Aplicații: Cum este distribuția datelor?

Dacă aceste condiții sunt îndeplinite

- media ≈ mediana ≈ modulul
- simetria ≈ 0
- boltirea ≈ 0
- cvartilele 1 și 3 simetrice față de media aritmetică
- În intervalul medie±abatere standard Ǝ minim 68,2% din observaţii;
- În intervalul medie±2*abatere standard Ǝ minim 95,4% din observaţii;
- În intervalul medie±3*abatere standard Ǝ minim 99,7% din observaţii,

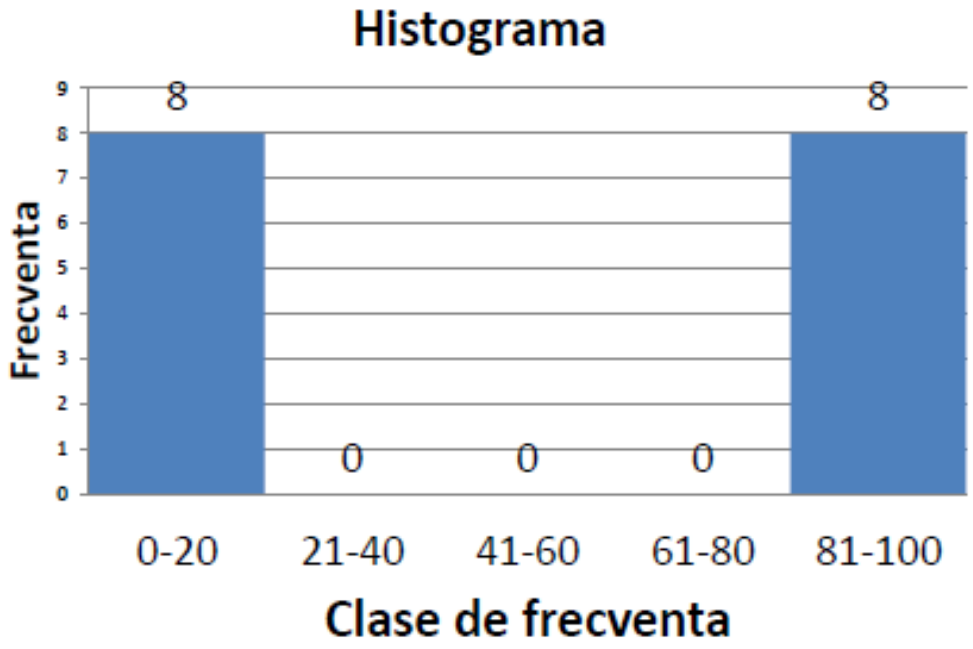
- atunci distribuția datelor obținute empiric se apropie de distribuția normală

- dacă oricare dintre condiții **nu** este îndeplinită
  - media ≈ mediana ≈ modulul
  - simetria ≈ 0
  - boltirea ≈ 0
  - cvartilele 1 și 3 simetrice față de media aritmetică
  - În intervalul medie±abatere standard Ǝ minim 68,2% din observaţii;
  - În intervalul medie±2*abatere standard Ǝ minim 95,4% din observaţii;
  - În intervalul medie±3*abatere standard Ǝ minim 99,7% din observaţii,
- atunci distribuția datelor obținute empiric nu se apropie de distribuția normală

Exemplu – Seria 1

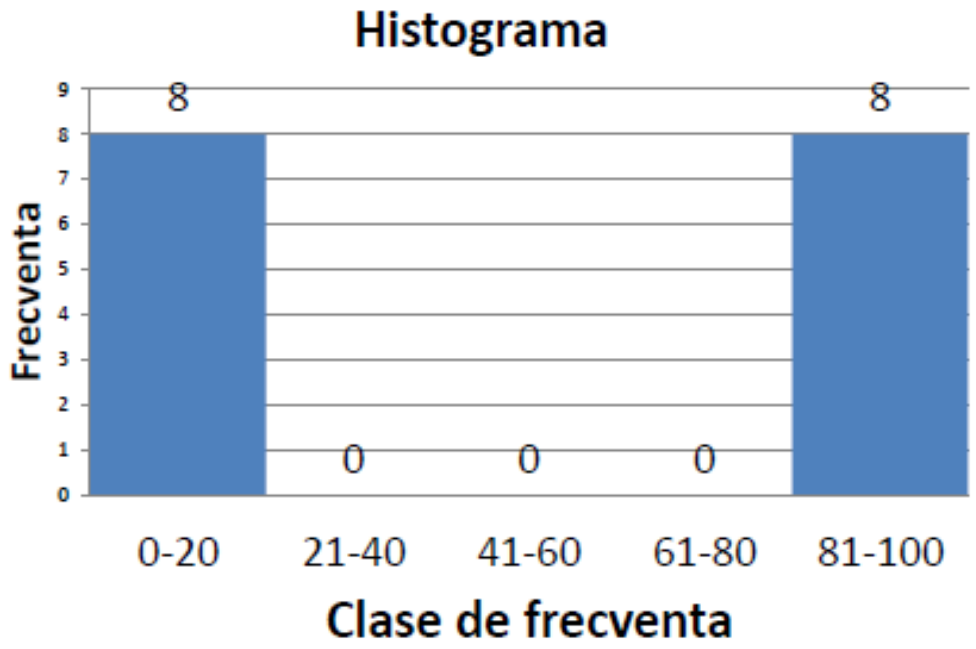
Seria 1
1
1
2
3
5
6
6
7
93
94
94
95
97
98
98
100

- Media aritmetică = 50
- Mediana = 50
- Modul – nu are
- Deviaţia standard = 47,70
- Cvartila 1 = 4,5
- Cvartila 3 = 95,5
- Simetria = 0,0002
- Boltirea = -2,29



Ne arată diferențe mari față de distribuția normală

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

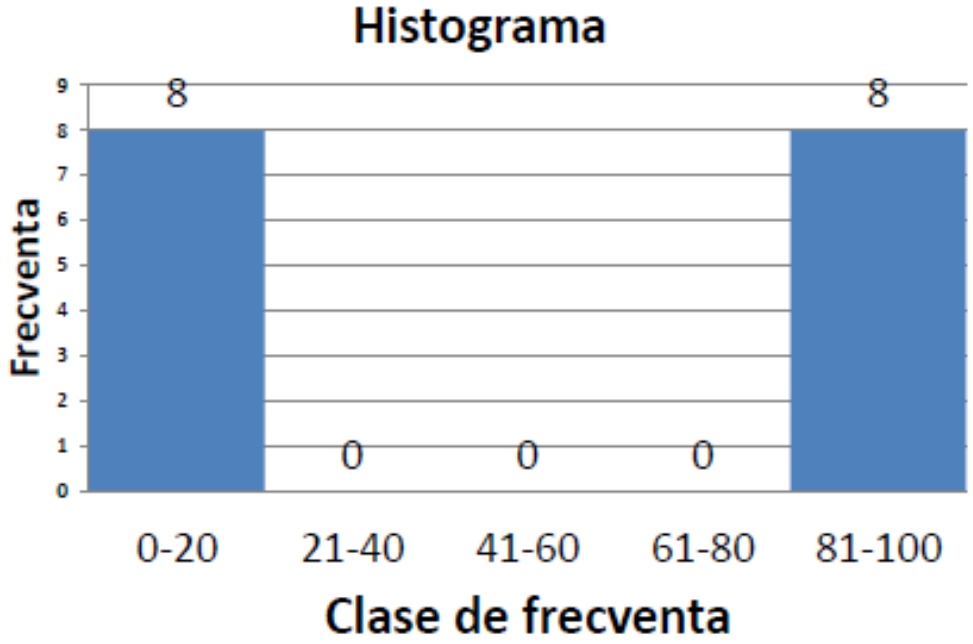


- Media aritmetică = 50
- Deviaţia standard = 47,70

Deviaţia standard foarte mare, concluzie: există date în cele două extreme

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

16



- Media aritmetică = 50
- Deviaţia standard = 47,70

- Media - deviaţia standard = 50 - 47,7 = 2,3
- Media + deviaţia standard = 50 + 47,7 = 97,7

- intervalul media ± deviaţia standard = [50 - 47,7; 50 + 47,7] = [2,3; 97,7]

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

16

- Media aritmetică = 50
- Deviaţia standard = 47,70

Media ± deviaţia standard = [50 - 47,7; 50 + 47,7] = [2,3; 97,7]

- In intervalul [2,3; 97,7] sunt 10 date, adica **62,5%** din date

10/16*100 = 62,5

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

Ca să fie distribuție normală:

in intervalul media ± deviaţia standard sunt minim 68,3% din date

in intervalul media ± 2*deviaţia standard sunt minim 95,4% din date

in intervalul media ± 3*deviaţia standard sunt minim 99,7% din date

- Media aritmetică = 50
- Deviaţia standard = 47,70

- Media ± deviaţia standard = [2,3; 97,7]
- In intervalul [2,3; 97,7] sunt 10 date, adica 62,5% din date

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

Ca să fie distribuție normală:

in intervalul media ± deviaţia standard sunt minim 68,3% din date

in intervalul media ± 2*deviaţia standard sunt minim 95,4% din date

in intervalul media ± 3*deviaţia standard sunt minim 99,7% din date

- Media aritmetică = 50
- Deviaţia standard = 47,70

- Media ± deviaţia standard = [2,3; 97,7]
- In intervalul [2,3; 97,7] sunt 10 date, adica 62,5% din date

62,5% < 68,3%, deci distributia nu este normala

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

16

-45,39 – date medicale negative ..... nu prea are sens

Medie ± 2*deviația standard = [50-2*47,7; 50+2*47,7] = **[-45,39**; 145,39]

in intervalul [-45,39; 145,39] sunt 16 valori, e.g. 16/16 = 100% dintre date

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

16

Ca să fie distribuție normală:

in intervalul media ± deviaţia standard sunt minim 68,3% din date

in intervalul media ± 2*deviaţia standard sunt minim 95,4% din date

in intervalul media ± 3*deviaţia standard sunt minim 99,7% din date

Medie ± 2*deviația standard = [50-2*47,7; 50+2*47,7] = [-45,39; 145,39]

in intervalul [-45,39; 145,39] sunt 16 valori, adica 16/16 = 100% dintre date

100% > 95,4 proprietatea e îndeplinită pentru acest interval

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

16

Ca să fie distribuție normală:

in intervalul media ± deviaţia standard sunt minim 68,3% din date

in intervalul media ± 2*deviaţia standard sunt minim 95,4% din date

in intervalul media ± 3*deviaţia standard sunt minim 99,7% din date

Media aritmetică = 50

Deviaţia standard = 47,70

Media ± deviaţia standard = [2,3; 97,7] cu 62,5% dintre date

Mean ± 2*st.dev = [-45,39; 145,39] sunt 16 valori, adica 100% dintre date

Mean ± 3*st.dev = [50-3*47,7; 50+3*47,7] = [-93,09; 193,09] sunt 16 valori, adică 16/16 = 100% dintre date

100% > 99,7 proprietatea e îndeplinită pentru acest interval

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

Ca să fie distribuție normală:

in intervalul media ± deviaţia standard sunt minim 68,3% din date

in intervalul media ± 2*deviaţia standard sunt minim 95,4% din date

in intervalul media ± 3*deviaţia standard sunt minim 99,7% din date

16 Media aritmetică = 50

Deviaţia standard = 47,70

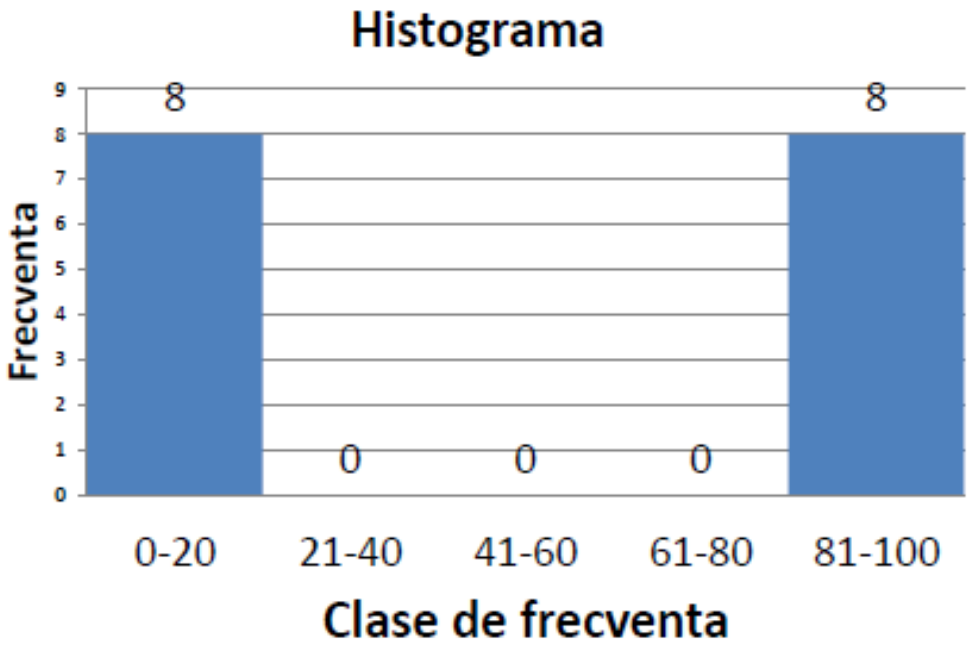
Media ± deviaţia standard = [2,3; 97,7] cu 10 valori, adică 62,5% dintre date

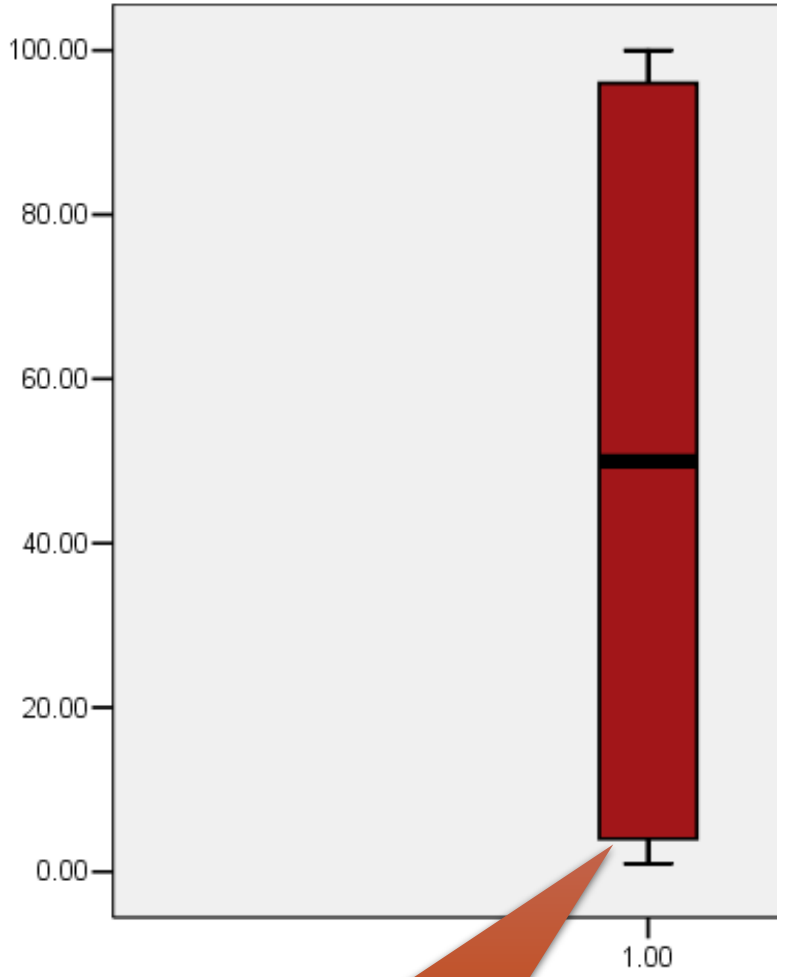
Mean ± 2*st.dev = [-45,39; 145,39] sunt 16 valori, adica 100% dintre date

Mean ± 3*st.dev = [-93,09; 193,09] sunt 16 valori, adică 16/16 = 100% dintre date

Distribuția nu este apropiată de cea normală

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





Între minim şi percentila 25 este o distanţă mică = în acest interval avem multe date, comparativ cu intervalul următor

Exemplu – Seria 2

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

Media aritmetică = 50

Mediana = 50

Modul = multimodală

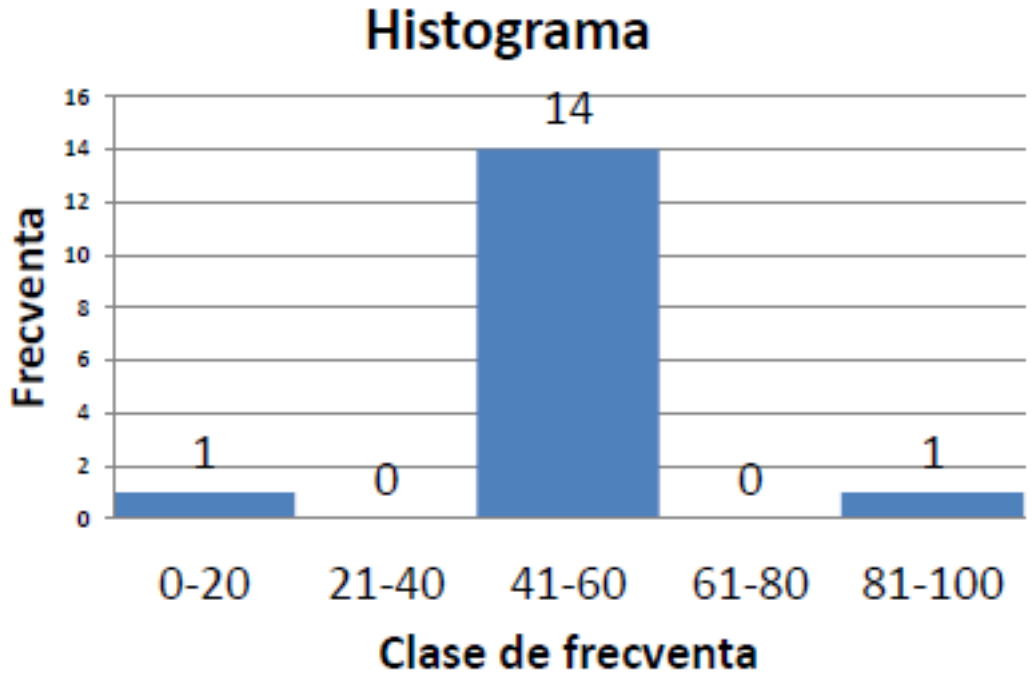
Deviaţia standard = 18,37

Cvartila 1 = 47,5

Cvartila 3 = 52,5

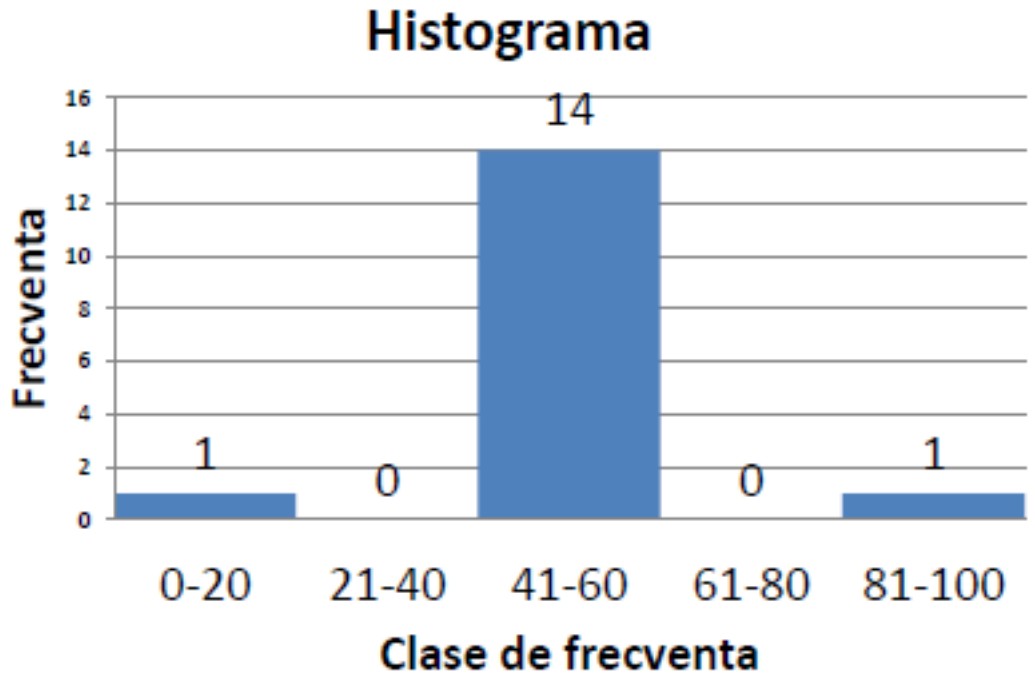
Simetria = 0,09

Boltirea = 6,81



Ne arată diferențe mari față de distribuția normală

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



Ca să fie distrib. normală:

Minim 68,3% din date

Minim 95,4% din date

Minim 99,7% din date

Deviaţia standard este mică, concluzie: cazurile sunt apropiate de medie

Media aritmetică = 50

Deviaţia standard = 18,37

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

16

Media aritmetică = 50
Deviaţia standard = 18,37

Media ± dev.st = [50-18,37; 50+18,37] = [31,63; 68,37]

in intervalul [31,63; 68,37] sunt 14 valori, adica 16/16 = 87,5% din date

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

16

Ca să fie distrib. normală:

Minim 68,3% din date

Minim 95,4% din date

Minim 99,7% din date

Media aritmetică = 50
Deviaţia standard = 18,37

Media ± dev.st = [50-18,37; 50+18,37] = [31,63; 68,37]
in intervalul [31,63; 68,37] sunt 14 valori, adica 16/16 = 87,5% din date

87,5 > 68,3, deci există minim 68,3% din date

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

16

Ca să fie distrib. normală:

Minim 68,3% din date

Minim 95,4% din date

Minim 99,7% din date

Media aritmetică = 50

Deviaţia standard = 18,37

Media ± dev.st = [31,63; 68,37] sunt 87,5% din date

Media ± 2*dev.st. = [50-2*18,37; 50+2*18,37] = [13,26; 86,74] sunt tot 14 date, adica 14/16 = **87,5%** din date, mai putine decat 95,4%
deci seria 2 nu este distribuita normal

Media ± 3*dev.st. = [-5,11; 105,11] sunt 100% din date

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

16

Ca să fie distrib. normală:

Minim 68,3% din date

Minim 95,4% din date

Minim 99,7% din date

Media aritmetică = 50

Deviaţia standard = 18,37

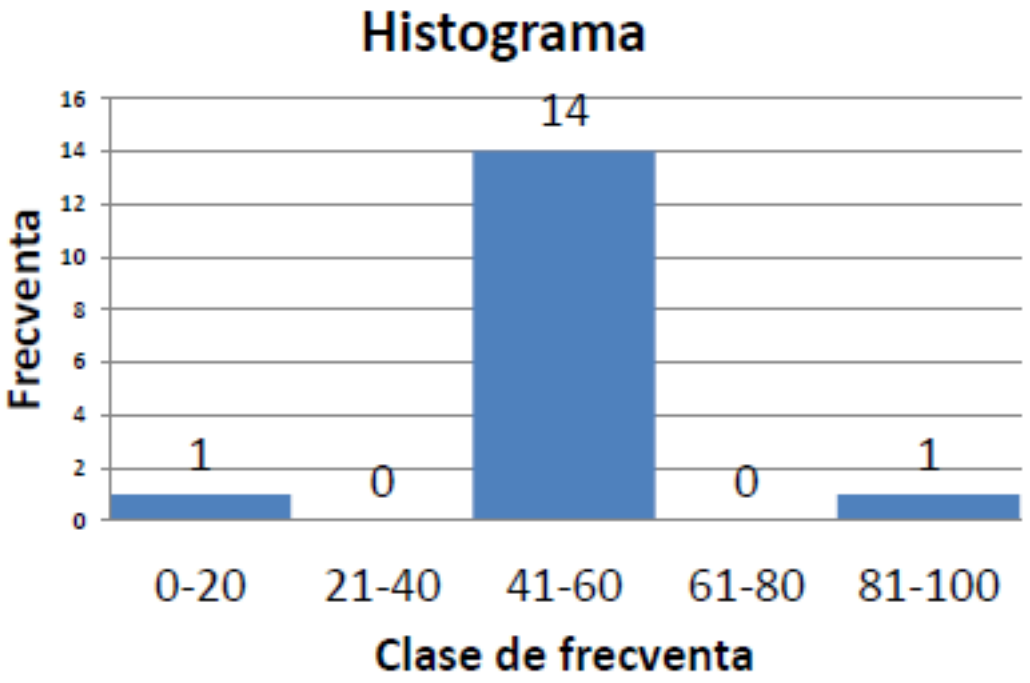
Media ± dev.st = [31,63; 68,37] sunt 14 valori - 87,5% din date

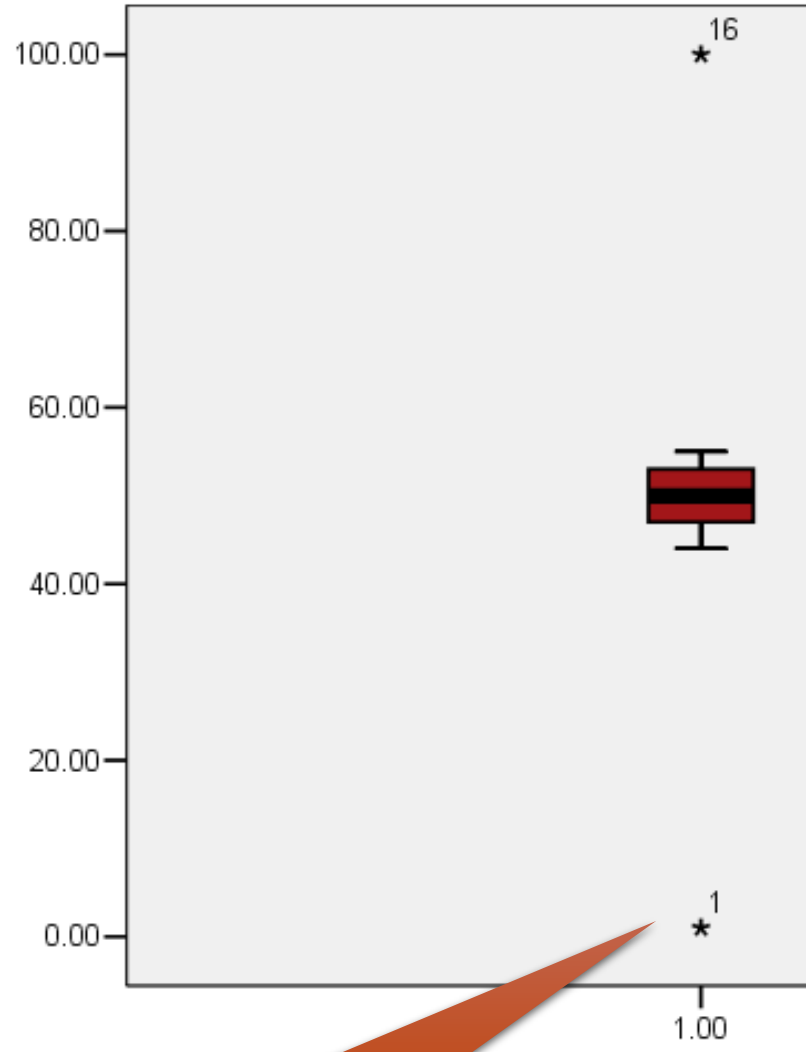
Media ± 2*dev.st. = [13,26; 86,74] sunt 14 valori - 87,5% din date

Media ± 3*dev.st. = [-5,11; 105,11] sunt 16 valori - 100% din date

Distribuția nu este apropiată de cea normală

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





Caz extrem

Exemplu – Seria 3

Seria 3
1
11
24
29
36
41
45
50
50
55
59
64
71
76
88
100

Media aritmetică = 50

Mediana = 50

Modul = 50

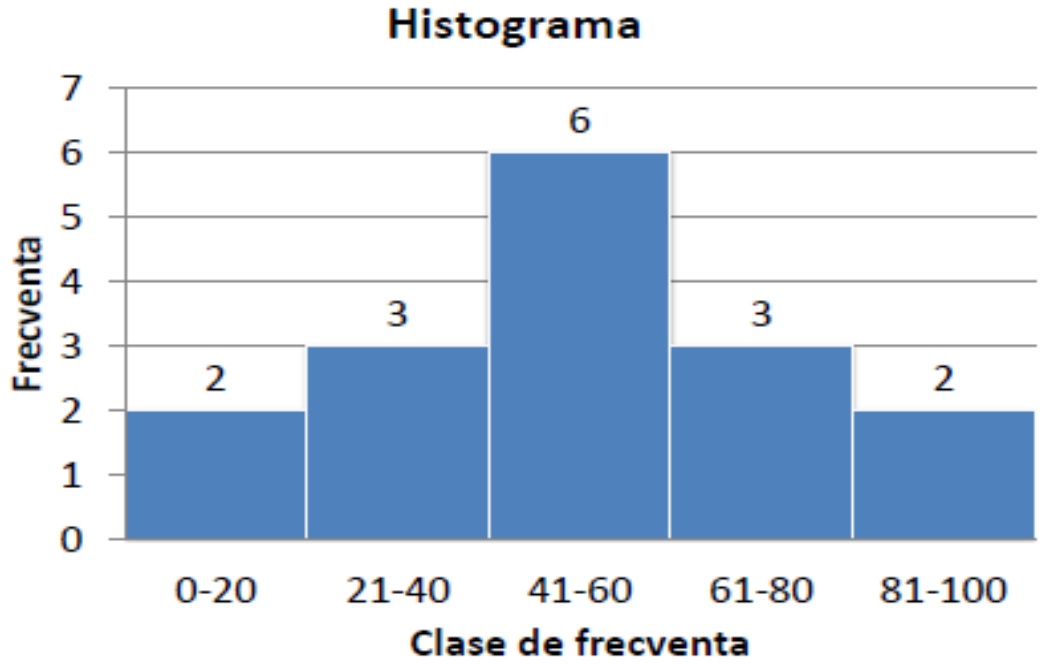
Deviaţia standard = 26,71

Cvartila 1 = 34,25

Cvartila 3 = 65,75

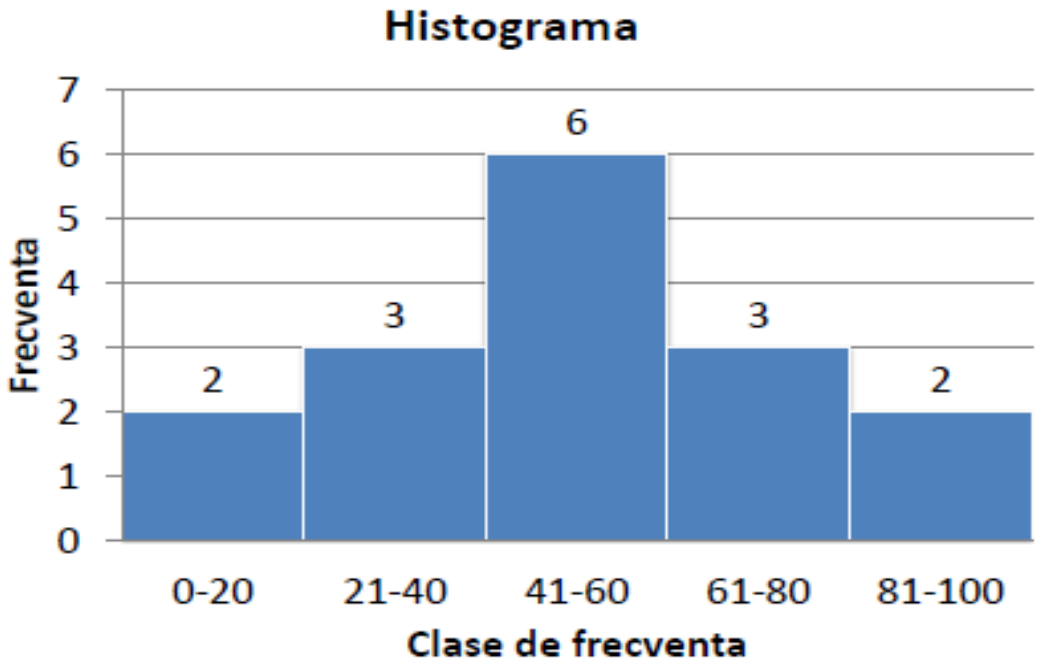
Simetria = 0,01

Boltirea = -0.23



Distribuția este apropiată de cea normală

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



Ca să fie distrib. normală:

Minim 68,3% din date

Minim 95,4% din date

Minim 99,7% din date

Distribuția este apropiată de cea normală

Media aritmetică = 50

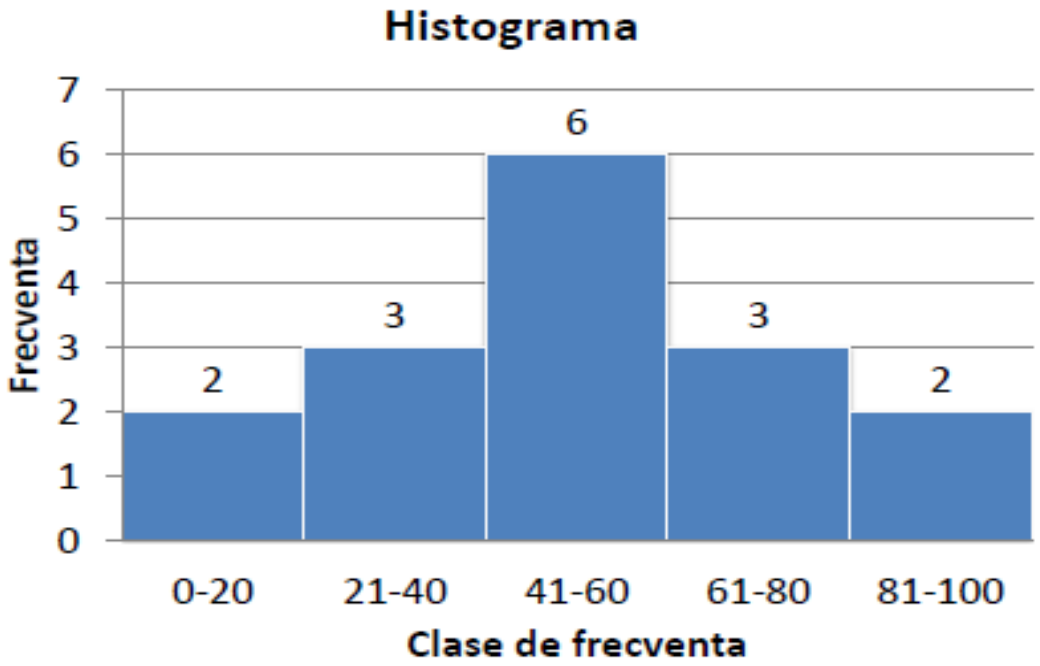
Deviaţia standard = 26,71

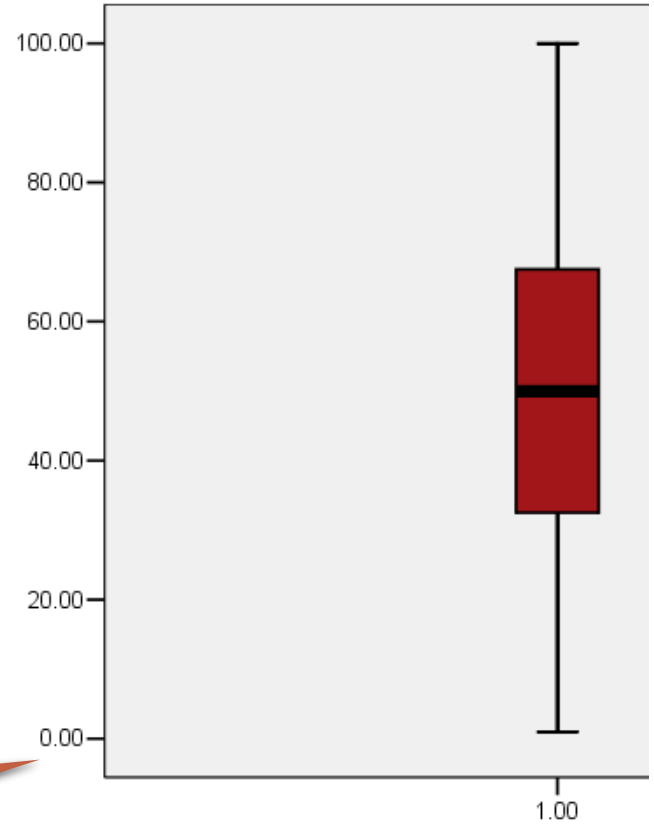
Media ± dev.st. = [23,28; 76,72] sunt 87,5% din date

Media ± 2*dev.st. = [-3,43; 103,43] sunt 100% din date

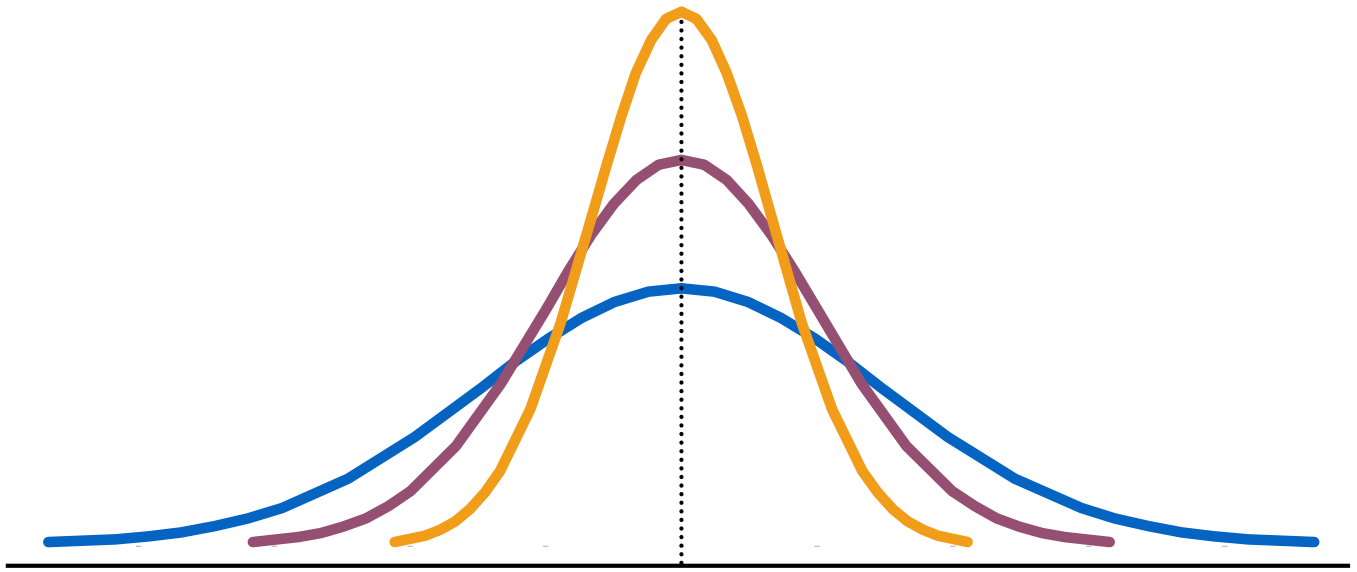
Media ± 3*dev. st. = [-30,15; 130,15] sunt 100% din date

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





Distribuția este apropiată de cea normală



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 sunt mai multe încercări

- cu atât rezultatele sunt mai apropiate de distribuția teoretică

- Ex.

Nașterea unui copil de sex feminin - 0,5 probabilitate teoretică

selectăm subiecți născuți în 2003

- din 10 – 6 de sex feminin — eroare 10%
- din 1000 – 510 de sex feminin — eroare 1%
- din 1.000.000 – 500.010 de sex feminin — eroare 0,001%

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

- Noi dorim un eșantion cât mai mic: rapiditate, cost, erori de măsurare etc.

- Câți subiecți trebuie să selectăm ca să avem un rezultat ce aproximează bine frecvența/media din populație?

talia eșantionului poate fi calculată

formule diferite pentru obiective diferite

! e nevoie de un statistician

Stabilirea taliei eșantionului

pentru comparea a două proporții http://statpages.org/proppowr.html

Eroarea 5%

Puterea studiului 80

Significance Level (alpha):	0.05	(Usually 0.05)
Power (% chance of detecting):	80	(Usually 80)
Group 1 Population Proportion:	.30	(Between 0.0 and 1.0)
Group 2 Population Proportion:	.50	(Between 0.0 and 1.0)
Relative Sample Sizes Required (Group 2 / Group 1):	1.0	(For equal samples, use 1.0)

diferența așteptată între proporții =20%

Compute

Sample Size Required

	Group 1	Group 2	Total
"Classical" Calculation:	93	93	186
With Continuity Correction:	103	103	206

Stabilirea taliei eșantionului
- pentru compararea a două medii (distributie normală)
http://sampsize.sourceforge.net/iface/s2.html#nm

diferența asteptată între medii = 230-210=20

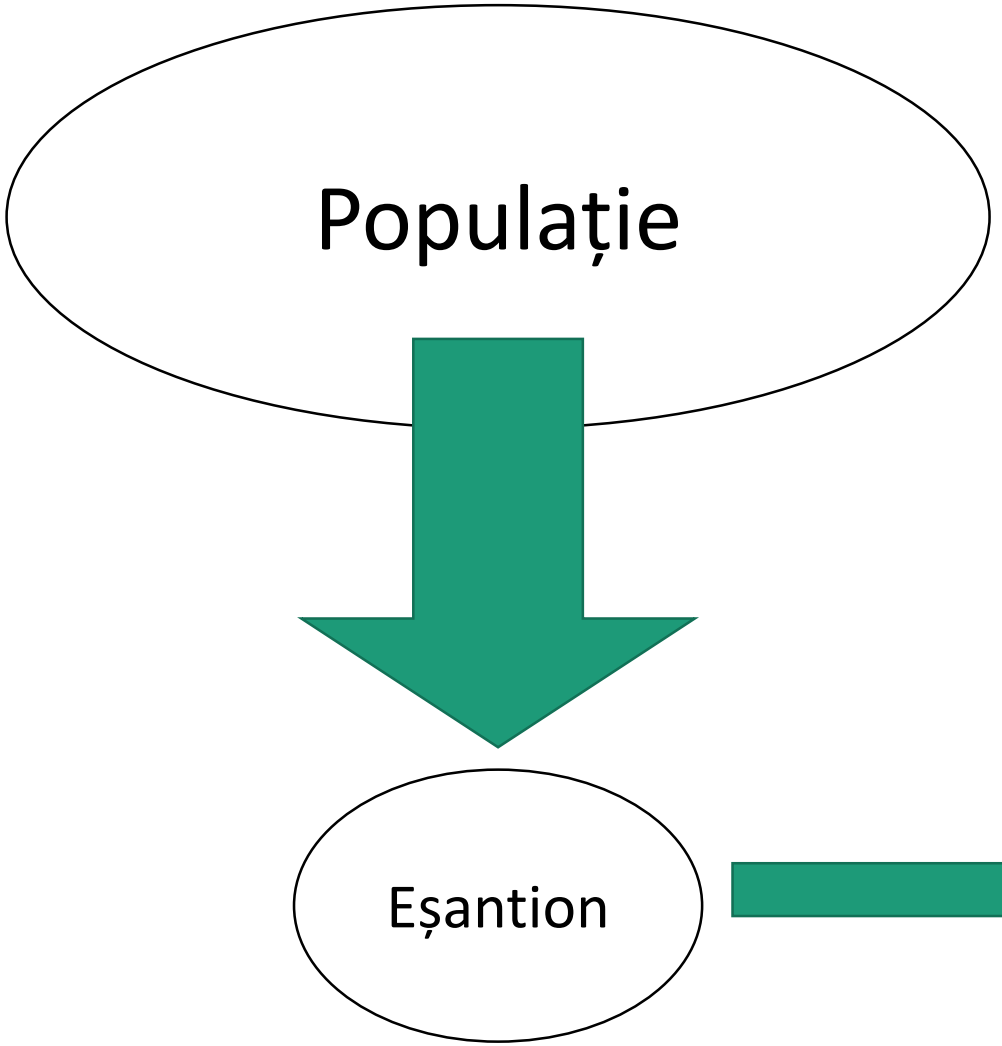
```
Assumptions:

    alpha =           5  (two-sided)
    power =          90
       m1 =         230
       m2 =         210
      sd1 =          26
      sd2 =          33
    n2/n1 =           1

Estimated sample size:

       n1 =          47
       n2 =          47
```

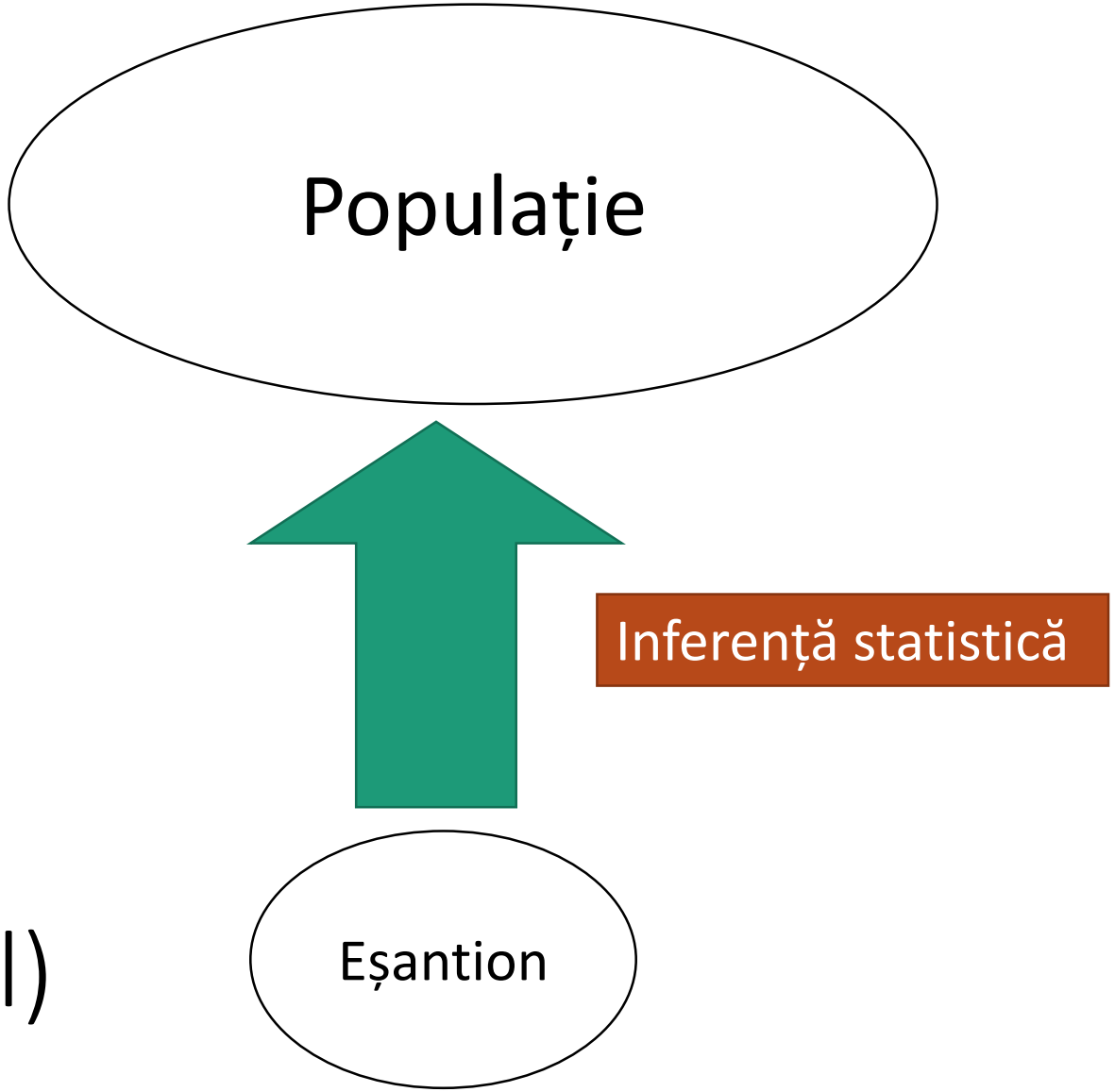
Distribuția de eșantionare



Necunoscută (inaccesibilă)

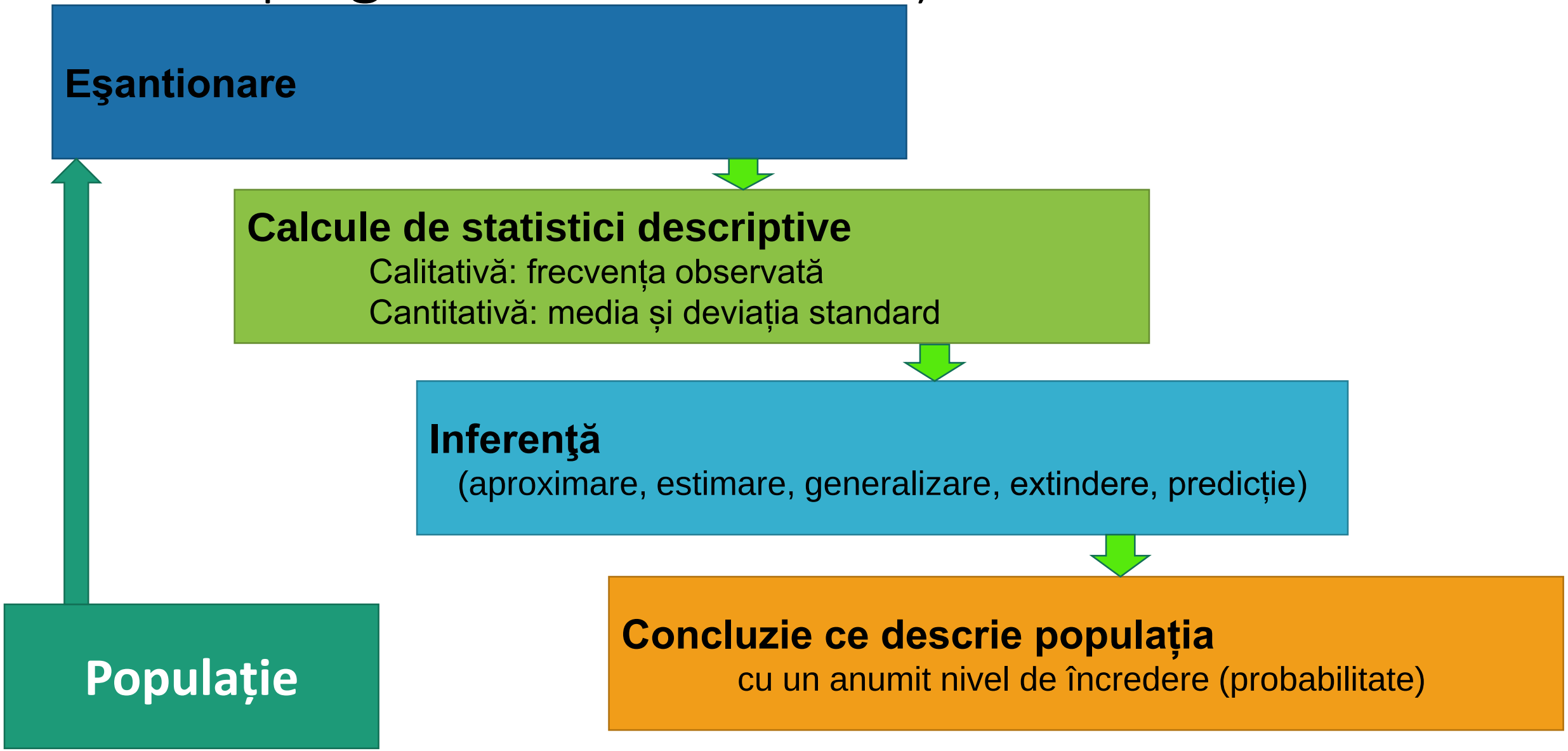
Cunoscut (accesibil)

Cunoscută prin estimare



Cunoscut (accesibil)

Principii generale în inferența statistică



Estimare

- culegem strugurii de pe un ar, cantitatea culeasă ne ajută să estimăm intreaga producție din acel an
- vânzările din ultimii 5 ani ne ajută să estimăm cât vor fi vănzările anul viitor

De ce estimăm? estimare = aproximare = predicție

Obiectiv: Dorim să cunoaștem efectul unui tratament antiviral Ivermectină asupra mortalității la bolnavii cu Covid-19 în comparație cu tratamentul standard

- În populația cu Covid-19 efectul Ivermectinei este necunoscut

Selectăm un eșantion de 200 pacienți. O parte dintre pacienți sunt alocați unui tratament, ceilalți celui de-al doilea tratament

Diferența de mortalitate găsită pe eșantion de -0,02% în favoarea Ivermectinei

Cum am putea să generalizăm la populație?

Cum am putea să generalizăm la populație?

O estimăm direct:

-0,02 va fi și în populație

Se numește

estimare punctuală

Definiţia unui estimator

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

  estimarea punctuală a parametrului populației

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

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

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

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

Scenariu

- Obiectiv
  - pentru băieții în vârstă de 2 ani

    media greutății = μ necunoscută

Populația
Caracteristica

Populație: băieți de 2 ani

- Selectăm în eșantion 100 de băieți de 2 ani aleși la întâmplare.
- Măsurăm greutatea.

Media greutății $\bar{X}_1$ = **12** kg

Repetăm studiul

- Încă odată selectăm 100 de băieți de 2 ani aleși la întâmplare.
- Măsurăm greutatea.

!!! Altă medie

- Media greutății $\bar{X}_2$ = **12.25** kg

Repetăm studiul pe <u>toate</u> eșantioanele de 100 de băieți posibile

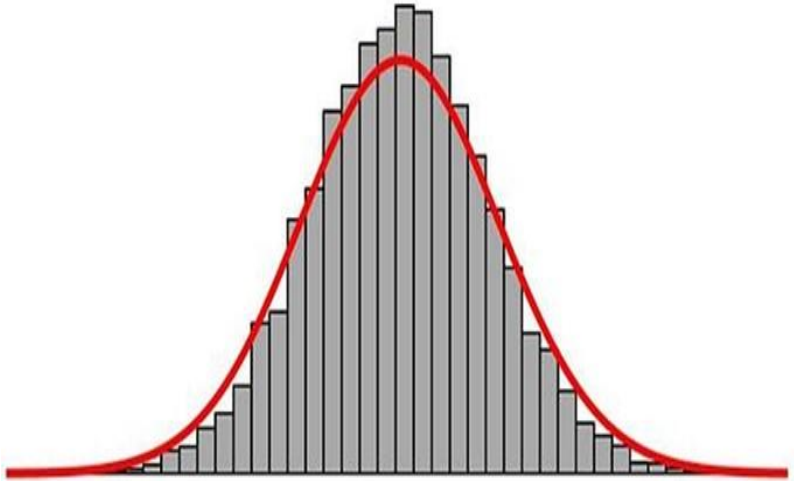
$\overline{X_1}$=14 kg

$\overline{X_2}$= 14,25 kg

$\overline{X_3}$= 14,5 kg

……

m ori $\overline{X_m}$= 14 kg



Distribuția mediilor provenite din studiile repetate

numită <u>distribuția de eșantionare</u>

Media $\bar{X}$
a mediilor $\overline{X_1}$, $\overline{X_2}$, …, $\overline{X_m}$

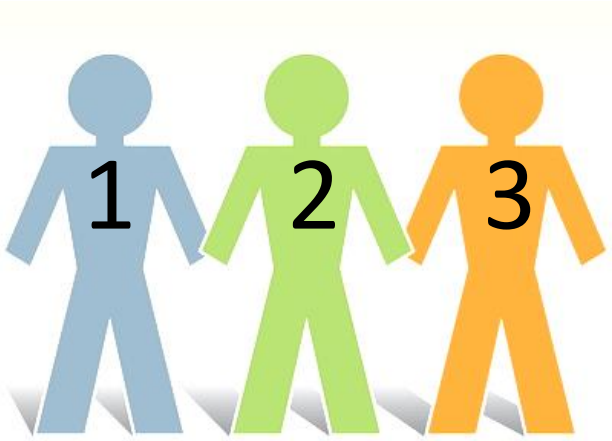
urmează distribuția normală

Repetarea studiului pe toate esantioanele posibile

- Ex. Populație alcătuită din **3 persoane:** 1, 2, 3

- Câte eșantioane de **2 persoane** putem alcătui?

1 cu 1	1 cu 2	1 cu 3
2 cu 1	2 cu 2	2 cu 3
3 cu 1	3 cu 2	3 cu 3



- 9 eșantioane

Repetarea studiului pe toate esantioanele posibile

- Ex. Populație alcătuită din **4 persoane** 1, 2, 3, 4

- Câte eșantioane de **2 persoane** putem alcătui?

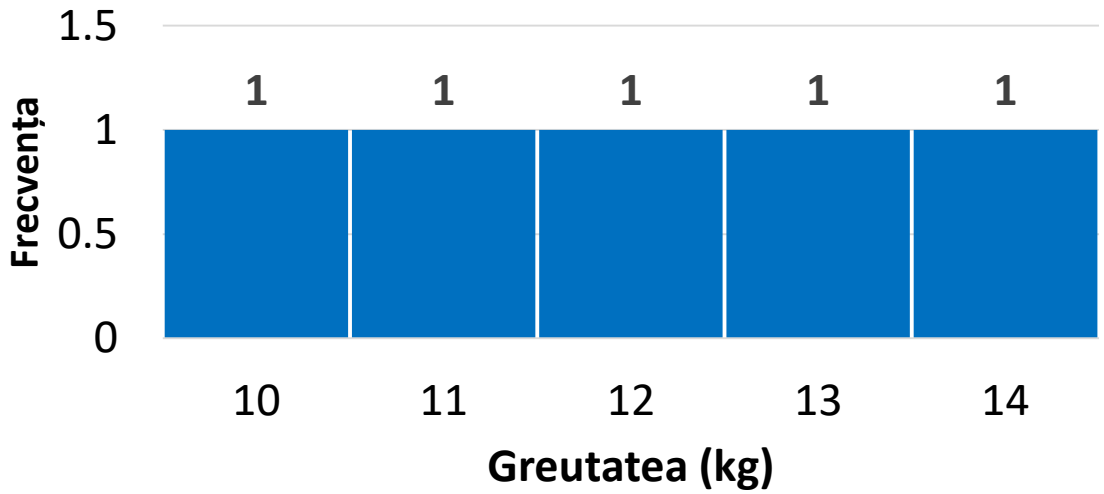
11	12	13	14
21	22	23	24
31	32	33	34
41	42	43	44

- 16 eșantioane

Scenariu

5 băieți de 2 ani: valorile greutății 10, 11, 12, 13, 14 kg

- Aceasta este <u>întreaga populație</u> 5 băieți



- Dacă luăm toate eșantioanele de 2 băieți:

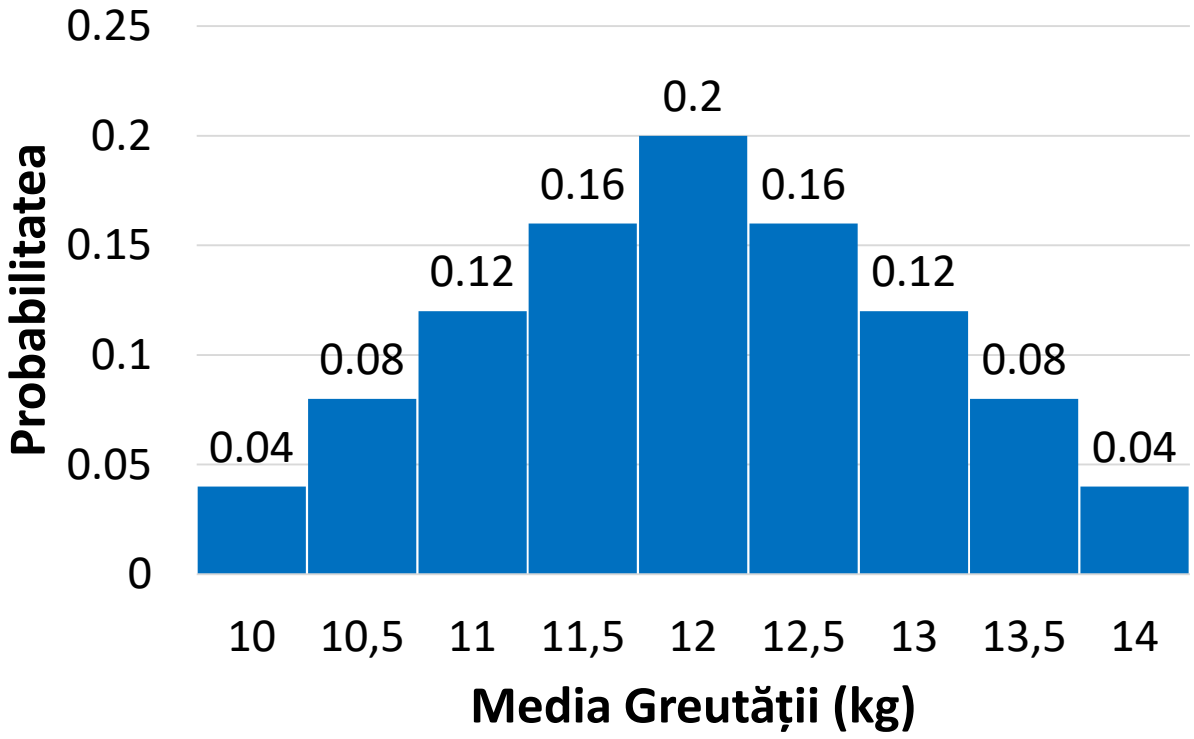
Dacă luăm toate eșantioanele de 2 băieți: **25 de eșantioane**

Primul băiat	Al doilea băiat
1	1
1	2
1	3
1	4
1	5
2	1
2	2
2	3
2	4
2	5
3	1
3	2
3	3
3	4
3	5
4	1
4	2
4	3
4	4
4	5
5	1
5	2
5	3
5	4
5	5

Primul băiat	Al doilea băiat	Greutatea pentru primul băiat	Greutatea pentru al doilea băiat	Media
1	1	10	10	10
1	2	11	10	10,5
1	3	12	10	11
1	4	13	10	11,5
1	5	14	10	12
2	1	10	11	10,5
2	2	11	11	11
2	3	12	11	11,5
2	4	13	11	12
2	5	14	11	12,5
3	1	10	12	11
3	2	11	12	11,5
3	3	12	12	12
3	4	13	12	12,5
3	5	14	12	13

Continuă pe slide-ul următor

Primul băiat	Al doilea băiat	Greutatea pentru primul băiat	Greutatea pentru al doilea băiat	Media
4	1	10	13	11,5
4	2	11	13	12
4	3	12	13	12,5
4	4	13	13	13
4	5	14	13	13,5
5	1	10	14	12
5	2	11	14	12,5
5	3	12	14	13
5	4	13	14	13.5
5	5	14	14	14

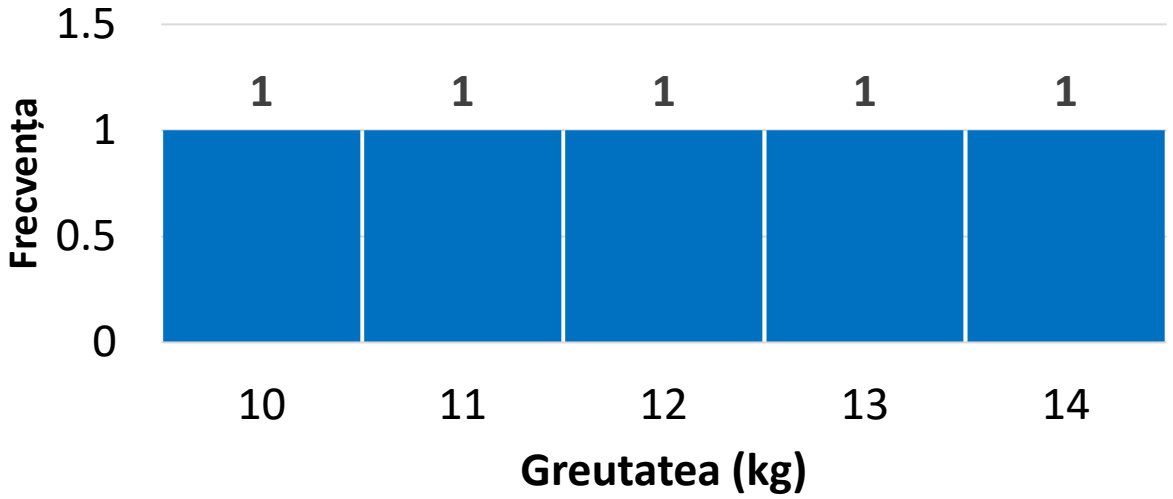


Distribuția de eșantionare (a mediilor)

Media de eșantionare = $\bar{X}$ = 12

Deviația standard de eșantionare = s = 1,02

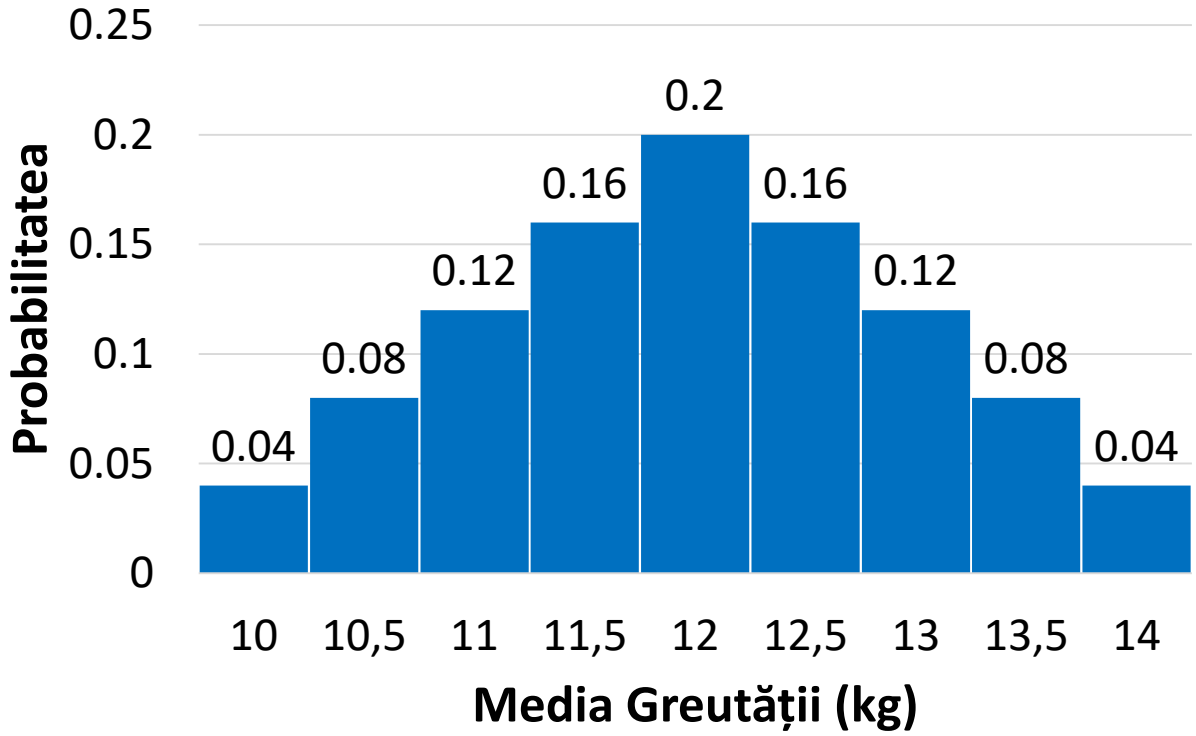
Urmează distribuția normală



Distribuția populației

Media populației = μ = 12 kg

Deviația standard = σ = 1,58

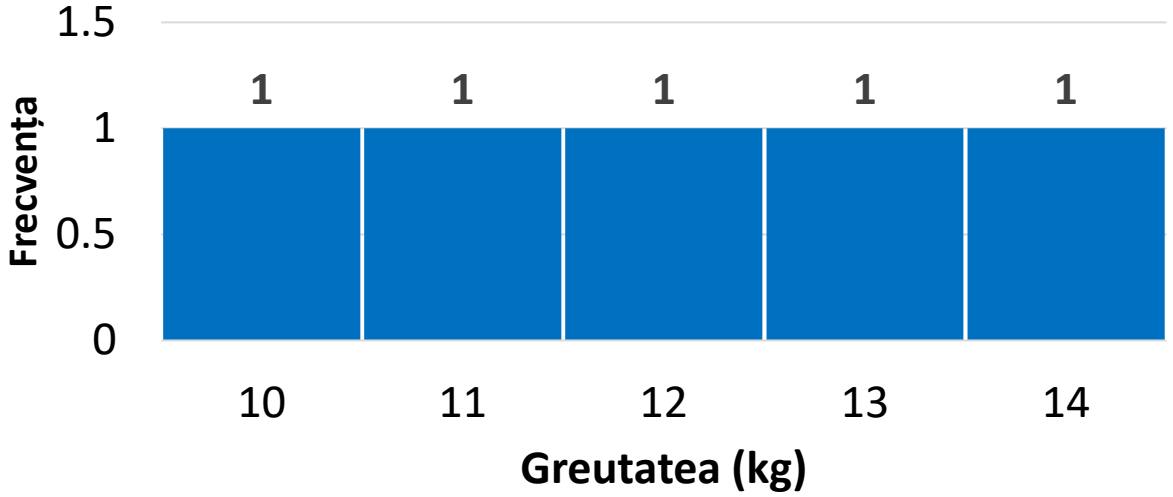


Distribuția de eșantionare (a mediilor)

Media de eșantionare = $\bar{X}$ = 12

Deviația standard de eșantionare = s = 1,02

Urmează distribuția normală

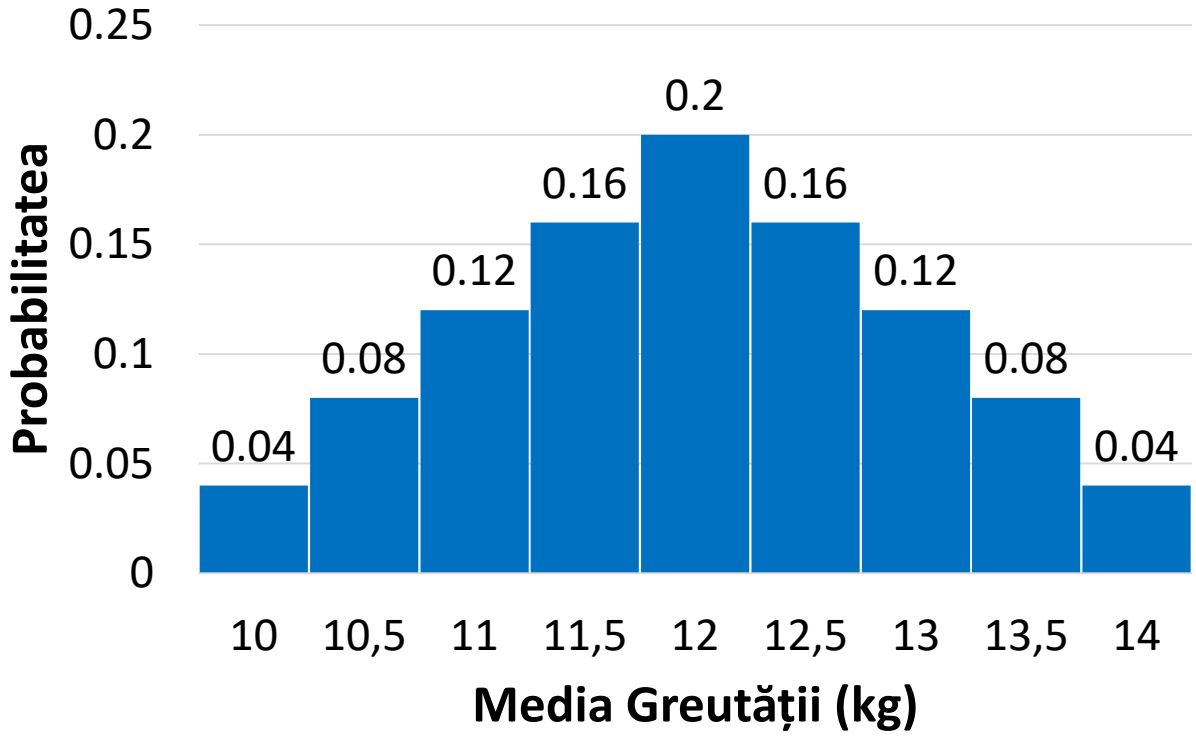


Distribuția populației

Media populației = $\mu = 12$ kg

Deviația standard = $\sigma = 1{,}58$

Surpriză au aceeași medie

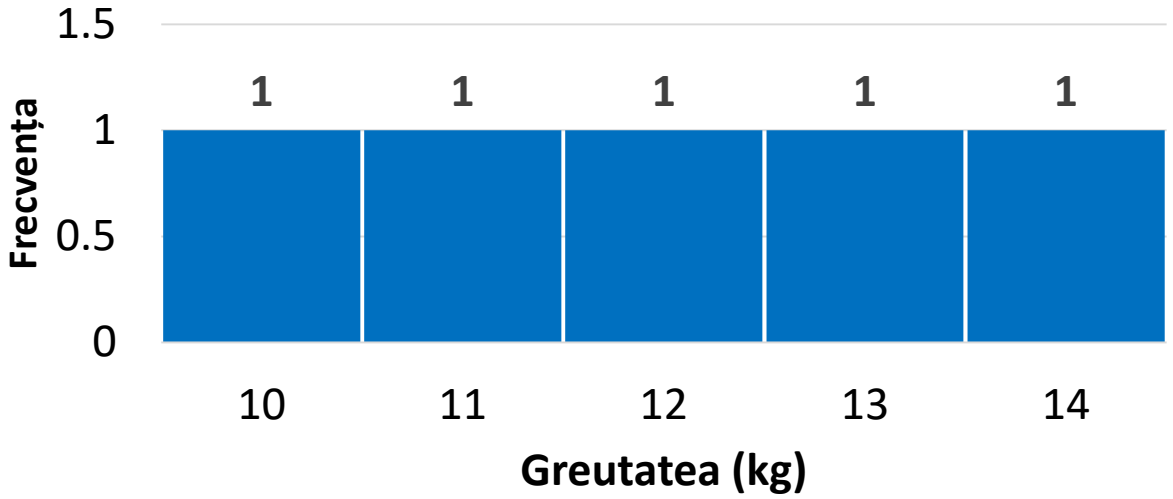


Distribuția de eșantionare (a mediilor)

Media de eșantionare = $\bar{X} = 12$

Deviația standard de eșantionare = s = 1,02

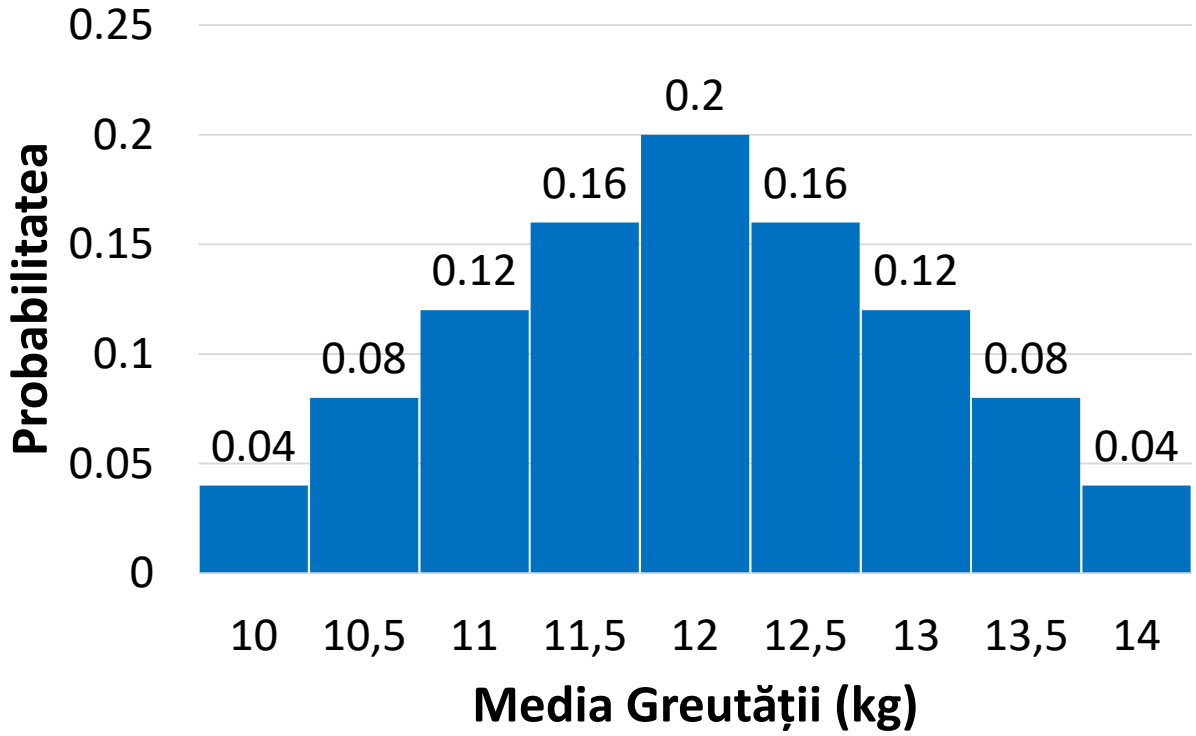
Urmează distribuția normală



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

Surpriză

$s=\sigma / \sqrt{n}$

Urmează distribuția normală

Teorema limitei centrale

Fie o populație cu media populației μ și deviația standard σ, atunci o distribuție de eșantionare (a mediilor) bazată pe *repetarea studiului* pe un eșantion de mărime n cu proprietățile:

Media distribuției de eșantionare $\overline{X} = \mu$ media populației

Deviația standard a distribuției de eșantionare este **eroarea standard** = $\sigma / \sqrt{n}$,

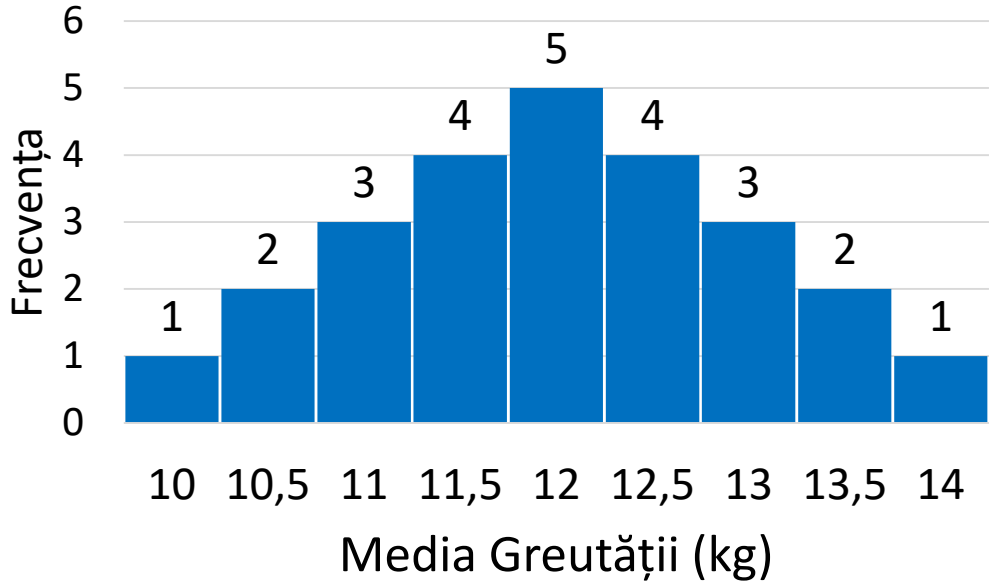
- Dacă distribuția populației este normală, atunci distribuția de eșantionare este normală.
- Dacă eșantionul este suficient de mare, atunci distribuția de eșantionare se apropie de distribuția normală indiferent de distribuția populației.

wooclap

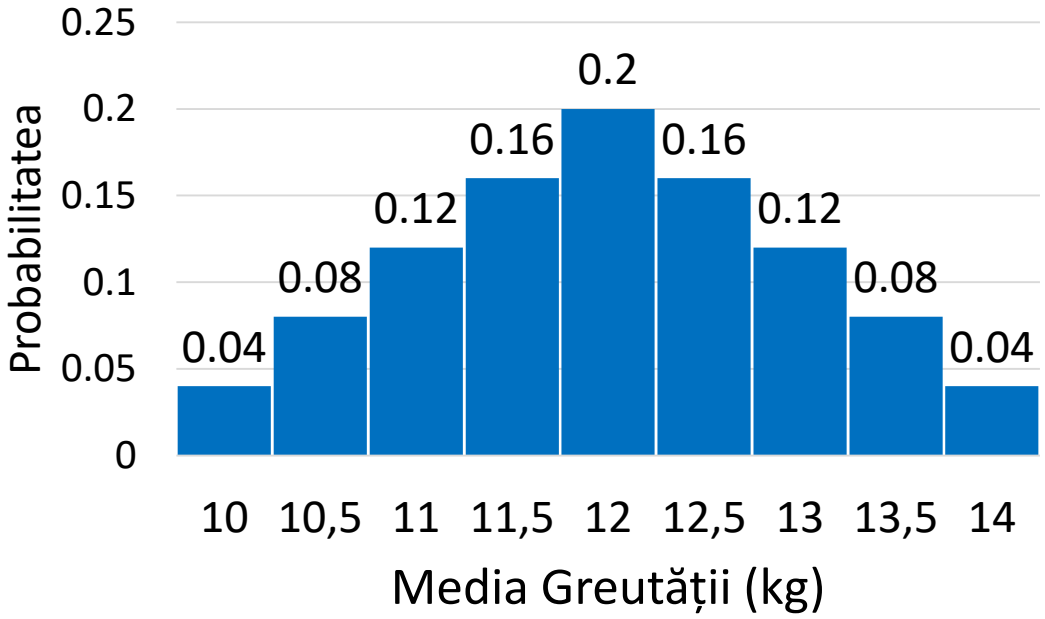
- https://app.wooclap.com/BFKRI06?from=event-page

Estimări

ex. Distribuția de eșantionare (a mediilor)



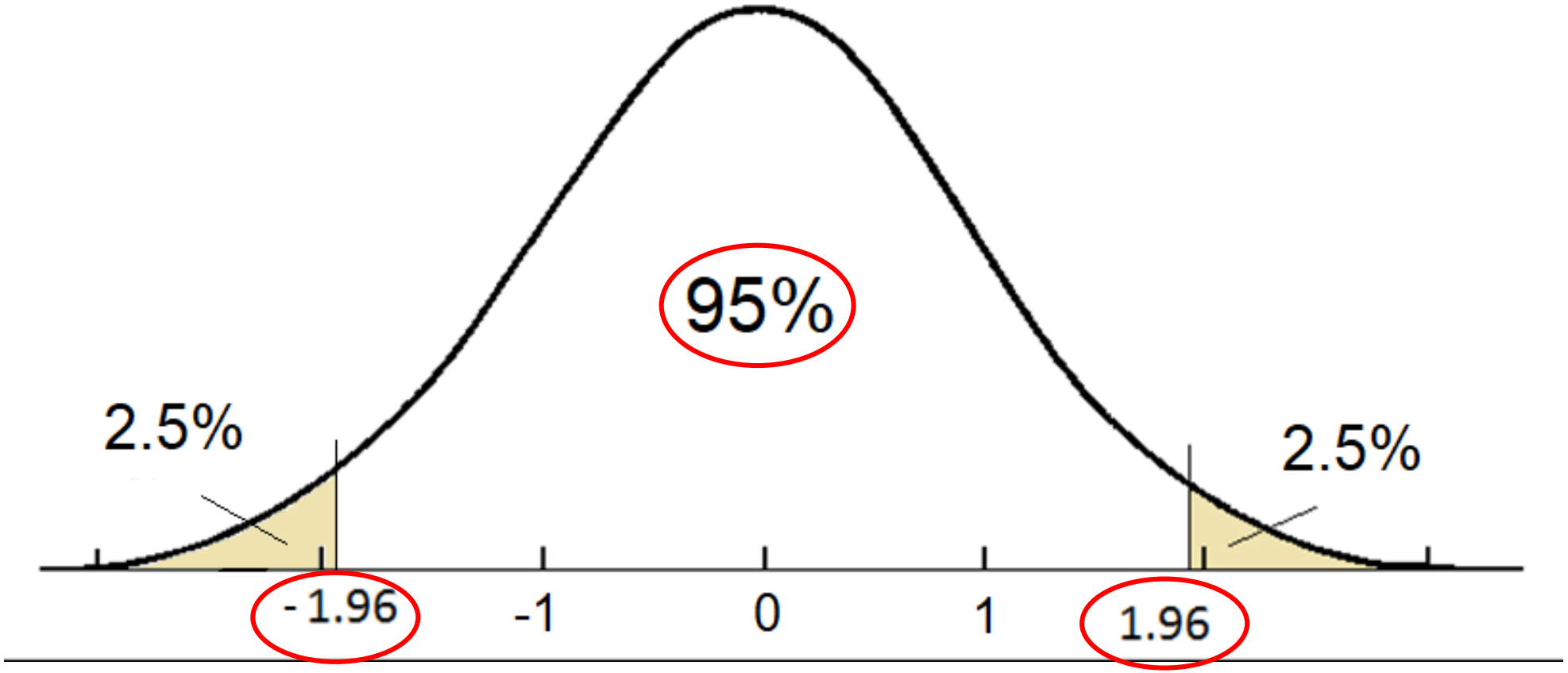
- frecvențe



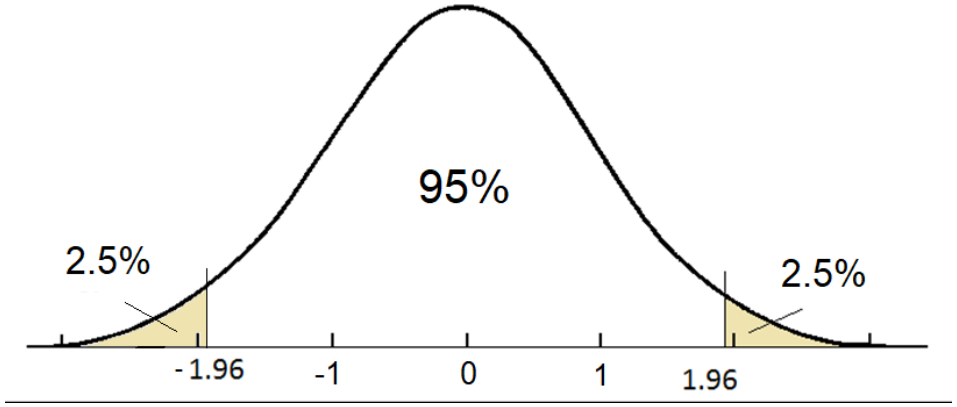
- probabilități

Proprietățile distribuției normale standard

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

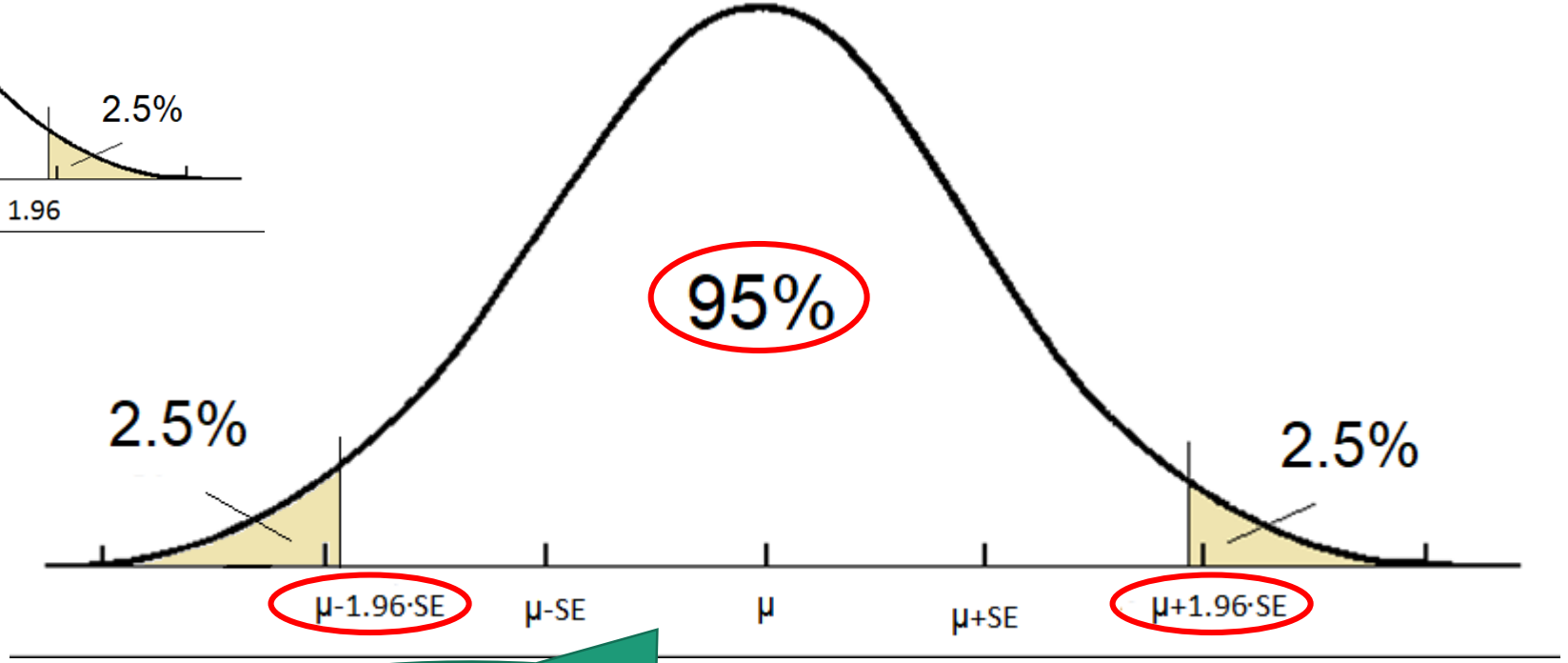


distribuția normală standard



distribuția mediilor dacă repetăm studiul

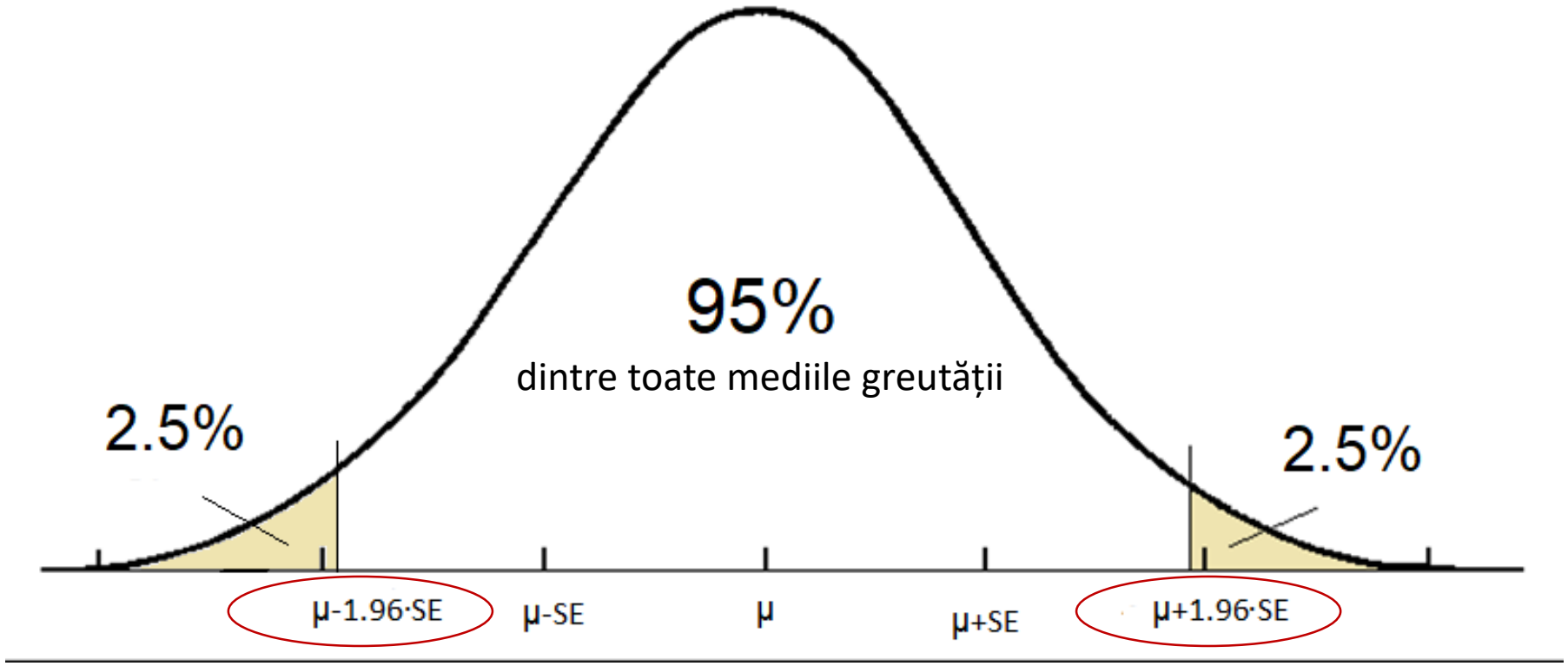
distribuția de eșantionare



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

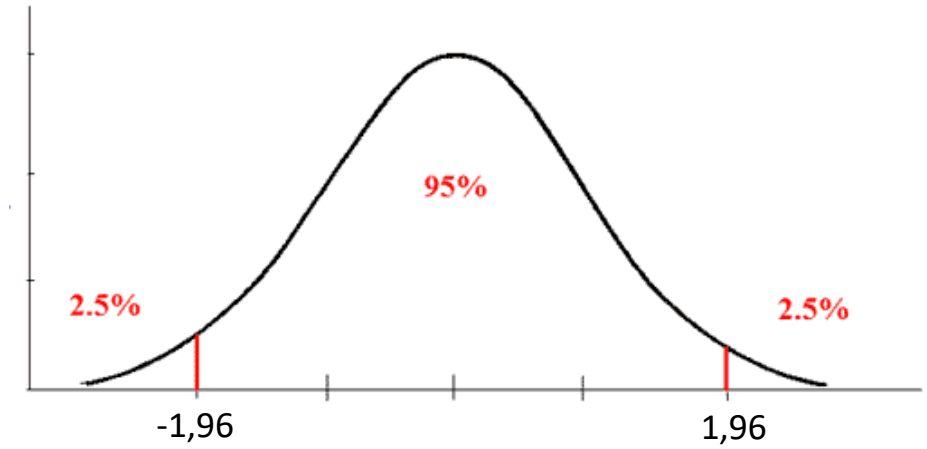
Intervalul de **încredere** de 95% pentru media μ

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

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

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

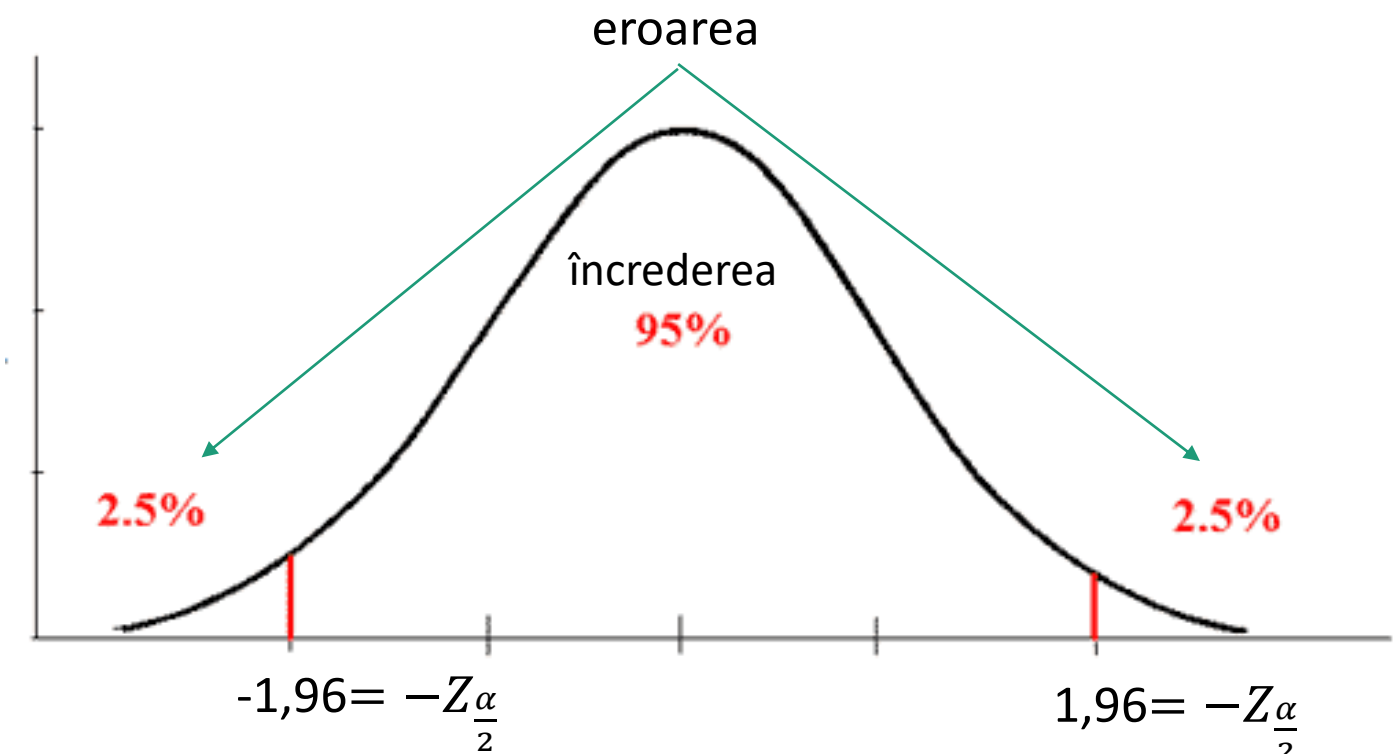
media $\bar{X}$ - media variabilei X pe eșantion, media μ – media variabilei X în populție, σ – deviația standard a variabilei X în populație, n – talia eșantionului, Z – valoare de pe axa Ox căreia îi corespunde în distribuția normală standardizată o arie egală cu nivelul de eroare



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 populție, 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

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

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

Corecția aplicată variației ca să fie un bun estimator

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

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

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

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

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

[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

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.

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

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

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

[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

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.

- esantion **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.

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

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

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

[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

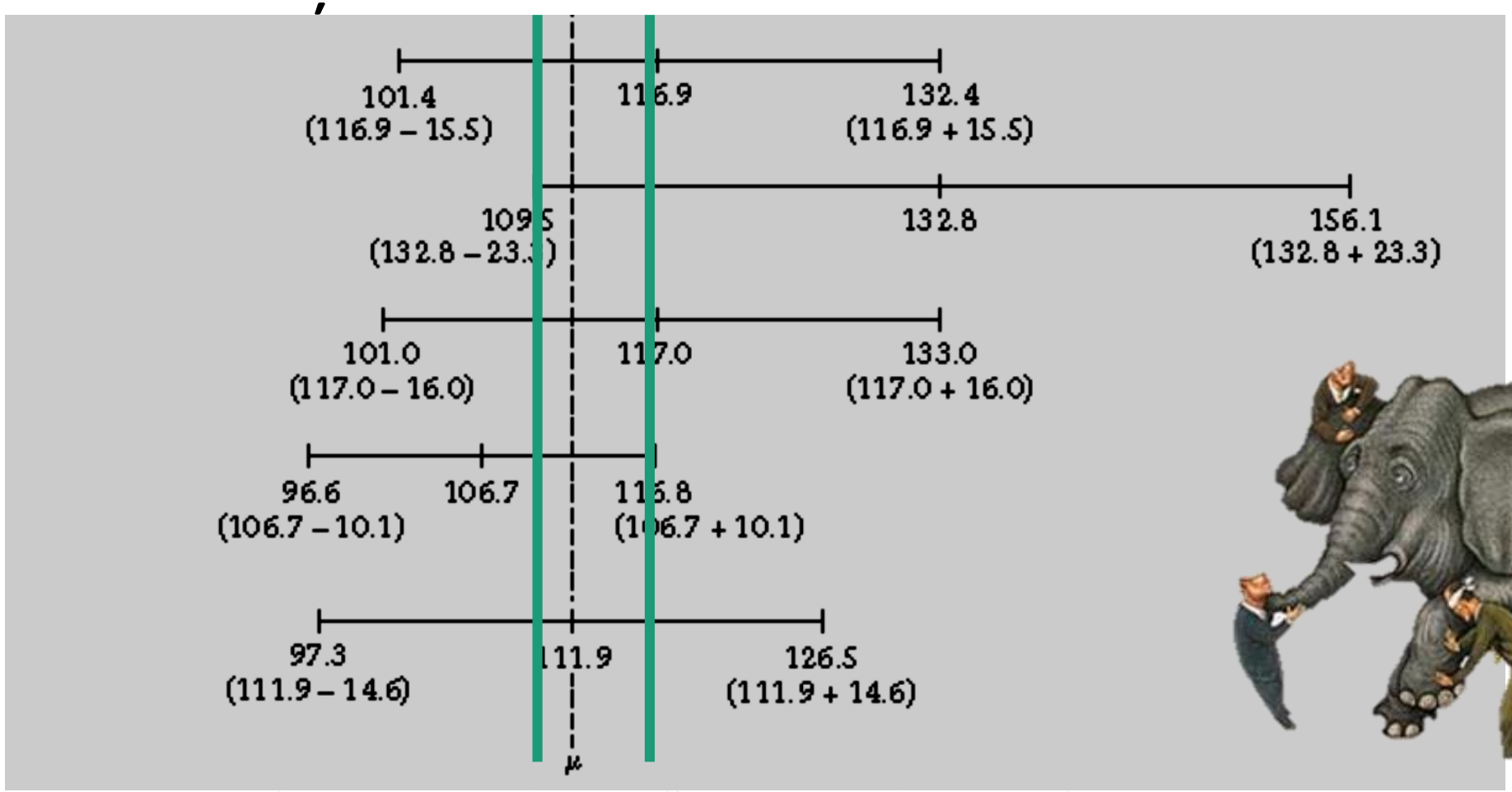
frecvența cancerului între 2,7% și 3,3% cu o probabilitate de 95%

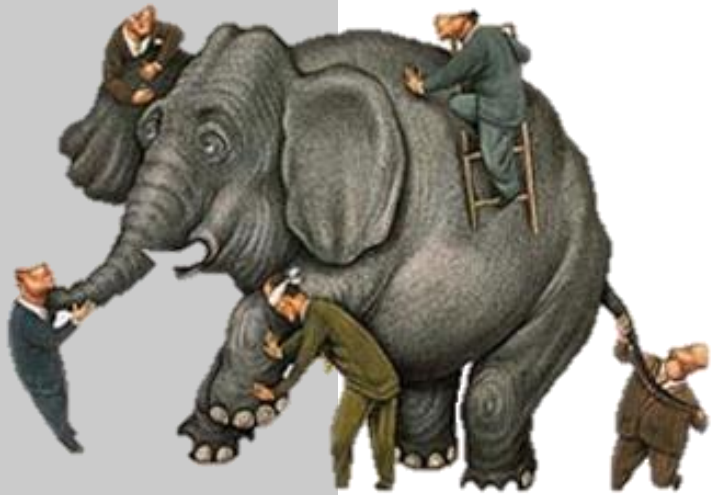
Răspuns: eșantion mai mic → interval mai mare

n la numitor are efect invers

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

Literatura de specialitate: mai multe studii care măsoară același lucru





Obținem un interval comun unde se găsește media populației μ

Systematic Review and Meta-Analysis

Medicine

OPEN

The prognostic values of estrogen receptor alpha and beta in patients with gastroesophageal cancer

A meta-analysis

Dongyun Zhang, MD[a,*], Jianwei Ku, MD[b], Yingjie Yi, MM[c], Junhui Zhang, MM[d], Rongzhi Liu, MM[a], Nianya Tang, MM[a]

Abstract

Background: Published studies have investigated the prognostic roles of estrogen receptor alpha (ERα) and estrogen receptor beta (ERβ) in gastroesophageal cancer patients with the controversial results. The aim of the study was to systematically evaluate the impacts of ERα and ERβ on the overall survival (OS) in patients.

Method: Relevant eligible studies were extracted from PubMed, Embase, Web of Science, CNKI and Wanfa start date to November 2018) following the Preferred Reporting Items for Systematic Reviews and Meta-Analyse HR (hazard ratio) with 95% confidence intervals (CIs) were used to assess the prognostic values of ERα and

Results: High ERα expression was associated with poor OS (HR = 1.58, 95% CI = 1.29–1.94, $P < .001$) and ERβ with better OS (HR = 0.56, 95% CI = 0.37–0.83, $P = .004$) in gastroesophageal cancer. Furthermore, unfavorable OS was found in Chinese gastroesophageal patients with higher ERα expression (HR = 1.57, 95% CI = 1.25–1.96, $P < .001$) and better OS with higher ERβ expression (HR = 0.51, 95% CI = 0.31–0.83, $P < .01$) in our subgroup analysis. Meanwhile, worse OS was found in esophageal squamous cell carcinoma (ESCC) patients with high ERα expression (HR = 1.74, 95% CI = 1.33–2.26, $P < .001$), and favorable OS in ESCC with ERβ overexpression (HR = 0.40, 95% CI = 0.31–0.52, $P < .001$). Besides, high ERα expression was associated with lower tumor differentiation in ESCC (OR = 1.64; 95% CI = 1.02–2.64, $P = .04$) and ERβ was linked with better tumor differentiation in gastric adenocarcinoma (GCA) (OR = 0.49; 95% CI = 0.26–0.94, $P = .03$).

HR=1.57, 95% CI =1.25 – 1.96
CI – confidence interval

Medicine (Baltimore). 2019 Nov;98(46):e17403. doi: 10.1097/MD.0000000000017403.

The association of endocannabinoid receptor genes (CNR1 and CNR2) polymorphisms with depression: A meta-analysis.

Kong X[1], Miao Q[1], Lu X[2], Zhang Z[1], Chen M[3], Zhang J[1], Zhai J[3].

Author information

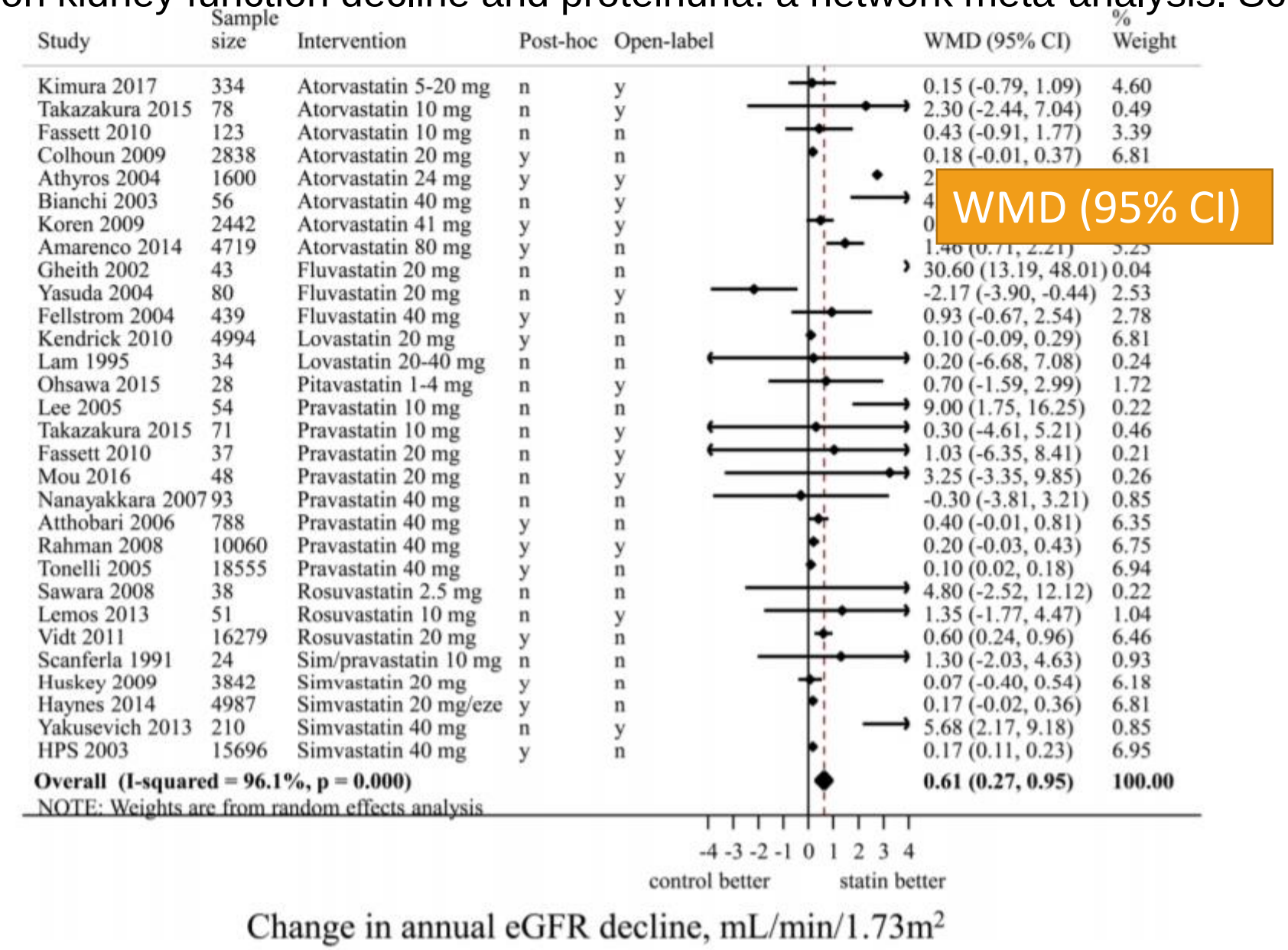
1 Department of Clinical Psychology, Jining Psychiatric Hospital.
2 Department of Clinical Psychology, Qindao Mental Health Center.
3 School of Mental Health, Jining Medical University, Shandong, China.

Abstract

Studies investigating the association between gene variants and depression susceptibility found inconsistent data. The present study aimed to clarify whether CNR1rs1049353, CNR1 AAT triplet repeat, and CNR2rs2501432 polymorphisms confer higher risk for depressive disorder.Literature from PubMed, Medline, Embase, Scopus, Cochrance Library, and Wanfang databases was searched (up to August 20, 2018). Seven case-control studies with various comorbidities were eligible. We targeted CNR single-nucleotide polymorphisms (SNPs) that have been reported by 2 or more studies to be involved in the current meta-analysis, resulting in a final list of 3 SNPs: CNR1rs1049353, CNR1 AAT triplet repeat polymorphism, and CNR2rs2501432. Odds ratios (ORs) and 95% confidence intervals (CIs) for allele and homozygote comparisons, dominant and recessive models, and triplet repeat polymorphism ((AAT)n≥5, ≥5 vs (AAT)n<5, <5 or <5, ≥5) were assessed using a random effect model as measures of association. Heterogeneity among in[illegible] test. Publication bias was also explored by Egger and rank correlation test.overall, no significant association was found between depression and CNR1rs1049353 (G vs A: OR [95% CI]=1.09 [0.61-1.95]; GG vs AA: 1.29 [0.73-2.26]; GG vs GA+AA: 1.10 [0.57-2.10]; GG+GA vs AA: 1.25 [0.72-2.18]; and AAT triplet repeat polymorphism ((AAT)n≥5, ≥5 vs (AAT)n<5, <5 or <5, ≥5): 1.92 [0.59-6.27]. In contrast, a significant association between CNR2rs2501432 and depression was detected, and the ORs and 95% CIs are as follows: allele contrast (OR=1.39, 95% CI=[1.12-1.72], P=.003); homozygous (OR=2.19, 95% CI=[1.34-3.59], P=.002); dominant (OR=1.93,95% CI=[1.23-3.04], P=.005); and recessive (OR=1.41, 95% CI=[1.04-1.92], P=.03).This meta-analysis revealed that CNR1rs1049353 or AAT triplet repeat polymorphism had no association with susceptibility to depression, while CNR2rs2501432 polymorphism was a remarkable mark for depression patients.

GG vs AA: OR [95% CI] =1.29 [0.73-2.26]

Esmeijer K, Dekkers OM, de Fijter JW, Dekker FW, Hoogeveen EK. Effect of different types of statins on kidney function decline and proteinuria: a network meta-analysis. Sci Rep. 2019 Nov 12;9(1):16632.



WMD (95% CI)

Pe un eșantion de 65 de persoane, s-a măsurat greutatea (kg) și s-a observat greutatea medie 75 cu o abatere standard 16. Pentru un grad de încredere de 95%, intervalul de încredere este (aproximativ):

A. [59; 91]

B. [43; 107]

C. [71; 79]

D. [50; 100]

E. [74; 76]

Raspuns: C

Într-un eșantion de 81 persoane s-a observat frecvența unei boli, aceasta având valoare de 35%. Cu un risc de eroare de 5%, în populația din care a fost extras eșantionul, frecvența este cuprinsă în intervalul:

A. [0,30; 0,40]
B. [0,246; 0,454]
C. [0,324; 0,376]
D. [0,274; 0,426]
E. [0,10; 0,60]

Raspuns: B

Suntem interesați să găsim intervalul de încredere asociat cu media colesterolului pe diferite probe de subiecți cu diferite boli. Dacă nivelul de încredere este mai mic cum este intervalul de încredere?

A. Mai mare
B. Îngust
C. De aceeași lungime
D. Eroare de calcul
E. Nu se poate calcula, precizia e tot timpul de 95%

Răspuns: B

Suntem interesați să găsim intervalul de încredere asociat cu media colesterolului pe diferite probe de subiecți cu diferite boli. Dacă dimensiunea eșantionului este mai mare cum este intervalul de încredere?

A. Mai mare
B. Îngust
C. De aceeași lungime
D. Eroare de calcul
E. Nu se poate calcula

Răspuns: B

Suntem interesați să găsim intervalul de încredere asociat cu media colesterolului pe diferite probe de subiecți cu diferite boli. Dacă abaterea standard este mai mare cum este intervalul de încredere?

A. Mai mare
B. Îngust
C. De aceeași lungime
D. Eroare de calcul
E. Nu se poate calcula

Răspuns: A

Suntem interesați să găsim intervalul de încredere asociat cu media colesterolului pe diferite probe de subiecți cu diferite boli. Dacă nivelul de încredere este mai mare cum este intervalul de încredere?

A. Mai mare
B. Îngust
C. De aceeași lungime
D. Eroare de calcul
E. Nu se poate calcula, precizia e tot timpul de 95%

Răspuns: A

Suntem interesați să găsim intervalul de încredere asociat cu media colesterolului pe diferite probe de subiecți cu diferite boli. Dacă media pentru care este calculat este mai mică cum este intervalul de încredere?

A. Mai mare
B. Îngust
C. De aceeași lungime
D. Eroare de calcul
E. Nu are sens întrebarea

Răspuns: C

Mulțumesc!