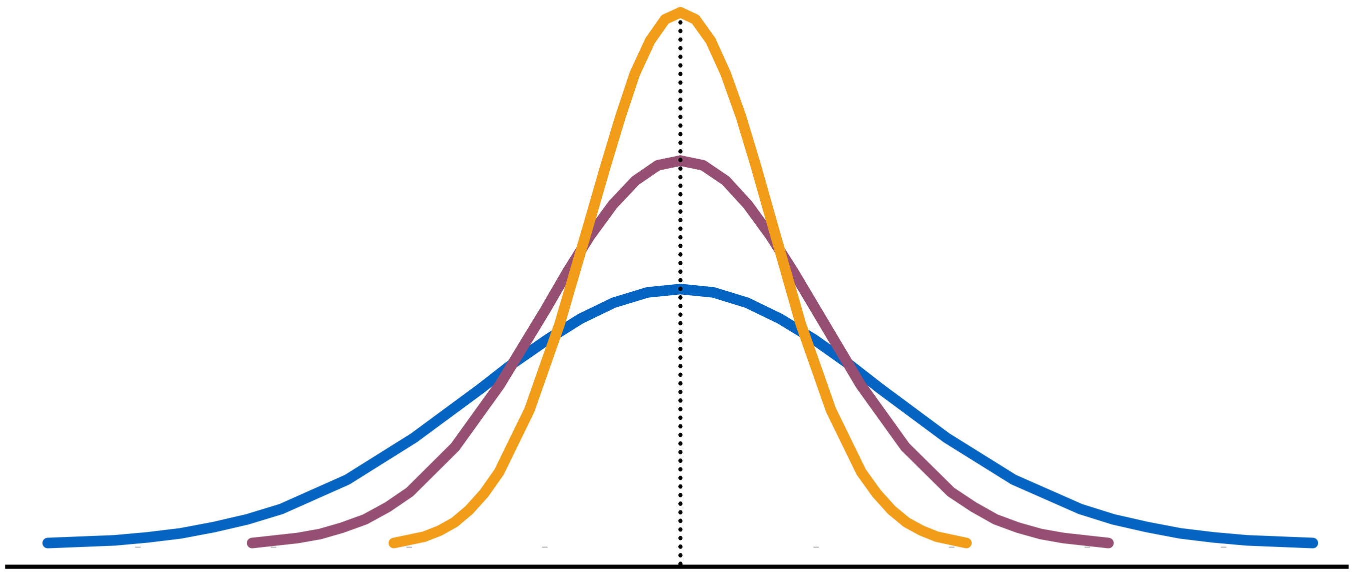


Bondor Cosmina, Tudor Drugan

Estimarea parametrilor statistici

A **ALWAYS**

S **SEEK**

K **KNOWLEDGE**

Obiective

Talia eșantionului. Legea numerelor mari

Distribuția de eșantionare

Deviația standard sau eroarea standard

Intervale de încredere

Exerciții

Talia eșantionului

Legea numerelor mari

Cu cât un eșantion e mai mare

cu atât rezultatul studiului este mai aproape de cel din populație

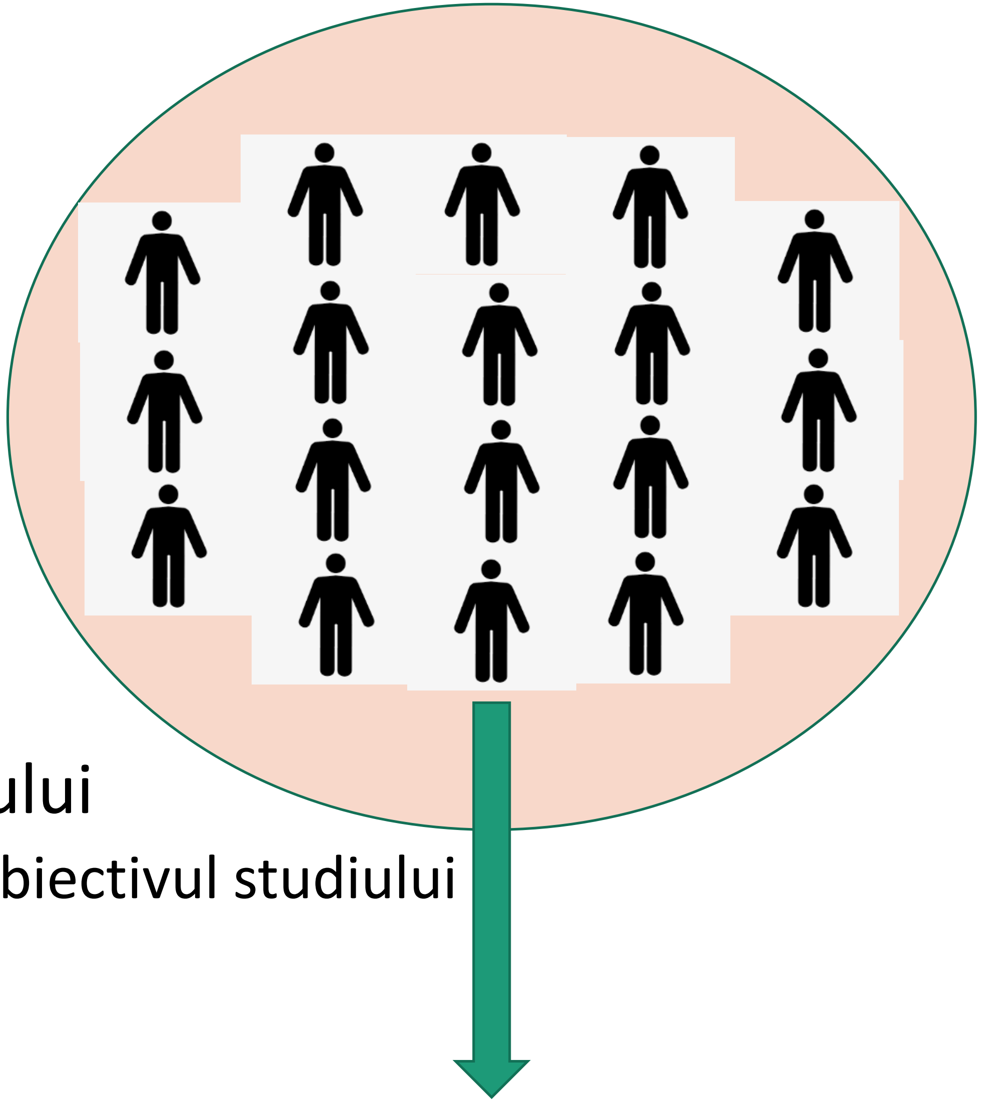
Ex. media de vârstă la persoanele cu diabet de tip 2 în populație = 65 ani

- selectăm subiecți cu diabet de tip 2
  - 10 – media 70 eroare 5 ani
  - 1000 – media 63 ani eroare 2 ani
  - 1.000.000 – media 66 ani eroare 1 an

!

Talia eșantionului

- numărul de indivizi
- eșantion
  - prea mic → erori de aproximare mari
  - prea mare → erori de măsurare, cost mare
- se calculează înainte de începerea studiului
  - cu formule statistice diferite în funcție de obiectivul studiului
- se mai numește și volum



Talia eșantionului = 18

Definiţia unui estimator

- Estimatorul unui parametru - o funcție care furnizează o valoare:

  estimarea punctuală a parametrului populației

- Ex: pe eșantion variabila X are valorile $x_1,\ldots,x_n$,

estimatorul punctual al mediei aritmetice μ a variabilei X pe populaţia din care a fost extras eșantionul

$\bar{X} = (x_1+x_2+\ldots+x_n)/n$

media aritmetică a valorilor variabilei X pe eșantion

Estimarea parametrului populației cu ajutorul intervalului de încredere

- cu ajutorul unui interval
  - în care parametrul (variabilei X în populație) se găseşte cu o probabilitate ridicată
- probabilitatea ca parametrul populației să se găsească în acest interval

  = încrederea

  = corectitudinea

  = precizia
- interval de încredere pentru orice parametru al populației
  - proporție
  - medie
  - coeficient de corelație
  - riscul relativ
  - etc.

Estimare

Media $\overline{X}$
pe eșantion

Media μ a întregii populații

în condițiile în care cunoaștem deviația standard σ a întregii populații

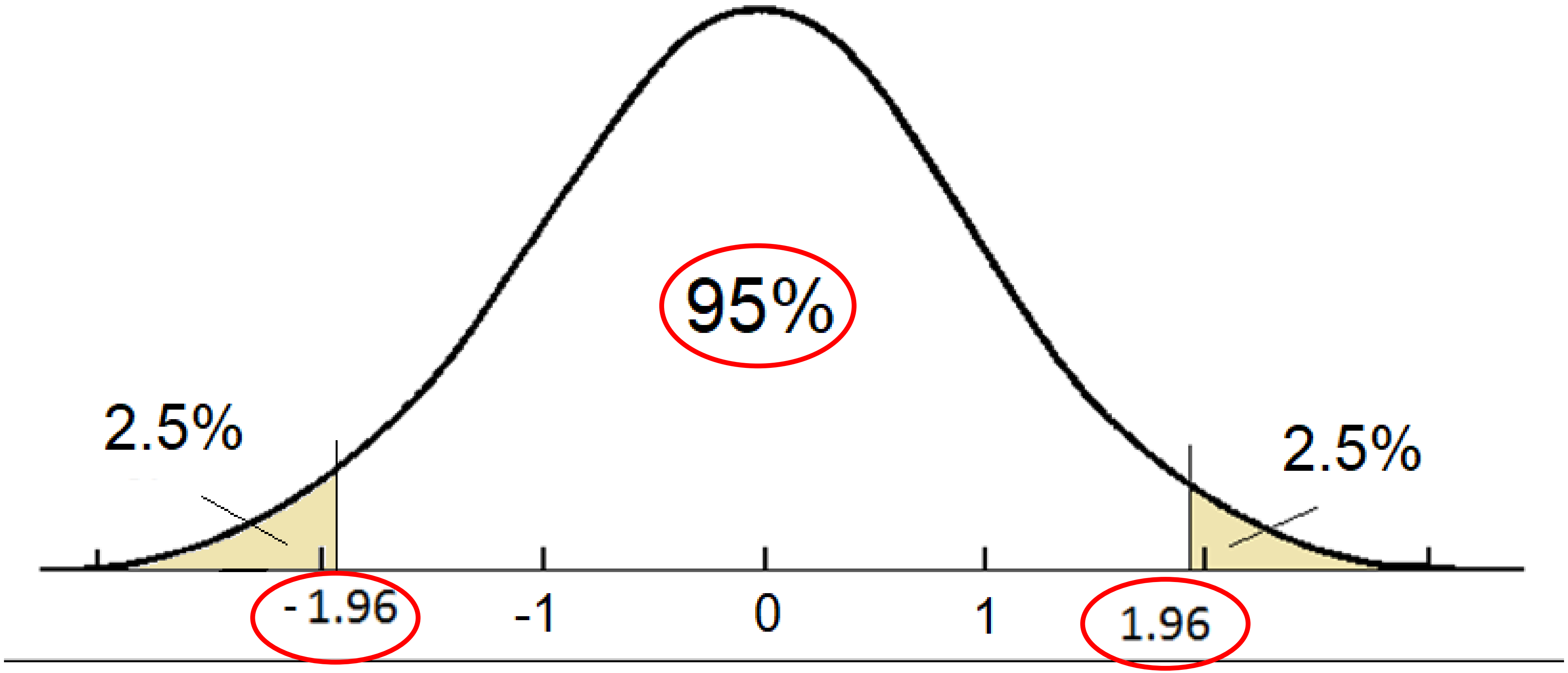
Frecvența f
pe eșantion

Frecvența π a întregii populații

mai sunt și alți parametrii ce pot fi estimați

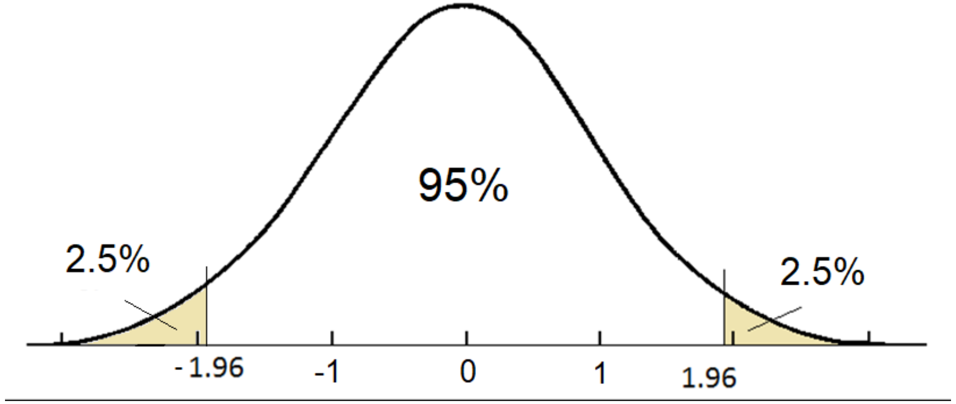
Proprietățile distribuției normale standard

- Exact 95% din aria de sub curbă este între -1.96 și 1.96

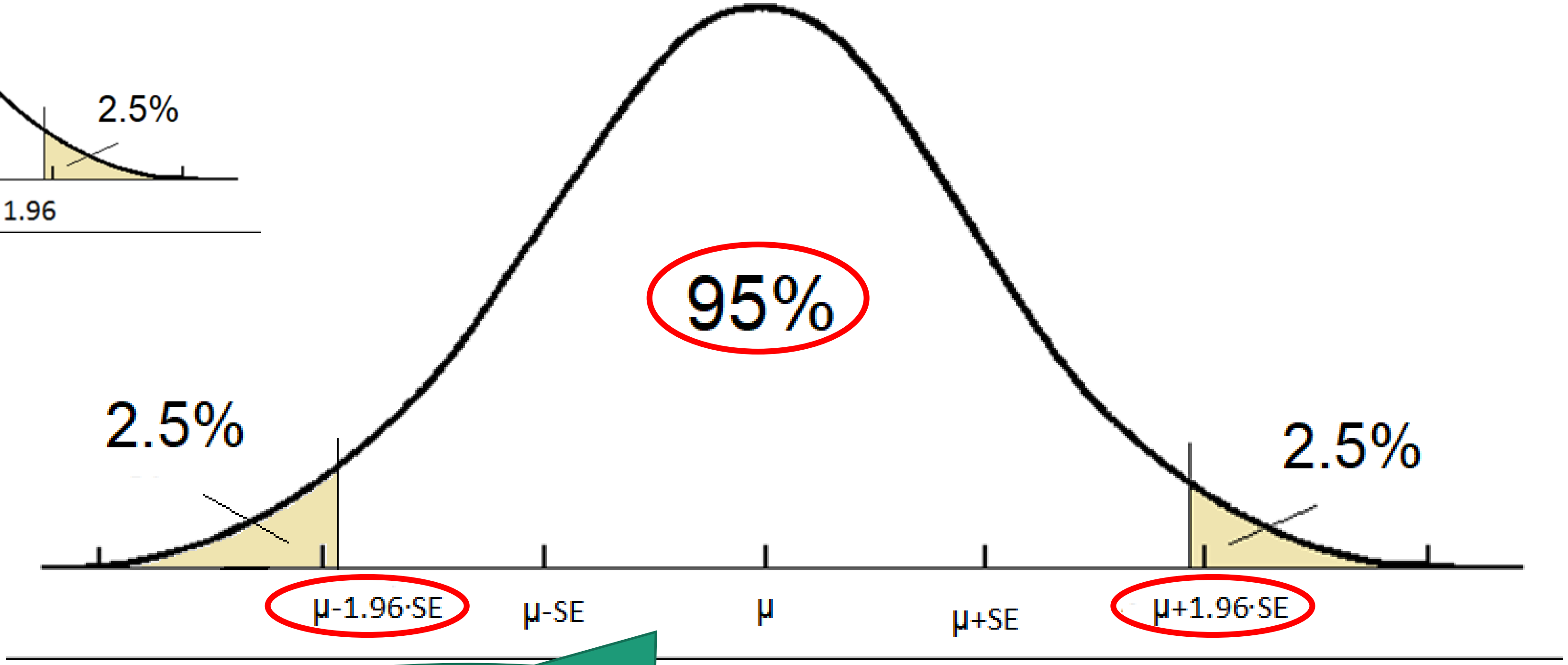


distribuția mediilor dacă repetăm studiul

distribuția normală standard



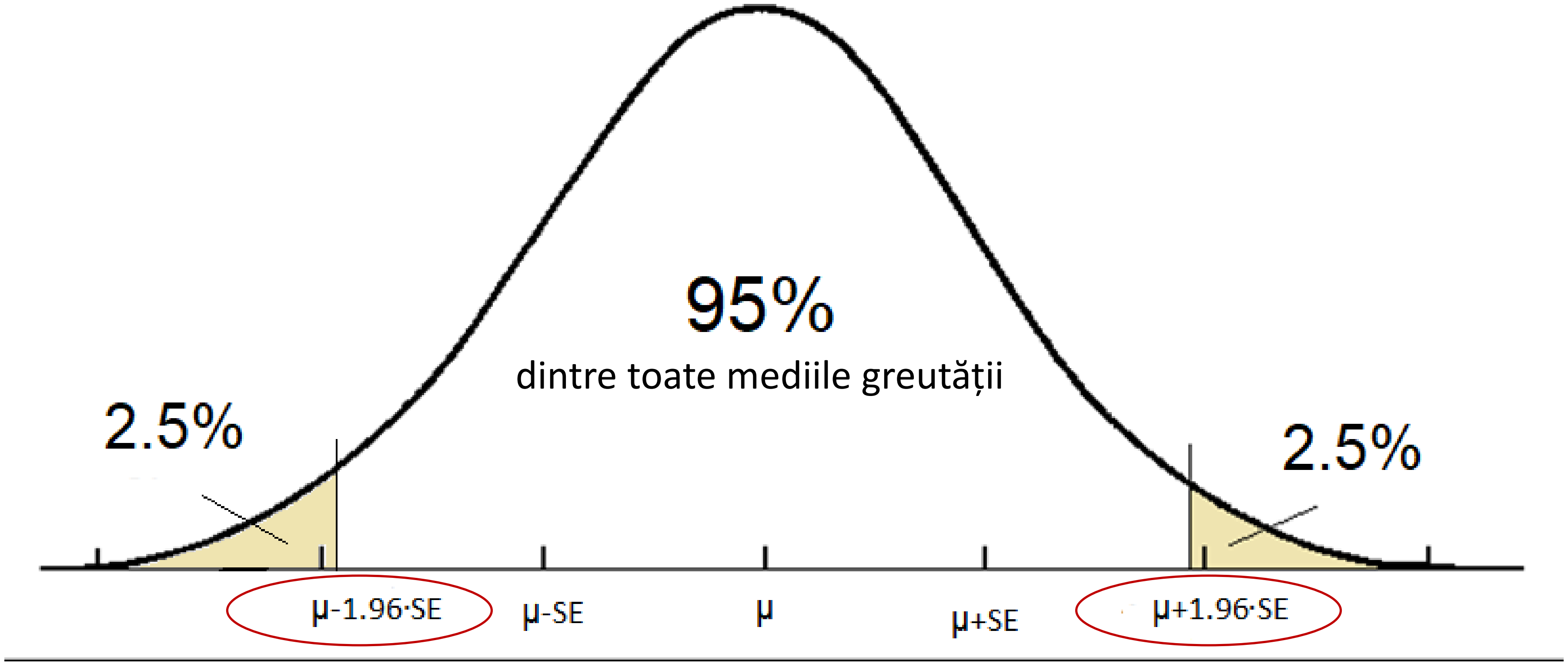
distribuția de eșantionare



s pentru distribuția de eșantionare = SE- eroarea standard

media μ – media variabilei X în populție, SE – eroarea standard a variabilei X în populație

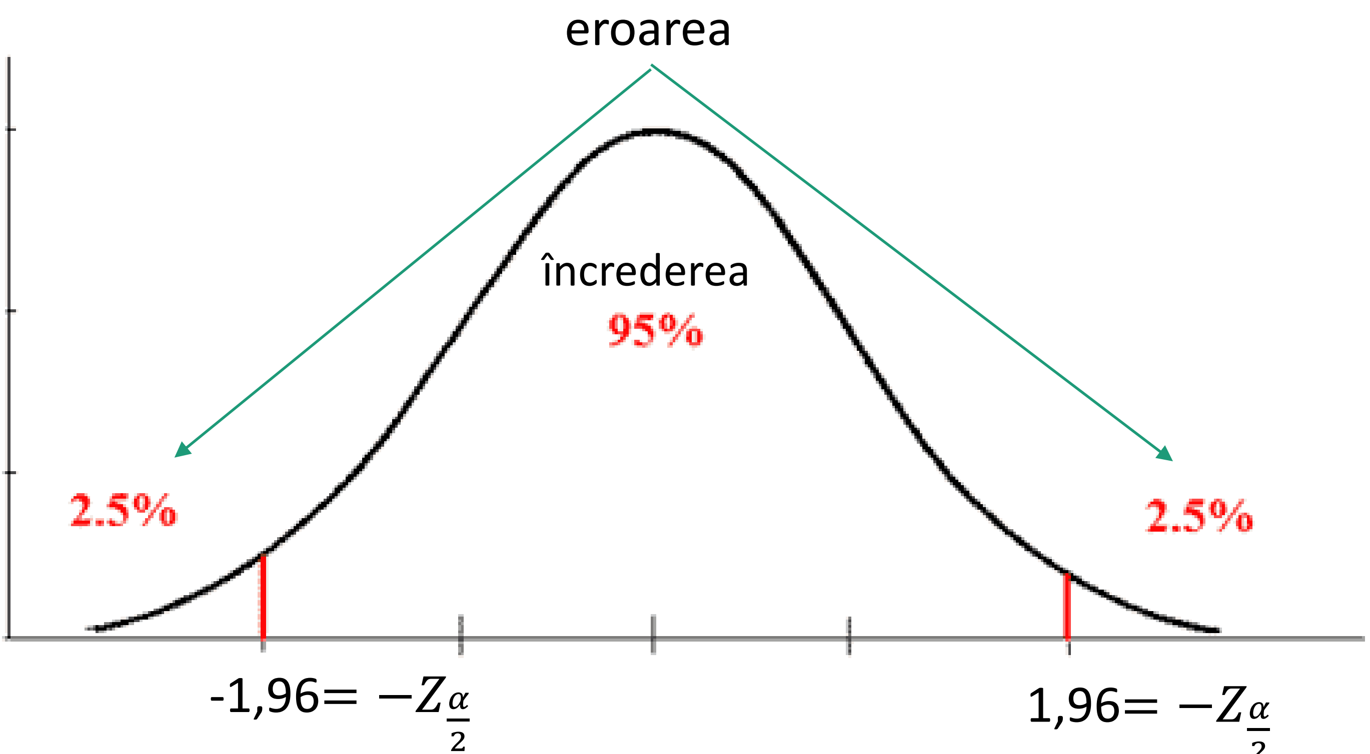
Exact 95% dintre mediile greutății pe toate eșantioanele de 2 băieți vor fi între µ-1.96SE și µ+1.96SE, unde µ este media greutății în populație și SE este eroarea standard



media µ – media variabilei X în populție, SE – eroarea standard a variabilei X în populație

Intervalul de **încredere** de 1-α pentru media μ în cazul σ cunoscută

$$[\bar{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}]$$



media $\bar{X}$ - media variabilei X pe eșantion, media μ – media variabilei X în popul̦ție, σ – deviația standard a variabilei X în populație, n – talia eșantionului, α – nivelul erorii, $Z_{\frac{\alpha}{2}}$ – valoare de pe axa Ox căreia îi corespunde în distribuția normală standardizată o arie egală cu nivelul de eroare

Intervalul de **încredere** de 95% pentru media μ în cazul eșantioanelor mari n≥30 și cu σ necunoscută

Dacă σ necunoscută o aproximăm cu s

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

Corecția aplicată variației ca să fie un bun estimator

unde

$\bar{X}$ – media aritmetică a variabilei X pe eșantion

s – deviația standard a lui X pe eșantion

n – numărul total de subiecți din eșantion

σ – deviația standard a variabilei X în populație

Exemplu

Dorim s-ă estimăm greutatea nou născuților. Pe un eşantion de n = 50 nou-născuţi

- media greutății $\bar{X}$= 3200g
- abaterea standard s = 300

- Să se calculeze intervalul de încredere de 95% pentru media greutății în populație

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

$$[3200 - 1,96 \frac{300}{\sqrt{50-1}}; 3200 - 1,96 \frac{300}{\sqrt{50-1}}]$$

$$[3200 - 84; 3200 + 84]$$

[3116; 3284]

– intervalul de încredere de 95%

Răspuns: Suntem 95% siguri că media μ a greutății întregii populații de nou-născuți se situează între 3116g și 3284g

Intervalul de **încredere** pentru proporție π pentru eșantioane mari (nf>10 și n(1-f)>10)

- Formula:

$$[f - Z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}; f + Z_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}]$$

unde

f – frecvența relativă a caracteristicii X în eșantion (!f<1)

n – nr. total de subiecți

π – proporția caracteristicii X în populație

α – nivelul erorii

Z – valoare de pe axa Ox căreia îi corespunde în distribuția normală standardizată o arie egală cu nivelul de eroare

Exemplu

Obiectiv: Vrem să estimăm frecvența **cancerului de esofag** la populația cu vârsta mai mare de 60 de ani.

- eșantion **n=10.000** de participanți observați timp de 10 ani
- 300 au făcut cancer de esofag
- Să se calculeze intervalul de încredere de 95% pentru frecvența cancerului de esofag la populația cu vârsta mai mare de 60 de ani

$$f = \frac{300}{10000} = 0{,}03$$

$$[f - Z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}; f + Z_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}]$$

$$[0{,}03 - 1{,}96\sqrt{\frac{0{,}03(1-0{,}03)}{10000}}; 0{,}03 + 1{,}96\sqrt{\frac{0{,}03(1-0{,}03)}{10000}}]$$

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

[0,027; 0,033] – intervalul de încredere de 95%

Răspuns: frecvența cancerului la populația peste 60 de ani este între 2,7% și 3,3% cu o probabilitate de 95%

Mulțumesc!