

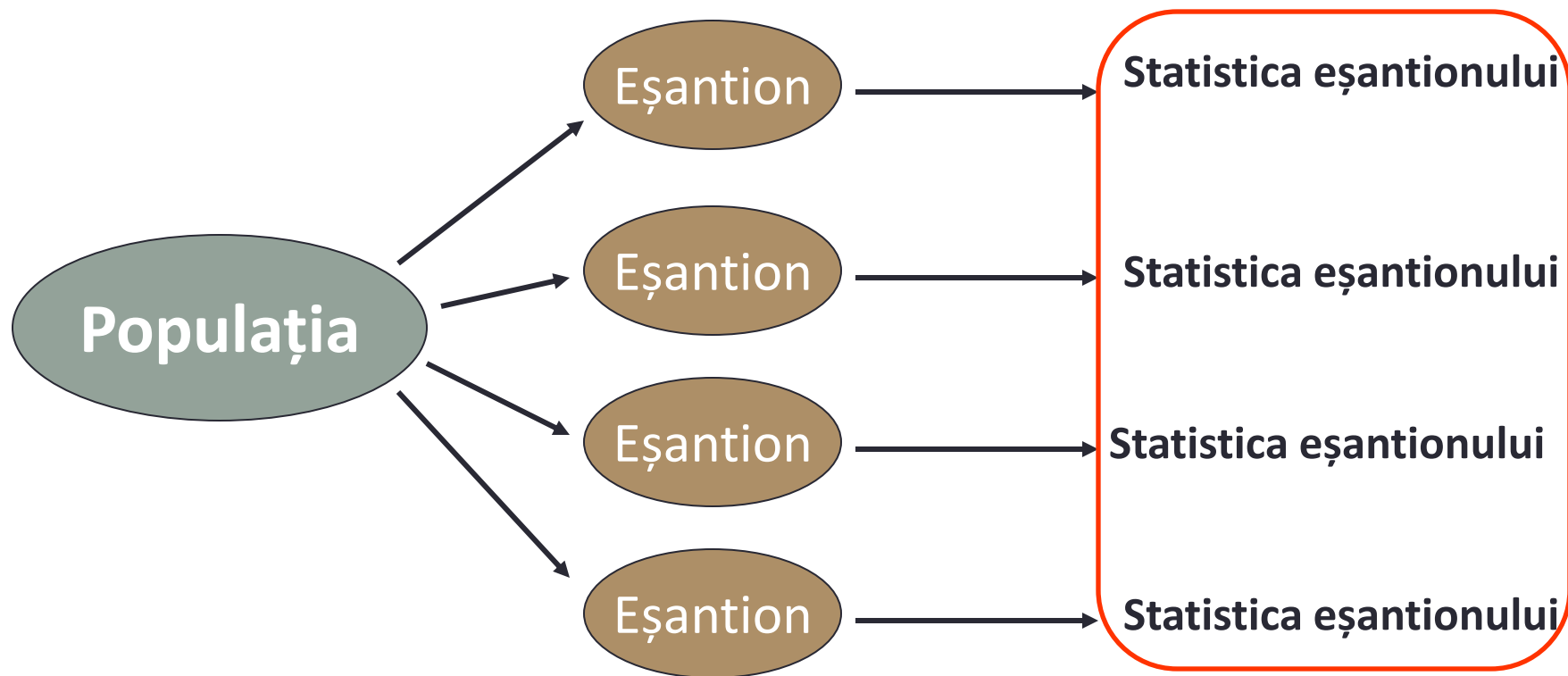
# VARIABILE ALEATOARE

# DISTRIBUȚII DE PROBABILITATE

# Cuprins

- Estimatorul punctual și intervalul de confidență sau încredere:
  - Intervalul de încredere pentru medie
  - Intervalul de încredere pentru frecvență
- Testarea ipotezelor statistice:
  - Concepte și practici generale

# ESTIMAOTRUL PUNCTUAL ȘI INTERVALUL DE ÎNCREDERE



**Distribuția  
eșantionului**

**≠**

**Distribuția  
de eșantionare**



**N = volumul  
populației**

**Copii de 1 an  
din Ro**

**AB:  $x_{AB,1}, x_{AB,2}, \dots, x_{AB,1000}$**

...

**GJ:  $x_{GJ,1}, x_{GJ,2}, \dots, x_{GJ,1000}$**

...

**VN:  $x_{VN,1}, x_{VN,2}, \dots, x_{VN,1000}$**

$\bar{x}_{AB}$

$\bar{x}_{GJ}$

$\bar{x}_{VN}$

$$\mu = \frac{x_{AB} + \dots + x_{GJ} + \dots + x_{VN}}{N}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}}$$

media ( $\bar{x}$ )  $\approx \mu$

$n \uparrow \Rightarrow s < \sigma$

# Estimarea punctuală

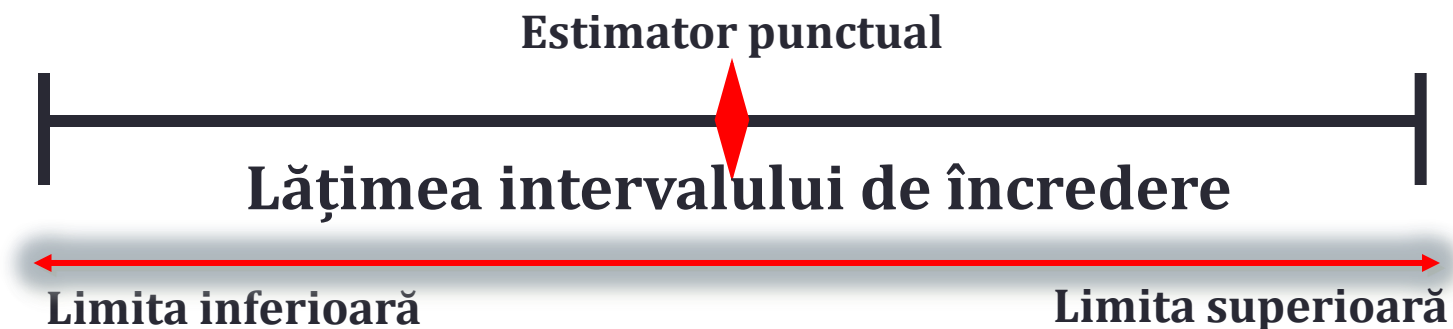
- O valoare a parametrului teoretic estimat
  - $\bar{m}$  (media eșantionului) este un estimator punctual al mediei populației ( $\mu$ )
- Este influențată de fluctuațiile de eșantionare
- Poate să fie foarte departe de valoarea reală a parametrului estimat
- Se recomandă ca estimarea unui parametru teoretic să se realizeze prin intermediul unui interval nu a unei singure valori
  - Acest interval se numește interval de confidență
  - Parametrul estimat aparține cu o probabilitate mare intervalului de confidență

# Definiție

- Un șir de valori al unui estimator de interes calculat astfel încât pentru o probabilitate de eroare aleasă să includă valorile adevărate ale variabilei.
- **$P[\text{valoarea critică inferioară} < \text{estimatorul} < \text{valoarea critică superioară}] = 1-\alpha$** 
  - unde  $\alpha$  = nivelul de semnificație
- Intervalul definit de valorile critice va cuprinde estimatorul populației cu o probabilitate de  $1-\alpha$

# Estimatorul punctual vs. intervalul de încredere

- Estimatorul punctual = valoarea unei statistici obținută pe un eșantion
  - Cât de multă incertitudine este asociată estimatorului punctual?
- Un interval oferă mai multe informații despre o caracteristică a populației decât un estimator punctual



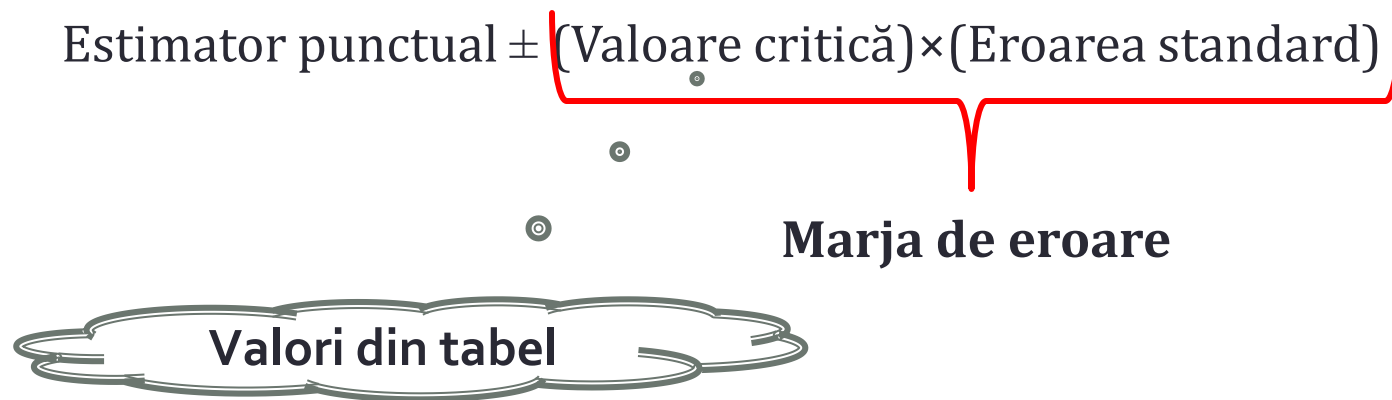
# Intervalul de încredere

- **Intervalul de încredere:**

- Ia în considerare variabilitatea de eșantionare → are valori diferite pentru fiecare eșantion
- Se calculează pe baza observării unui singur eșantion
- Oferă informații despre apropierea parametrului necunoscut al populației

- **Formula generală:**

Estimator punctual  $\pm$  (Valoare critică)  $\times$  (Eroarea standard)



# Intervalul de încredere

Estimator punctual  $\pm$  (Valoarea critică)  $\times$  (Eroarea standard)

**Marja de eroare**

- Marja de eroare și respectiv lățimea intervalului de încredere este cu atât mai mică cu cât volumul eșantionului e mai mare.
- Marja de eroare și respectiv lățimea intervalului de încredere variază cu valoarea nivelului de semnificație.

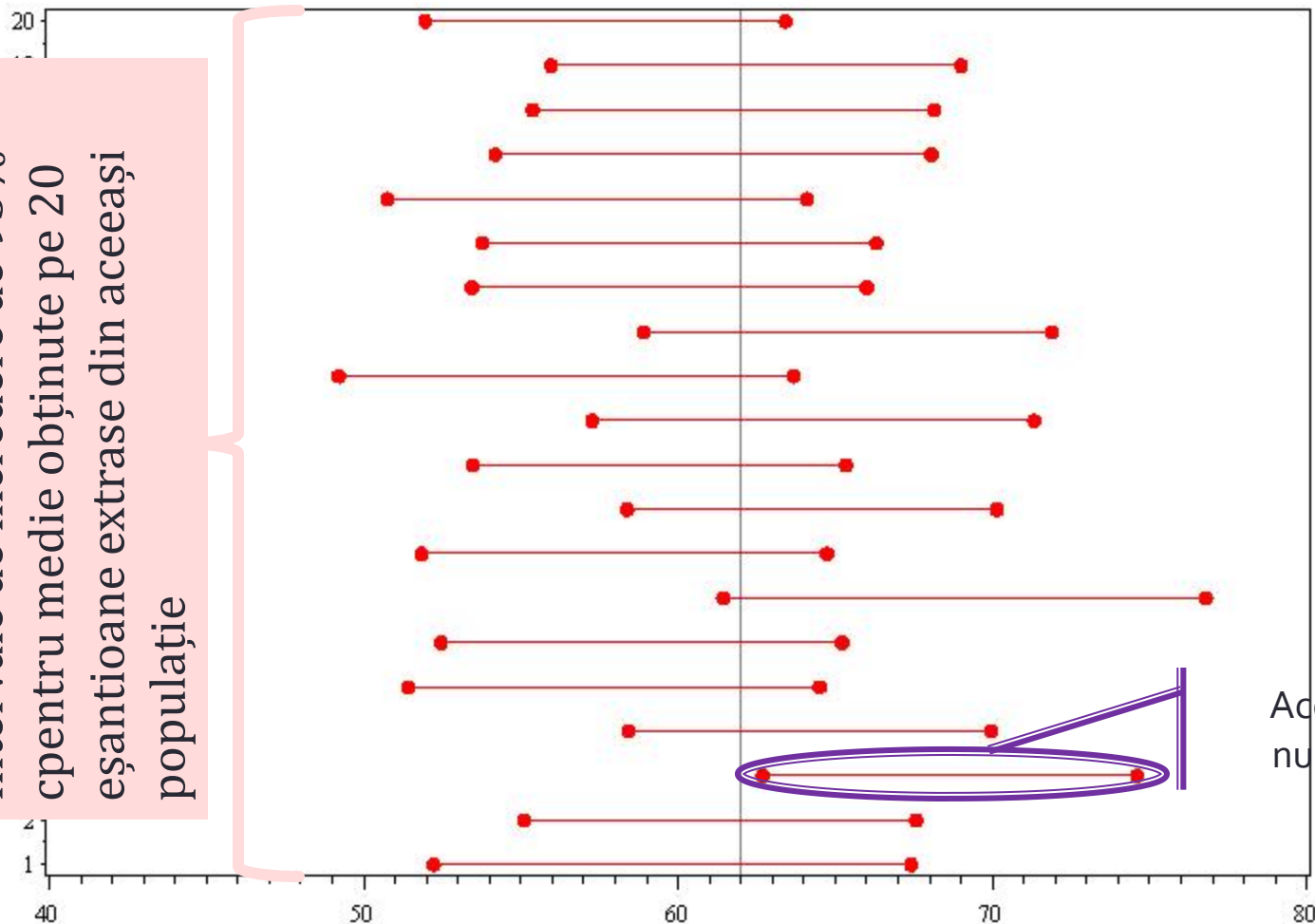
# Intervalul de încredere

- Nivel de semnificație  $\alpha = 5\% \rightarrow$  interval de încredere de 95% (IC95%)
- $IC = (1 - \alpha) = 0.95$
- Interpretare:
  - Dacă toate eșantioanele posibile de volum  $n$  s-ar extrage din populație și mediile și intervalele de încredere asociate ar fi calculate, 95% din intervalele de încredere vor conține valoarea reală a parametrului populației
  - Un interval de încredere poate să conțină sau poate să nu conțină valoarea reală a parametrului (datorită riscului de 5%)

# Intervalul de încredere

Valoarea reală  
(62)

Intervale de încredere de 95%  
pentru medie obținute pe 20  
eșantioane extrase din aceeași  
populație



Acest interval de încredere  
nu cuprinde valoarea reală

# Intervalul de încredere

- Se calculează în funcție de:
  - Talia eșantionului sau a populației
  - Tipul de variabilă (calitativă SAU cantitativă)
- Formula de calcul cuprinde 2 părți
  - Un estimator al calității eșantionului pe baza căruia estimatorul populației s-a calculat (eroarea standard)
    - Eroarea standard:
      - Cu cât  $n$  este mai mare cu atât eroarea standard este mai mică.
      - Este întotdeauna mai mică decât deviația standard
  - Gradul de încredere (confidență) al intervalului specificat (scorul  $Z_\alpha$ )
- Se poate calcula pentru orice estimator

# Intervalul de încredere pentru medie

- Eroarea standard a mediei este egală cu deviația standard împărțită la radicalul volumului eșantionului
  - Dacă deviația standard este mare, șansa de eroare în estimator este mare
  - Dacă volumul eșantionului este mare, șansa erorii în estimator este mică

Deviația standard a populației  
cunoscută ( $\sigma$ )

$$\left[ m - Z_{\alpha} \frac{\sigma}{\sqrt{n}}, m + Z_{\alpha} \frac{\sigma}{\sqrt{n}} \right]$$

Deviația standard a populației  
necunoscută

$$\left[ m - Z_{\alpha} \frac{s}{\sqrt{n}}, m + Z_{\alpha} \frac{s}{\sqrt{n}} \right]$$

# Intervalul de încredere pentru medie

- Media glicemiei la un eșantion de 121 pacienți este de 105 iar variația de 36. Care este intervalul de încredere al mediei glicemiei în populația din care s-a extras eșantionul cu un prag de semnificație  $\alpha=0,05$ , considerând că glicemia este normal distribuită și pentru acest prag  $Z = 1,96$ .

- $n = 121$
- $s^2 = 36$
- $s = 6$

$$\bar{X} = 105$$

- $[105-1.07, 105+1.07]$
- $[103.93 - 106.07]$
- $[104-106]$

$$\left[ 105 - 1,96 \frac{6}{\sqrt{121}} ; 105 + 1,96 \frac{6}{\sqrt{121}} \right]$$

- Parametrul populației: nivelul colesterolului la femeile din România cu HTA și obezitate
- Media populației necunoscută ( $\mu$ ), deviația standard a populației ( $\sigma$ ) egală cu 30 mg/dl
- S-a extras din această populație un eșantion randomizat de volum 20 și s-a determinat o statistică a estimatorului punctual medie egală cu 220 mg/dl
  - Media de 220 mg/dl este estimatorul punctual al parametrului necunoscut al populației
  - Datorită variabilității de eșantionare, media va fi însoțită de intervalul de încredere asociat pentru estimarea corectă a parametrului populației

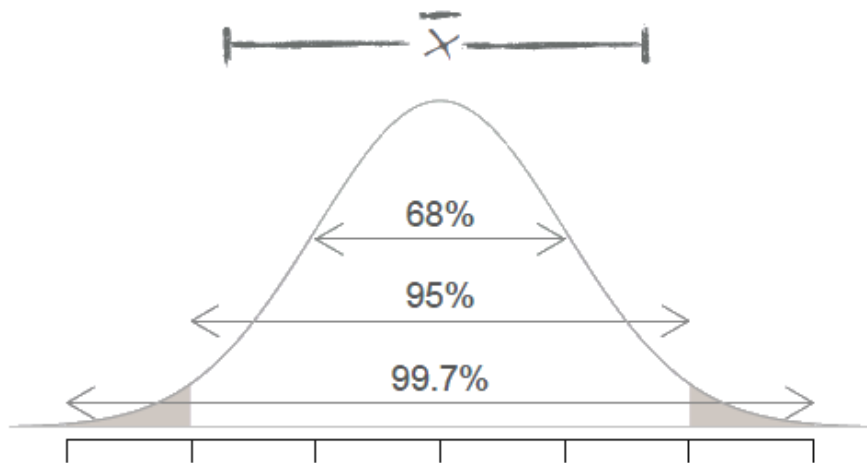
$$IC_{95\%} = \left[ 220 - 1.96 \frac{30}{\sqrt{20}}; 220 + 1.96 \frac{30}{\sqrt{20}} \right] = [207; 233]$$

$$\text{Lățimea} = 233 - 207 = 26$$

$$IC_{95\%} = \left[ 220 - 2.58 \frac{30}{\sqrt{20}}; 220 + 2.58 \frac{30}{\sqrt{20}} \right] = [203; 237]$$

$$\text{Lățimea} = 237 - 203 = 34$$

- Dacă dorim să fim siguri că media populației este cuprinsă în IC acesta trebuie să fie **larg** sau îngust?
  - 99% (↑): lărgimea (↑) + acuratețe (↑) + precizie (↓)
- Cum putem avea precizie mare și acuratețe mare?
  - $n \uparrow$



$$ME = Z_{\alpha} \frac{s}{\sqrt{n}} \rightarrow n = \left( \frac{Z_{\alpha} \times s}{ME} \right)^2$$

- Se dorește testarea efectului unui medicament folosit în tratamentul epilepsiei la mamă asupra dezvoltării cognitive a copilului. Dezvoltarea cognitivă se testează prin estimarea indicelui de inteligență a copilului de 3 ani născut de femei care au urmat în timpul sarcinii tratament cu medicamentul de interes.
- Studii anterioare au arătat că deviația standard a indicelui de inteligență a copilului de 3 ani este egală cu 18 puncte.
- Care este numărul de copii în vârstă de 3 ani care trebuie incluși în studiu pentru a obține un interval de confidență de 90% cu o margine a erorii mai mică sau egală cu 4 puncte?

ME ≤ 4  
IC = 90%  
Z = 1,65  
σ = 18  
n=56

ME ≤ 2  
IC = 90%  
Z = 1,65  
σ = 18  
n=4×56=224

ME ≤ 6  
IC = 90%  
Z = 1,65  
σ = 18  
n=24,50

ME ≤ 4  
IC = 95%  
Z = 1,96  
σ = 18  
n=77,79=78

ME ≤ 2  
IC = 95%  
n=311,17=312  
  
ME ≤ 5  
IC = 95%  
n=49,79=50

- Se dorește testarea efectului unui medicament folosit în tratamentul epilepsiei la mamă asupra dezvoltării cognitive a copilului. Dezvoltarea cognitivă se testează prin estimarea indicelui de inteligență a copilului de 3 ani născut de femei care au urmat în timpul sarcinii tratament cu medicamentul de interes.
- Studii anterioare au arătat că deviația standard a indicelui de inteligență a copilului de 3 ani este egală cu 18 puncte.
- Care este numărul de copii în vârstă de 3 ani care trebuie incluși în studiu pentru a obține un interval de confidență de 90% cu o margine a erorii mai mică sau egală cu 4 puncte?

$$ME \leq 4$$

$$IC = 90\%$$

$$Z = 1,65$$

$$\Sigma = 18$$

$$ME = Z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

$$n = \left( \frac{Z_{\alpha} \times \sigma}{ME} \right)^2 = \left( \frac{1,65 \times 18}{4} \right)^2$$

$$n = 55,13 \rightarrow n = 56$$

- Un eșantion de 49 studenți au fost întrebați în câte relații exclusive au fost implicați până la data studiului. Studenții din eșantion au avut în medie 3 relații exclusive, cu o deviație standard de 1,2. estimați media adevărată a numărului de relații exclusive bazată pe rezultatele acestui eșantion utilizând intervalul de confidență de 95%. Distribuția de eșantionare a fost aproximativ normală.
- Pasul 1: verificarea condițiilor.
  - Numărul de relații exclusive ale unui student este independent de al altui student.  $n = 49 < 10\% \times N$  (numărul de studenți din universitate)
  - $n > 30 \rightarrow$  distribuția de eșantionare a numărului de relații exclusive dintr-un eșantion de volum egal cu 49 este aproximativ normală.

- Un eșantion de 49 studenți au fost întrebați în câte relații exclusive au fost implicați până la data studiului. Studenții din eșantion au avut în medie 3 relații exclusive, cu o deviație standard de 1,2. estimați media adevărată a numărului de relații exclusive bazată pe rezultatele acestui eșantion utilizând intervalul de confidență de 95%. Distribuția de eșantionare a fost aproximativ normală.

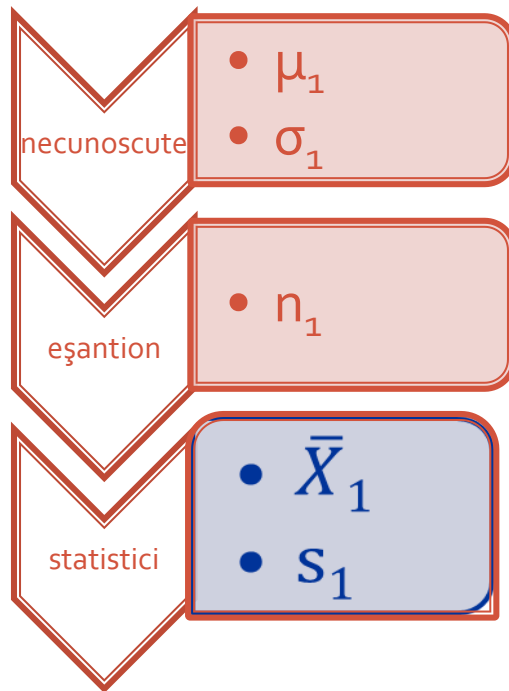
- Pasul 2: calculăm IC95%

$$\begin{array}{ll} n = 49 & ES = \frac{1,2}{\sqrt{60}} \approx 0,1714 \\ \bar{x} = 3 & ME = 3 \pm 1,96 * 0,1714 \\ s = 1,2 & IC95\% [2,66; 5,66] \end{array}$$

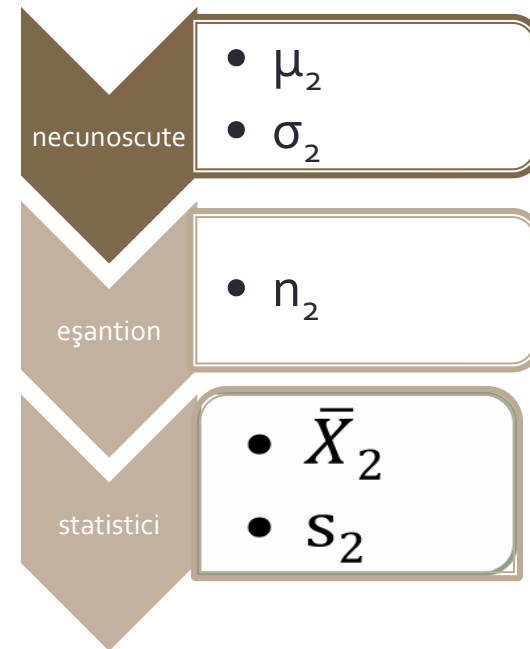
- Suntem 95% siguri că studenții au fost implicați în medie în 2,66 - 5,66 relații exclusive.

# Intervale de încredere pentru diferența dintre medii

Populația 1



Populația 2



Estmăm  $(\mu_1 - \mu_2)$  cu  $\bar{X}_1 - \bar{X}_2$

# Intervale de încredere pentru diferența dintre medii

$$(\bar{X}_1 - \bar{X}_2) \pm t_{df} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1} \left(\frac{s_2^2}{n_2}\right)^2}$$

|          |   |    |   |   |    |   |   |   |   |
|----------|---|----|---|---|----|---|---|---|---|
| Grupa 10 | 7 | 7  | 8 | 8 | 8  | 6 | 9 | 6 | 5 |
| Grupa 3  | 8 | 10 | 9 | 6 | 10 | 8 | 9 | 7 | 8 |

|                   | Grupa 1 | Grupa 2 |
|-------------------|---------|---------|
| Media             | 7,11    | 8,33    |
| Variația          | 1,27    | 1,32    |
| Deviația standard | 1,61    | 1,75    |

df=15,97

pentru  $\alpha = 0,05 \rightarrow t_{15,97} = 2,13$

$(7,11 - 8,33) \pm 2,13 \sqrt{(0,18 + 0,19)}$

$-1,22 \pm 2,13 * 0,61$

$-1,22 \pm 1,30 \rightarrow [-2,52; 0,08]$

# Intervale de încredere pentru diferența dintre medii

- Interpretare:

- Dacă valoarea zero este în intervalul de încredere → diferența dintre medii nu este semnificativ diferită de zero
- Dacă valoarea zero este în intervalul de încredere → diferența dintre medii este semnificativ diferită de zero

# Intervalul de încredere pentru frecvențe

- Se calculează dacă:

- $n \cdot f > 10$ , unde  $n$  = talia eșantionului,  $f$  = frecvența

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

# Intervalul de încredere pentru frecvențe

- Suntem interesați în estimarea frecvenței cancerului de sân la femeile între 50 și 54 de ani care au antecedente familiale pozitive. Într-un studiu randomizat la care au participat 10000 de femei, s-a constatat că 400 dintre acestea au fost diagnosticate cu cancer de sân.
- Care este intervalul de încredere de 95% asociat frecvenței observate?

$$\blacksquare f = 400/10000 = 0.04$$

$$\left[ 0,04 - 1,96 \sqrt{\frac{0,04 \cdot 0,96}{10000}}; 0,04 + 1,96 \sqrt{\frac{0,04 \cdot 0,96}{10000}} \right]$$

$$\blacksquare [0,04 - 0,004; 0,04 + 0,004]$$

$$\blacksquare [0,036; 0,044]$$

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

# COMPARING MEANS USING CONFIDENCE INTERVALS

<http://www.biomedcentral.com/content/pdf/1471-2458-12-1013.pdf>

**Table 1** Living conditions of the MS-MV and the immigrant population (CASEN survey 2006)

|                     | <b>IMMIGRANT POPULATION 1% total sample, n = 154 431 weighted population (1877 real observations)</b> |               | <b>MS-MV GROUP 0.67% total sample, n = 108 599 weighted population (1477 real observations)</b> |               |
|---------------------|-------------------------------------------------------------------------------------------------------|---------------|-------------------------------------------------------------------------------------------------|---------------|
|                     | <b>% or mean</b>                                                                                      | <b>95% CI</b> | <b>% or mean</b>                                                                                | <b>95% CI</b> |
| <i>DEMOGRAPHICS</i> |                                                                                                       |               |                                                                                                 |               |
| Mean age**          | X = 33.41                                                                                             | 31.81–35.00   | X = 26.13                                                                                       | 23.41–28.26   |
| Age categories:     |                                                                                                       |               |                                                                                                 |               |
| <16 years old**     | 13.60                                                                                                 | 11.29–16.28   | 45.25                                                                                           | 39.53–51.10   |
| 16-65 years old**   | 79.08                                                                                                 | 75.92–81.93   | 47.26                                                                                           | 41.64–52.94   |
| >65 years old       | 7.32                                                                                                  | 5.33–9.97     | 7.49                                                                                            | 5.31–10.46    |
| Sex (female = 1)    | 45.21                                                                                                 | 41.74–48.72   | 51.27                                                                                           | 47.99–55.41   |
| Marital status:     |                                                                                                       |               |                                                                                                 |               |
| Single**            | 45.81                                                                                                 | 42.06–49.62   | 64.30                                                                                           | 59.36–68.95   |
| Married**           | 45.49                                                                                                 | 41.66–49.36   | 29.39                                                                                           | 25.09–34.10   |

# De reținut!

- Estimarea corectă a unui parametru statistic se face cu ajutorul intervalului de încredere.
- Intervalul de încredere depinde de volumul eșantionului și de eroarea standard.
- Cu cât eroarea standard este mai mare cu atât intervalul de încredere este mai larg.
- Cu cât volumul eșantionului este mai mic cu atât intervalul de încredere este mai larg.

$$\left[ \bar{X} - Z_{\alpha} \frac{s}{\sqrt{n}}; \bar{X} + Z_{\alpha} \frac{s}{\sqrt{n}} \right]$$

$$((m_1 - m_2) - t_{critic} \times ES; (m_1 - m_2) + t_{critic} \times ES)$$

$$SE(m_1 - m_2) = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

$$(f_1 - f_2) \pm Z_{critic} \times ES$$

$$ES = \sqrt{(f_1 * (1 - f_1) / n_1) + (f_2 * (1 - f_2) / n_2)}$$

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

# TESTAREA IPOTEZELOR STATISTICE

# Definiții

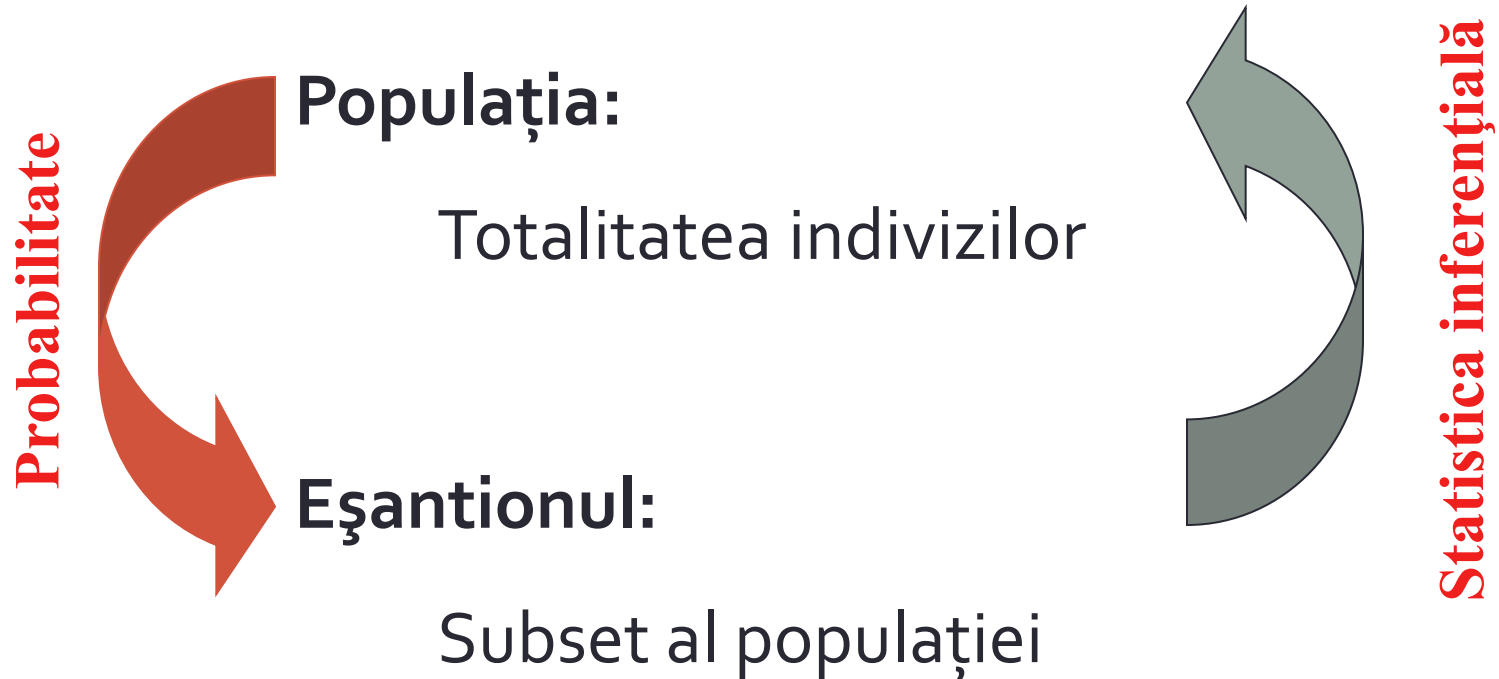
- **Test statistic** = metodă a deciziei medicale prin utilizarea datelor experimentale.
- Un rezultat se numește semnificativ statistic dacă este puțin probabil să apară datorită întâmplării
- Ipoteza statistică = asumție asupra parametrului populației. Această asumție poate sau nu să fie adevărată.
- Ipoteza clinică = o idee explicativă care permite structurarea datelor cu privire la un pacient în așa fel încât să ducă la o mai bună înțelegere a patologiei sau respectiv la o decizie medicală corectă.

[Lazare A. The Psychiatric Examination in the Walk-In Clinic: Hypothesis Generation and Hypothesis Testing. Archives of General Psychiatry 1976;33:96-102.]

## ■ Ipoteza clinică:

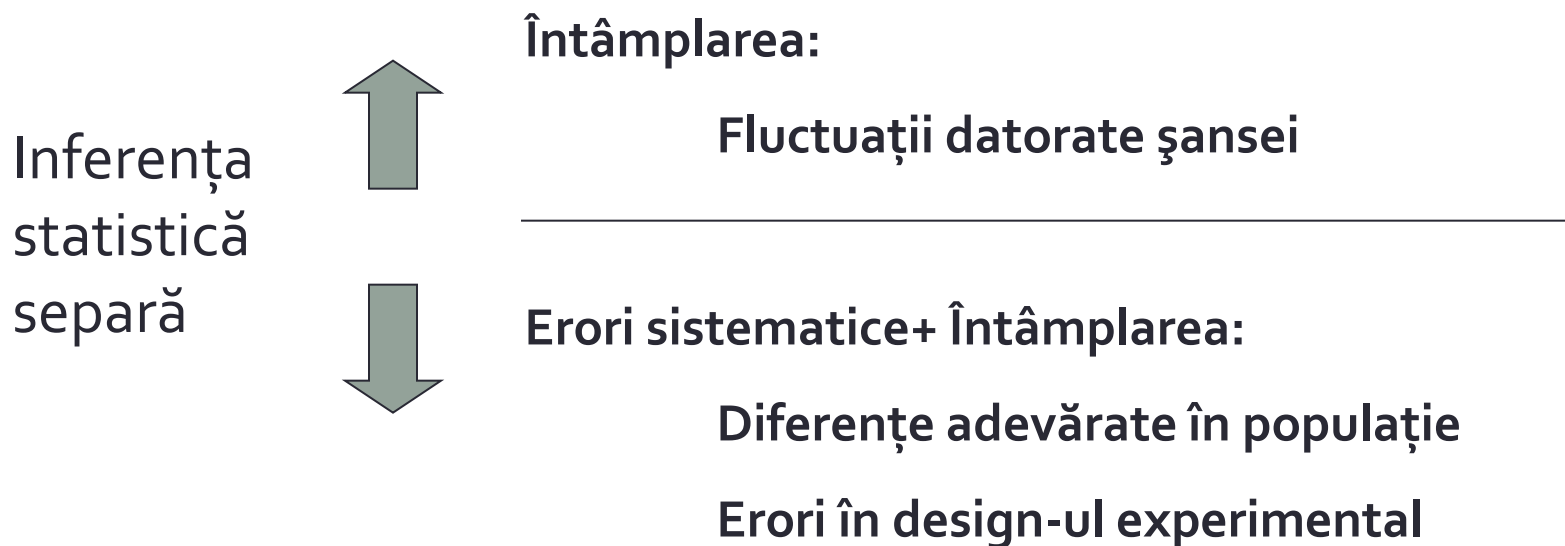
- O propoziție sau un set de propoziții, prezentate ca explicație a apariției unui grup de fenomene; această explicație poate să fie o ipoteză de lucru sau o ipoteză foarte probabilă în lumina faptelor stabilite.
- O explicație posibilă a unei observații sau a unui fenomen sau o problemă care necesită investigații
- O asumție

# Testarea ipotezelor



# Statistica inferențială

- Realizăm un studiu pe un eșantion
- Întrebarea cheie în statistica inferențială este:
  - Ar putea ca întâmplarea singură să producă un eșantion ca al nostru?
- 2 interpretări ale tiparelor în date:



# Etape ale testării ipotezelor

1. Formulează ipoteza cu privire la un parametru necunoscut al populației de interes.
2. Culege datele.
3. În asumptia că ipoteza nulă este adevărată, care este probabilitatea de a obține rezultate ca și ale noastre? (aceasta este valoarea “p”).
4. Dacă probabilitatea este mică nivelul de semnificație ( $< 0,05$ ) atunci respinge ipoteza nulă.

# Testarea Ipotezelor: Pasul 1

- Transpune problema de cercetat în termeni statistici
  - Ipoteza nulă (ipoteza statistică care urmează a fi testată):  
abreviată ca  $H_0$ 
    - “Nimic interesant nu se întâmplă”
  - Ipoteza alternativă (ipoteza care într-un anumit sens contrazice ipoteza nulă): abreviată ca  $H_1$  sau  $H_a$ 
    - Ceea ce cercetătorul crede că se întâmplă
    - Poate să fie unilaterală sau bilaterală
- Ipotezele statistice se referă la parametrii populației

| Unilateral                                                   | Bilateral                               |
|--------------------------------------------------------------|-----------------------------------------|
| $H_0: \mu = 110$<br>$H_1: \mu < 110$ SAU<br>$H_1: \mu > 110$ | $H_0: \mu = 110$<br>$H_1: \mu \neq 110$ |

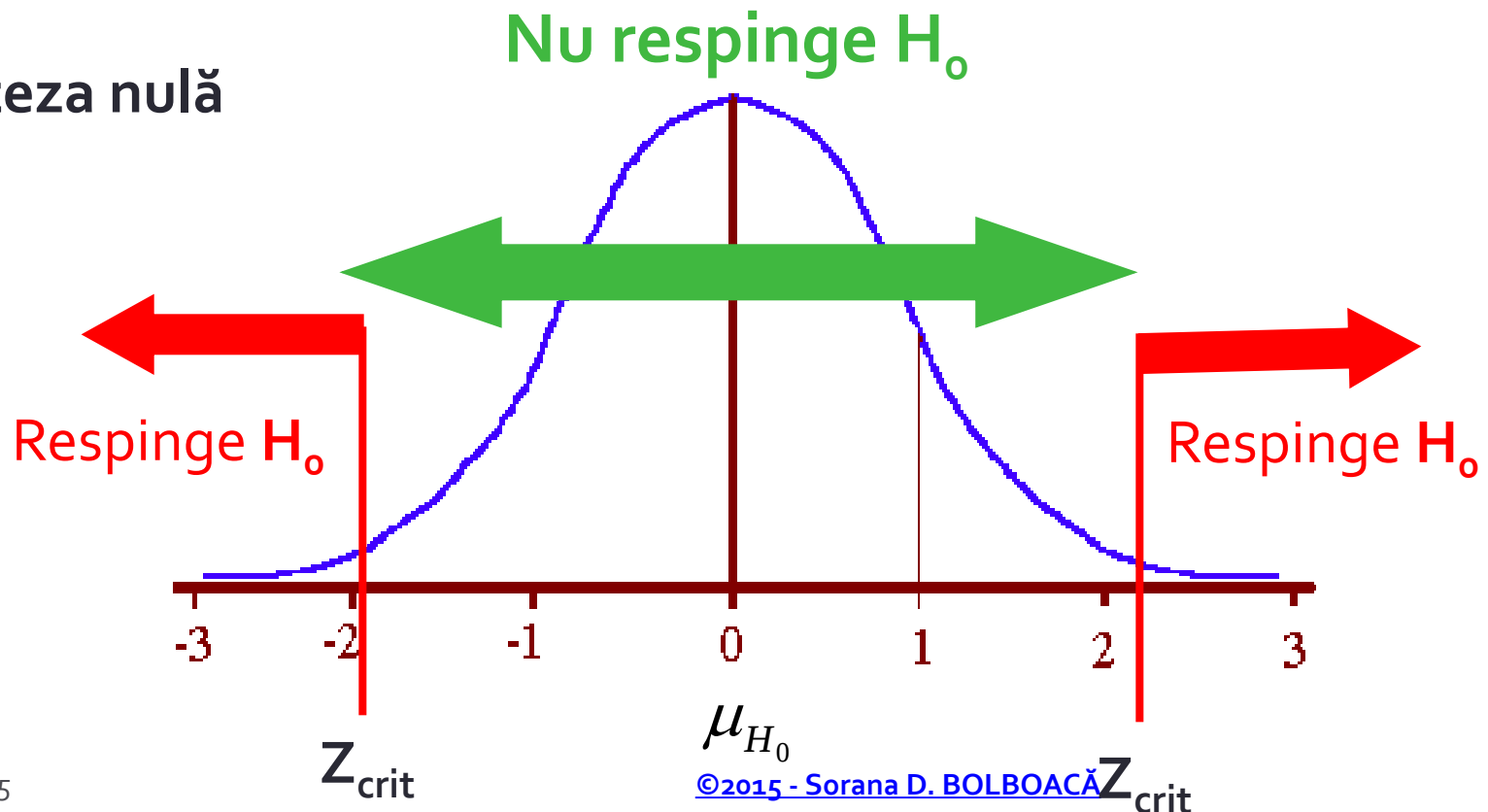
# Testarea Ipotezelor: Pasul 2

- Definiți regiunea critică:
  - Decideți care valoare  $p$  ar fi “mai puțin probabilă”
  - Această valoare prag se numește nivel de semnificație sau prag alfa
  - Atunci când probabilitatea asociată parametrului eșantionului este mai mică decât această valoare prag se spune că rezultatul este semnificativ statistic
  - De obicei nivelul alfa are valoare de 0,05 sau 0,01
- Nivelul alfa (nivelul de semnificație) = probabilitatea erorii de tip I (probabilitatea de a respinge ipoteza nulă în condițiile în care  $H_0$  este adevărată)
- Probabilitatea erorii de tip II este probabilitatea de a accepta ipoteza nulă în condițiile în care ipoteza alternativă este adevărată. Probabilitatea erorii de tip II se abreviază cu  $\beta$ .

# Testarea Ipotezelor: Pasul 3

- Regiunea critică:
  - Dacă valoarea parametrului statistic aparține regiunii critice, ipoteza nulă  $H_0$  va fi respinsă și va fi acceptată ipoteza alternativă  $H_1$ .
  - Dacă valoarea parametrului statistic nu aparține regiunii critice, ipoteza nulă  $H_0$  va fi acceptată.

Ipoteza nulă



# Testarea Ipotezelor: Pasul 4

- Calculează parametrul testului
- Parametrul statistic al testului aplicat (ex.  $Z_{\text{test}}$ ,  $T_{\text{test}}$ , or  $F_{\text{test}}$ ) este informația care se va utiliza pentru a decide dacă respingem sau nu ipoteza nulă.

# Testarea Ipotezelor: Pasul 5

- Concluzia statistică: În principiu nu acceptăm niciodată ipoteza nulă; ipoteza nulă o respingem sau nu o respingem

# Testarea ipotezelor statistice

1. Scrieți ipotezele statistice ( $H_0$  și  $H_1$ )
2. Alegeți nivelul de semnificație
3. Stabiliți regiunea critică
4. Calculați statistica testului și valoarea  $p$  asociată
5. Stabiliți concluzia statistică a testului

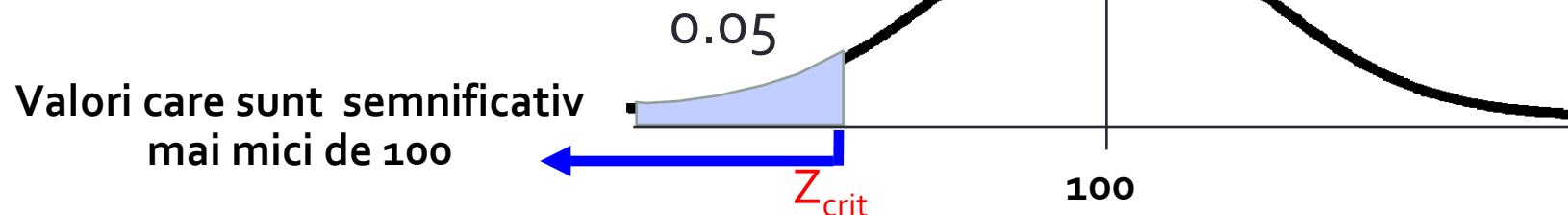
# Testul unilateral sau bilateral

- Testul unilateral se folosește când:
  1. Modificările în direcția opusă este lipsită de sens
  2. Modificările în direcția opusă nu este de interes
  3. Nici o teorie nu prezice schimbarea în direcția opusă
- Prin convenție în științele sociale și medicale se folosește testul bilateral
- De ce? Testul este mai conservativ.

## Test unilateral

Respinge  $H_0$

Nu respinge  $H_0$

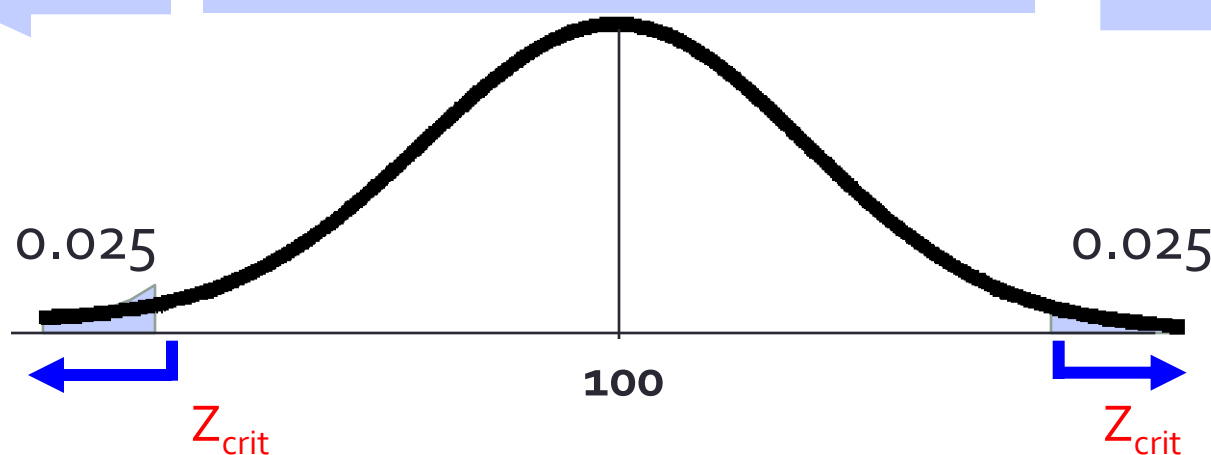


## Test bilateral

Respinge  $H_0$

Nu respinge  $H_0$

Res



Valori care diferă semnificativ de 100

# Diferența între valoarea $p$ și intervalul de confidență

- Valoarea  $p$  măsoară puterea evidenței împotriva ipotezei nule.
- $P$  este probabilitatea de a obține un rezultat extrem dacă ipoteza nulă este adevărată.
- Permite compararea mai multor studii.
- Valoarea  $p$  măsoară semnificația statistică
- Intervalul de confidență oferă un interval de valori care permite interpretarea clinică a rezultatelor

# Intervalul de confidență vs test statistic

- Un eșantion de 50 studenți a fost întrebat câte ore învață în medie pe zi. Studenții din eșantion învață în medie 1,2 ore pe zi cu o deviație standard de 0,6. Datele sunt ușor asimetrice spre stânga. Estimați numărul adevărat de ore pe zi dedicate studiului utilizând intervalul de confidență de 95%?

$$n=50, \text{ media} = 1,20, s=0,6$$

$$\text{media} \pm z * ES \text{ (} ES = s/\sqrt{n} \text{)} \rightarrow 1,20 \pm 1,96 * 0,07 \rightarrow [1,06; 1,34]$$

→ Suntem 95% siguri că studenții învață în medie între 1,06 și 1,34 ore pe zi

# Intervalul de confidență vs test statistic

- Un eșantion de 50 studenți a fost întrebat câte ore învață în medie pe zi. Studenții din eșantion învață în medie 1,2 ore pe zi cu o deviație standard de 0,6. Datele sunt ușor asimetrice spre stânga. Este această valoare semnificativ diferită de zero?
  - IC95% [1,06; 1,34] → Valoarea este semnificativ diferită de 0 deoarece valoarea 0 nu este cuprinsă în intervalul de încredere.
  - Test:  $H_0: \mu = 0$  vs.  $H_1: \mu \neq 0$
  - $Z = (\text{media}-0)/ES = (1,20-0)/0,07 = 17,14 - p < 0.00001$ 
    - Respingem ipoteza nulă cu un risc de eroare de 5%

# De reținut!

- Structură generală de urmat în testarea ipotezelor statistice.
- Alegerea testului statistic se face în funcție de ipoteza de cercetare.
- Test statistic vs. Interval de confidență
  - → ambele acceptă sau nu ipoteza nulă

# Temă!

|    | A  | B                             | C               | D               | E                        | F                                | G                                       |
|----|----|-------------------------------|-----------------|-----------------|--------------------------|----------------------------------|-----------------------------------------|
| 1  | ID | Durata spitalizării<br>(zile) | Vârsta<br>(ani) | Sex<br>1=M, 2=F | Antibiotic<br>1=da, 2=nu | Cultura bacteriană<br>1=da, 2=nu | Serviciu<br>1=medical,<br>2=chirurgical |
| 2  | 1  | 5                             | 30              | 2               | 2                        | 2                                | 1                                       |
| 3  | 2  | 10                            | 73              | 2               | 2                        | 1                                | 1                                       |
| 4  | 3  | 6                             | 40              | 2               | 2                        | 2                                | 2                                       |
| 5  | 4  | 11                            | 47              | 2               | 2                        | 2                                | 2                                       |
| 6  | 5  | 8                             | 25              | 1               | 2                        | 2                                | 2                                       |
| 7  | 6  | 14                            | 82              | 1               | 1                        | 2                                | 2                                       |
| 8  | 7  | 20                            | 60              | 2               | 1                        | 1                                | 1                                       |
| 9  | 8  | 11                            | 56              | 2               | 2                        | 2                                | 1                                       |
| 10 | 9  | 12                            | 43              | 1               | 2                        | 2                                | 1                                       |
| 11 | 10 | 3                             | 50              | 2               | 2                        | 1                                | 2                                       |
| 12 | 11 | 9                             | 59              | 1               | 2                        | 1                                | 1                                       |
| 13 | 12 | 3                             | 4               | 2               | 2                        | 2                                | 2                                       |
| 14 | 13 | 8                             | 22              | 2               | 1                        | 2                                | 2                                       |
| 15 | 14 | 8                             | 33              | 2               | 1                        | 1                                | 2                                       |
| 16 | 15 | 5                             | 20              | 1               | 2                        | 1                                | 2                                       |
| 17 | 16 | 5                             | 32              | 1               | 2                        | 2                                | 2                                       |
| 18 | 17 | 7                             | 36              | 1               | 1                        | 2                                | 2                                       |
| 19 | 18 | 4                             | 69              | 1               | 2                        | 2                                | 2                                       |
| 20 | 19 | 3                             | 47              | 1               | 1                        | 2                                | 1                                       |
| 21 | 20 | 7                             | 22              | 1               | 2                        | 2                                | 2                                       |
| 22 | 21 | 9                             | 11              | 1               | 2                        | 2                                | 2                                       |
| 23 | 22 | 8                             | 19              | 1               | 1                        | 2                                | 2                                       |
| 24 | 23 | 11                            | 67              | 2               | 2                        | 2                                | 1                                       |
| 25 | 24 | 9                             | 43              | 2               | 2                        | 2                                | 2                                       |
| 26 | 25 | 4                             | 41              | 2               | 2                        | 2                                | 1                                       |

# Temă!

1. Calculați media, mediana, variația, amplitudinea, deviația standard, coeficientul de variație pentru durata spitalizării și vârstă
2. Care din parametrii medie, mediană, modul caracterizează cel mai bine durata spitalizării? Dar vârsta? Argumentați răspunsul
3. Durata de spitalizare este legată de faptul că au primit antibiotic?
4. Durata de spitalizare este legată de rezultatul antibiogrammei?
5. Durata de spitalizare este legată de faptul că au suferit o manoperă chirurgicală?

# Temă!

- Rezolvați cerințele într-un fișier Word folosind toate elementele de formatare învățate.
- Expediați fișierul Word și Excel atașate pe adresa [sbolboaca@umfcluj.ro](mailto:sbolboaca@umfcluj.ro) până cel târziu în 3 Decembrie ora 9,00 p.m. E-mail-urile primite ulterior deadline-ului nu vor fi luate în considerare.
- Scrieți la subiectul e-mailului: Nume, prenume – Grupa – Tema 27.11.2015

