



Autor: PhD. MsC. Bondor Cosmina-loana

## Statistică inferențială: intervale de confidență



ALWAYS



SEEK



KNOWLEDGE

# Objective

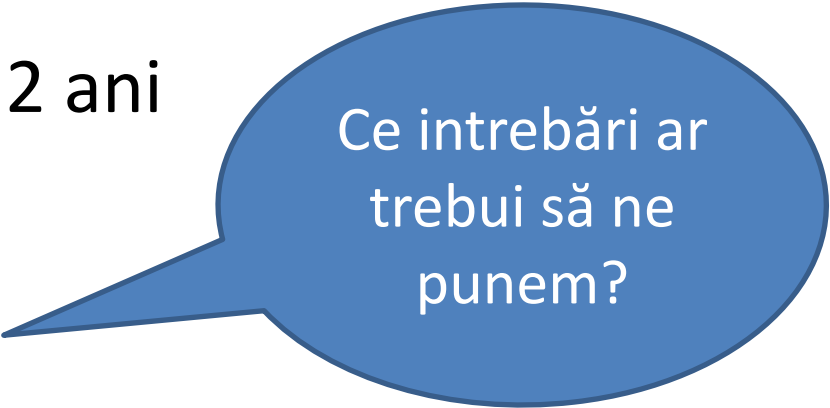
Distribuția de eșantionare

Deviația standard versus eroarea standard

Intervale de confidență

# Scenariu

- Vrem să caracterizăm greutatea băieților la 2 ani
  - Populația țintă?
  - Populația accesibilă?
  - Cât de mare ar trebui să fie eșantionul?
  - Cum selectăm pacienții?
- 
- Am inclus în studiu 200 de băieți de 2 ani aleși la întâmplare. Am făcut măsurarea greutății.



Ce întrebări ar  
trebui să ne  
punem?

# Scenariu

- Media greutății  $\bar{X} = 15\text{kg}$  (băieți)
- În populație media greutății  $= \mu$  (la băieții de 2 ani) este **necunoscută**

# Scenariu

- Cum am putea să găsim media greutateii băieților de 2 ani în întreaga populație?

A. Să-i luăm pe toți băieții de 2 ani (toată populația)

B. Să o aproximăm.

Cum?

Am putea să repetăm studiul.

- Repetăm studiul:

Prima dată  $\overline{X_1} = 15,75$  kg

A doua oară  $\overline{X_2} = 15,25$  kg

A treia oară  $\overline{X_2} = 15,5$  kg

..... = 16 kg

..... = 14,75 kg

..... = 15 kg

..... = 15,25 kg

..... = 14,5 kg

..... = 14,5 kg

..... .....

m time  $\overline{X_m} = 15$  kg

- $\bar{X}: \bar{X}_1, \bar{X}_2, \dots, \bar{X}_m$

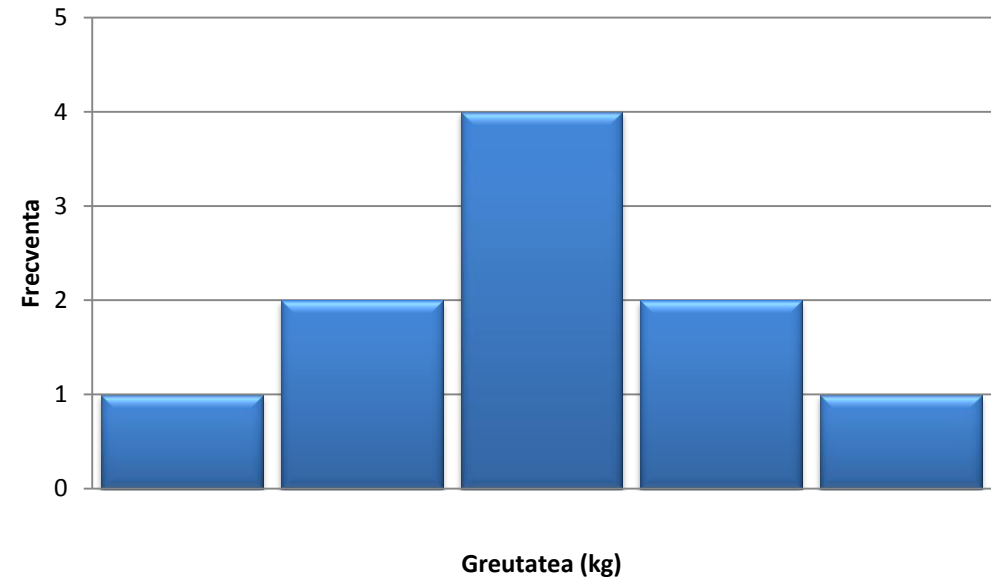
Aceste  $m$  medii alcătuiesc o altă distribuție: distribuția mediilor

# Scenariu

- Dacă repetăm studiul m eșantioane de băieți de 2 ani  $\bar{X}: \bar{X}_1, \bar{X}_2, \dots, \bar{X}_n$

Ex. 15,75, 15,25, 15,5, 16, 14,75, 15, 15,25, 14,5, 14,5, ..., 15 kg

Aceste m medii alcătuiesc o altă distribuție: **distribuția mediilor**



Va fi distribuită normal

- Populația 3 băieți
- Câte eșantioane de 2 băieți putem lua?



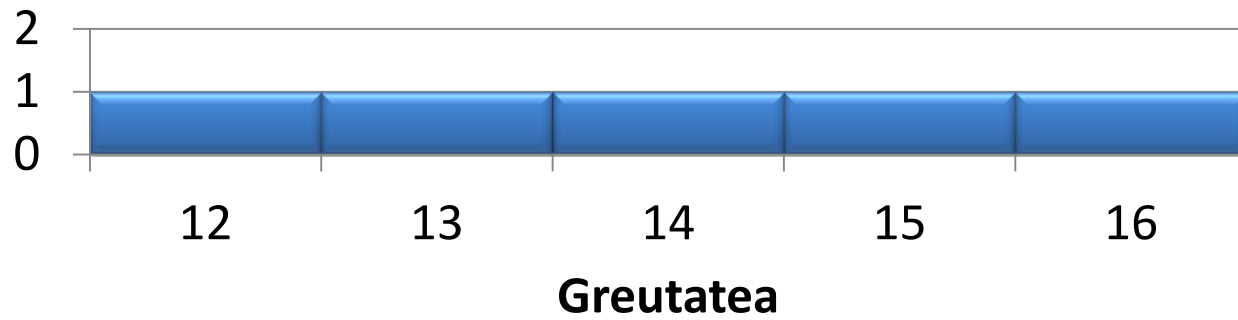
- Populația 3 băieți
- Câte eșantioane de 2 băieți putem lua?

# Distribuția de eșantionare

- Scenariu: 5 băieți de 2 ani cu greutatea: 12, 13, 14, 15, 16 kg
- Aceasta este întreaga populație
- Selectăm toate eșantioanele de 2 subiecți posibile:

| Primul băiat | Al doilea băiat | Greutatea pentru<br>primul băiat | Greutatea pentru<br>al doilea băiat | Media |
|--------------|-----------------|----------------------------------|-------------------------------------|-------|
| 1            | 1               | 12                               | 12                                  | 12    |
| 1            | 2               | 12                               | 13                                  | 12,5  |
| 1            | 3               | 12                               | 14                                  | 13    |
| 1            | 4               | 12                               | 15                                  | 13,5  |
| 1            | 5               | 12                               | 16                                  | 14    |
| 2            | 1               | 13                               | 12                                  | 12,5  |
| 2            | 2               | 13                               | 13                                  | 13    |
| 2            | 3               | 13                               | 14                                  | 13,5  |
| 2            | 4               | 13                               | 15                                  | 14    |
| 2            | 5               | 13                               | 16                                  | 14,5  |
| 3            | 1               | 14                               | 12                                  | 13    |
| 3            | 2               | 14                               | 13                                  | 13,5  |
| 3            | 3               | 14                               | 14                                  | 14    |
| 3            | 4               | 14                               | 15                                  | 14,5  |
| 3            | 5               | 14                               | 16                                  | 15    |

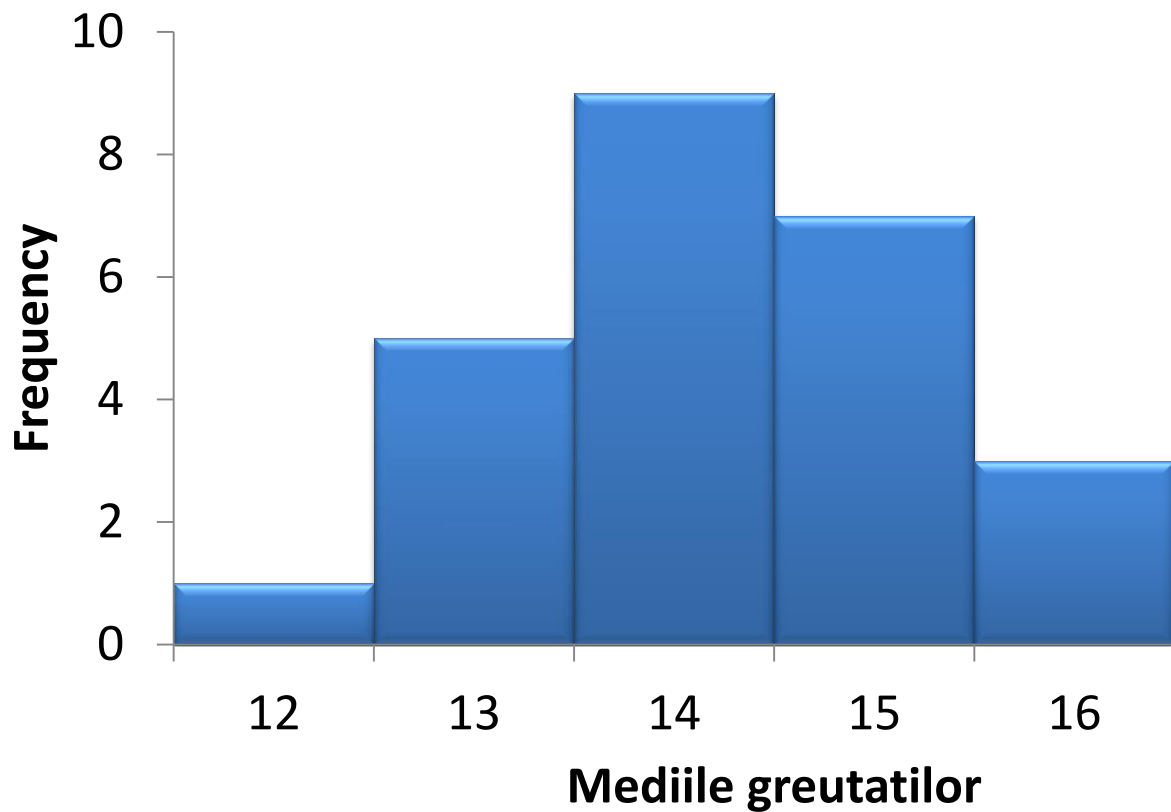
| Primul băiat | Al doilea băiat | Greutatea pentru<br>primul băiat | Greutatea pentru<br>al doilea băiat | Media |
|--------------|-----------------|----------------------------------|-------------------------------------|-------|
| 4            | 1               | 15                               | 12                                  | 13,5  |
| 4            | 2               | 15                               | 13                                  | 14    |
| 4            | 3               | 15                               | 14                                  | 14,5  |
| 4            | 4               | 15                               | 15                                  | 15    |
| 4            | 5               | 15                               | 16                                  | 15,5  |
| 5            | 1               | 16                               | 12                                  | 14    |
| 5            | 2               | 16                               | 13                                  | 14,5  |
| 5            | 3               | 16                               | 14                                  | 15    |
| 5            | 4               | 16                               | 15                                  | 15,5  |
| 5            | 5               | 16                               | 16                                  | 16    |



În populație

Media greutății = 14 kg,

Deviația standard = 1,58



Distribuția de eșantionare (a mediilor)

Media mediilor = 14 kg

Deviația standard = 1,02

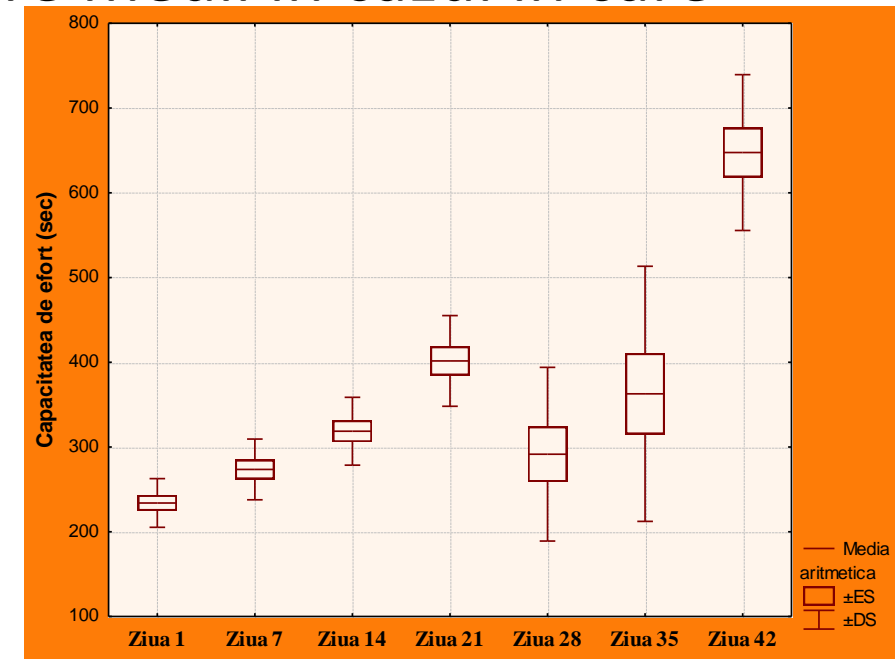
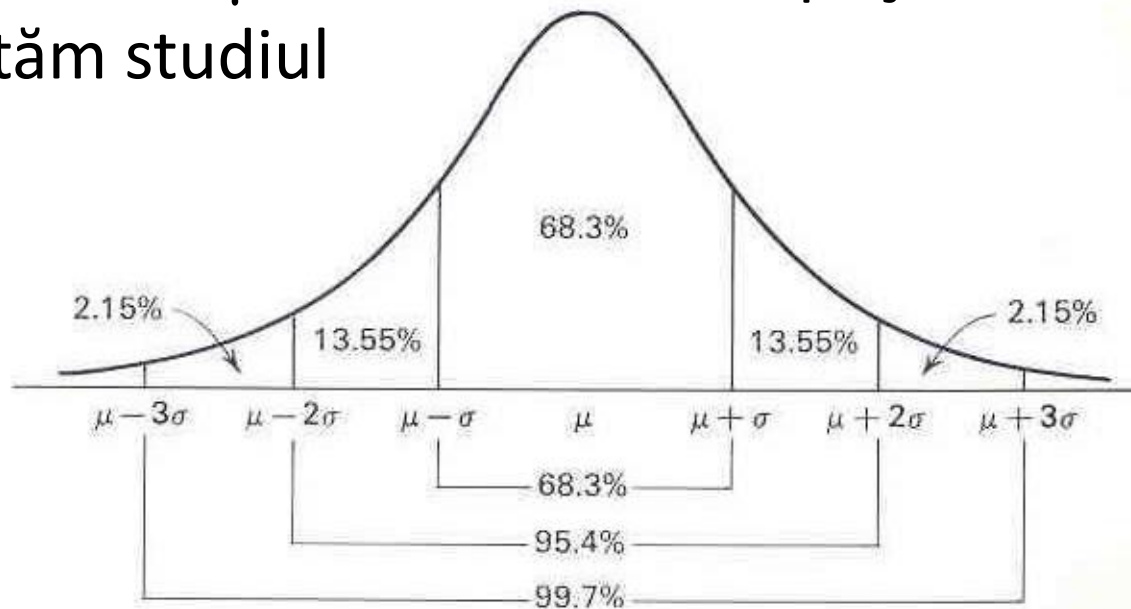
# Teorema limitei centrale

Fie o populație cu media  $\mu$  și deviația standard  $\sigma$ , o distribuție de eșantionare bazată pe eșantionarea repetată de mărime  $n$  are proprietățile:

- Media distribuției de eșantionare  $= \mu$  media populației
- Deviația standard a distribuției de eșantionare  $= \sigma / \sqrt{n}$ , adică eroarea standard
- Dacă distribuția populației este normală, distribuția de eșantionare este normală.
- Dacă eșantionul este suficient de mare, distribuția de eșantionare se apropie de cea normală.

# Deviația standard versus eroarea standard

- $\sigma$  = deviația standard în populație
- În intervalul  $\mu \pm 2 \cdot \sigma$  există cel puțin 95% dintre date în cazul unei distribuții normale
- SE = eroarea standard
- În intervalul  $\mu \pm 2 \cdot SE$  există cel puțin 95% dintre medii în cazul în care repetăm studiul



# Observație

- Media  $\mu$  și deviația standard  $\sigma$  sunt necunoscute
- Ele sunt approximate (estimate) de  $\bar{X}$ ,  $s$  și  $s / \sqrt{n}$  - unde  $s$  este deviația standard de pe eșantion



# Intervale de confidență (de încredere)

Următoarele date reprezintă perioada de incubație în zile a unui eșantion aleatoriu de 43 de cazuri de hepatită A dintr-o sursă epidemică din orașul X din 1992:

34, 23, 28, 31, 25, 30, 29, 35, 22, 32, 27, 28, 33, 24, 27, 26, 31,  
26, 28, 36, 30, 23, 28, 34, 25, 29, 30, 28, 25, 35, 29, 27, 31, 24,  
29, 33, 27, 32, 30, 37, 29, 26, 28

Vrem să estimăm variabilitatea mediei  $\bar{X} = 28,9$

Vom determina un interval de încredere: 27,8 - 30,0 zile, unde 95% din medii sunt incluse (dacă se repetă studiul)

# Intervale de confidență (de încredere) pentru medie

$$Z_{\alpha} \leq z \leq Z_{\alpha}$$

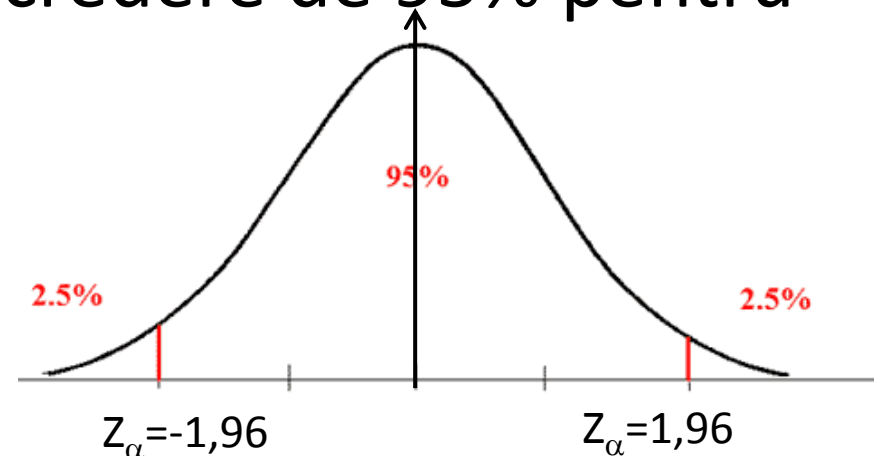
$Z_{\alpha}$  astfel încât **aria de sub curbă = 95%**       $Z_{\alpha}=1,96$

$$P(-1.96 \leq z \leq 1,96) = P\left(-1,96 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 1,96\right) = 0,95$$

$$P\left(\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}\right) = 0,95$$

$\left[\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}\right]$  – intervalul de încredere de 95% pentru medie

Legea  
normală  
redușă



# Intervale de confidență (de încredere) pentru medie

Dacă  $X$  este o variabilă cantitativă

$[\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}]$  - intervalul de încredere de 95% al mediei  $\mu$

Estimat prin:

$$[\bar{X} - 1,96 \frac{s}{\sqrt{n-1}}, \bar{X} + 1,96 \frac{s}{\sqrt{n-1}}]$$

unde  $\bar{X}$  - media aritmetică a variabilei  $X$  pe eșantion,  $s$  – deviația standard a variabilei  $X$  pe eșantion,  $n$  – numărul total de subiecți de pe eșantion

# Exemplu

Media perioadei de incubatie hepatitei a unui esantion aleatoriu de 43 de cazuri de hepatita A dintr-o sursa epidemica din oraşul X din 1992 a fost  $\bar{X} = 28,93$  şi deviaţia standard  $s=3,68$

$$[\bar{X} - 1,96 \frac{s}{\sqrt{n-1}}, \bar{X} + 1,96 \frac{s}{\sqrt{n-1}}]$$

$$\left[ 28,93 - 1,96 \frac{3,68}{\sqrt{43-1}}; 28,93 + 1,96 \frac{3,68}{\sqrt{43-1}} \right]$$

$$[28,93 - 1,11; 28,93 + 1,11]$$

$$[27,82; 30,04]$$

# Intervale de confidență pentru proporții

- Formula:

$$\left[ f - Z_{\alpha} \sqrt{\frac{f(1-f)}{n}}; f + Z_{\alpha} \sqrt{\frac{f(1-f)}{n}} \right]$$

unde  $f$  – frecvența relativă,  $f < 1$ ,  $n$  – numărul total de subiecți

# Exemplu

- Vrem să estimăm frecvența cancerului gurii în populația cu vârsta mai mare de 60 de ani.
- Într-un studiu cu 10.000 de participanți, 300 au avut cancer la nivelul guriipe parcursul vieții

$$f = \frac{300}{10.000} = 0,03$$

$$\left[ 0,03 - 1,96 \sqrt{\frac{0,03(1 - 0,03)}{10.000}}; 0,03 + 1,96 \sqrt{\frac{0,03(1 - 0,03)}{10.000}} \right]$$

$$[0,03 - 0,003; 0,03 + 0,003]$$

$$[0,027; 0,033]$$

$$\left[ f - Z_{\alpha} \sqrt{\frac{f(1-f)}{n}}; f + Z_{\alpha} \sqrt{\frac{f(1-f)}{n}} \right]$$

Multumesc!!!