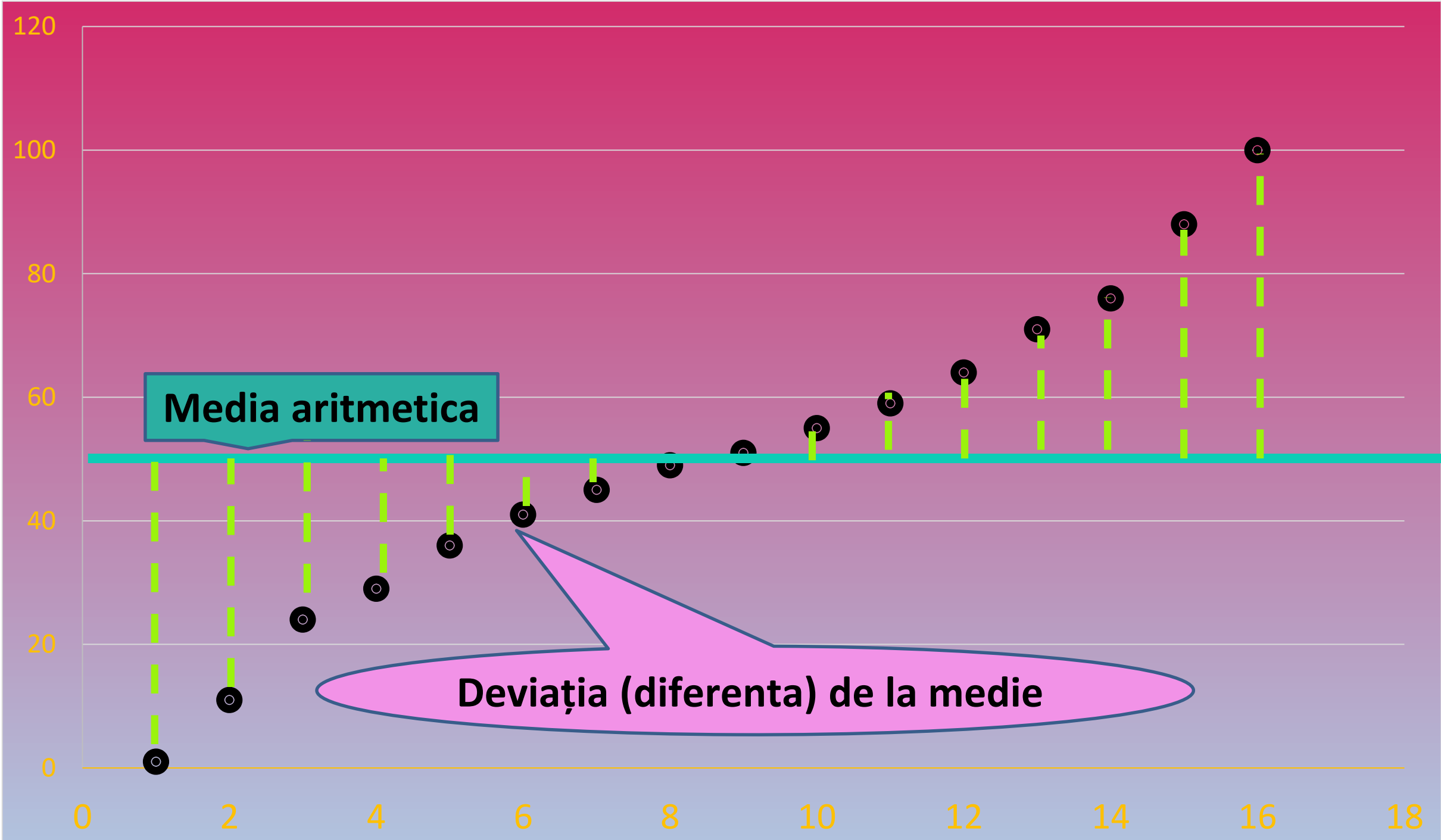
$N$  – numărul total de observații

$\bar{X}$  - media aritmetică

$x_1, \dots, x_N$  - observațiile





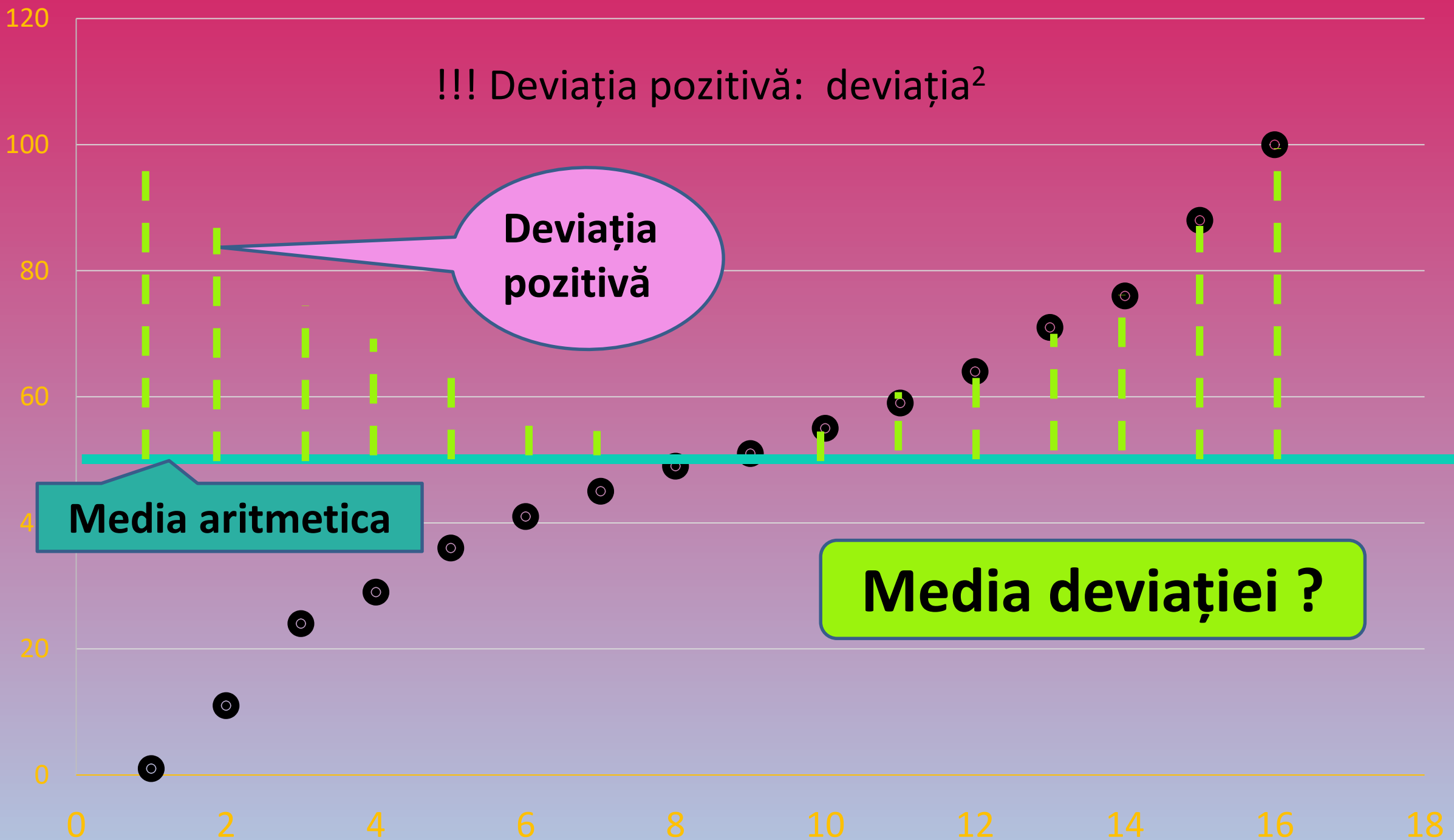


!!! Deviația pozitivă: deviația<sup>2</sup>

Deviația  
pozitivă

Media aritmetica

Media deviației ?



Media deviației





Defapt în loc de linii sunt pătrate ( $\text{deviația}^2$ )



# 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:

$$S = \sqrt{\frac{(20-25)^2 + (24-25)^2 + (26-25)^2 + (25-25)^2 + (25-25)^2 + (30-25)^2}{6-1}} =$$
$$= \sqrt{\frac{(-5)^2 + (-1)^2 + 1^2 + 0^2 + 0^2 + 5^2}{5}} = \sqrt{\frac{25+1+1+25}{5}} = \sqrt{\frac{52}{5}} = 10,4$$

Deviația  
pozitivă de  
la medie

Media deviației de la medie :-P

# Exemplu

Ex. Vârsta pe un eșantion de 6 persoane: 24, 25, 29, 29, 30, 31 ani.

$X = 24, 25, 29, 29, 30, 31$  – datele individuale,  $\bar{X} = 28$

1.  $X - \bar{X} = -4, -3, 1, 1, 2, 3.$

2.  $(X - \bar{X})^2 = 16, 9, 1, 1, 4, 9.$

3.  $\Sigma(X - \bar{X})^2 = 16+9+1+1+4+9 = 40$

4.  $\Sigma(X - \bar{X})^2 / (n-1) = 40/5 = 8.$

5.  $\sqrt{\Sigma(X - \bar{X})^2 / (n-1)} = 2,83$  Această valoare este deviația standard

# Deviația standard - exemplu

Colesterolul pentru 10 pacienți: 200, 180, 140, 160, 180, 150, 170, 110, 230, 170

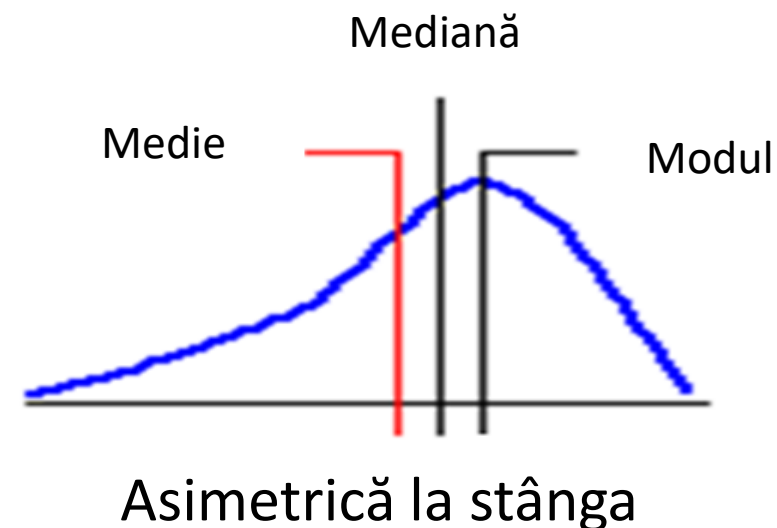
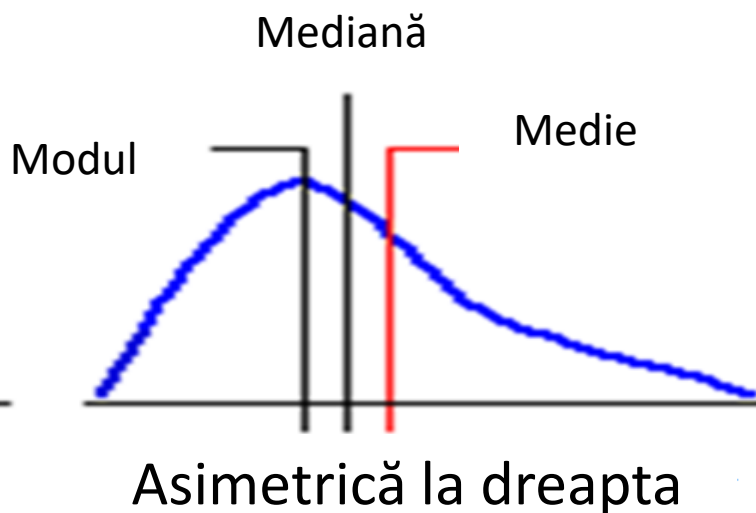
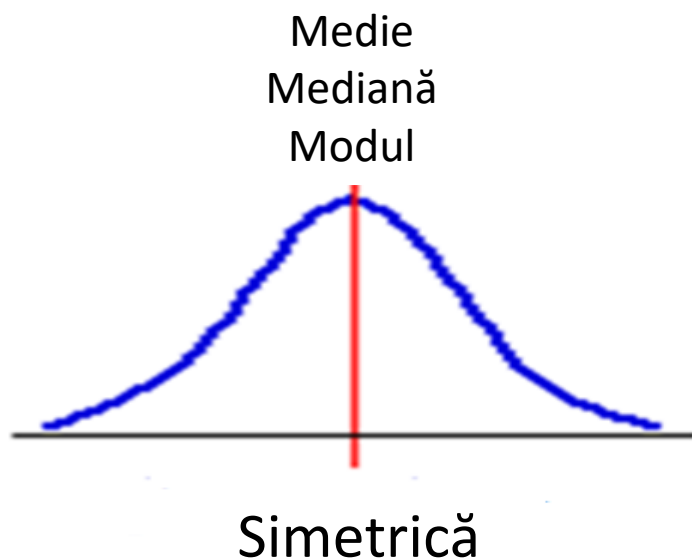
	Colesterol ( $X$ )	$X - \bar{X}$	$(X - \bar{X})^2$
	200	31	961
	180	11	121
	140	-29	841
	160	-9	81
	180	11	121
	150	-19	361
	170	1	1
	110	-59	3481
	230	61	3721
	170	1	1
Suma	1690	0	9690
$\bar{X}$	169	$(X - \bar{X})^2 / (n-1)$	1076.6
		Dev.St.	32.8

Deviația de la medie

Deviația pozitivă de la medie

# Măsuri de centralitate - aplicabilitate

- Cum să interpretăm media, mediana și modulul:



# Asimetria

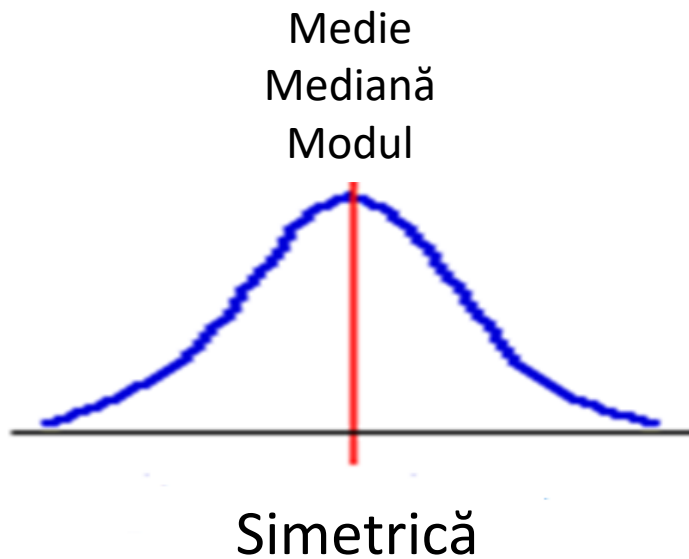
## Asimetria – Asimetria datelor

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

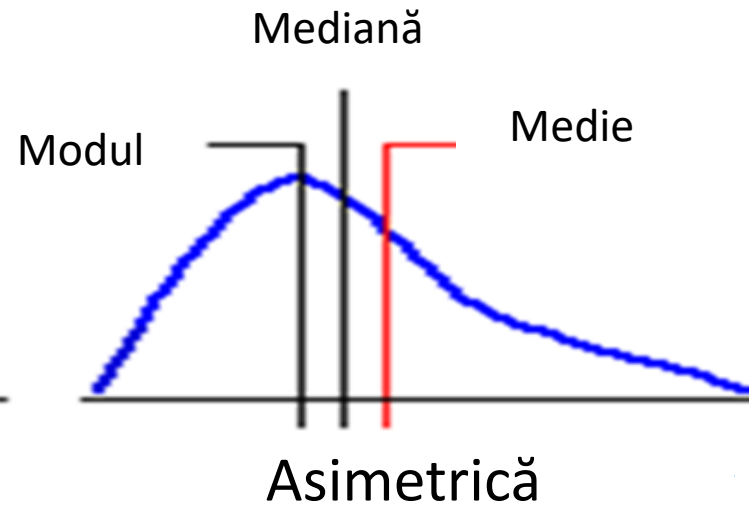
*Unde  $s$  – deviația standard,  $n$  nr. de indivizi,  
 $X_1, \dots, X_n$  – valorile seriei,  $\bar{X}$  media aritmetică*

# Asimetria

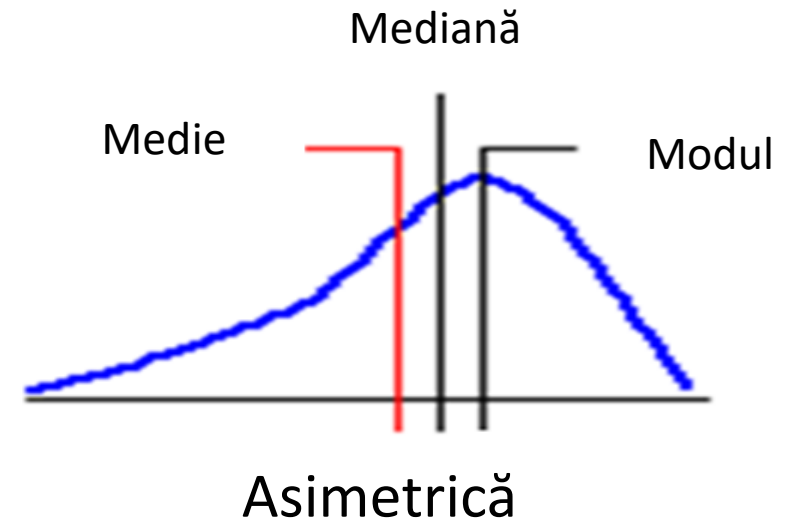
**Asimetria** este o măsură care descrie distribuția unei variabile cantitative continue



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



$\text{Simetria} > 1$   
asimetrică



$\text{Simetria} < -1$   
asimetrică

# Boltirea

## Boltirea – Platitudinea

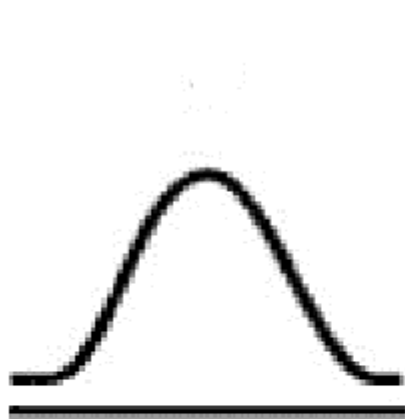
$$K_E \approx \frac{1}{n} \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{s^4} - 3$$

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

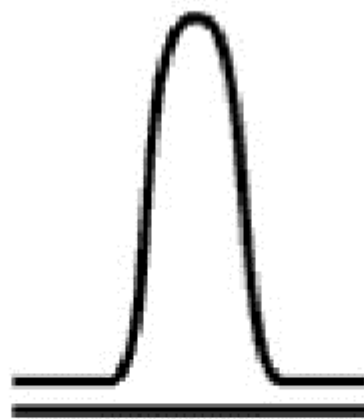


# Boltirea (kurtosis)

**Boltirea** este o măsură care descrie distribuția unei variabile cantitative continue



Mesocurtică  
Boltirea  $\approx 0$



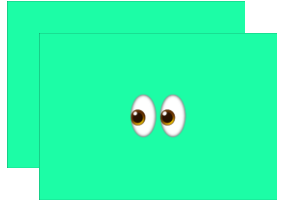
Leptocurtică  
Boltirea  $> 1$



Platicurtică  
Boltirea  $< -1$



# întrebări



<https://app.wooclap.com/BFKRI03?from=event-page>

# 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

# Date despre 1099 de pacienti cu Coronavirus internati in China in ianuarie 2020

## Clinical Characteristics of Coronavirus Disease 2019 in China

Wei-jie Guan, Ph.D., Zheng-yi Ni, M.D., Yu Hu, M.D., Wen-hua Liang, Ph.D., Chun-quan Ou, Ph.D., Jian-xing He, M.D., Lei Liu, M.D., Hong Shan, M.D., Chun-liang Lei, M.D., David S.C. Hui, M.D., Bin Du, M.D., Lan-juan Li, M.D., [et al.](#), for the China Medical Treatment Expert Group for Covid-19\*

[Article](#) [Figures/Media](#)

[Metrics](#)

[24 References](#) [12 Citing Articles](#)

### Abstract

#### BACKGROUND

Since December 2019, when coronavirus disease 2019 (Covid-19) emerged in Wuhan city and rapidly spread throughout China, data have been needed on the clinical characteristics of the affected patients.

#### METHODS

We extracted data regarding 1099 patients with laboratory-confirmed Covid-19 from 552 hospitals in 30 provinces, autonomous regions, and municipalities in mainland China through January 29, 2020. The primary composite end point was admission to an intensive care unit (ICU), the use of mechanical ventilation, or death.

February 28, 2020

DOI: [10.1056/NEJMoa2002032](#)

[Chinese Translation](#) [中文翻译](#)

#### Related Articles

**EDITORIAL** FEB 28, 2020

Covid-19 — Navigating the Uncharted

A.S. Fauci and Others

**NEJM**  
CareerCenter

**PHYSICIAN JOBS**

**MARCH 11, 2020**

Activate Windows

**Table 1.** Clinical Characteristics of the Study Patients, According to Disease Severity and the Presence or Absence of the Primary Composite End Point.\*

Characteristic	All Patients (N = 1099)	Disease Severity		Presence of Primary Composite End Point†	
		Nonsevere (N = 926)	Severe (N = 173)	Yes (N = 67)	No (N = 1032)
Age					
Median (IQR) — yr	47.0 (35.0–58.0)	45.0 (34.0–57.0)	52.0 (40.0–65.0)	63.0 (53.0–71.0)	46.0 (35.0–57.0)
Distribution — no./total no. (%)					
0–14 yr	9/1011 (0.9)	8/848 (0.9)	1/163 (0.6)	0	9/946 (1.0)
15–49 yr	557/1011 (55.1)	490/848 (57.8)	67/163 (41.1)	12/65 (18.5)	545/946 (57.6)
50–64 yr	292/1011 (28.9)	241/848 (28.4)			16 (28.6)
≥65 yr	153/1011 (15.1)	109/848 (12.9)			16 (12.8)

45.0 (34.0–57.0)

52.0 (40.0–65.0)

Date despre 1099 de pacienti cu Coronavirus internati in China in ianuarie 2020

**Table 1.** Clinical Characteristics of the Study Patients, According to Disease Severity and the Presence or Absence of the Primary Composite End Point.\*

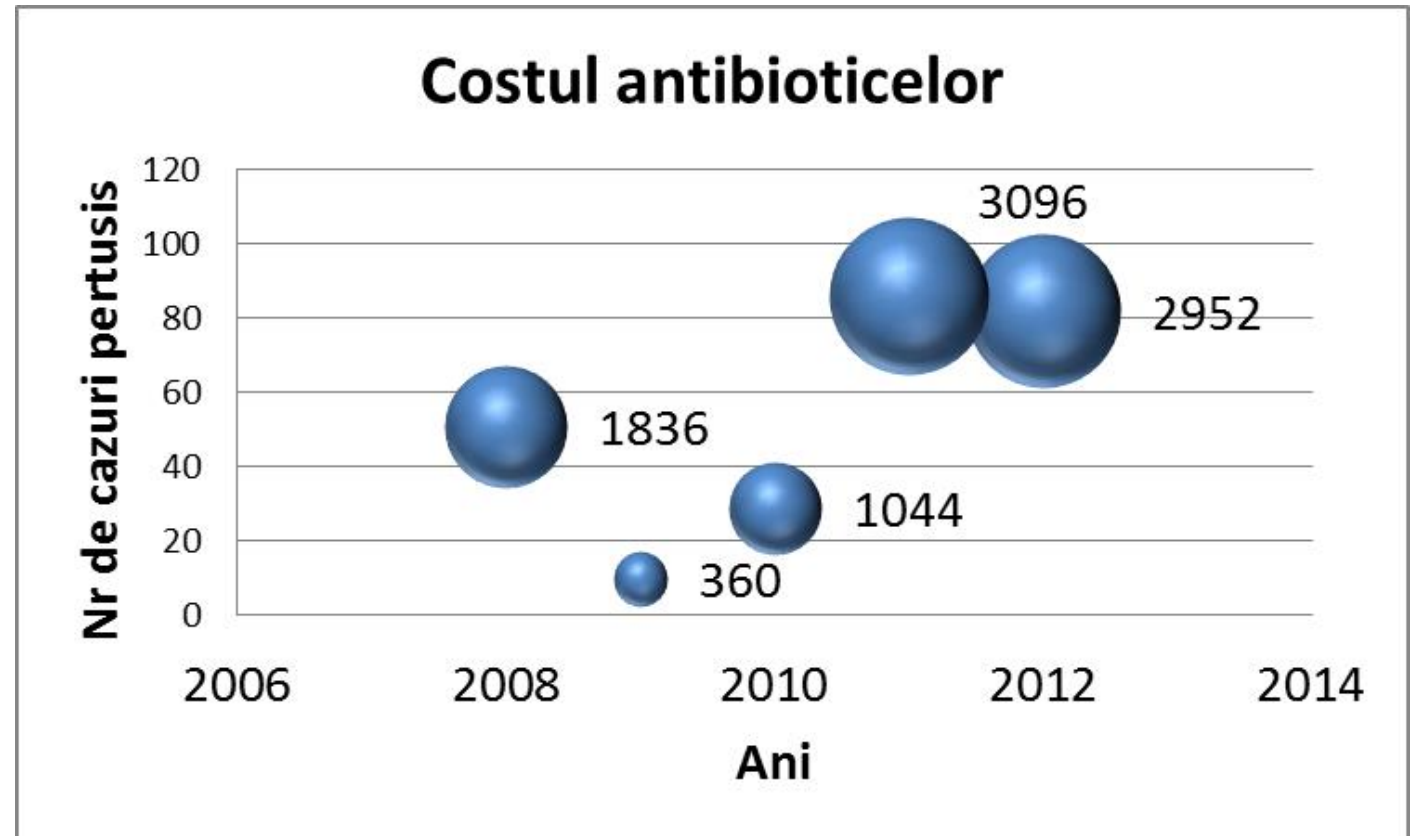
Characteristic	All Patients (N = 1099)	Disease Severity		Presence of Primary Composite End Point†	
		Nonsevere (N = 926)	Severe (N = 173)	Yes (N = 67)	No (N = 1032)
Median incubation period (IQR) — days‡	4.0 (2.0–7.0)	4.0 (2.8–7.0)	4.0 (2.0–7.0)	4.0 (1.0–7.5)	4.0 (2.0–7.0)
Fever on admission					
Patients — no./total no. (%)	473/1081 (43.8)	391/910 (43.0)	82/171 (48.0)	24/66 (36.4)	449/1015 (44.2)
Median temperature (IQR) — °C	37.3 (36.7–38.0)	37.3 (36.7–38.0)	37.4 (36.7–38.1)	36.8 (36.3–37.8)	37.3 (36.7–38.0)

Median incubation period (IQR) — days‡

4.0 (2.0–7.0)

# Exemple de grafice ... mai puțin bune

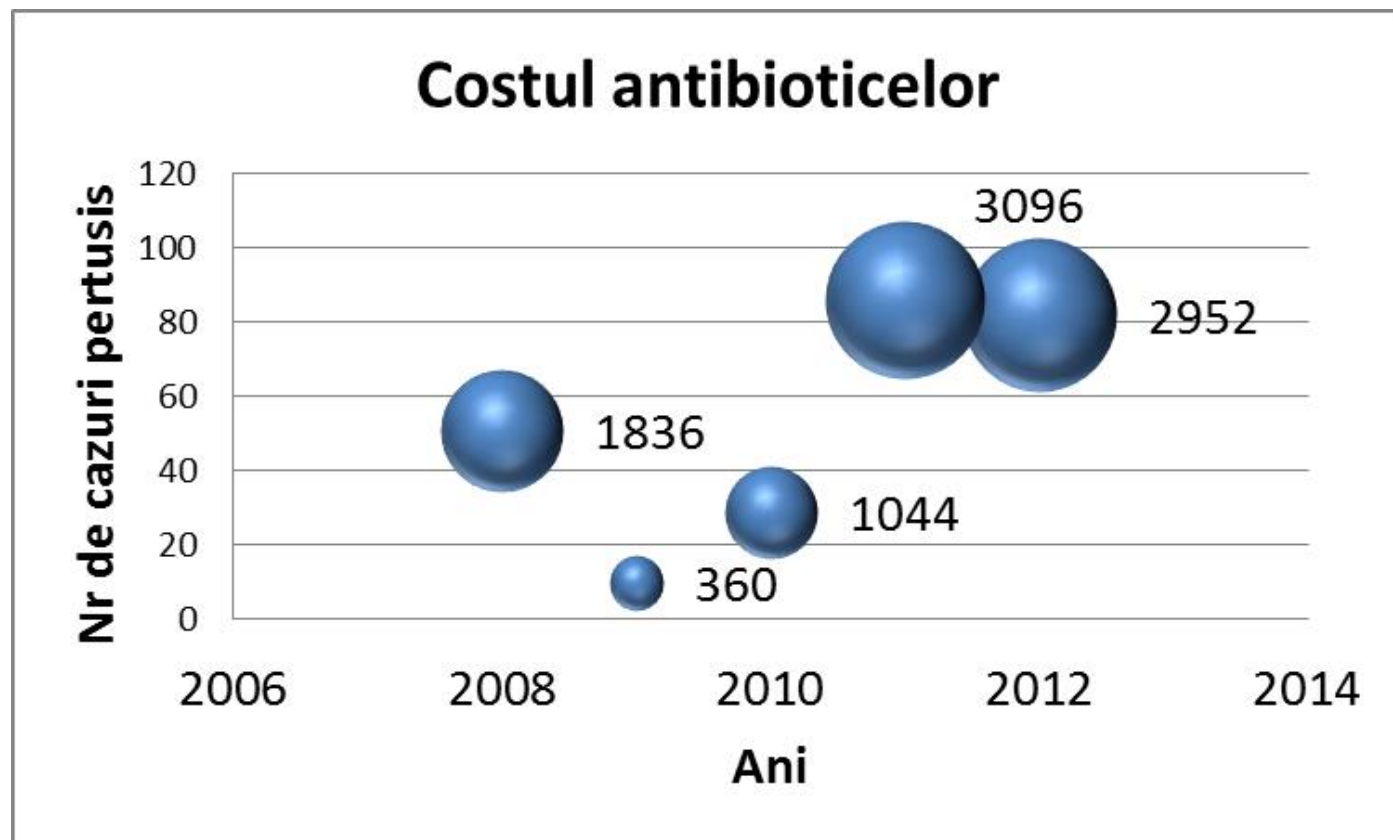
Un grafic trebuie sa fie simplu!



# Exemple de grafice ... care induc in eroare

Un grafic trebuie sa fie simplu!

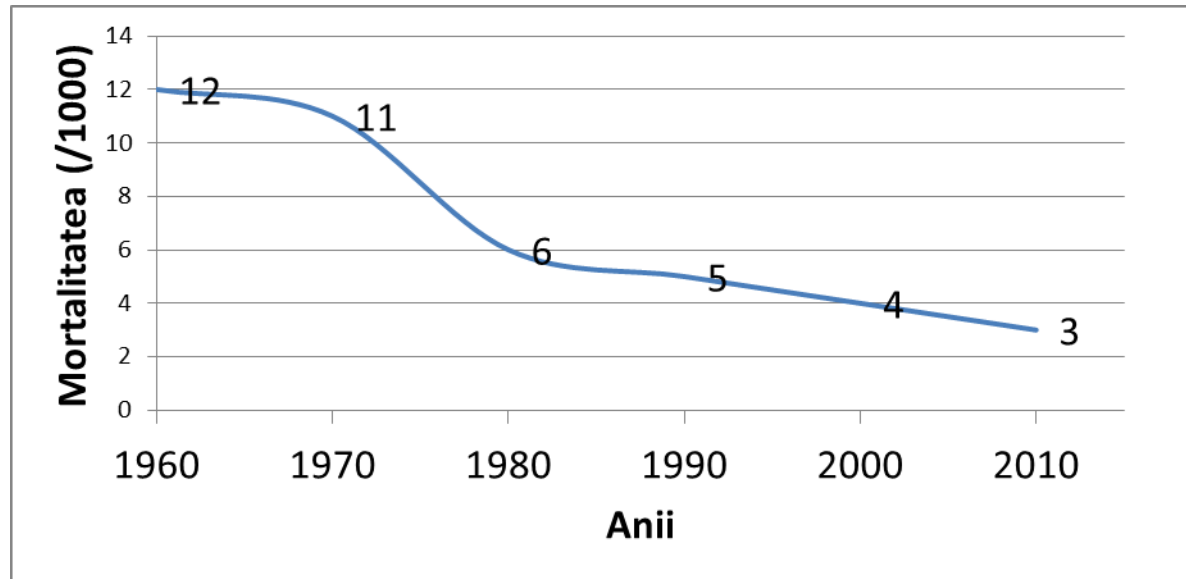
An	Pertussis nr. de cazuri	Costul mediu al antibioticelor
2012	82	2952
2011	86	3096
2010	29	1044
2009	10	360
2008	51	1836



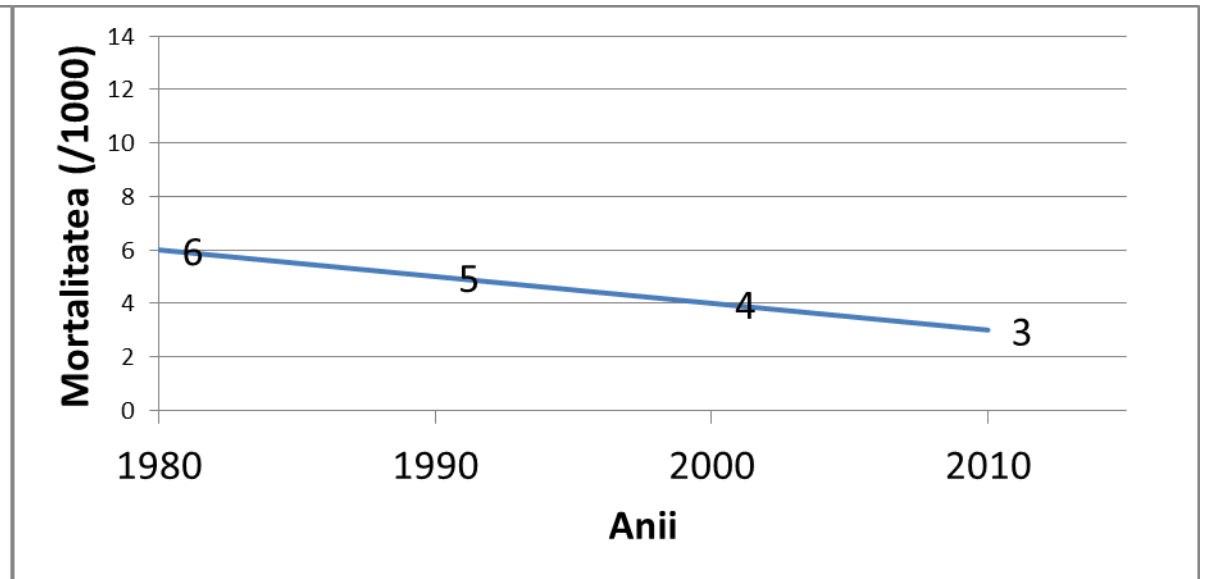


# Exemple de grafice care induc in eroare

Din ce an să reprezentăm datele?



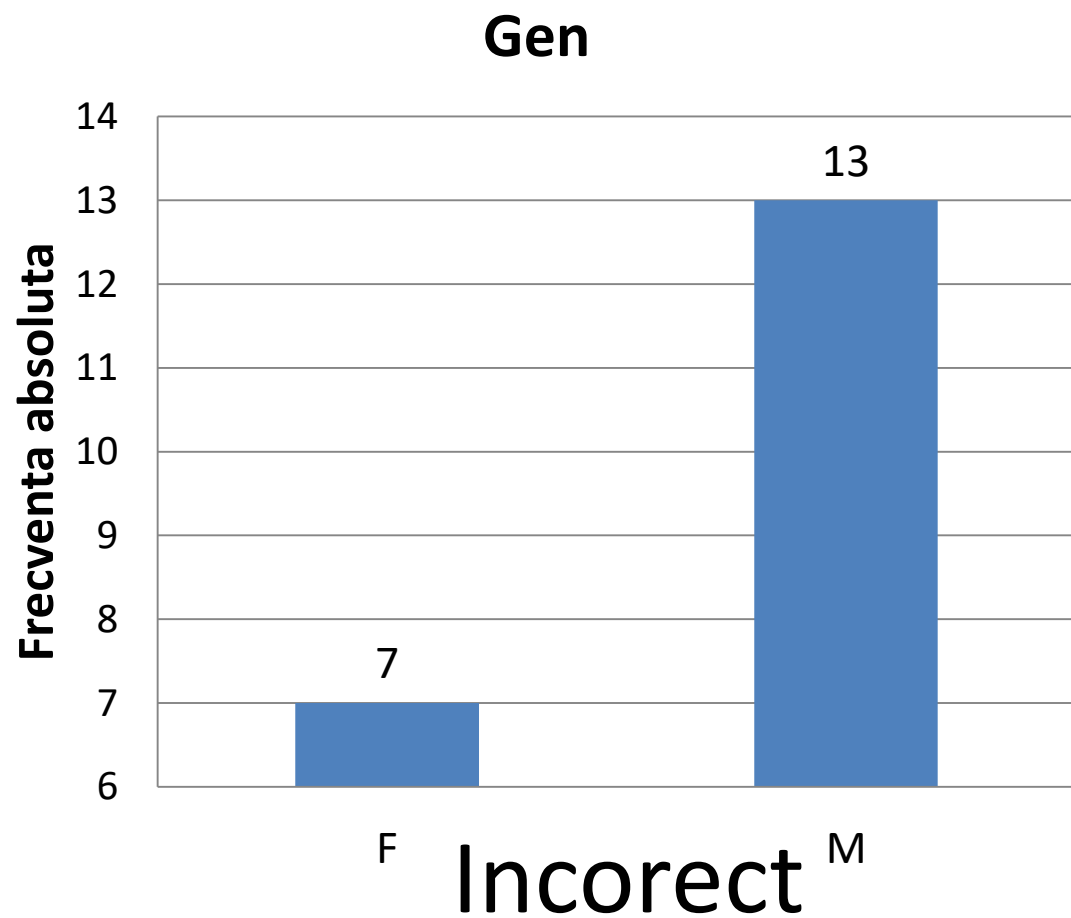
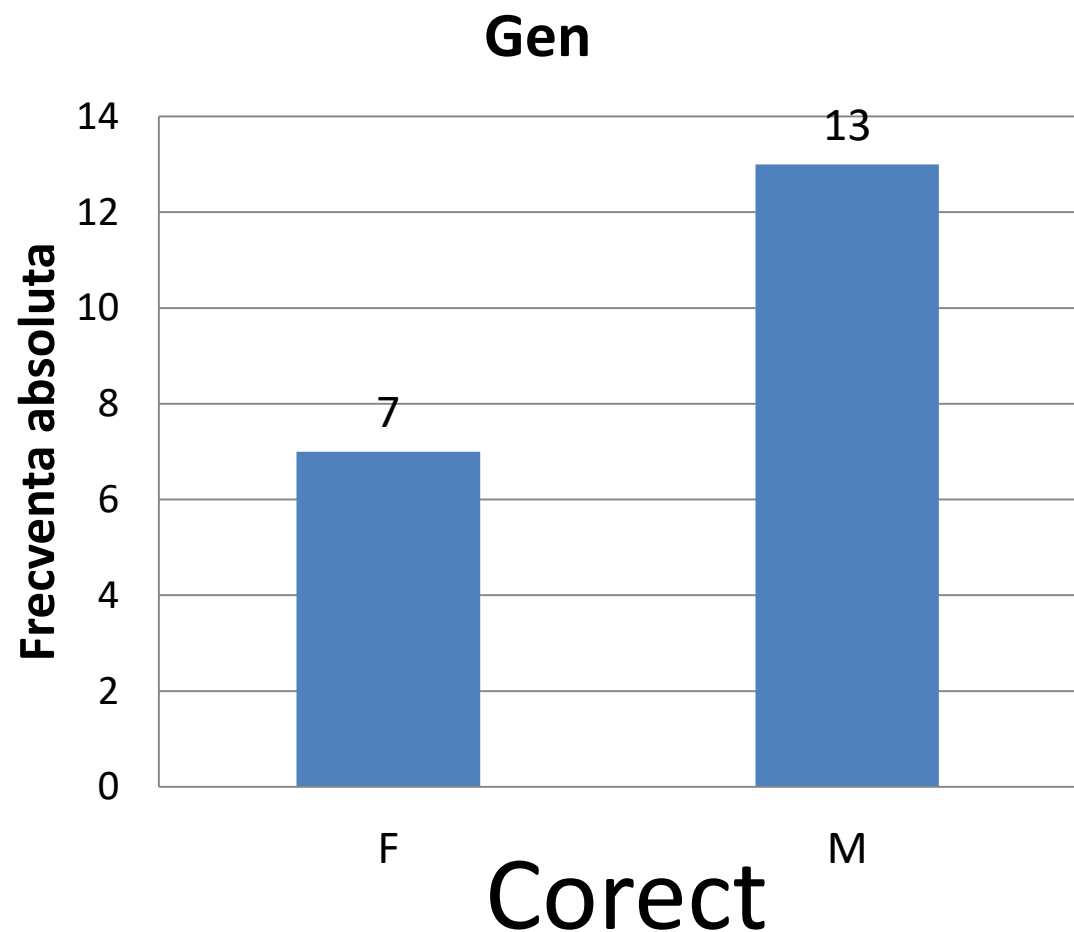
Corect



Incorect

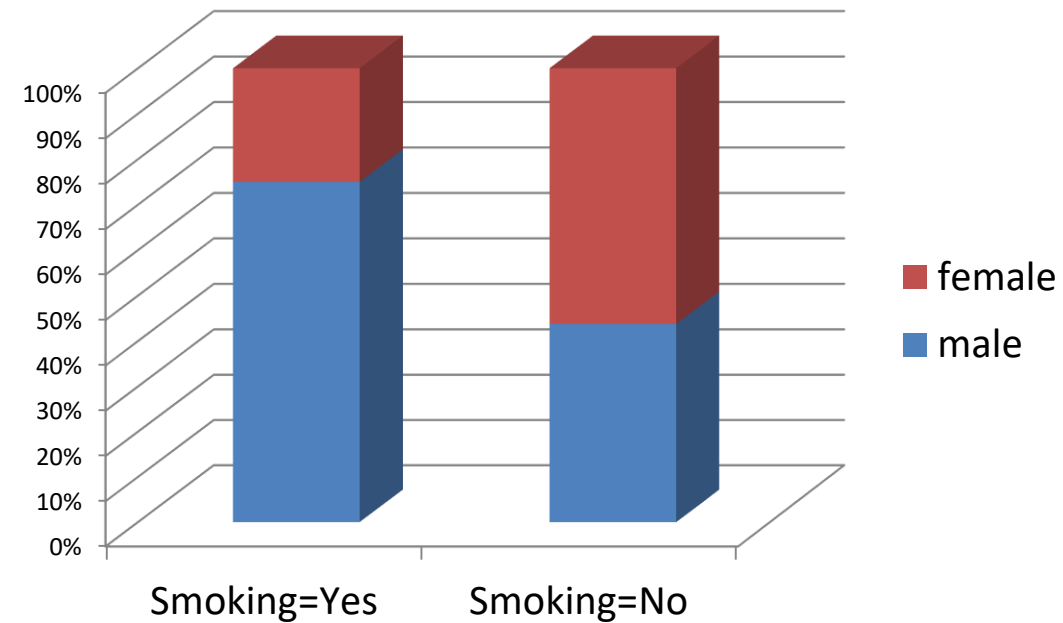
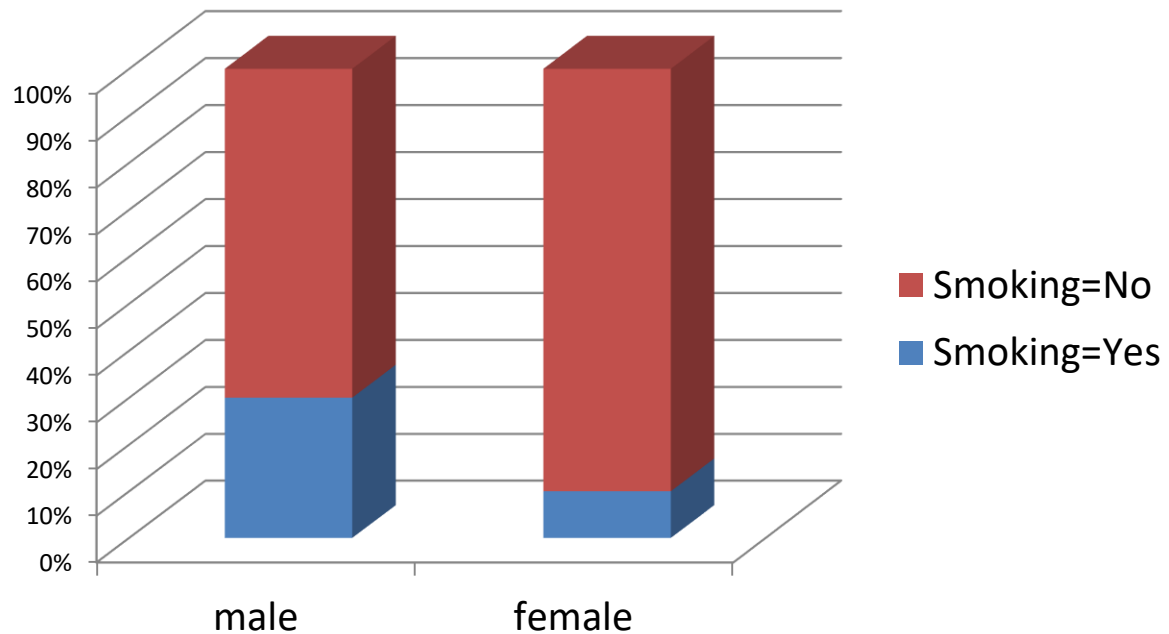
# Exemple de grafice ... care induc in eroare

Diferenta poate fi marita in mod fals daca nu incepem la 0



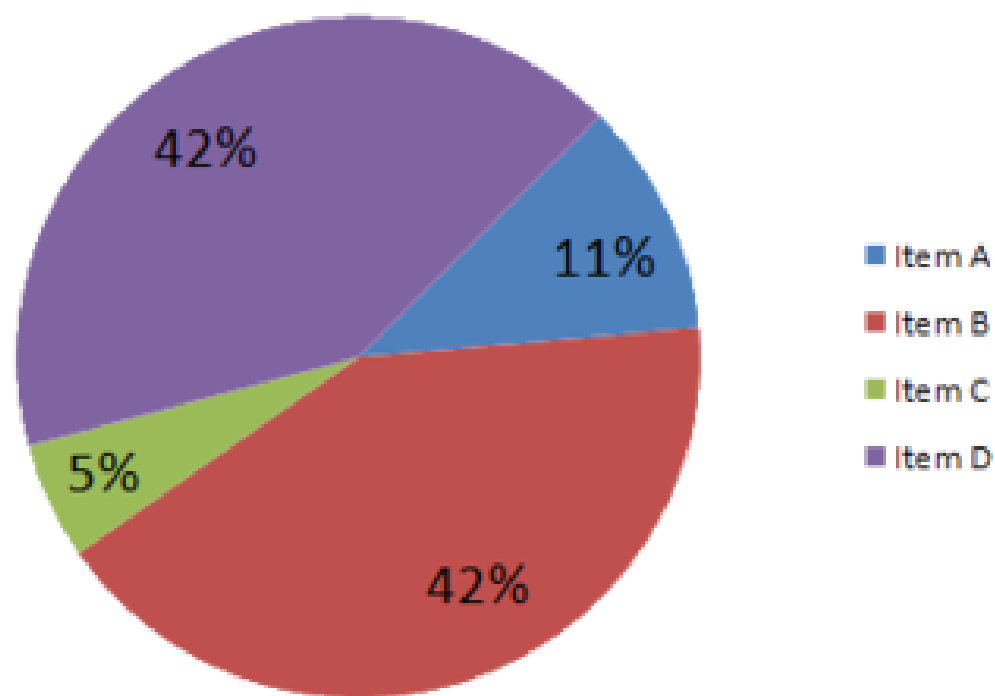
# Exemple de grafice ... care induc in eroare

Procente pe linii sau pe coloane?

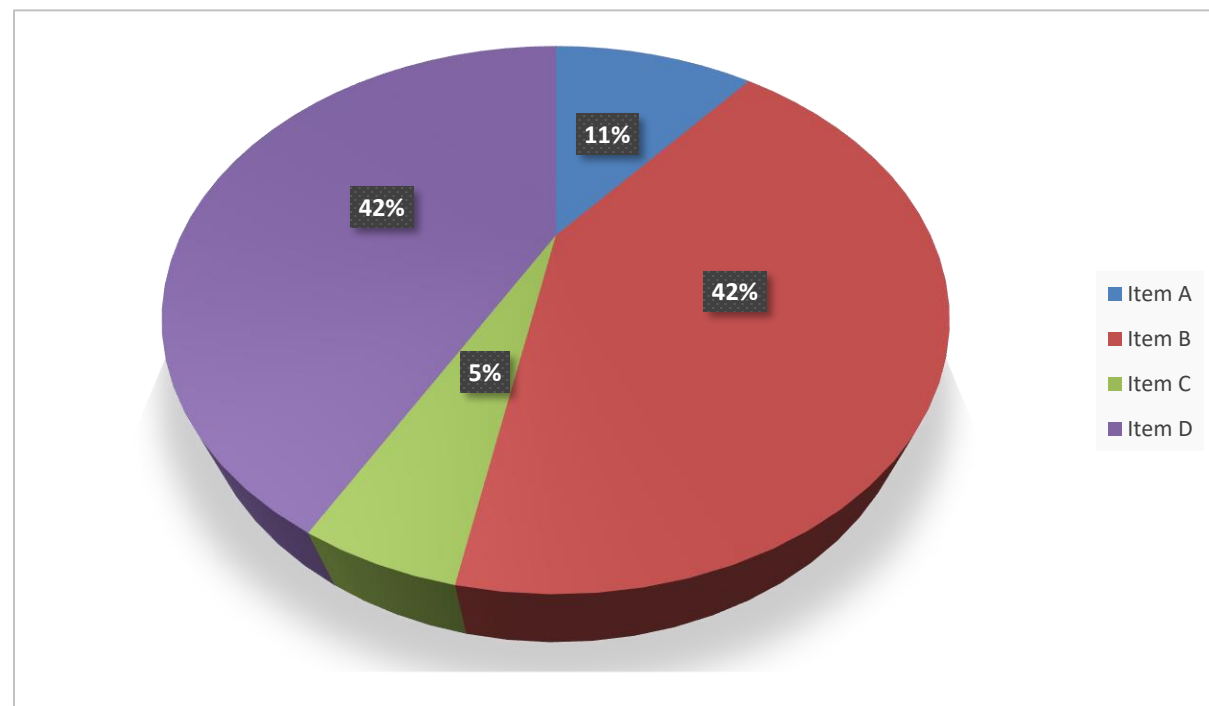


# Exemple de grafice ... care induc in eroare

## Graficele 3D

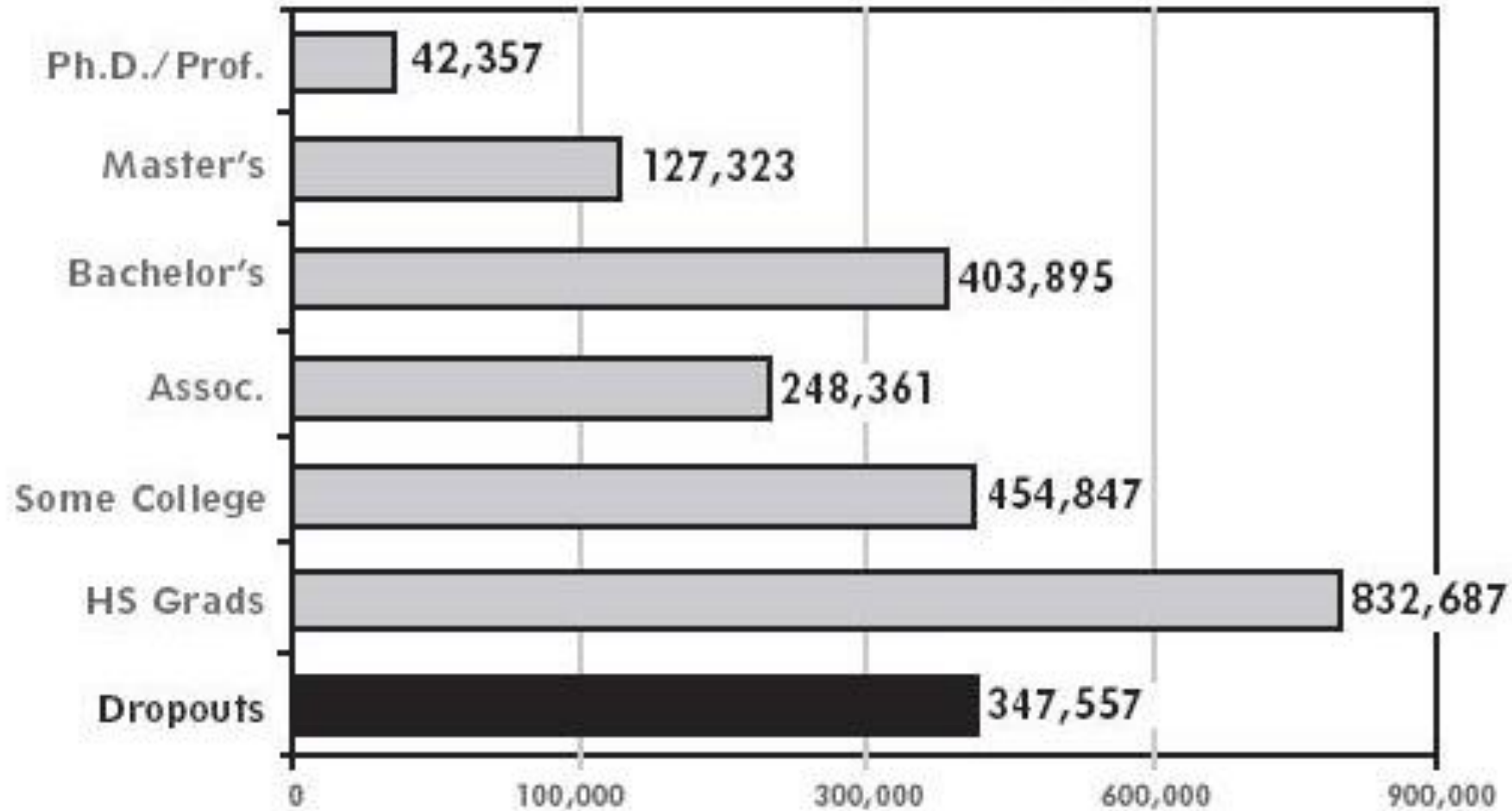


Corect

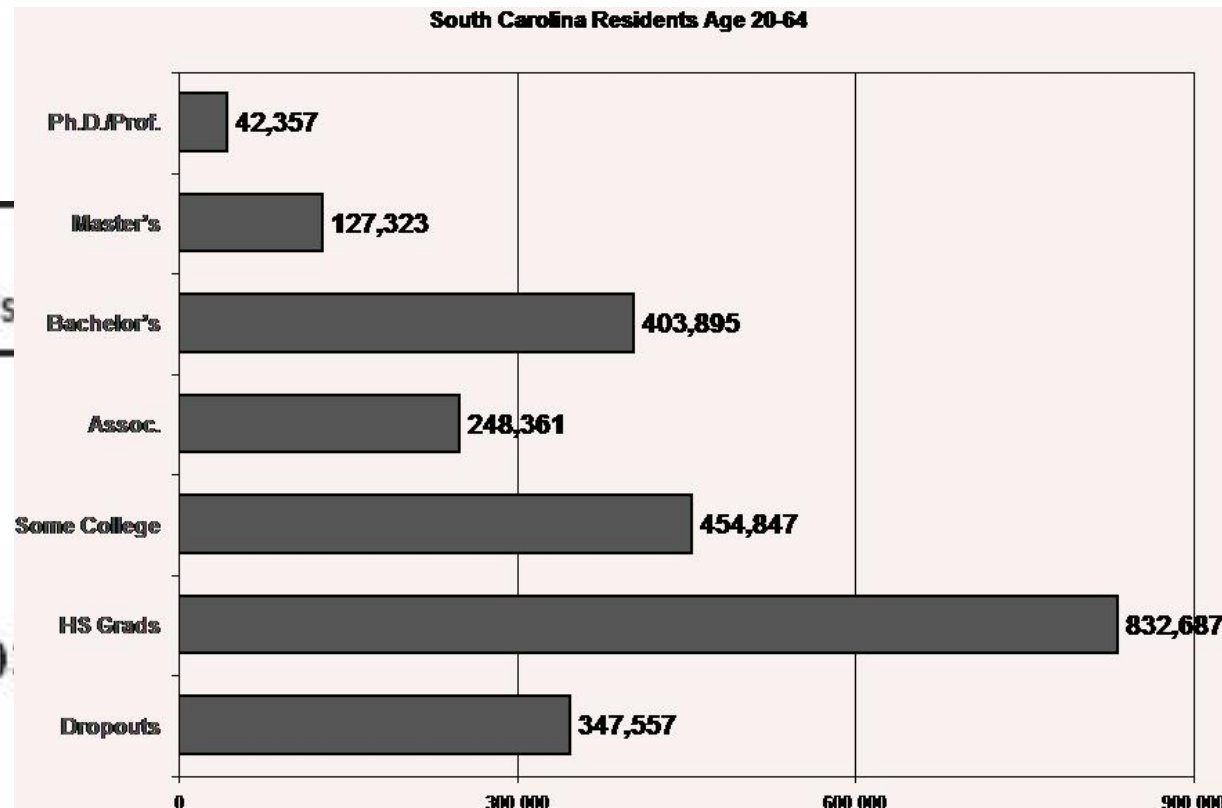
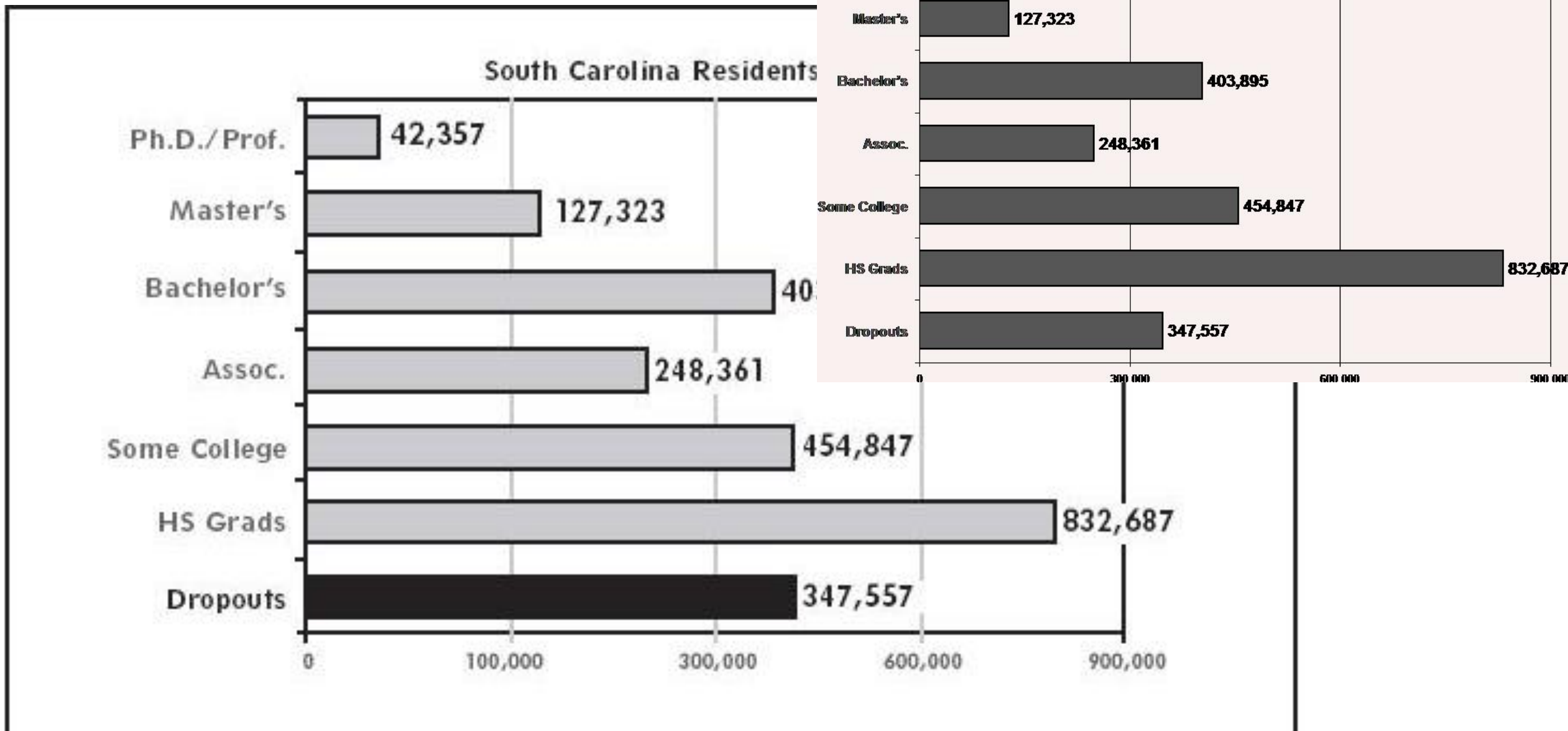


Incorrect

### South Carolina Residents Age 20-64



Source: PolEcon analysis of U.S. Census Bureau's Current Population Survey March 2004-06 Supplement data for South Carolina







Pestera lui Iosif- Muntii Buzau





Pestera lui Iosif- Muntii Buzau



Muțumesc!