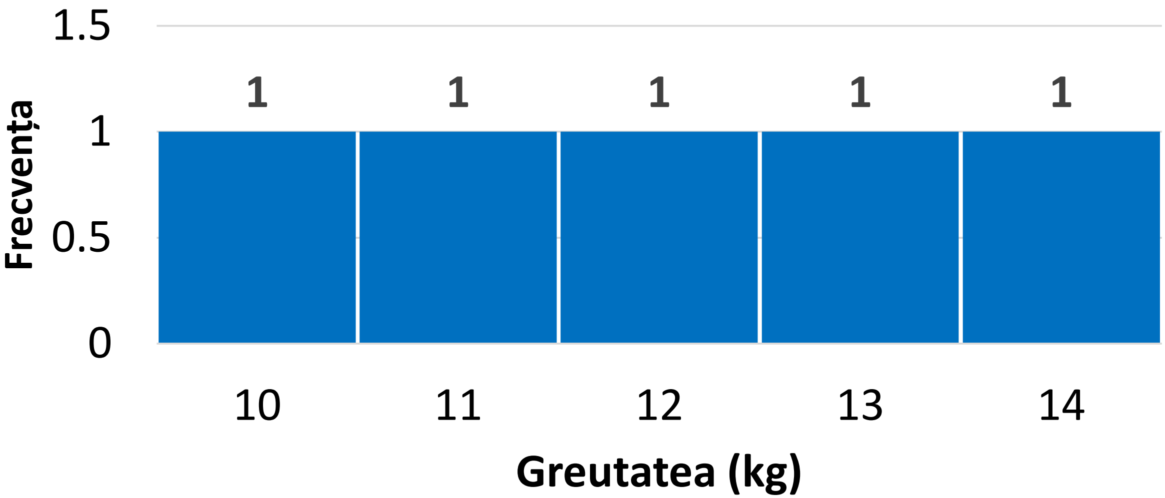
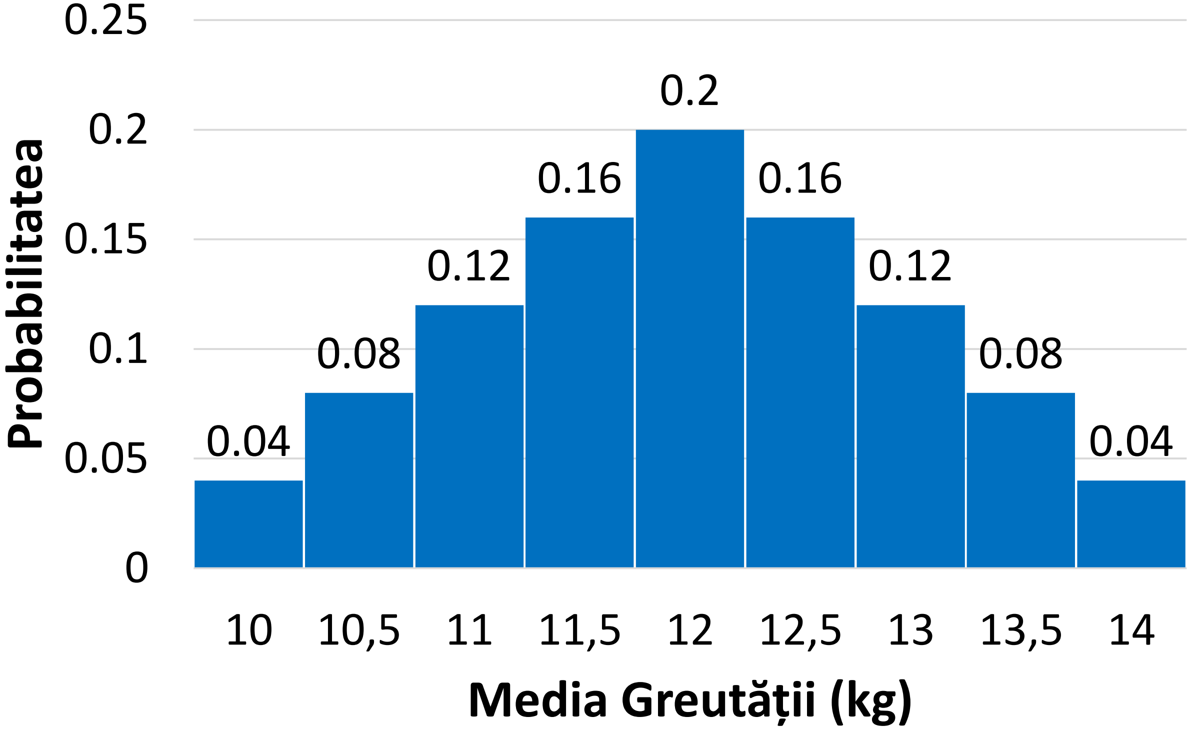
Estimări



Distribuția populației

Media populației = μ = 12 kg

Deviația standard = σ = 1,58



Distribuția de eșantionare (a mediilor)

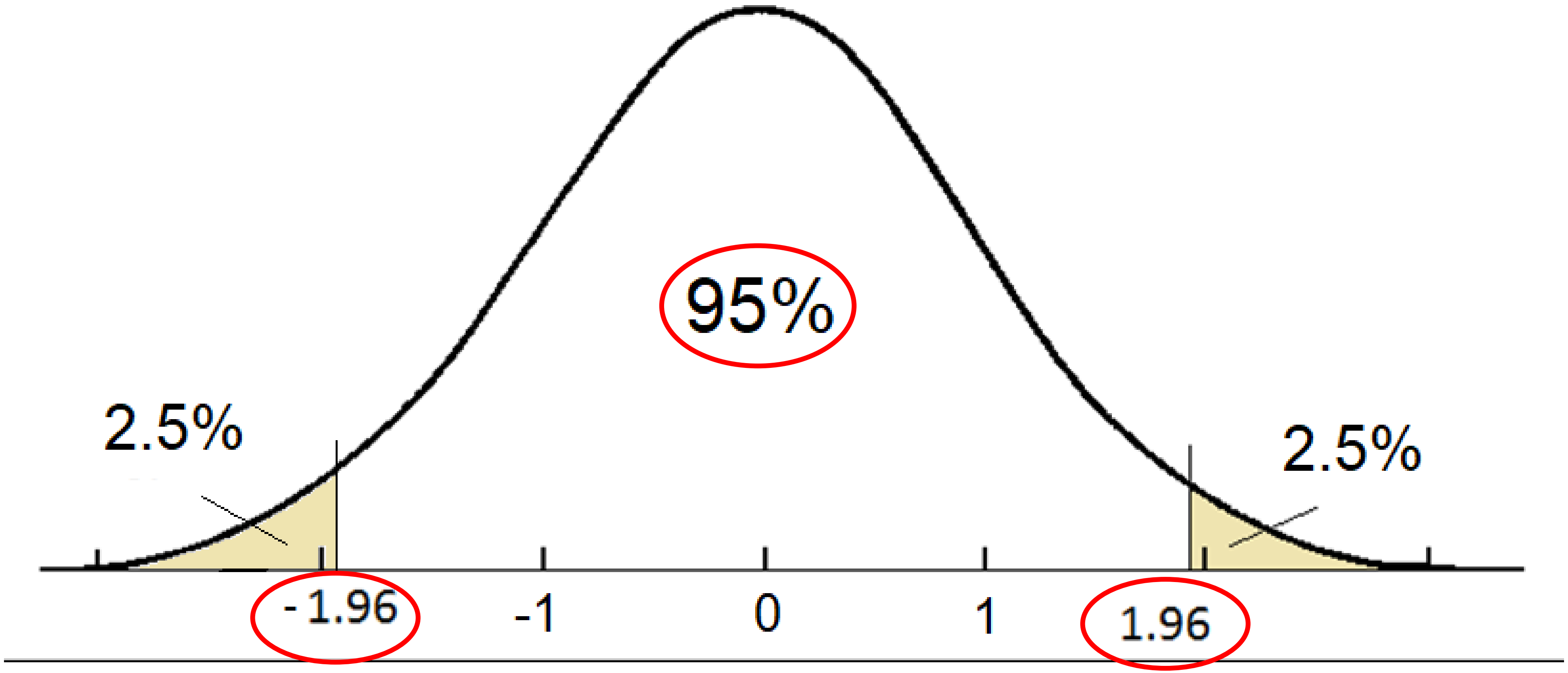
Media de eșantionare = $\bar{X}$ = 12

Deviația standard de eșantionare = s = 1,02

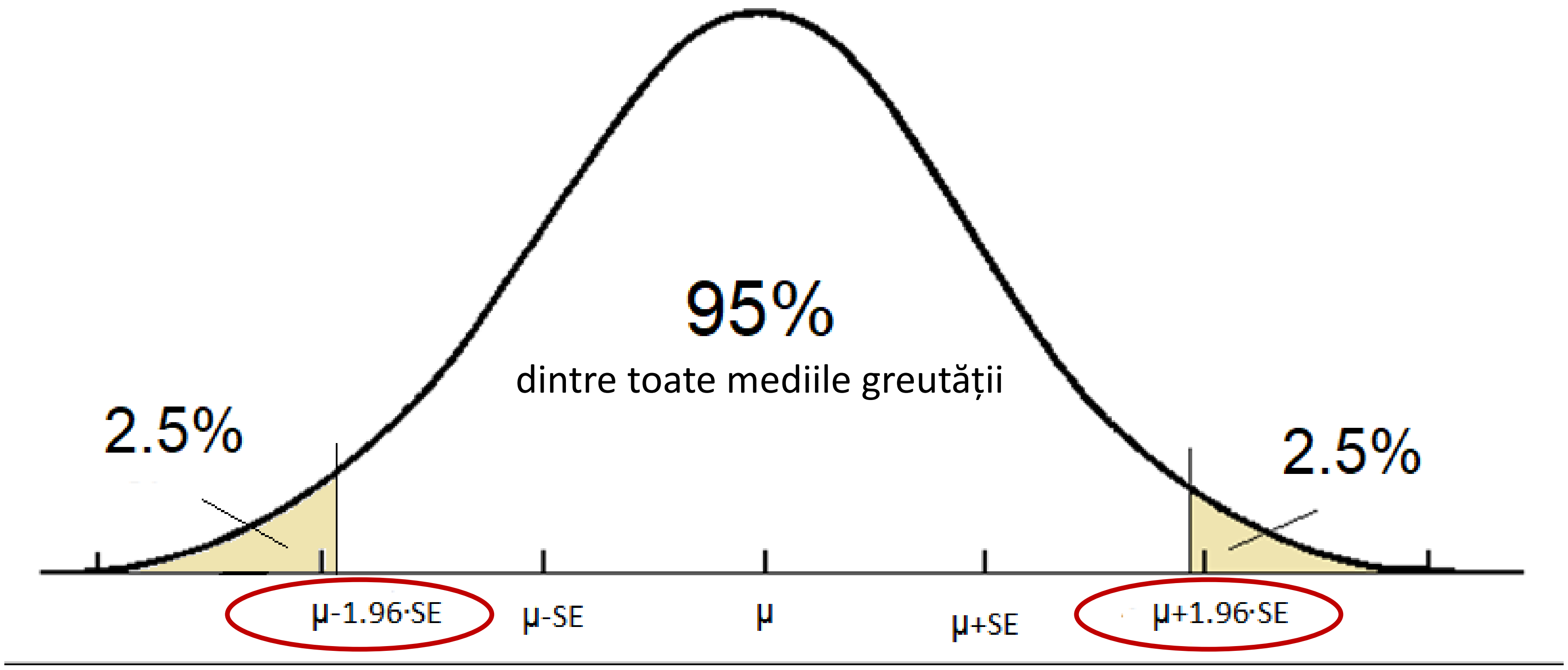
Urmează distribuția normală

Proprietățile distribuției normale standard

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



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

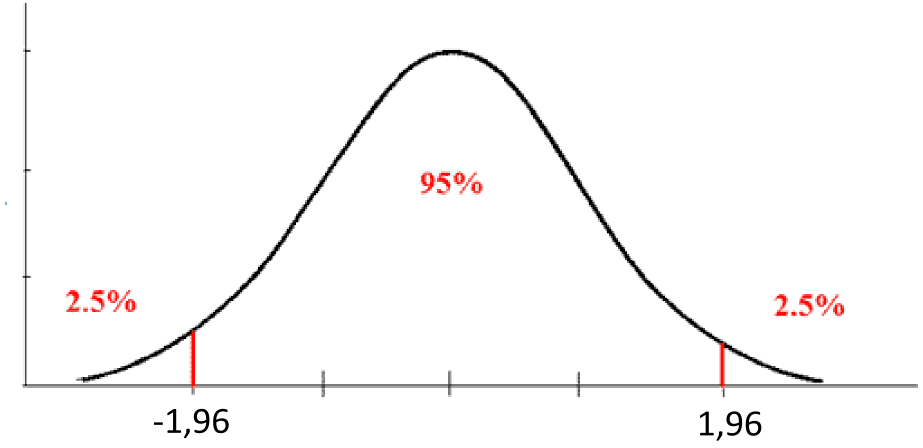
Intervalul de încredere de 95% pentru media μ

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

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

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

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



cazul α =0,05, n ≥ 30, σ necunoscută

Dacă σ necunoscută o aproximăm cu s, atunci **intervalul de încredere de 95% pentru media μ a populației** este

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

unde

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

s – deviația standard a lui X pe eșantion

n – numărul total de subiecți din eșantion

σ – deviația standard a variabilei X în populație

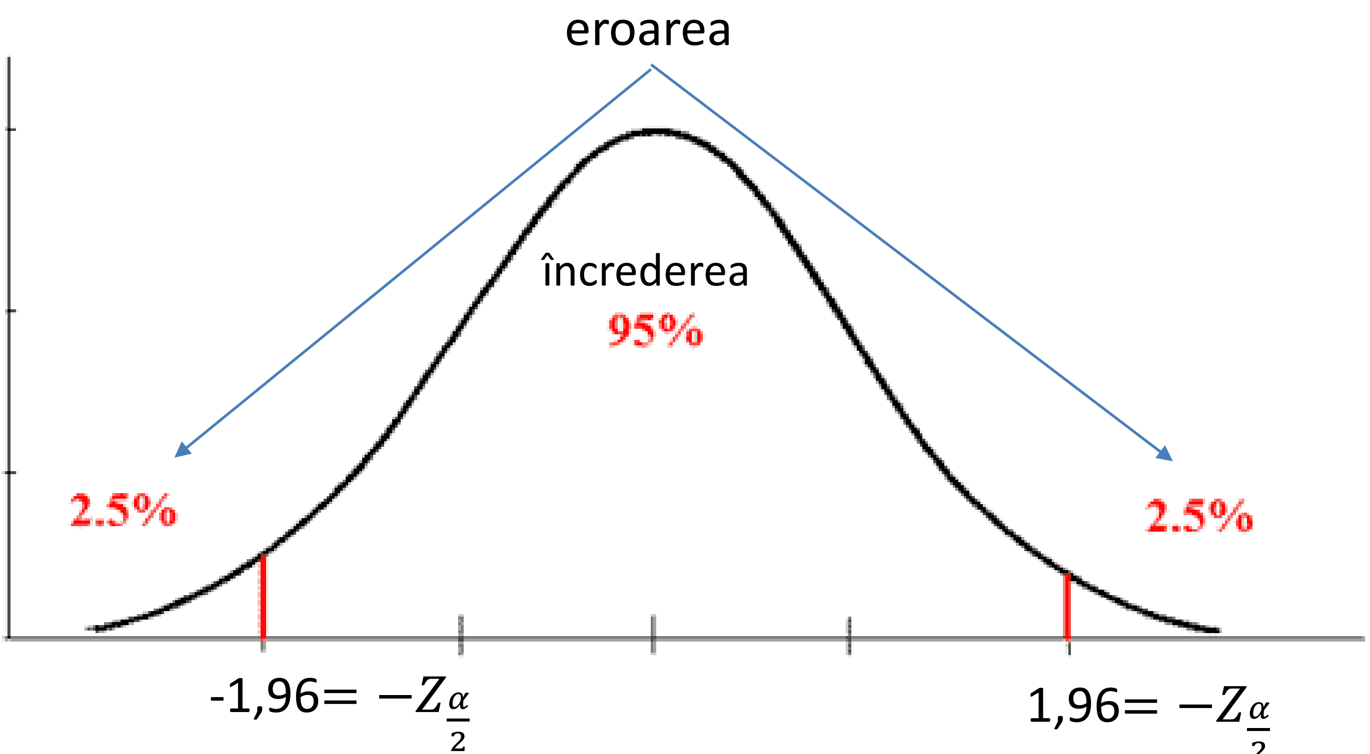
μ – media aritmetică a variabilei X în populație

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

cazul orice α, n ≥ 30, σ cunoscută

Intervalul de încredere de 1-α pentru media μ a populației

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



media $\bar{X}$ - media variabilei X pe eșantion, media μ – media variabilei X în popultție, 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

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	α =0,05, n ≥ 30, σ necunoscută , deci folosim formula: $[\bar{X} - 1{,}96 \frac{s}{\sqrt{n-1}}, \bar{X} + 1{,}96 \frac{s}{\sqrt{n-1}}]$ $[3200 - 1{,}96 \frac{300}{\sqrt{50-1}}; 3200 - 1{,}96 \frac{300}{\sqrt{50-1}}]$ $[3200 - 84; 3200 + 84]$ [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

Exemplu

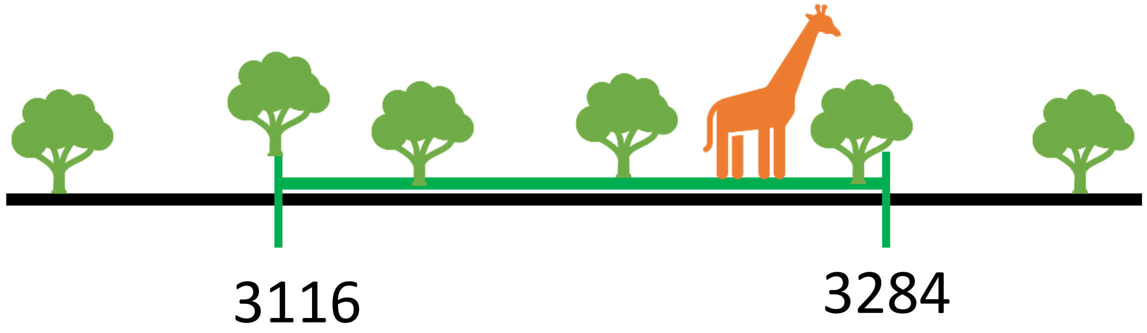
Dorim s-ă estimăm greutatea nou născuților. Pe un eşantion de n = 50 nou-născuţi

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

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

[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



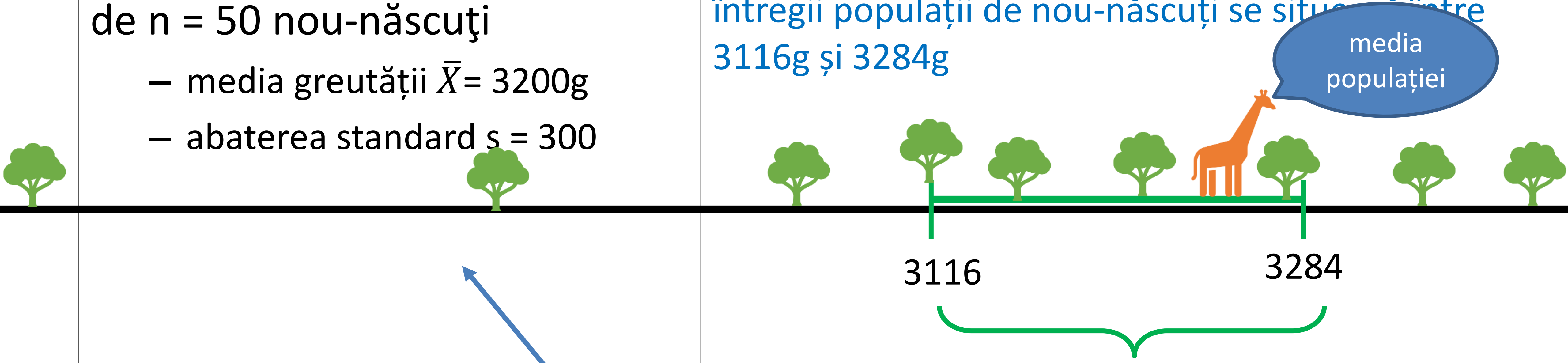
Exemplu

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

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

[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



5% siguri că e aici

Exemplu

Obiectiv: să estimăm media μ a colesterolului în populație

- un eșantion de **n=101**
- media **colesterolului**

$\bar{X}$ = 120 mg/dl

- deviația standard

s=16

Să se calculeze intervalul de încredere de 95% pentru **media** colesterolului în populație

α =0,05, n ≥ 30, σ necunoscută , deci folosim formula:

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

$$[120 - 1{,}96 \frac{16}{\sqrt{101-1}}; 120 - 1{,}96 \frac{16}{\sqrt{101-1}}]$$

$$[120 - 3{,}14; 120 + 3{,}14]$$

[116,86; 123,14] – intervalul de încredere de 95%

Răspuns: media μ a colesterolului pe întreaga populație se situează între 116,86 și 123,14 mg/dl cu o probabilitate de 95%

Exemplu

Obiectiv: să estimăm media μ a colesterolului în populație

- un eșantion de **n=101**
- media **colesterolului**

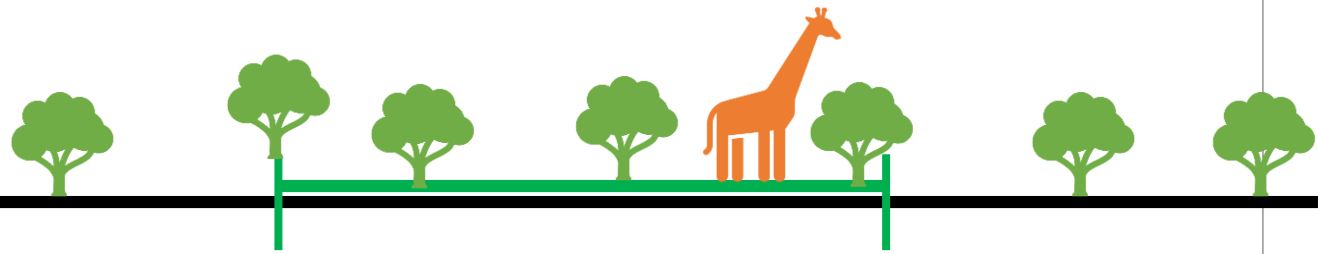
$$\bar{X} = 120 \text{ mg/dl}$$

- deviația standard

s=16

[116,86; 123,14] – intervalul de încredere de 95%

media μ a colesterolului pe întreaga populație se situează între 116,86 și 123,14 mg/dl cu o probabilitate de 95%



116,86

123,14

95% siguri că e aici

5% siguri că e aici

Exemplu

Studiul Fitzgerald al mobilității prin extensie a coloanei lombare la indivizi de vârste cuprinse între 30 şi 39 de ani

n=37, media=40° şi s=1,36°

Să se calculeze intervalul de încredere de 95% pentru medie.

α =0,05, n ≥ 30, σ necunoscută, deci folosim formula:

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

$$[40 - 1{,}96\,\frac{1{,}36}{\sqrt{36}}; 40 + 1{,}96\,\frac{1{,}36}{\sqrt{36}}]$$

$$[40 - 0{,}44; 40 + 0{,}44]$$

[39,66°; 40,44°]

Răspuns: mobilitatea coloanei lombare la indivizi tineri este între 39,66° și 40,44° cu o eroare de 5%

wooclap

- https://app.wooclap.com/BFKRI06?from=event-page

cazul orice α, nf>10 și n(1-f)>10, π necunoscută

Intervalul de încredere de 1-α pentru proporția π a populației

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

unde

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

n – nr. total de subiecți

π – proporția lui 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

cazul α =0,05, nf>10 și n(1-f)>10, π necunoscută

Intervalul de încredere de 95% pentru proporția π a populației

$$[f - 1,96\sqrt{\frac{f(1-f)}{n}}; f + 1,96\sqrt{\frac{f(1-f)}{n}}]$$

unde

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

n – nr. total de subiecți

π – proporția lui X în populație

α – nivelul erorii

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%

Ce se întâmplă dacă eșantionul este mai mic?

- Vrem să estimăm frecvența cancerului de esofag la populația cu vârsta mai mare de 60 de ani.
- Într-un studiu cu 1000 de participanți, 30 au avut 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.

$$f = \frac{30}{1000} = 0{,}03$$

$$\left[0{,}03 - 1{,}96\sqrt{\frac{0{,}03(1-0{,}03)}{1000}}; 0{,}03 + 1{,}96\sqrt{\frac{0{,}03(1-0{,}03)}{1000}}\right]$$

$$[0{,}03 - 0{,}011; 0{,}03 + 0{,}011]$$

[0,019; 0,041] – intervalul pentru n=1000

frecvența cancerului între 1,9% și 4,1% cu o probabilitate de 95%

[0,027; 0,033] – intervalul pentru n=10000

frecvența cancerului între 2,7% și 3,3% cu o probabilitate de 95%

Răspuns: eșantion mai mic → interval mai mare

n la numitor are efect invers

Creste eșantionul → crește precizia de măsurare prin scăderea intervalului necesar estimării

Autor: Conf. Dr. Bondor Cosmina-Ioana

Testarea ipotezelor

A ALWAYS

S SEEK

K KNOWLEDGE

Obiective

Testarea ipotezelor

Testul Student T

Testul hi-pătrat

Erori în testarea ipotezelor și puterea testului

Intervale de încredere sau testarea ipotezelor

Exerciții

Compararea a două grupuri cu metoda intervalelor de încredere

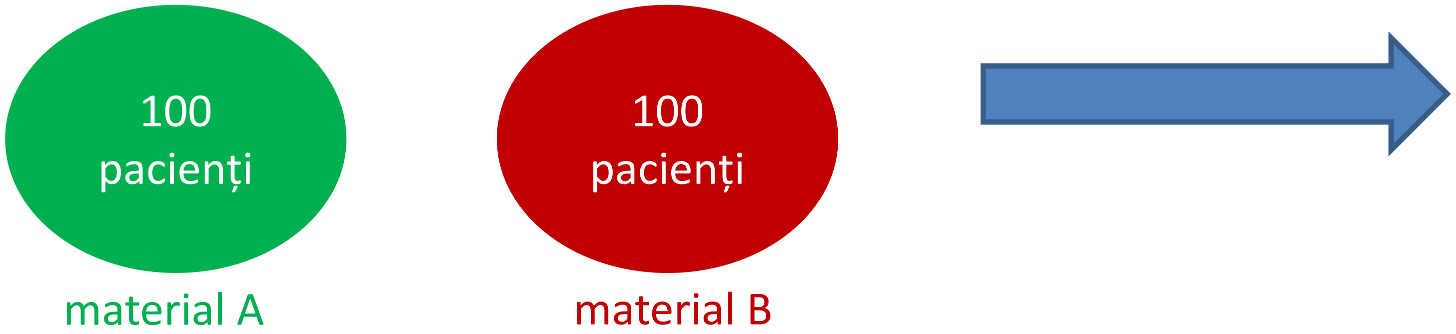
- Obiectiv: Compararea timpului de supraviețuire a protezei de șold în artroplastia totală de şold efectuată utilizând materialul A versus B versus C

când începem cercetarea nu cunoaștem supraviețuirea medie a protezei de șold realizată cu materialele A, B sau C.

scopul nostru: să găsim care material are cea mai îndelungată supraviețuire în artroplastia totală de şold

- ne propunem o diferență importantă clinic de 8 luni

- Metodă: realizăm un studiu unde să demonstrăm că există o diferență de minim 8 luni între supraviețuirea protezei de șold realizată cu materialul A versus B:
  - Calculăm dimensiunea necesară a eșantionului:
    - selectăm aleator 2 grupuri de 100 de pacienți



Rezultate:

100
pacienți

100 pacienți

material A:
media: 250 luni
s=50,8

material B:
media: 290 luni
s=101,55

Calculăm intervalele de încredere ale mediilor

- cu formula:

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

unde

$\bar{X}$ - media aritmetică

s – deviația standard

n – numărul de indivizi luați în studiu

Materialul A

Să calculăm pentru întreaga populație tratată cu artroplastia totală de şold efectuată utilizând proteză din materialul A

timpul **mediu** de supraviețuire a protezei de șold

n=100, media de supraviețuire=250 luni şi s=50,8

Calculăm intervalul de încredere de 95% pentru medie.

α =0,05, n ≥ 30, σ necunoscută, deci folosim formula:

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

$$[250 - 1{,}96\frac{50{,}8}{\sqrt{100-1}}; 250 + 1{,}96\frac{50{,}8}{\sqrt{100-1}}]$$

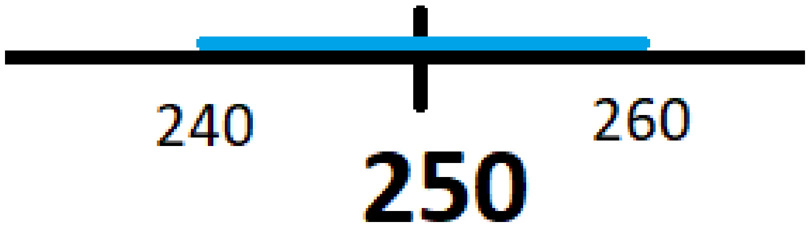
$$[250 - 10; 250 + 10]$$

[240; 260]

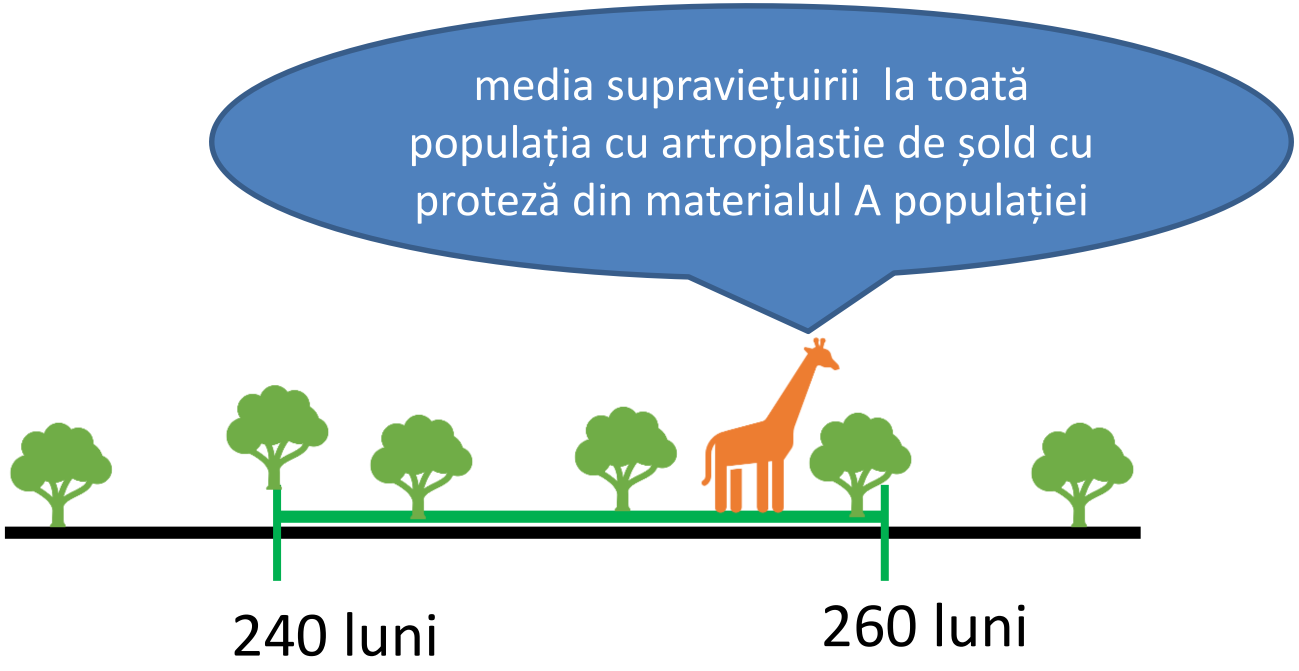
Răspuns: timpul mediu de supraviețuire a protezei de șold la materialul A este între 240 și 260 de luni cu o eroare de 5%

intervalul de încredere de 95% pentru medie [240; 260]

Material A



- timpul mediu de supraviețuire: 250 luni
- 95% Interval de încredere 240 – 260 luni



Interpretare: la întreaga populație timpul mediu de supraviețuire al protezei de șold executată cu materialul A va fi între 240 luni și 260 luni cu 5% eroare

Se poate ca timpul mediu de supraviețuire al protezei de șold executată cu materialul A la toată populația să fie <u>240 luni (cel mai rău caz</u>)

Se poate ca timpul mediu de supraviețuire al protezei de șold executată cu materialul A la toată populația să fie <u>260 luni (cel mai bun caz</u>)

Materialul B

Să calculăm pentru întreaga populație tratată cu artroplastia totală de şold efectuată utilizând proteză din materialul B

timpul **mediu** de supraviețuire a protezei de șold

n=100, media de supraviețuire=290 luni şi s=101,55

Calculăm intervalul de încredere de 95% pentru medie.

α =0,05, n ≥ 30, σ necunoscută, deci folosim formula:

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

$$[290 - 1{,}96 \frac{101{,}55}{\sqrt{100-1}}; 290 + 1{,}96 \frac{101{,}55}{\sqrt{100-1}}]$$

$$[290 - 20; 290 + 20]$$

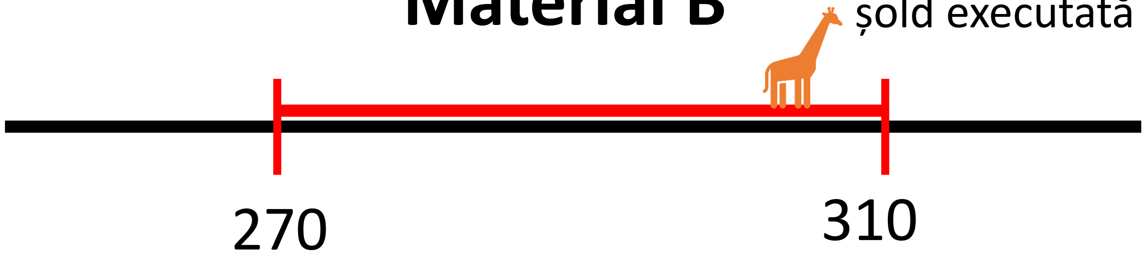
[270; 310]

Răspuns: timpul mediu de supraviețuire a protezei de șold la materialul B este între 270 și 310 de luni cu o eroare de 5%

intervalul de încredere de 95% pentru medie [270; 310]

Material B

timpul mediu de supraviețuire al protezei de șold executată cu materialul B în populație

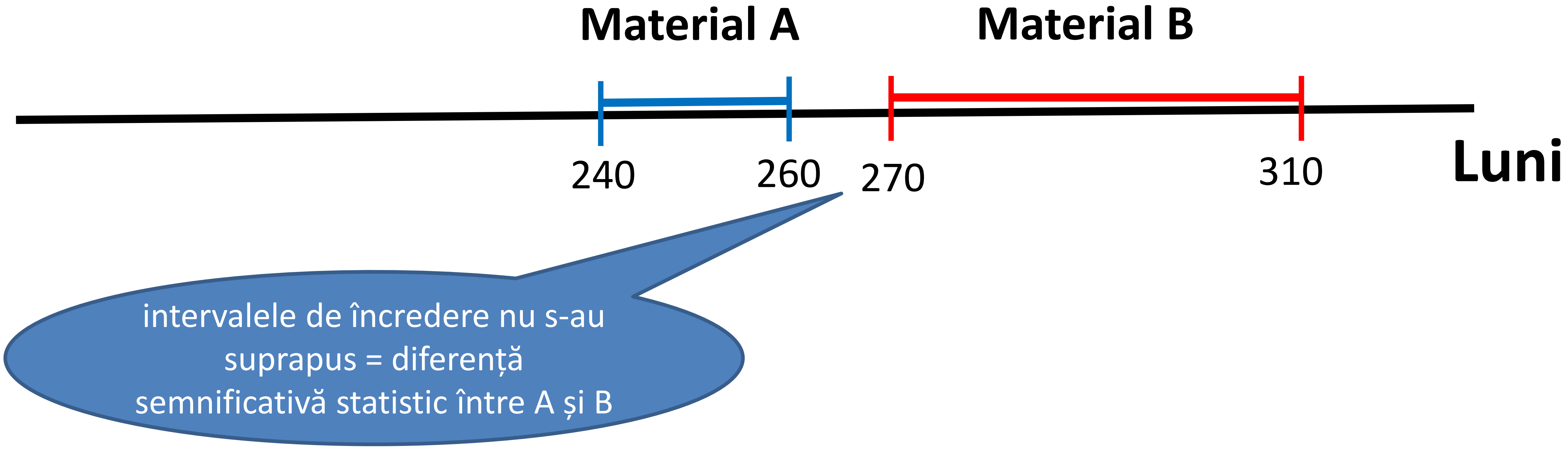


- timpul mediu de supraviețuire: 290 luni
- 95% Interval de încredere 270 – 310 luni

Interpretare: la întreaga populație timpul mediu de supraviețuire al protezei de șold executată cu materialul B va fi între 270 luni și 310 luni cu 5% eroare (nu știm exact cât va fi)

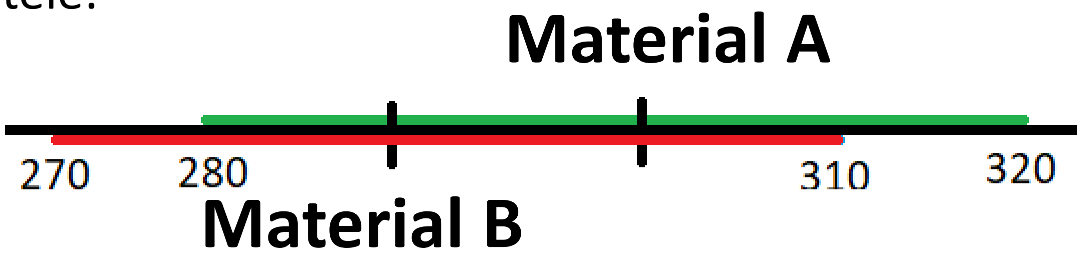
În cel mai rău caz, la toată populația timpul mediu de supraviețuire al protezei de șold executată cu materialul B poate fi 270 luni

În cel mai bun caz, la toată populația timpul mediu de supraviețuire al protezei de șold executată cu materialul B poate fi 310 luni



- <u>Concluzie:</u> Acest studiu a demonstrat că materialul B are o supraviețuire a protezei de șold semnificativ statistic mai mare decât materialul A (95% IC pentru materialul A nu se suprapune cu 95% IC pentru materialul B. Diferența de supraviețuire între materialul A și B este egală sau mai mare de 10 luni.)

Dacă acestea erau rezultatele:



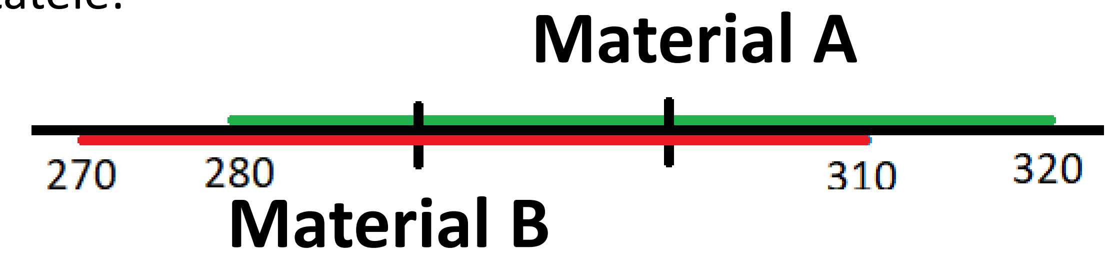
- 95% Interval de încredere pentru materialul A: 280 – 320 luni
- 95% Interval de încredere pentru materialul B: 280 – 310 luni

Interpretare:

la întreaga populație timpul mediu de supraviețuire al protezei de șold executată cu materialul A va fi între 280 luni și 320 luni cu 5% eroare

la întreaga populație timpul mediu de supraviețuire al protezei de șold executată cu materialul B va fi între 270 luni și 310 luni cu 5% eroare

Dacă acestea erau rezultatele:



intervalele de încredere s-au suprapus = nu există diferență semnificativă statistic între A și B

- <u>Concluzie:</u> Acest studiu a demonstrat că materialul A nu are o supraviețuire a protezei de șold semnificativ statistic diferită față de materialul B (95% IC pentru materialul A se suprapune cu 95% IC pentru materialul B. Nu există diferența de supraviețuire între materialul A și B.)

Comparația mediilor aritmetice a două eșantioane cu ajutorul testelor statistice

pe scurt

Metodă:

presupunem prin absurd că mediile populațiilor nu diferă (sunt egale)

calculăm probabilitatea p ca această afirmație să fie adevărată

dacă **$p < 0{,}05$**, atunci șansele ca afirmația să fie adevărată sunt foarte mici, o să spunem că afirmația este falsă, și **avem diferență semnificativă statistic între mediile populațiilor**

Dacă **$p \geq 0{,}05$**, atunci șansele ca afirmația să fie adevărată sunt mari, o să spunem că nu am putut demonstra că afirmația este falsă, **nu am putut demonstra că există diferență semnificativă statistic între mediile populațiilor**

Obiectiv:

- **există o diferență** semnificativă statistic între media timpului de supraviețuire în populație a protezei de șold executată cu materialul A și media timpului de supraviețuire în populație a protezei de șold executată cu materialul B

Pas 1. Formularea ipotezelor

- Obiectiv: **există o diferență** semnificativă statistic între media timpului de supraviețuire în populație a protezei de șold executată cu materialul A și media timpului de supraviețuire în populație a protezei de șold executată cu materialul B

- H0 - <u>ipoteza nulă:</u> **nu există o diferență** semnificativă statistic între media timpului de supraviețuire a protezei de șold executată cu materialul A și media timpului de supraviețuire a protezei de șold executată cu materialul B
- H1 – <u>ipoteza alternativă</u>: **există o diferență** semnificativă statistic între media timpului de supraviețuire a protezei de șold executată cu materialul A și media timpului de supraviețuire a protezei de șold executată cu materialul B

Alegerea testului statistic

- distribuția normală
  - de obicei ar trebui să verificăm distribuția,
  - dar acum facem această presupunere că există în ambele grupuri,
- variațe egale
  - de obicei ar trebui să verificăm,
  - facem această presupunere că sunt egale
- sunt comparate două medii aritmetice,
- grupurile sunt independente (persoane diferite)
- vom alege:
  - **Testul t Student pentru eșantioane independente în cazul variațelor egale**

Pas 3. Efectuarea testului

- Rezultat test p=0,0004

Concluzia:

presupunem prin absurd că mediile populațiilor nu diferă (sunt egale)

am calculat probabilitatea p=0,0004 ca această afirmație să fie adevărată

dacă **p < 0,05**, atunci șansele ca afirmația să fie adevărată sunt foarte mici, o să spunem că afirmația este falsă, și **avem diferență semnificativă statistic între mediile populațiilor**

~~Dacă **p ≥ 0,05**, atunci șansele ca afirmația să fie adevărată sunt mari, o să spunem că nu am putut demonstra că afirmația este falsă, **nu am putut demonstra că există diferență semnificativă statistic între mediile populațiilor**~~

Concluzia

- Deoarece p=0,0004 < 0,05
  - respingem H0, acceptăm H1:
    - **există o diferență** semnificativă statistic între media timpului de supraviețuire a protezei de șold executată cu materialul A și media timpului de supraviețuire a protezei de șold executată cu materialul B

Multumesc!!!