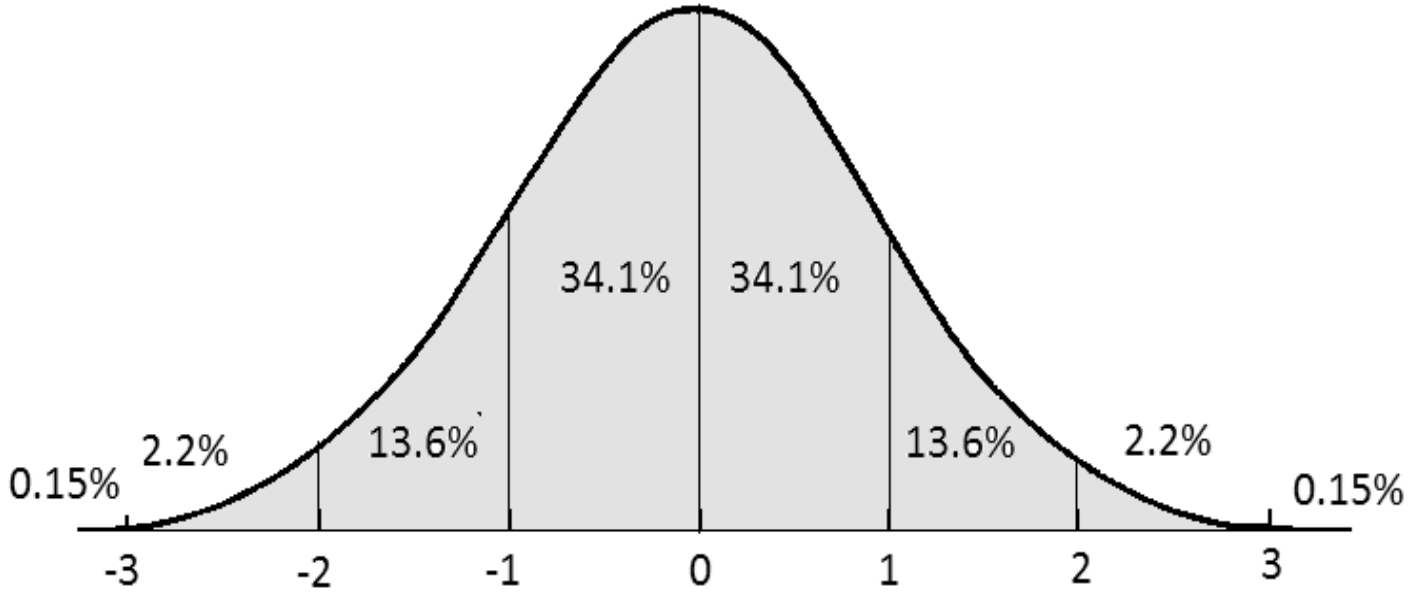


Bondor Cosmina

Variabile aleatoare, distributii de probabilitate

A ALWAYS

S SEEK

K KNOWLEDGE

Obiectivele cursului

După parcurgerea acestui curs studenții vor putea:

- să realizeze o eșantionare în scopul de a obține un eșantion reprezentativ al populației țintă într-un studiu
- să identifice posibile erori de selecție
- să determine distribuția de probabilitate a unei variabile aleatoare finite pe baza unui scenariu dat;
- să identifice caracteristicile distribuției normale și vor putea calcula diferite valori numerice asociate acestei distribuții.

Legendă  de ținut minte  pentru pasionați  important pentru înțelegerea noțiunilor ce urmează a fi prezentate

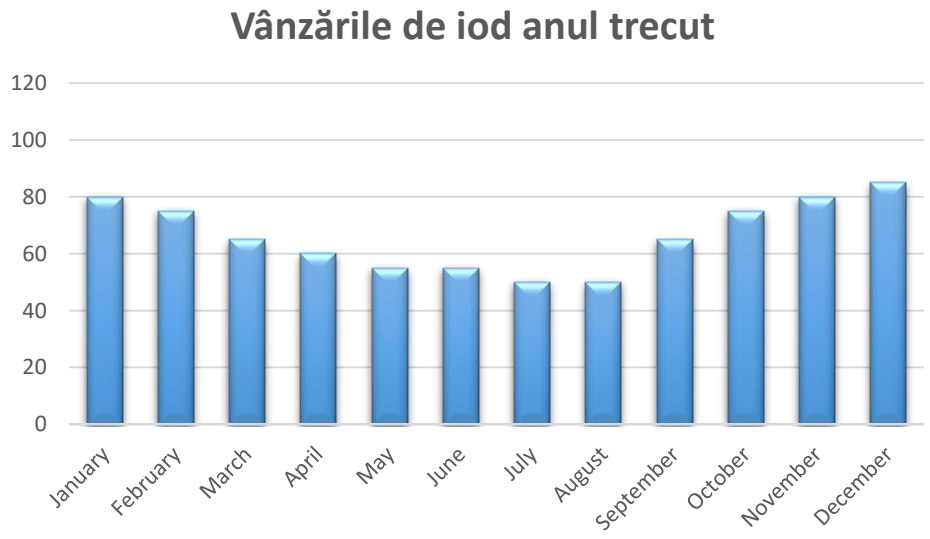
Obiectiv: Prevalența obezității în populația România

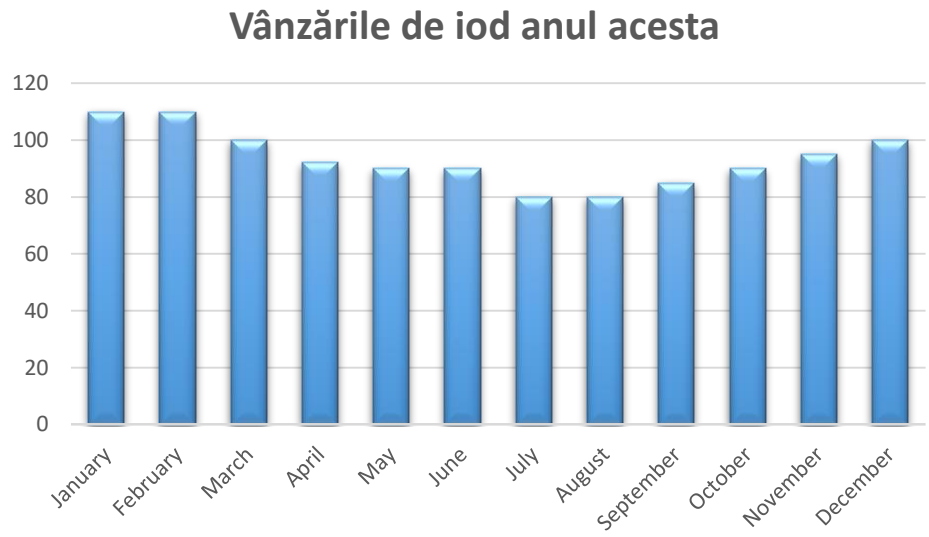
- Cum realizăm studiul? cântărim toată populația țării
  - nu putem, costuri mari, timp îndelungat, lipsă de personal etc.
  - dar putem să cântărim 2000 de persoane

Eșantion

Eșantion

- ne dă o idee asupra întregii populații
  - ne dă o idee despre cum arată dispersia și forma datelor
  - de ce e importantă forma?
    - compararea a două grupuri





De ce să studiem eșantioane în locul întregii populații?

Se studiază eșantioane în locul întregii populații

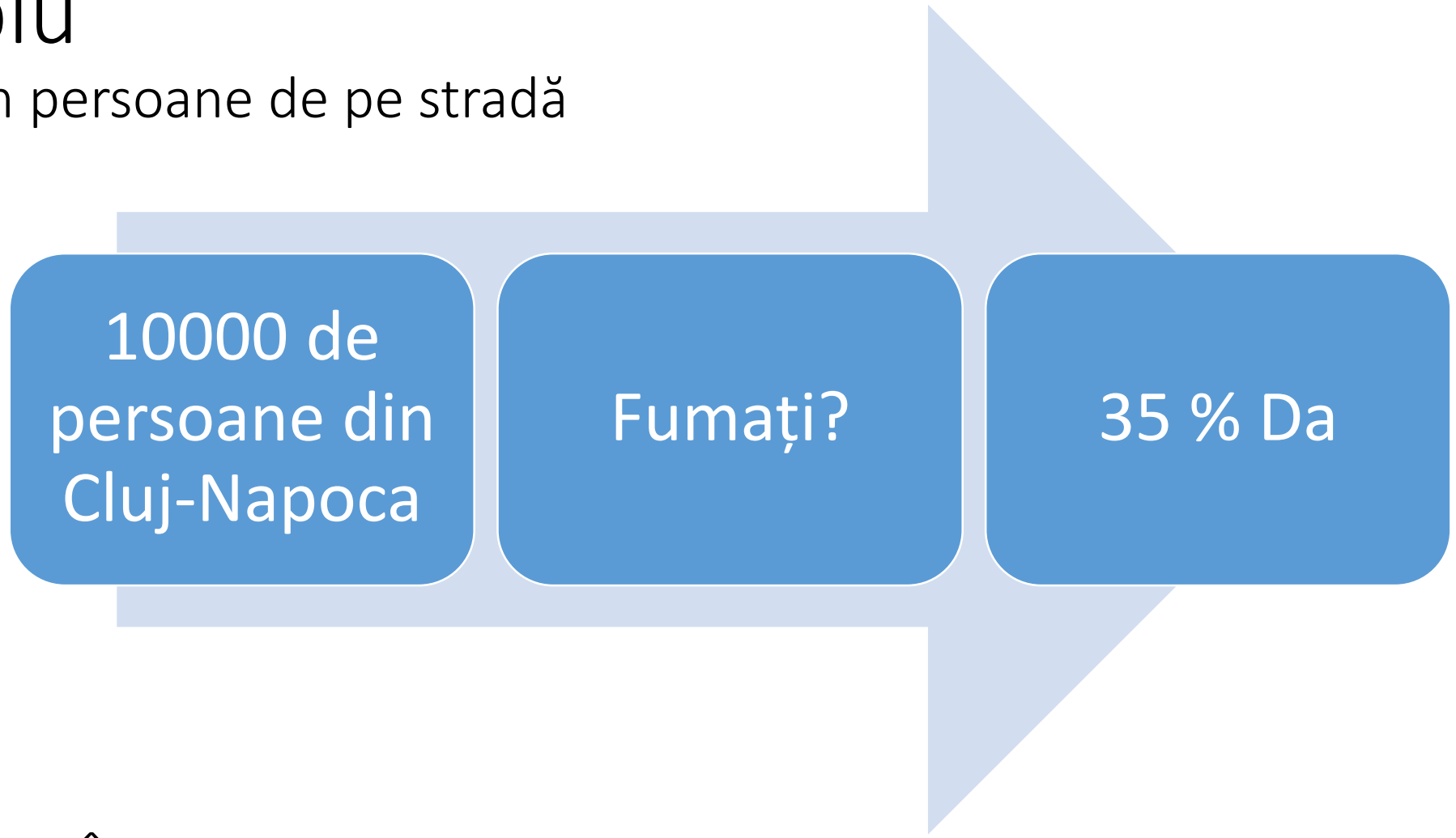
- Mai rapid
  - epidemie COVID-19, e nevoie rapidă de soluții
- Mai puțin costisitor
- Mai puțin periculos
  - întreaga populație primește un tratament nou netestat înainte
- Concluzii mai precise
  - mulți investigatori, multe măsurători – probabilitate mare de eroare

Populația

- din punct de vedere
  - statistic
    - o colecție de elemente care au aceeași caracteristică
  - in domeniul sănătății
    - pacienți
    - unități spitalicești

Exemplu

– selectăm persoane de pe stradă



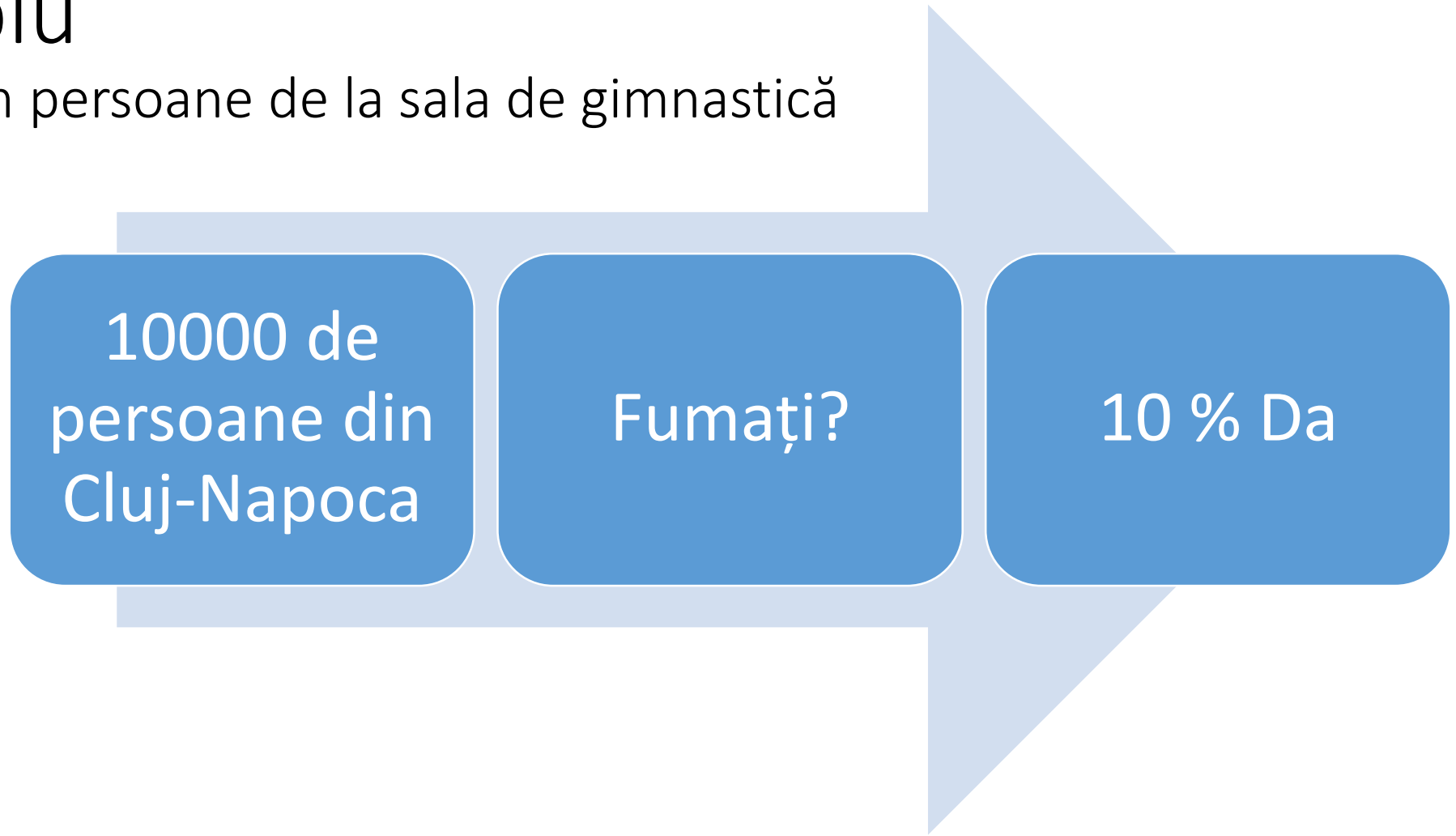
Generalizarea: În Cluj-Napoca probabilitatea ca o persoană să fumeze este 0,35

Avem 35% fumători?

Când se poate realiza?

Exemplu

– selectăm persoane de la sala de gimnastică



Selecția influențează rezultatul!

Ca să realizăm o aproximație corectă a frecvenței fumatului în populația țintă - selectăm un eșantion reprezentativ pentru populația din Cluj-Napoca

Eroare (bias) de selecție

- Ex.
- obiectivul - numărul de fracturi în populația generală într-un an
  - selecție de indivizi de la clubul de ski
- obiectivul – numarul de persoane cu infertilitate masculină
  - selecție de bărbați care vin la laborator să testeze infertilitatea
    - aceștia suspectează că sunt infertili

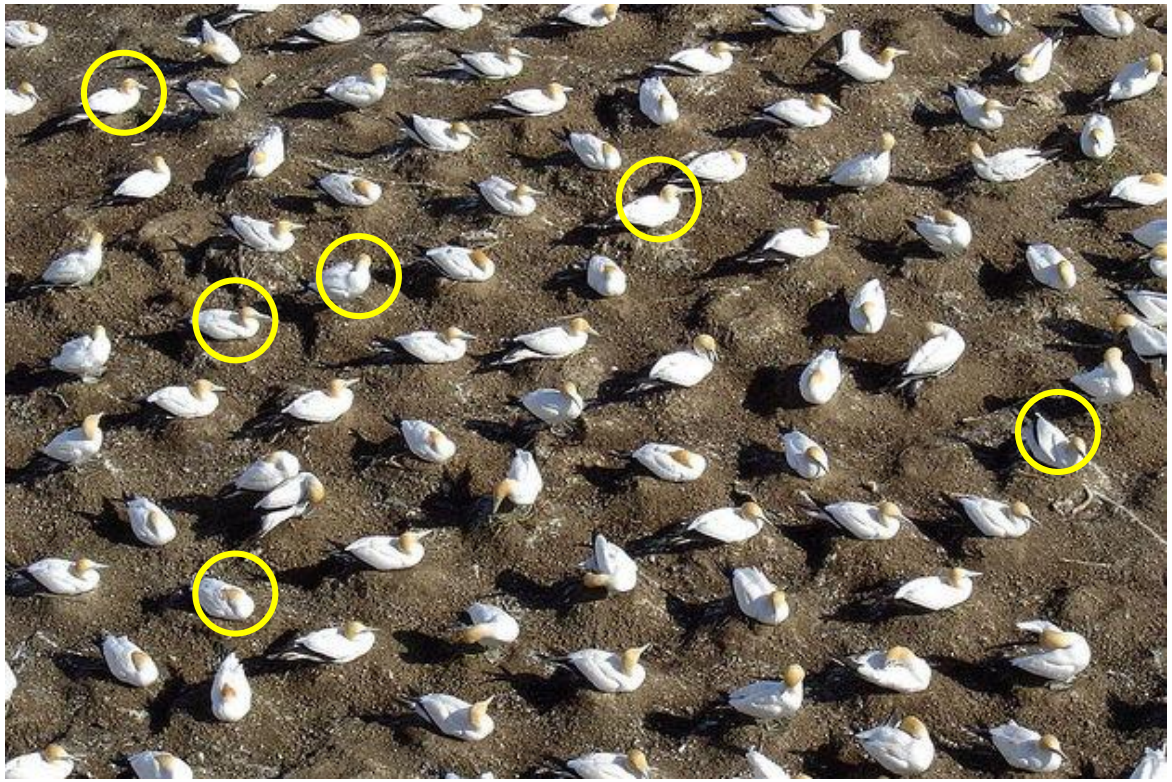
Avem nevoie de un rezultat fiabil și valid

Cum facem selecţia?

Metode de eșantionare

Eşantion reprezentativ pentru populație

Selecție aleatoare



Selecție aleatoare

- Fiecare individ din populație are aceeași pobabilitate de a fi selectat în eșantion

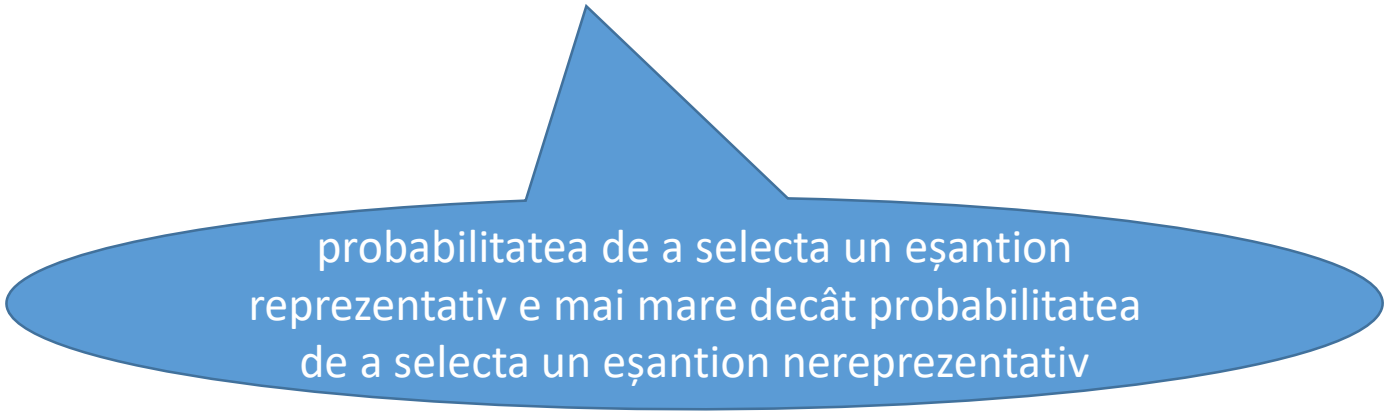
- Ex. Mergeți la primărie. Luați toate CNP-urile – extrageți aleator

De ce selecție aleatoare?

- Micșorarea/eliminarea erorilor experimentale = micșorarea/eliminarea **biasului de selecție**

- la selecția aleatoare

  P (eșantion reprezentativ) > P (eșantion nereprezentativ)

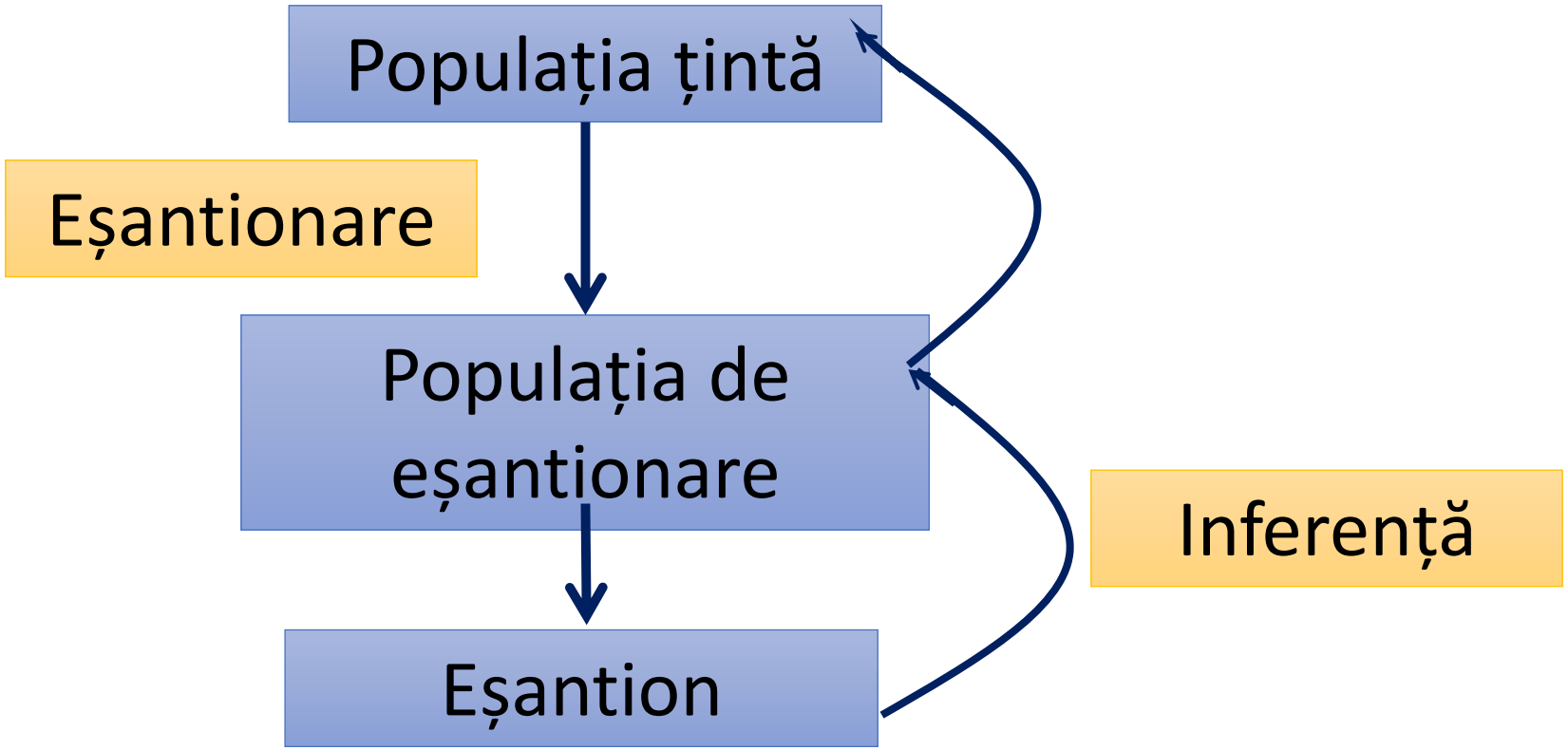


Chiar dacă selecția este aleatoare

- există o probabilitate mică să avem la selecție un eșantion nereprezentativ

- ex. prevalența fumatului – sunt șanse mici, dar există să selectăm numai persoane care fac sport de performanță

- Ce putem face?
  - studiile sunt replicate
  - dacă rezultatele sun consecvente în mai multe studii → evidențe medicale



Populație țintă – populația la care se dorește generalizarea rezultatelor studiului
Populația de eșantionare – populația din care a fost extras eșantionul

Condiţia inferenței eşantion → populaţie

Selecţia aleatoare

ne asigurăm că

diferențele constatate în studiile de cohortă se vor datora mai degrabă factorilor ce intervin decât erorilor de eșantionare

prevalența bolii rezultată din studiu va reflecta mai degrabă prevalența din populație decât erorile de eșantionare

- Când folosim termenul „eșantion” în contextul cercetării medicale
  - vom presupune că eșantionul a fost selectat aleator într-un mod corect

- Preferabil să citiți
  - rezultate ale unor studii realizate pe eșantioane selectate aleator

Metode de eșantionare

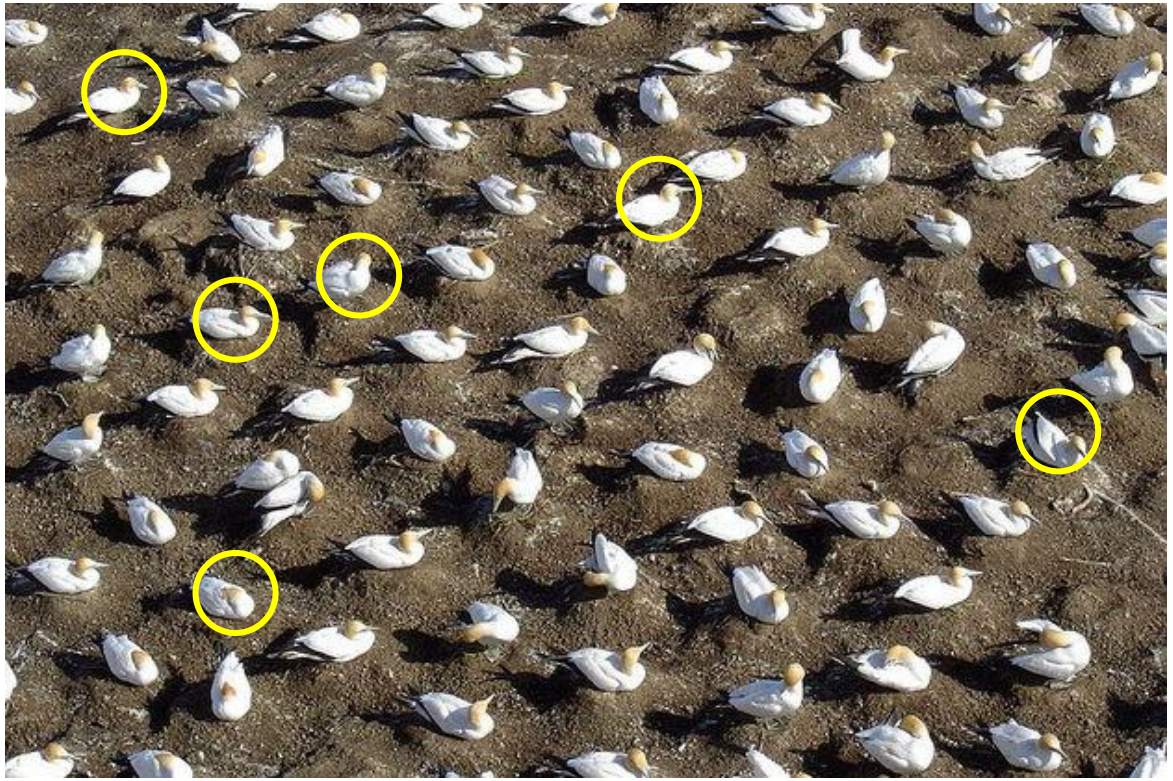
Probabilistică: fiecare subiect din populaţie are o probabilitate cunoscută de a fi selectat

- Eșantionare simplu randomizată
  - Subiecților li se atribuie un număr
  - Se extrag numere aleatorii din listă
- Eșantionare sistematică
  - tot al k-lea individ se alege pentru a fi inclus în eșantion
- Eșantionare stratificată
  - Populația este împărțită în straturi după însușiri care nu sunt echiprobabile, dar care pot influența obiectivul studiului, se extrage aleator din fiecare strat
- Eșantionare de tip cluster
  - Cluster= arie delimitată geografic
  - Delimitarea clusterelor, selectarea aleatorie a clusterelor
  - Selectare aleatorie a subiecților din fiecare cluster selectat

Eșantionare aleatorie simplă

- fiecare subiect are o probabilitate egală de a fi selectat

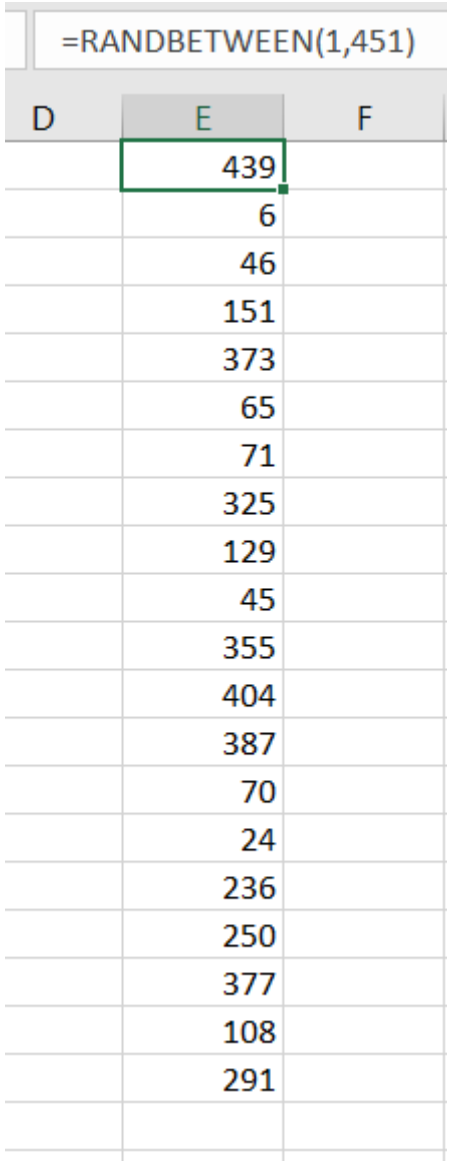
- Cum se realizează?
- Subiecților li se atribuie un număr
- Se extrag numere aleatorii din listă



Eșantionare aleatorie simplă

- Dorim sa luăm la întâmplare 20 de studenți din anul I medicină, total 450 de studenți
- Numerotăm studenții cu numere de la 1 la 451
- Folosim în Excel funcția RANDBETWEEN

- ce facem cu numerele duble?
  - eșantionare cu/fără înlocuire



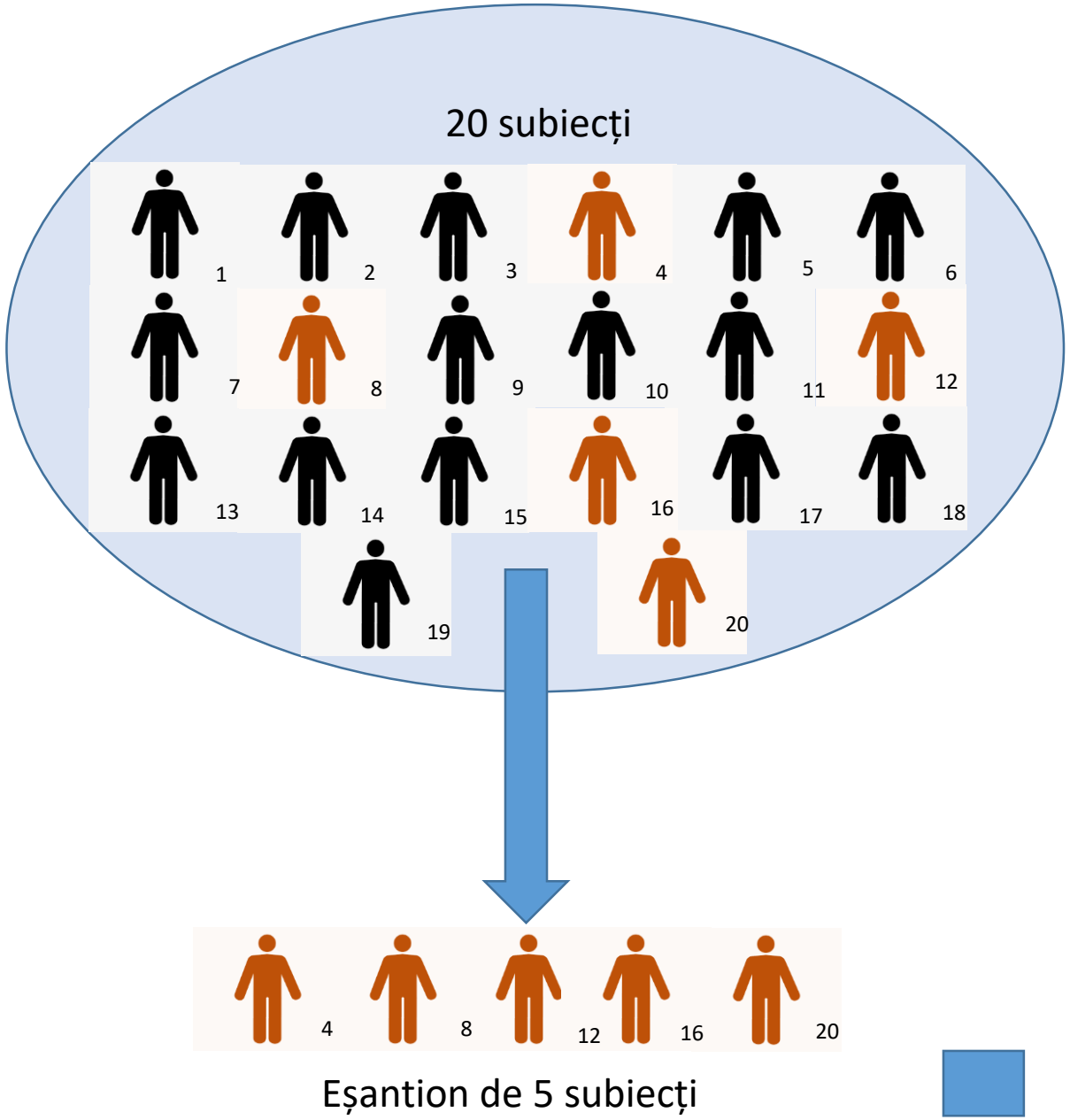
Eșantionare sistematică

- tot al k-lea individ se alege pentru a fi inclus în eșantion.
- Pentru N = 20, eșantion de 5 indivizi,

k= 20/5 = 4

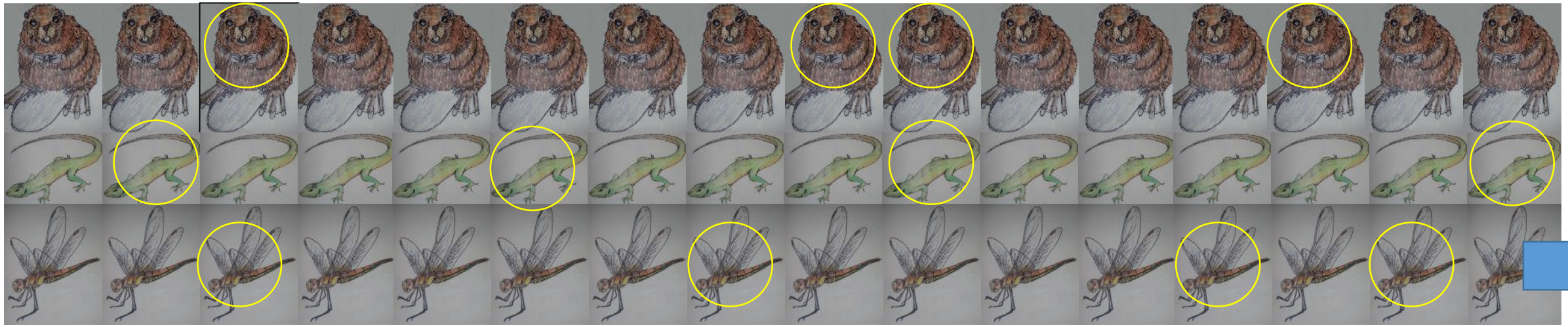
1. Se extrage aleator punctul de start
2. Tot al 4-lea individ va fi selectat

- Nu se recomandă
  - în cazul datelor ciclice
- ex. cazuri din clinica de ginecologie
  - sarcinile ectopice apar în general primăvara



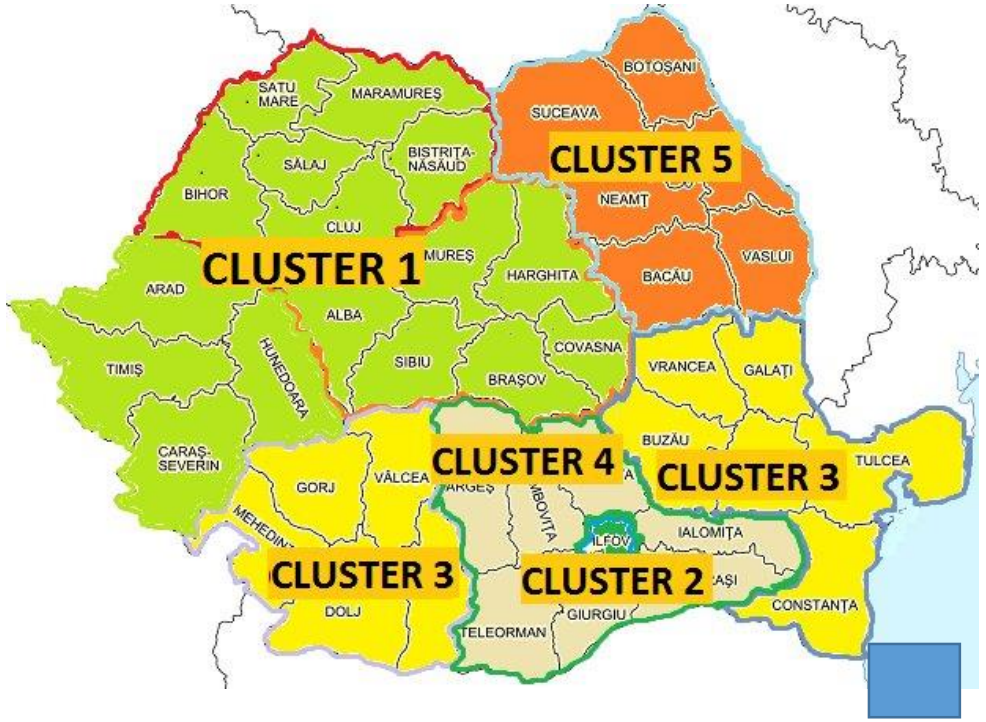
Eșantionare stratificată

- Populația este împărțită în straturi după însușiri care nu sunt echiprobabile, dar care pot influența obiectivul studiului
- ex.
  - categorii de vârstă
  - gen
- Se extrage aleator din fiecare grup câte un eșantion în funcție de cât de reprezentativ este stratul respectiv



Eșantionare de tip cluster

- Cluster= arie delimitată geografic
- Pas 1. Selectarea aleatorie a clusterelor
  - Pas 2. Selectare aleatorie a subiecților din fiecare cluster
- Ex. în studii multicentrice
  - cluster
    - fiecare unitate spitalicească
    - fiecare cabinet medical



Ulman S Particularitati regionale ale saraciei rurale in Romania

Metode de eșantionare

Non-probabilistică: probabilitatea unui individ de a fi selectat este necunoscută

- Convenient:
    - Participanţii sunt selectaţi doarece sunt accessibili
- Bulgărele de zăpadă:
    - Subiecţii incluşi în studiu vor aduce alţi potenţiali participanţi
    - ex.
        - membrii ai aceluiaşi grup
        - activităţi commune
- Deliberat
    - Grup de tehnici de eşantionare care au la bază gândirea cercetătorului
    - ex.
        - eșantionarea cu variație maximă,
        - eșantionarea cazurilor extreme,
        - eșantionarea realizată de experți

Eșantionarea non-probabilistică

- de multe ori - Reflectă erorile de gândire ale cercetătorului
- poate duce la rezultate
  - subiective
  - eronate
- au un posibil bias de selecție
- nu putem estima erorile de eșantionare

Recensământ

- Recensământ
  - participă toată populația – nu necesită inferență statistică
  - se aplică metode ale statisticii descriptive

Variabila aleatoare, distribuția de probabilitate

Eșantioane aleatorii

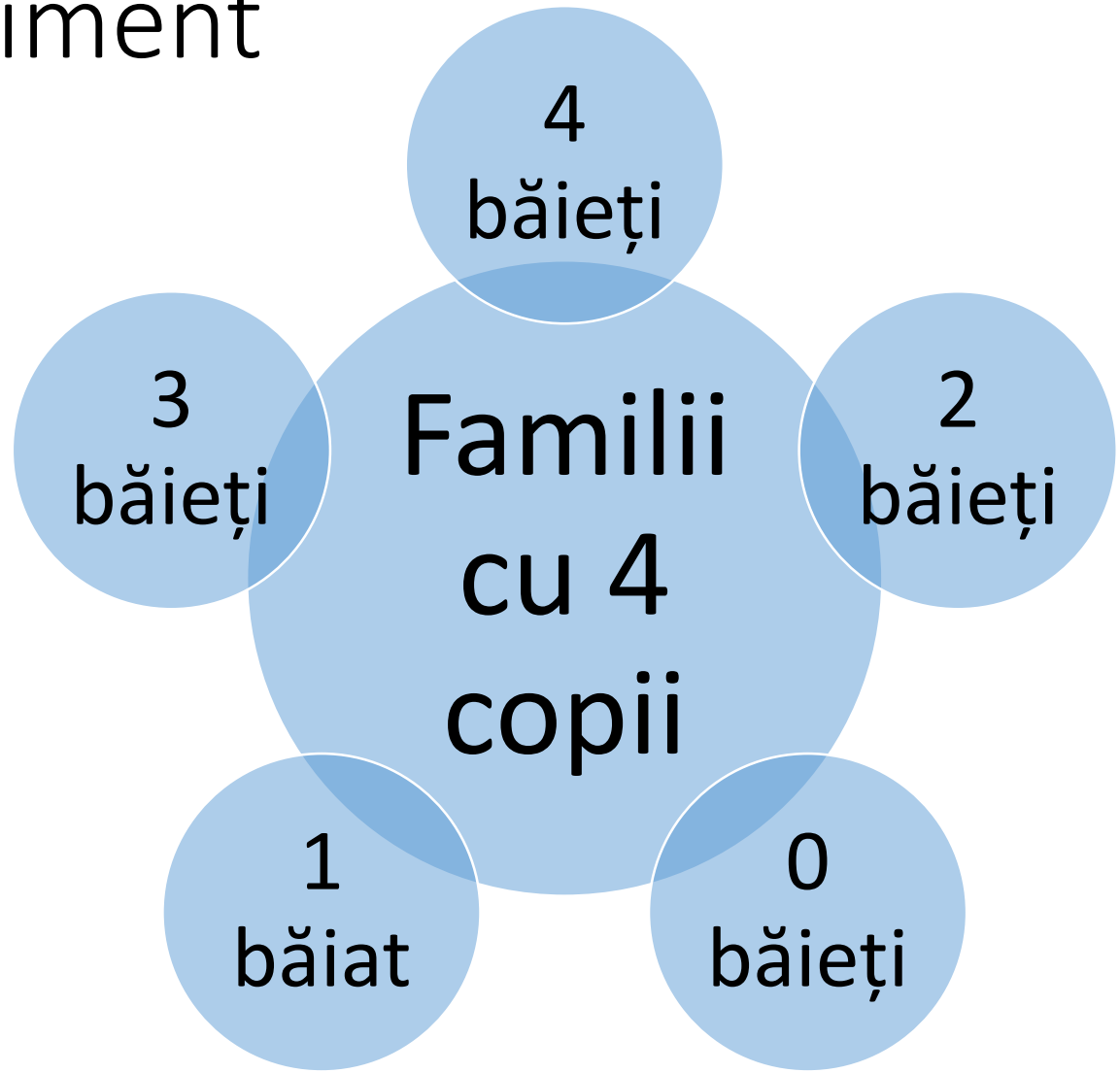
Măsurători

Rezultatul imprevizibil

Rezultatul = variabilă aleatoare

Probabilitatea ca să se nască un copil de sex masculin≈0.5 (50% dintre cazuri)

Experiment



În **100 de familii cu 4 copii**?

In 100 de familii cu 4 copii?

Selectăm aleator **100 de familii cu 4 copii**:

Număr de băieți	**0**	**1**	**2**	**3**	**4**	**Total**
Nr. de familii	4	29	40	24	9	100

- Numarul de băieți într-o familie – **Variabila aleatoare**
- 0 băieți în 4 familii
- 1 băiat în 29 de familii
- 2 băieți în 40 familii
- 3 băieți în 24 de familii
- 4 băieți în 9 familii

Distribuția de probabilitate obținută empiric

Numim **distribuție de probabilitate a variabilei X** numărul de apariții a valorilor posibile a variabilei X

Număr de băieți	**0**	**1**	**2**	**3**	**4**	**Total**
~~Nr. de familii~~	~~4~~	~~29~~	~~40~~	~~24~~	~~9~~	~~100~~
Probabilitatea	0,04	0,29	0,40	0,24	0,09	1

- Numărul de băieți într-o familie – **Variabila aleatoare**

Distribuție de probabilitate

Frecvențele de apariție a valorilor unei variabile aleatoare

→sumarizate într-o **distribuție de frecvențe** =

= distribuție de probabilitate

Cum calculăm distribuția de probabilitate?

- Empiric
  - eșantion, apoi inferență
- Teoretic
  - Formulă
  - Regulă
  - Aproximare cu o distribuție de probabilitate teoretică cunoscută

Ex. **Distribuția de probabilitate teoretică** a numărului de baieți în familiile cu 4 copii

Număr băieți	0	1	2	3	4	Total
Probabilitatea	0,0625	0,25	0,375	0,25	0,0625	1,00

- Cum a fost calculată?
  - modelată după așteptări,
  - un comportament “normal” = neinfluențat de diverși factori

Dacă ne interesează același obiectiv la clinica de infertilitate

- Familii cu gemeni:

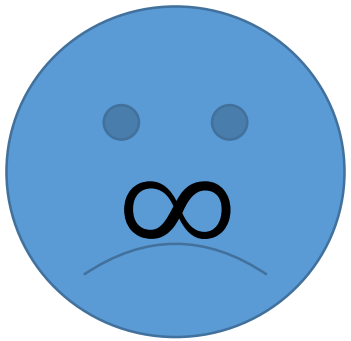
Numarul de baieti	0	1	2	3	4	Total
	0,50	0	0	0	0,50	100

Cazul discret – Cazul continuu

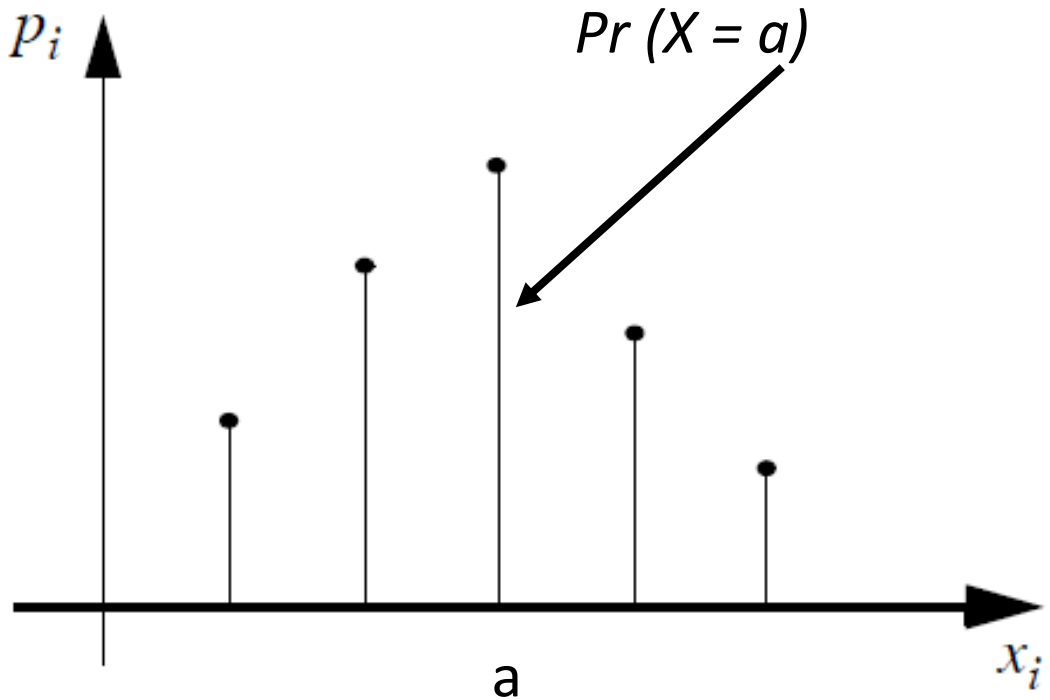
- Analog, dar ∞

$$\sum_{i=1}^{n}$$

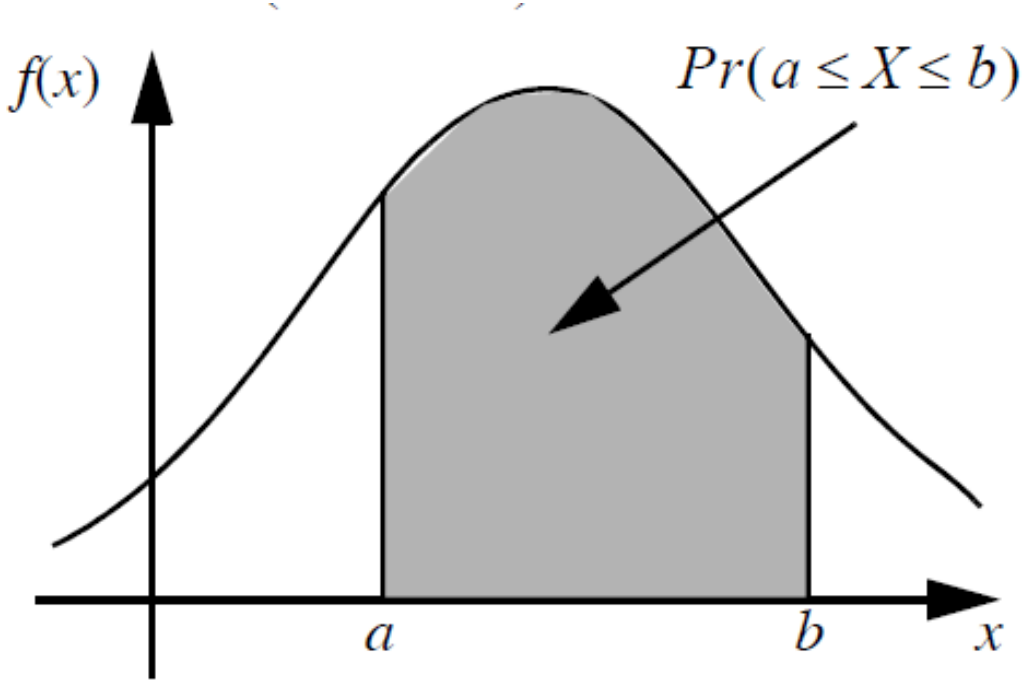
$$\sum_{i=1}^{\infty}$$



Cazul discret – Cazul continuu



Aria de sub curbă



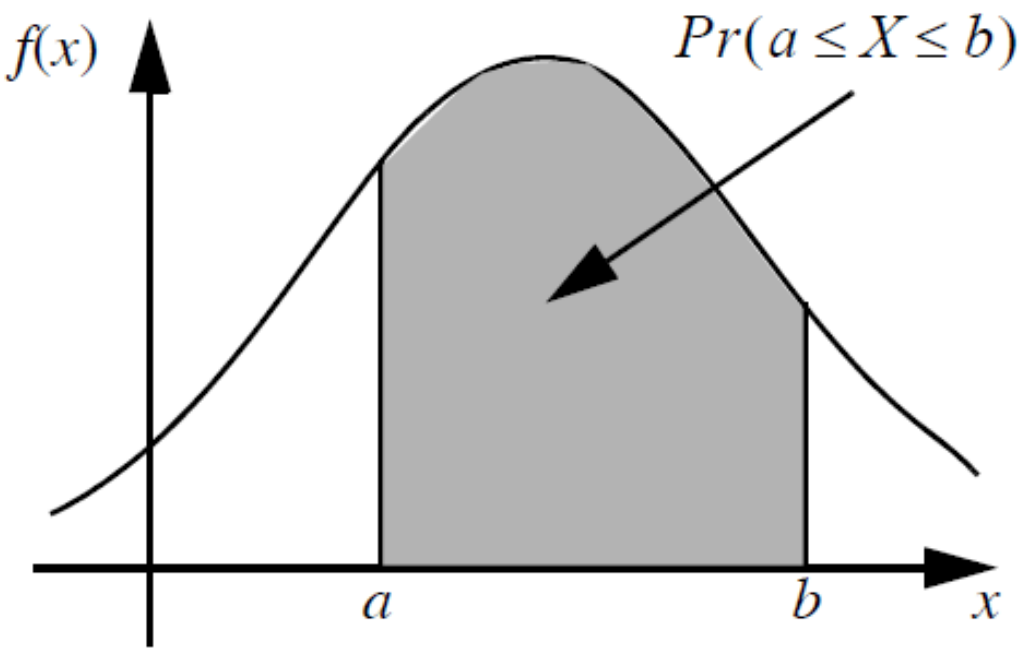
Cazul continuu

- In cazul unei variabile aleatoare continue X, se consideră o funcţie f:R→R numită *densitate de probabilitate*, care are proprietăţile:

$$\Pr(a \le X \le b) = \int_a^b f(x)dx$$

$f(x) \ge 0, \forall\ x \in R$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$



Cum aflăm distribuția de probabilitate?

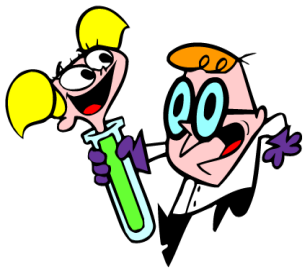
Dacă variabila e finită

- Empiric – experiment – distribuție de frecvențe

Dacă e ∞?

! suntem norocoși - găsim

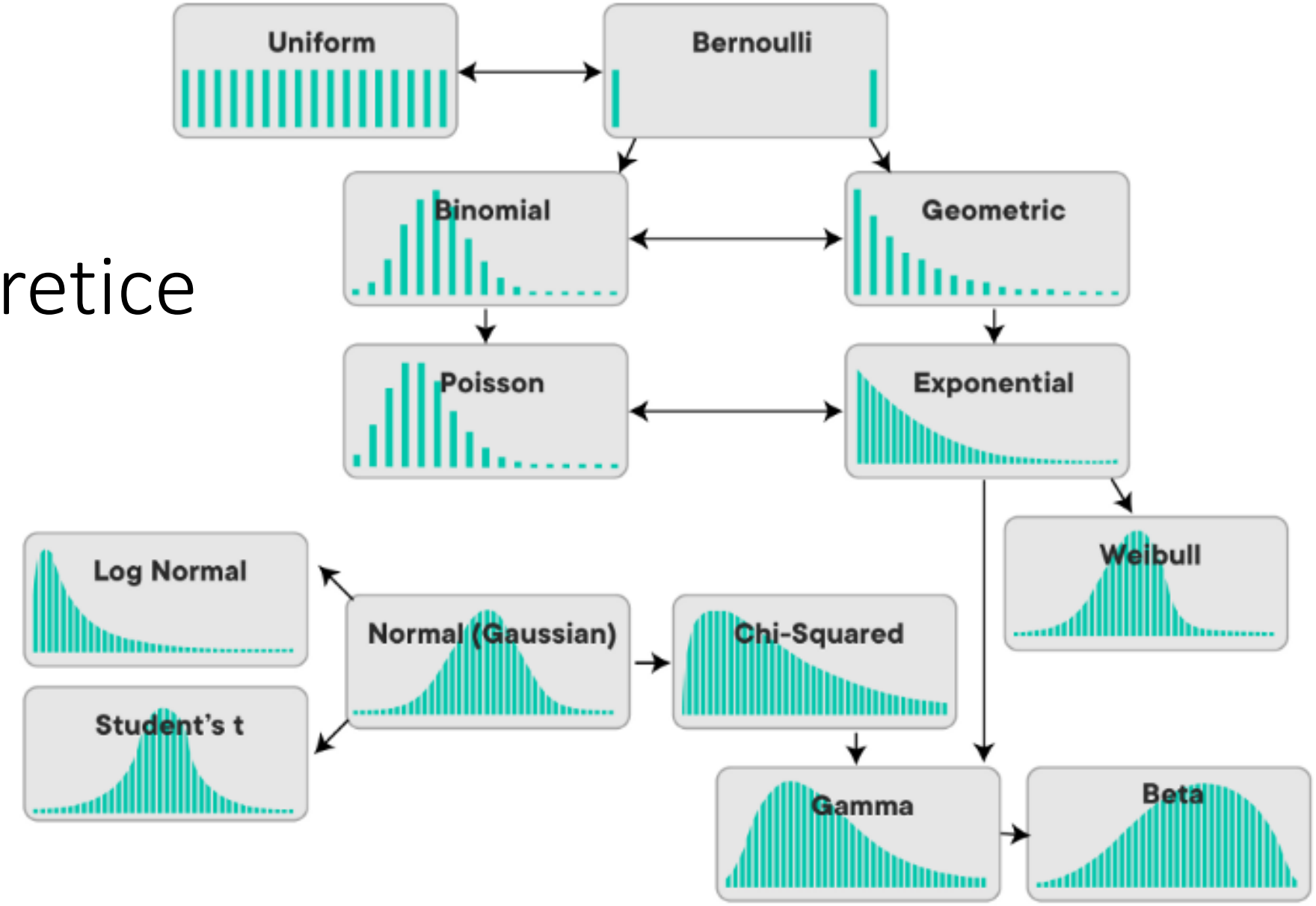
- Formulă
- Regulă



Dacă nu suntem norocoși:

- Modelăm (aproximăm) după o distribuție teoretică de probabilitate (cunoscută – una la care am fost norocosi)

Distribuţii teoretice



- Au o funcție cunoscută, medie și deviaţie standard deductibilă

Akshay S. Understanding Different Types of Distributions You Will Encounter As A Data Scientist

Principalele legi de distribuţie

legea BINOMIALĂ (BERNOULLI)	• probabilitatea unei variabile de tip succes/eșec. De câte ori apare evenimentul intr-un număr dat de trialuri
legea POISSON	• probabilitatea evenimentelor rare
legea normală sau legea LAPLACE-GAUSS	• probabilitatea unui eveniment în cazul variabilelor continue
legea STUDENT (t)	• probabilitatea unui eveniment în cazul variabilelor continue
legea $\chi 2$ a lui PEARSON	• sume de pătrate a unor variabile independente normal distribuite
legea F a lui FISHER.	• comportarea câtului a două variabile cu distribuţie Hi-pătrat

Simboluri utilizate frecvent în inferența statistică

	Simbolul pentru parametru calculat pe populaţie	Simbolul pentru statistică calculată pe eşantion
Media	μ	$\bar{X}$
Deviația standard	σ	S
Proporția	π	p

LEGEA NORMALĂ

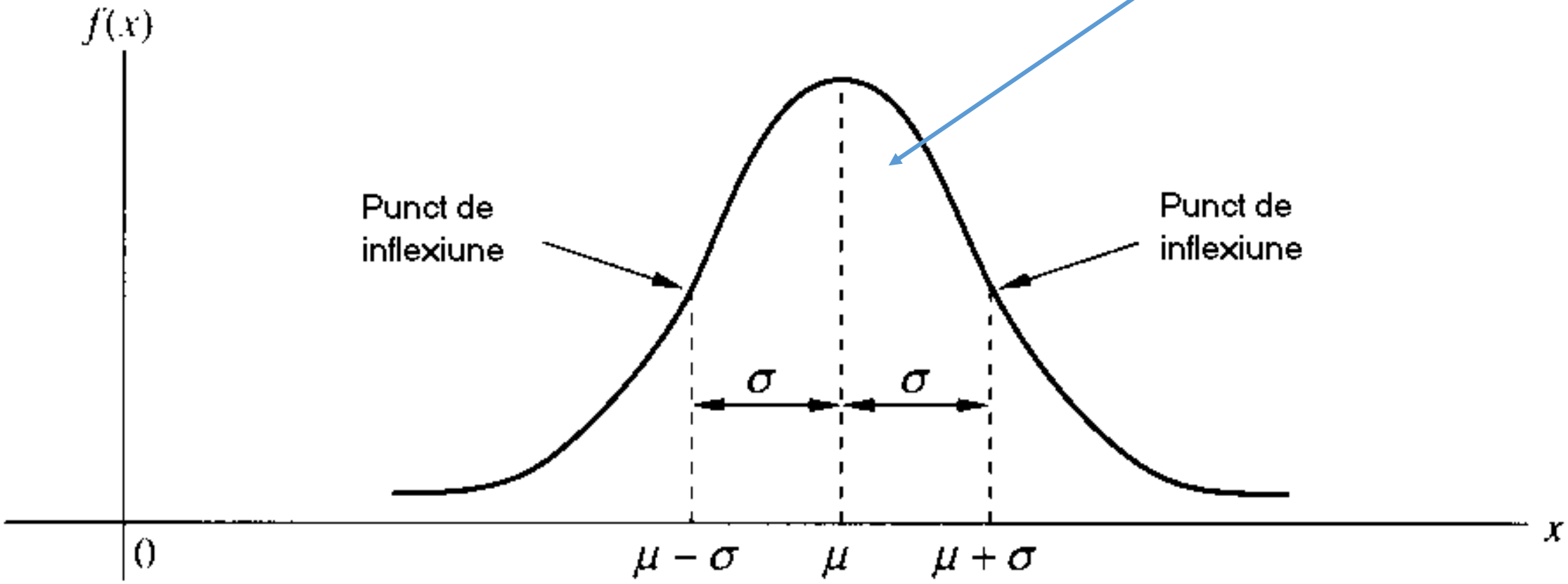
1777–1855

- variabilă aleatoare continuă
- funcţie de probabilitate - alură de clopot
  - curba normală
  - curba lui Gauss
- Această distribuţie depinde de doi parametri:
  - media aritmetică μ
  - abaterea standard (varianța) σ
- densitate de probabilitate:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

LEGEA NORMALĂ

Aria de sub curbă este 1, ca la orice distribuție de probabilitate



σ deviația standard (varianța) este distanța dintre medie și punctul de inflexiune (acolo unde curba se schimbă din concavă în convexă)

Transformarea Z

- când media aritmetică a unei distribuții ≠ 0 și varianța ≠ 1

Pasul 1. Mutăm distribuția în sus sau în jos pe linia numerică, astfel încât media să fie 0, adică $X - \mu$

Pasul 2. Ajustăm distribuția fie mai largă, fie mai îngustă /σ

$$Z = \frac{X - \mu}{\sigma}$$

numit și scor z, o abatere normală

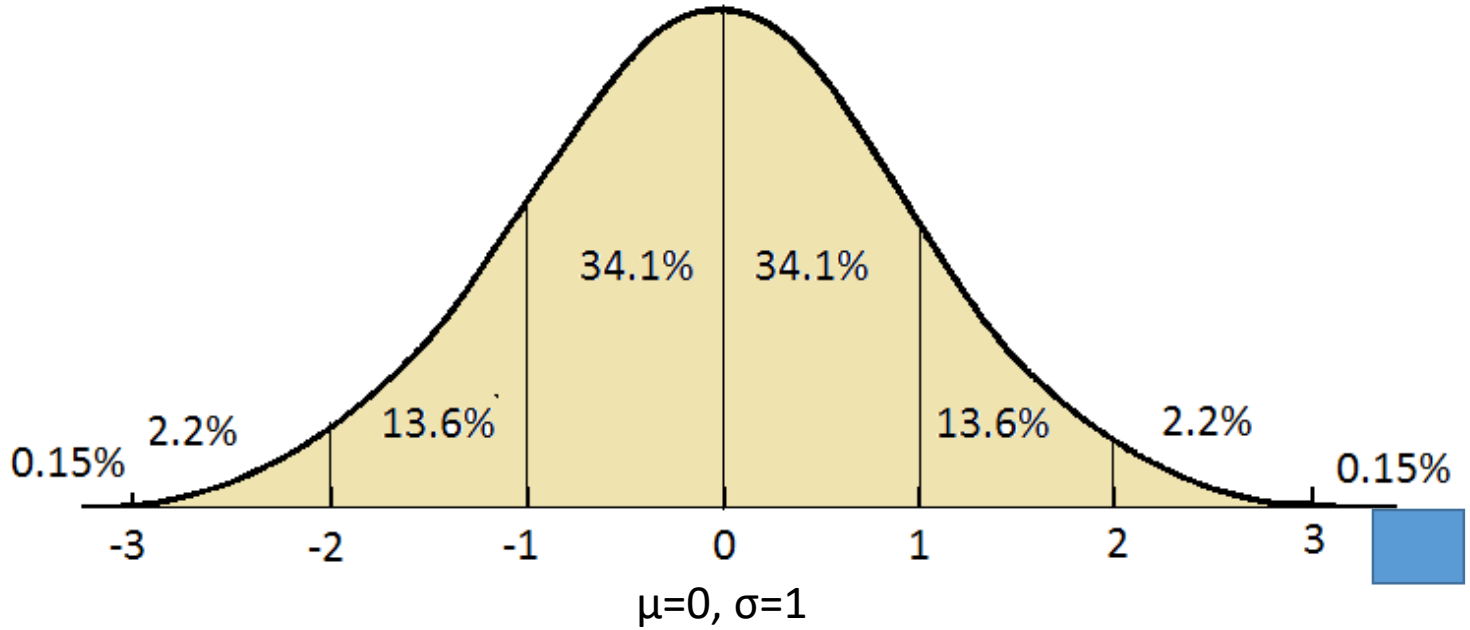
Distribuția normală standardizată

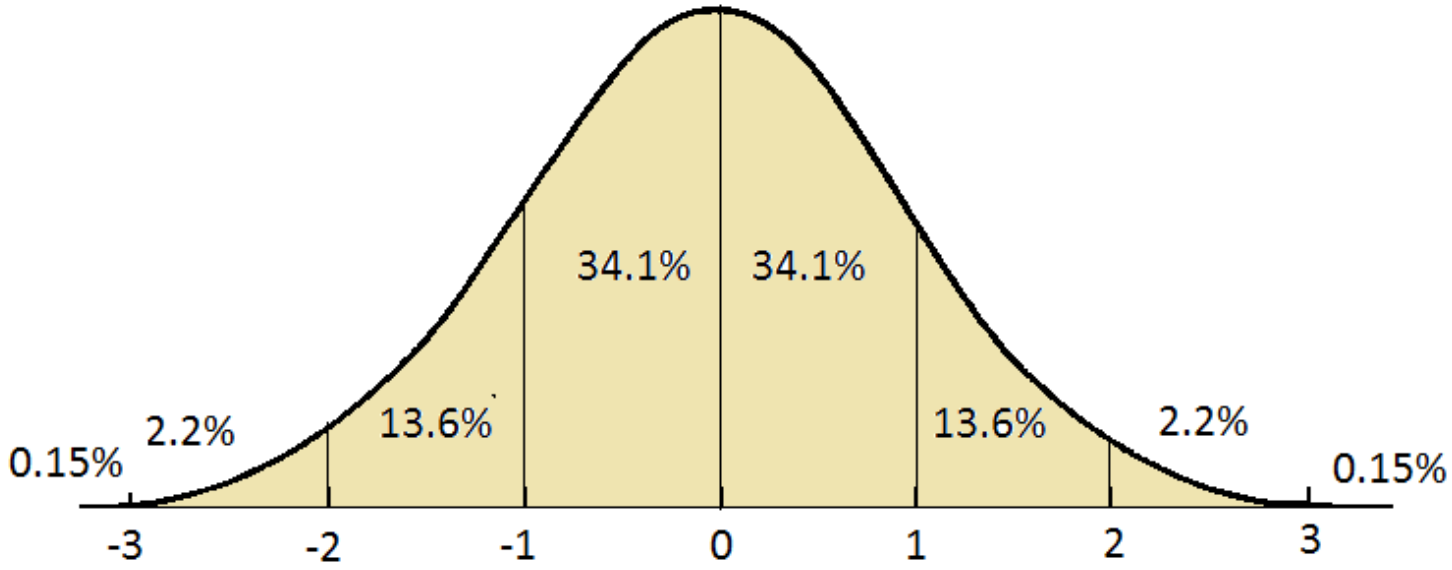
O distribuție normală cu μ=0 si σ=1.

$$Z = \frac{X - \mu}{\sigma}$$

Formula distribuției z:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$





- Câți subiecți sunt peste 0?
  - 50%
- Câți subiecți sunt între 0 și 1?
  - 34,1%
- Câți subiecți sunt peste 2 (două deviații standard)?
  - 2,35%

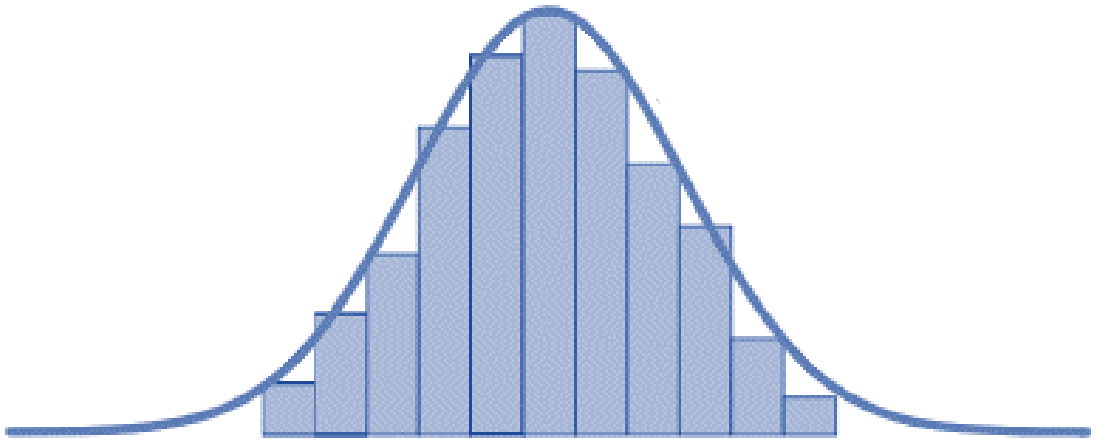
Proprietăți:

- punct de maxim este media aritmetică

  media aritmetică = modul

- simetrică față de media aritmetică

  media aritmetică = mediana

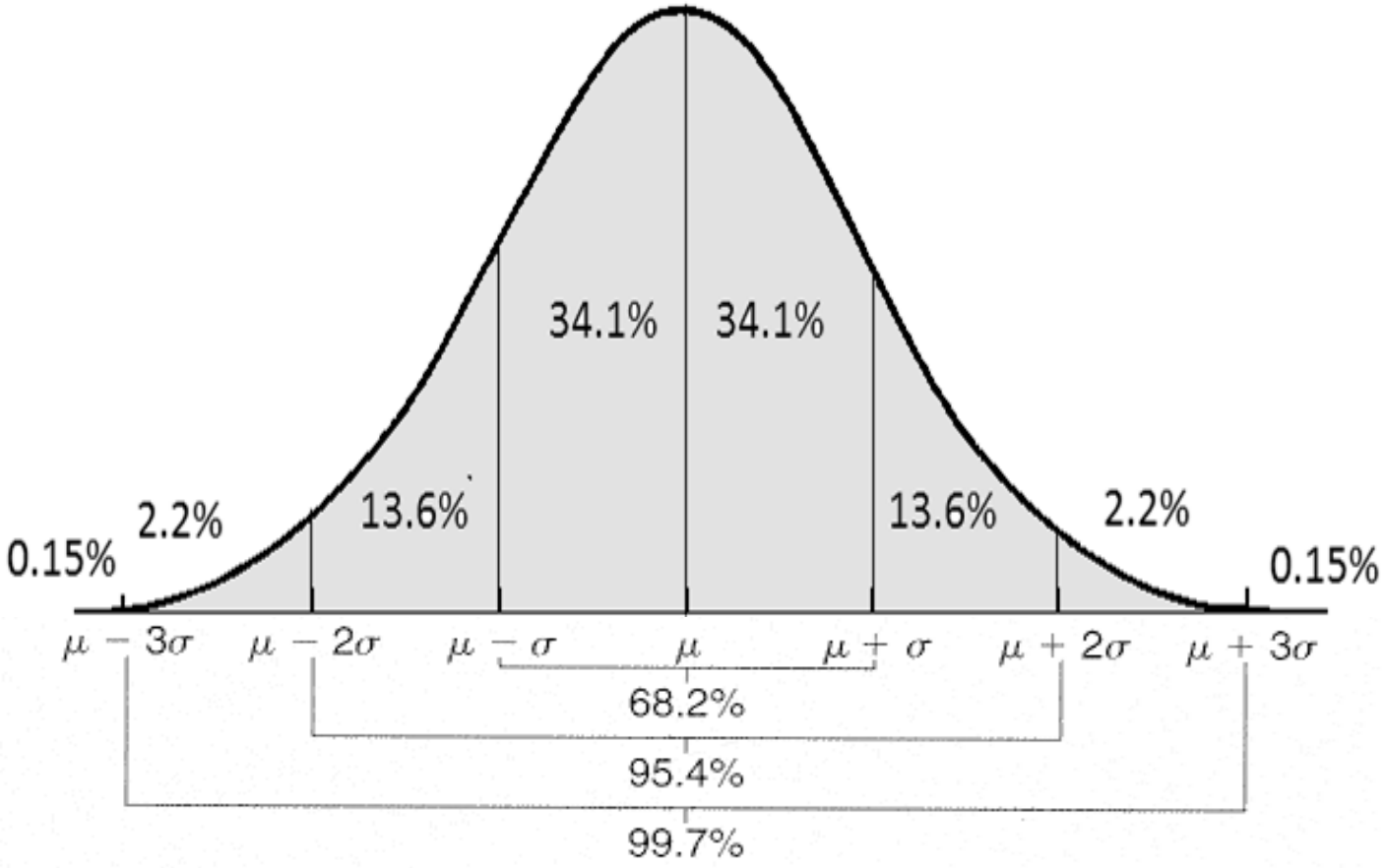


Proprietăți:

În intervalul medie ± abatere standard - minim 68,2% din observaţii;

În intervalul medie ± 2 * abatere standard - minim 95,4% din observaţii;

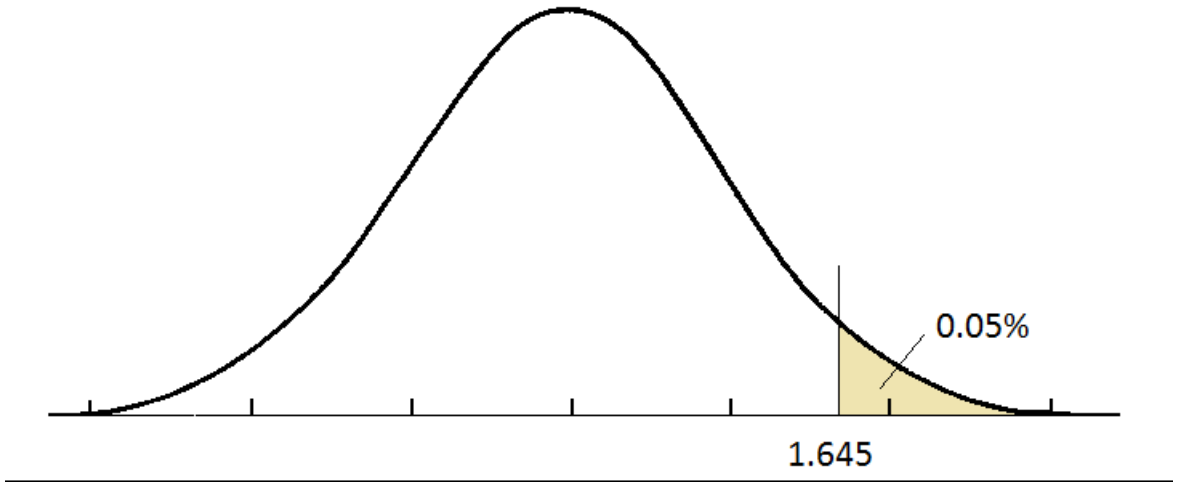
În intervalul medie ± 3 * abatere standard - minim 99,7% din observaţii.

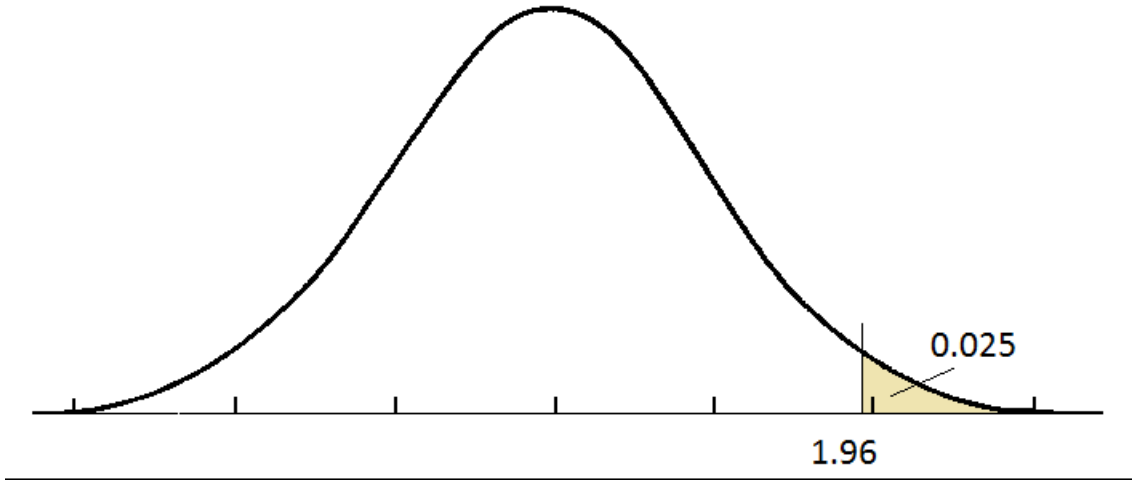


Intrebări

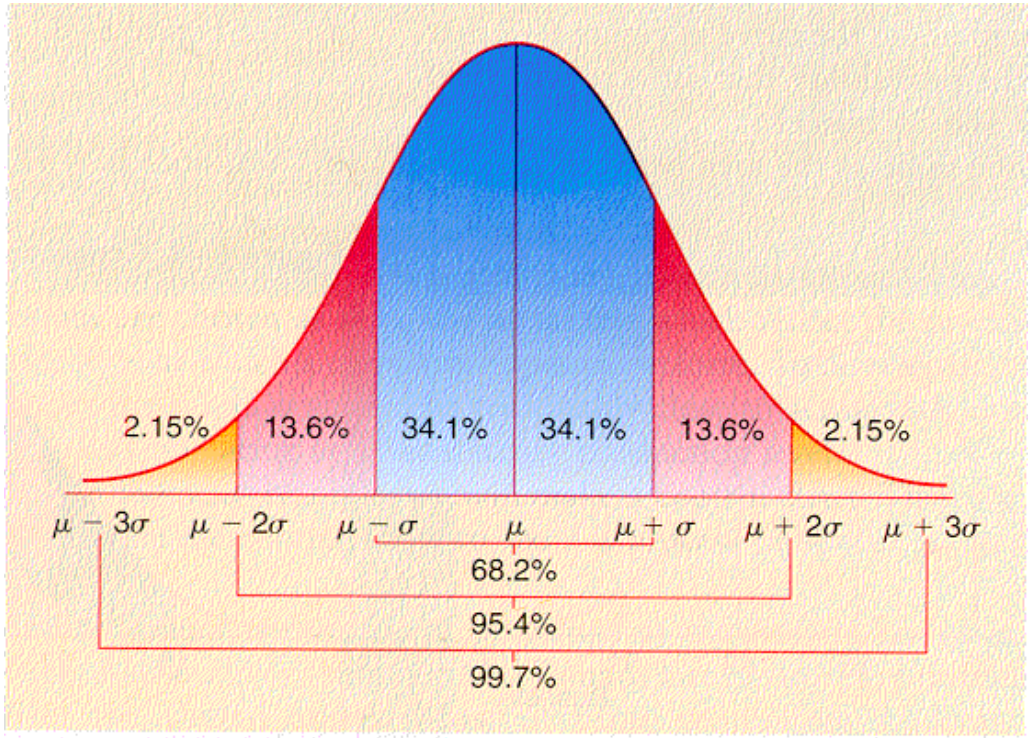
Există tabele cu aceste valori, există funcție în Excel care întoarce aceste valori, nu este necesar să știm să le calculăm, însă aceste două valori sunt de reținut, vom face estimări cu 5% eroare (vezi cursurile viitoare)

- Care valoare a lui Z divide aria in 95% și 5%? $Z_\alpha = 1{,}645$
- Care valoare a lui Z divide aria in 97,5% și 2,5%? $Z_\alpha = 1{,}96$





In mod normal în populație

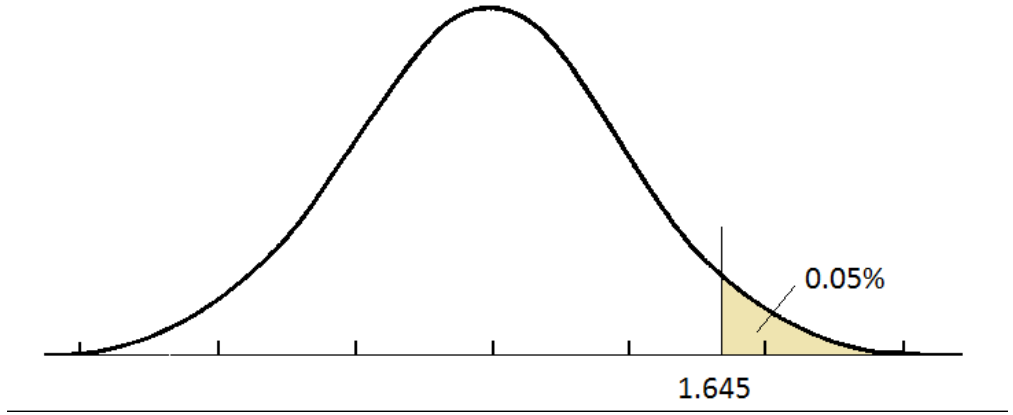


- Aproximativ 95,4% din distribuția de probabilitate - cuprinsă în intervalul medie +/- 2 σ (abaterea standard-populație)
- 95% din distribuția de probabilitate - cuprinsă în intervalul medie +/- 1,96 σ

Exerciții

Colesterolul - distribuit normal cu μ=160 si σ=15 dL/mg.

1. Care valoare a Colesterolului divide aria de sub curbă în 95% si 5%?



$$Z = \frac{X - \mu}{\sigma}$$

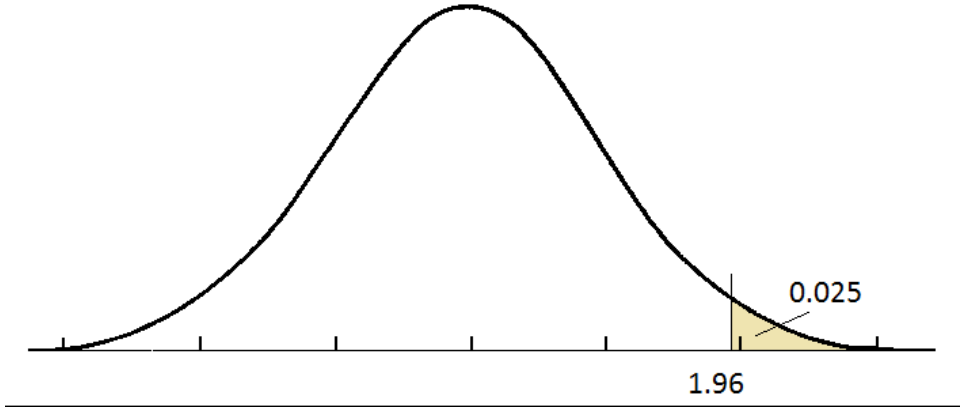
$$1{,}645 = \frac{X - 160}{15}$$

1,645*15=X-160

24,675=X-160

X=184,675

2. Care valoare a Colesterolului divide aria de sub curbă în 97.5% si 2.5%?



$$1{,}96 = \frac{X - 160}{15}$$

1,96*15=X-160

29,4=X-160

X=189,4

Aplicații: Cum este distribuția datelor?

Dacă aceste condiții sunt îndeplinite

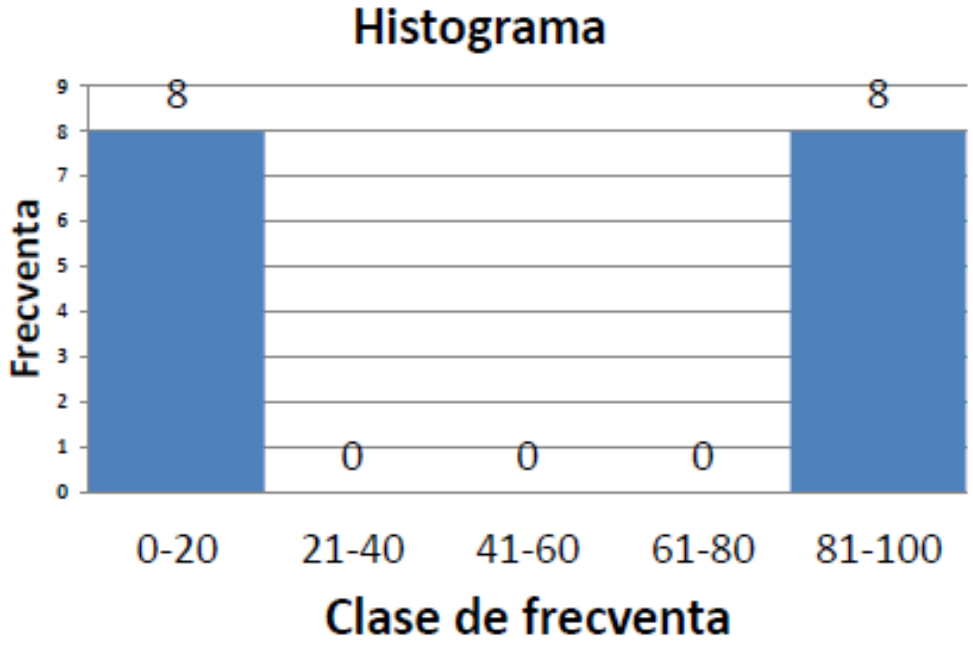
- media ≈ mediana ≈ modulul
- simetria ≈ 0
- boltirea ≈ 0
- cvartilele 1 și 3 simetrice față de media aritmetică
- În intervalul medie±abatere standard Ǝ minim 68,2% din observaţii;
- În intervalul medie±2*abatere standard Ǝ minim 95,4% din observaţii;
- În intervalul medie±3*abatere standard Ǝ minim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

- atunci distribuția datelor obținute empiric se apropie de distribuția normală

Exemplu – Seria 1

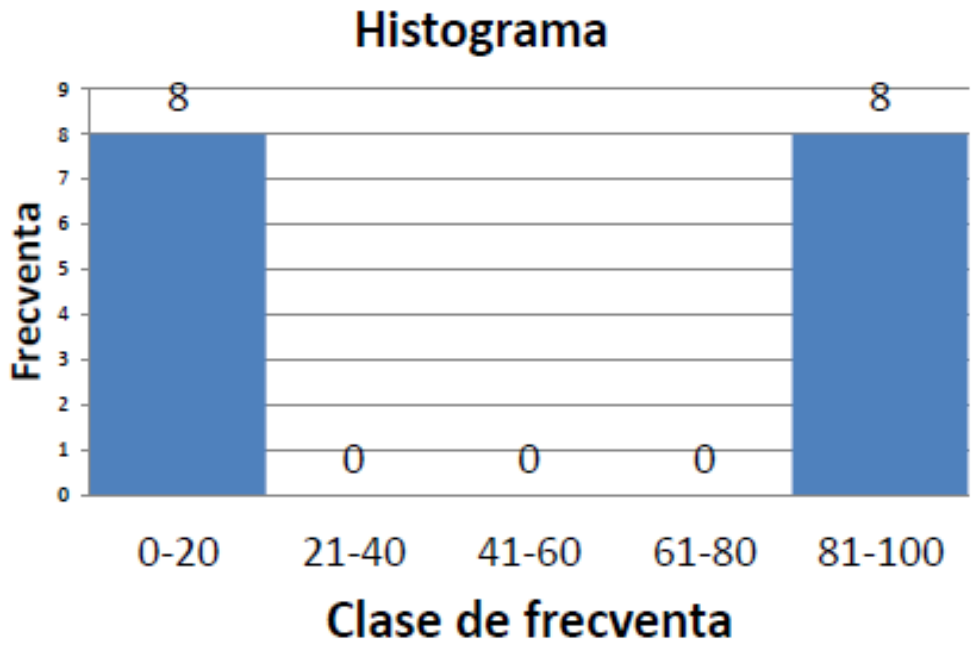
Seria 1
1
1
2
3
5
6
6
7
93
94
94
95
97
98
98
100

- 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

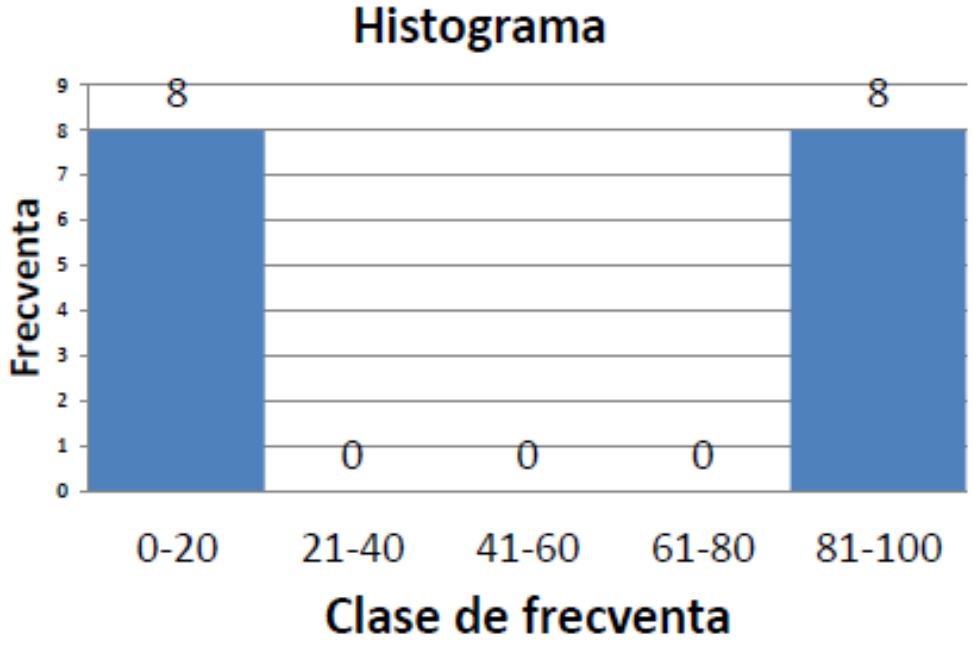


- Media aritmetică = 50
- Deviaţia standard = 47,70

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

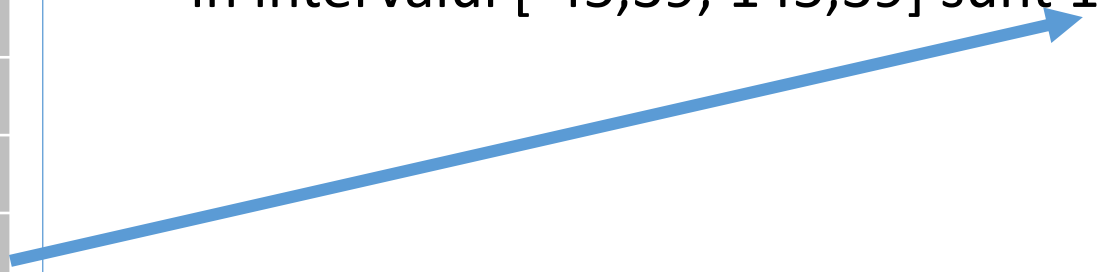
Seria 1
1
1
2
3
5
6
6
7
93
94
94
95
97
98
98
100

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

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

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

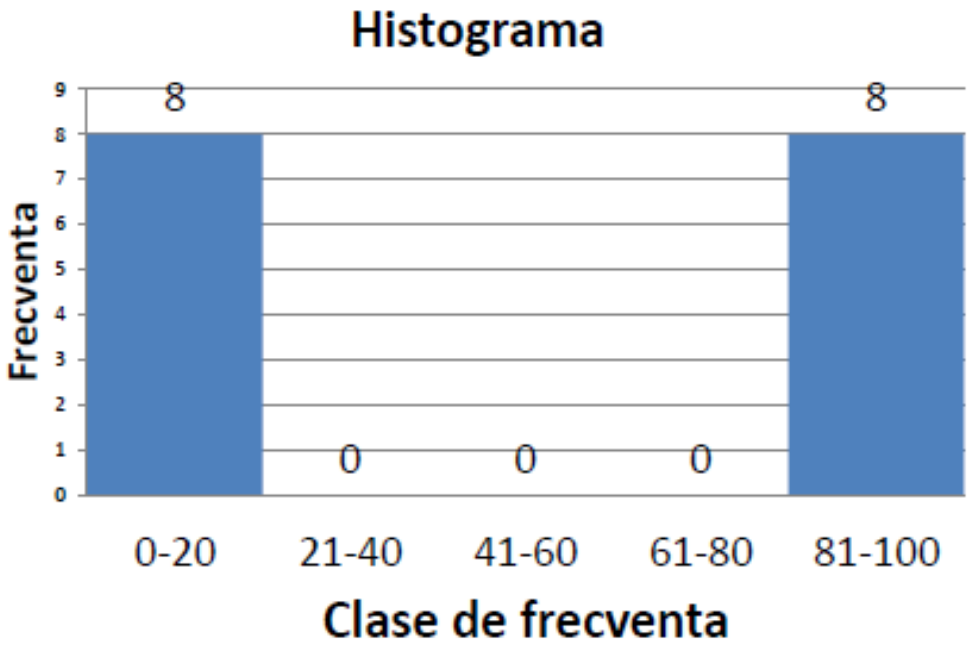
Media ± deviaţia standard = [2,3; 97,7] cu 10 valori, adică 62,5% dintre date

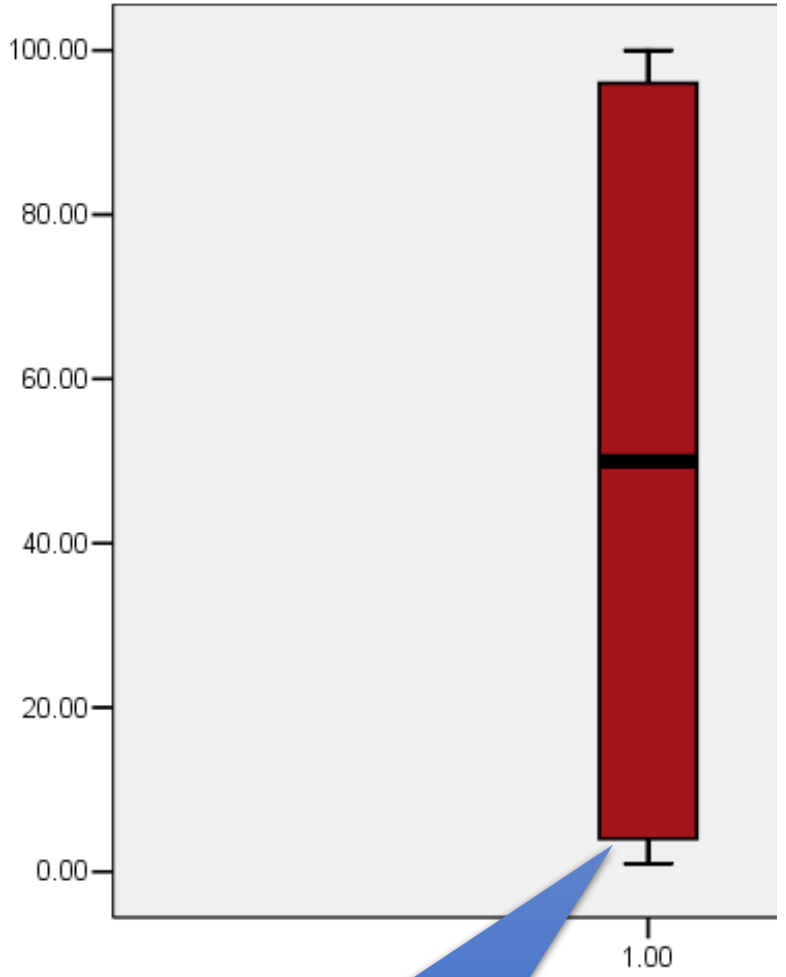
Mean ± 2*st.dev = [-45,39; 145,39] sunt 16 valori, adica 100% dintre date

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

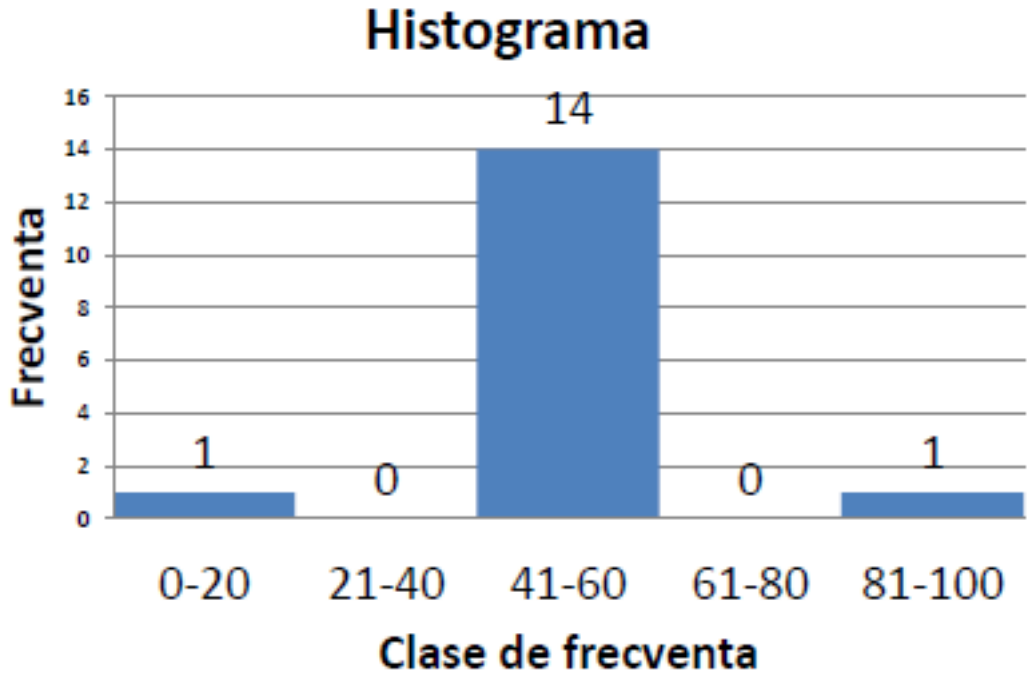
Deviaţia standard = 18,37

Cvartila 1 = 47,5

Cvartila 3 = 52,5

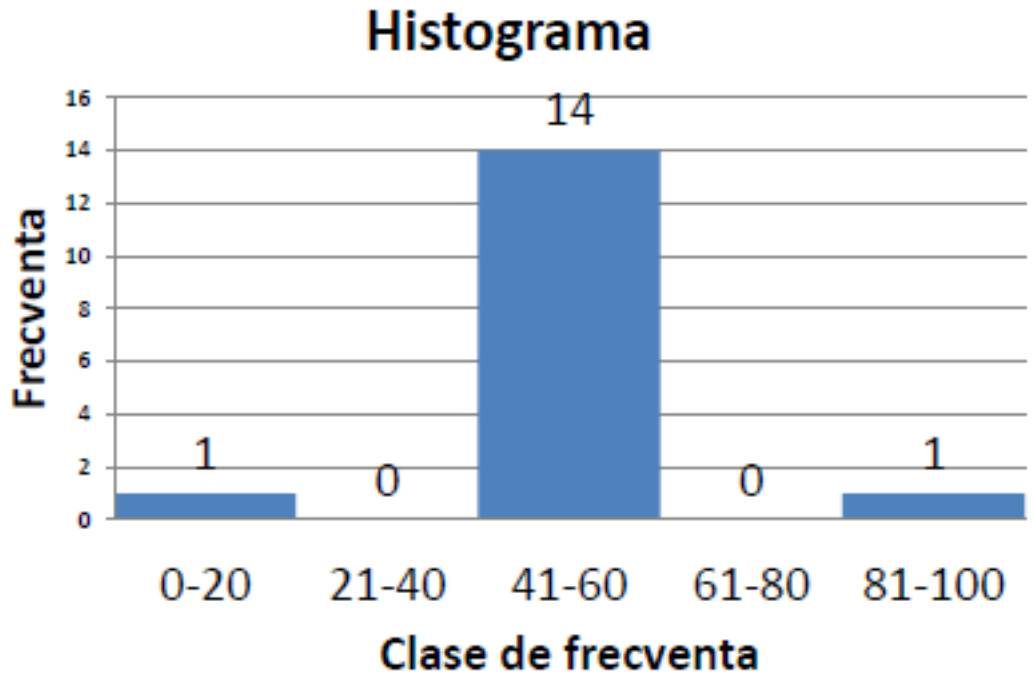
Simetria = 0,09

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+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

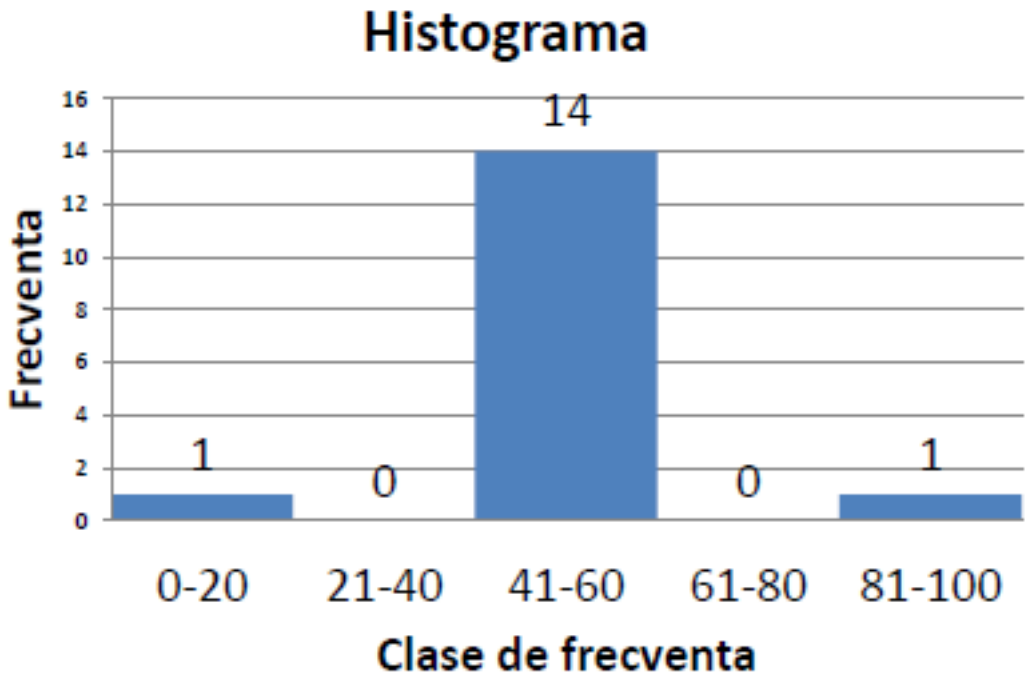
Media ± dev.st = [31,63; 68,37] sunt 14 valori - 87,5% din date

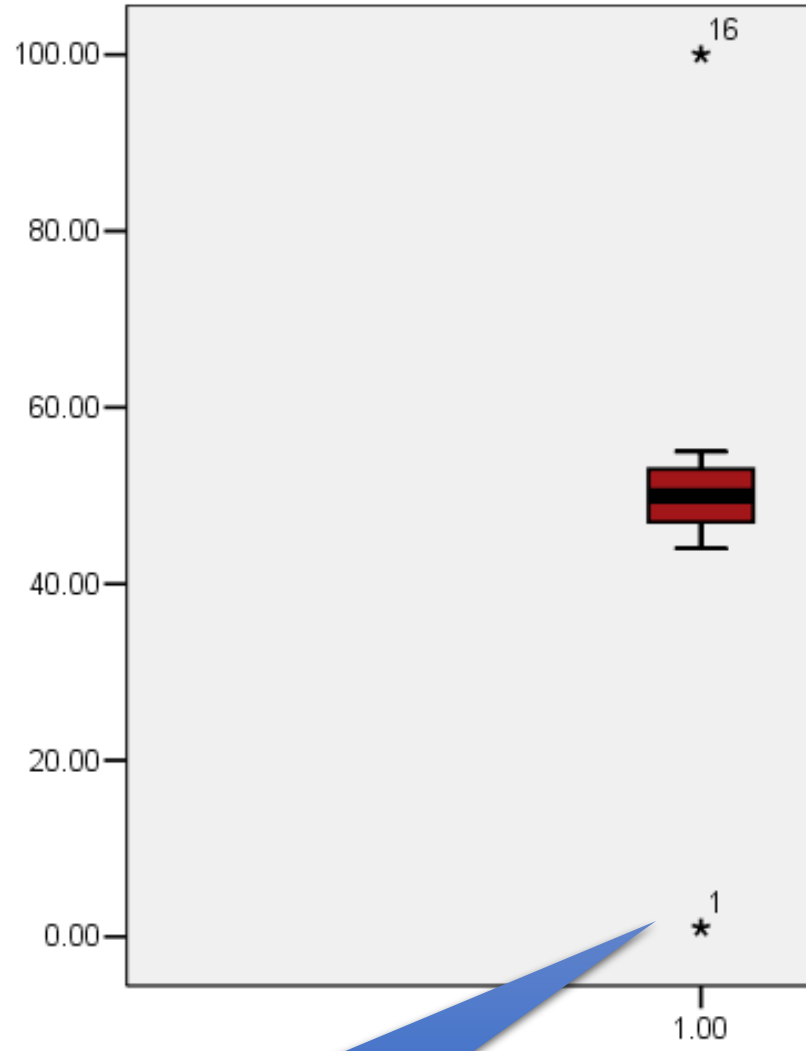
Media ± 2*dev.st. = [13,26; 86,74] sunt 14 valori - 87,5% din date

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

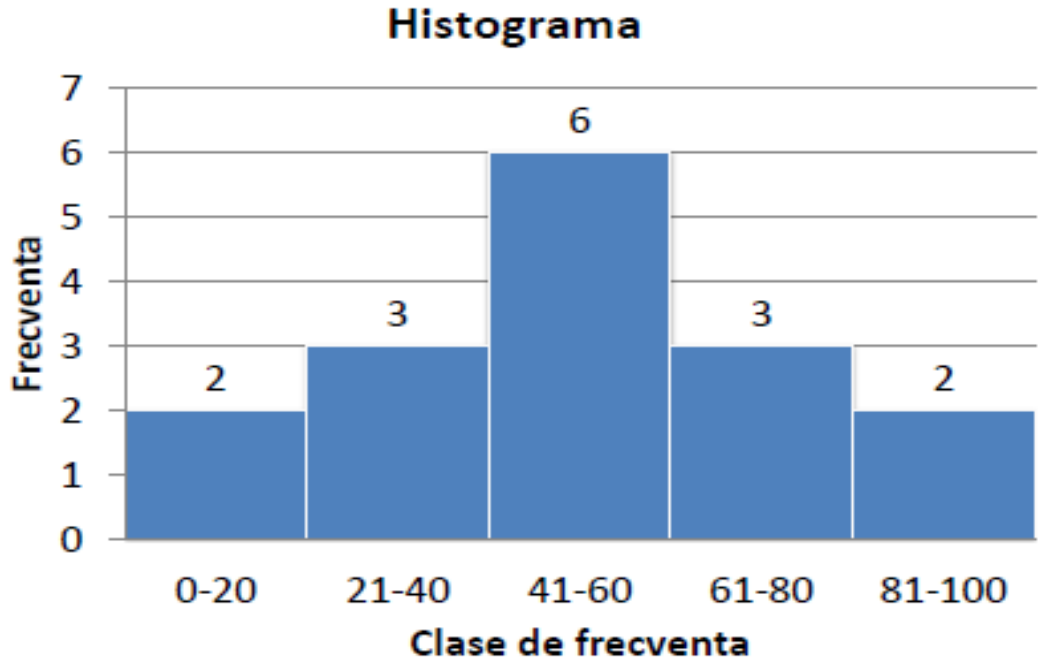
Deviaţia standard = 26,71

Cvartila 1 = 34,25

Cvartila 3 = 65,75

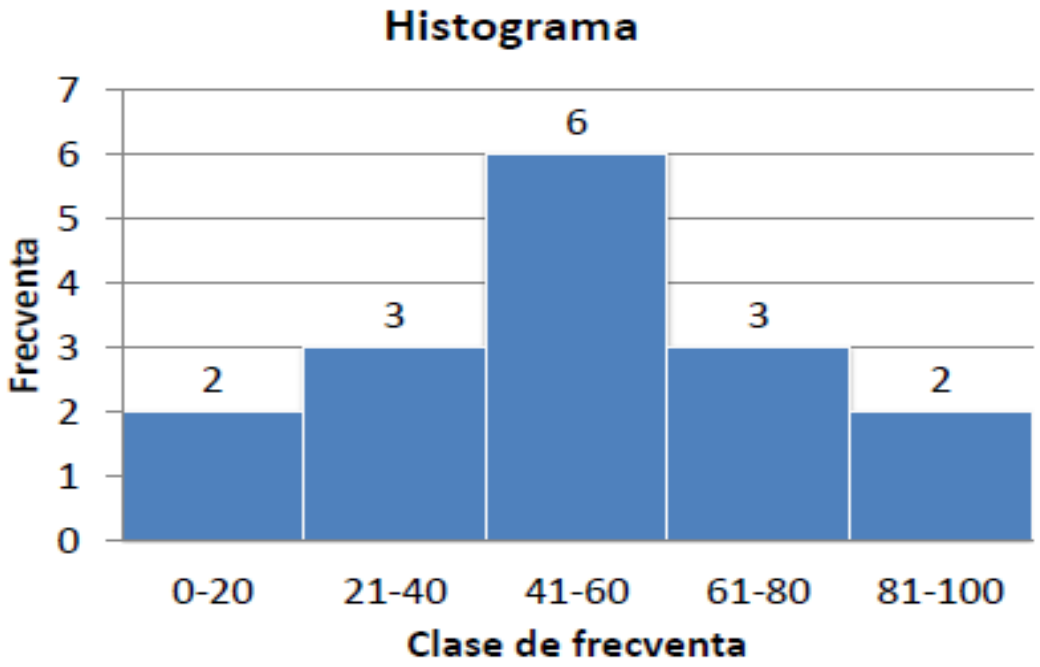
Simetria = 0,01

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

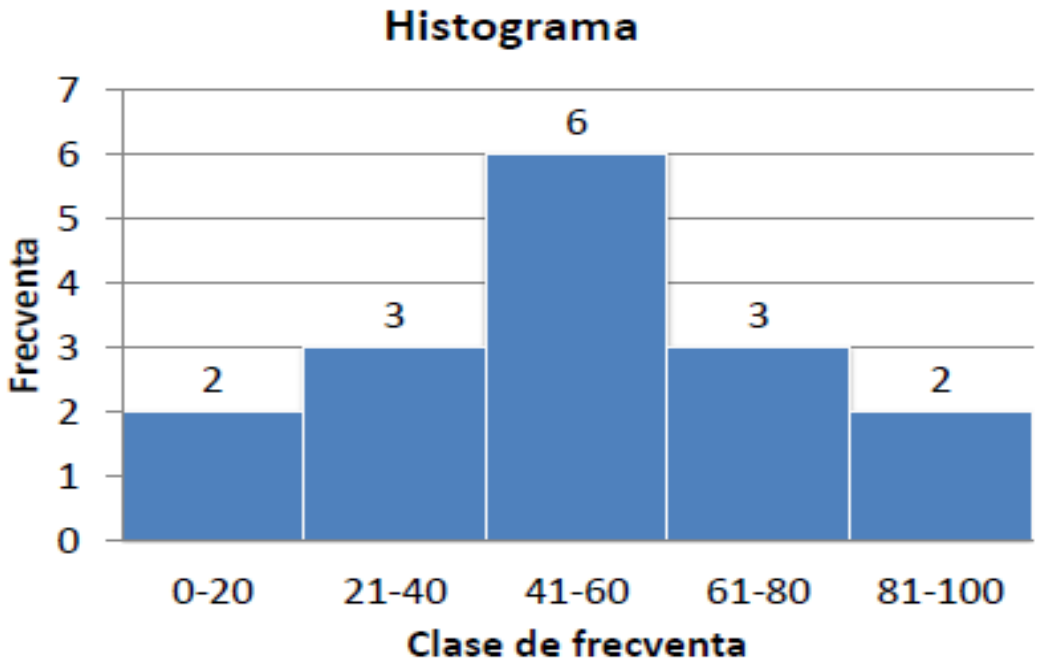
Deviaţia standard = 26,71

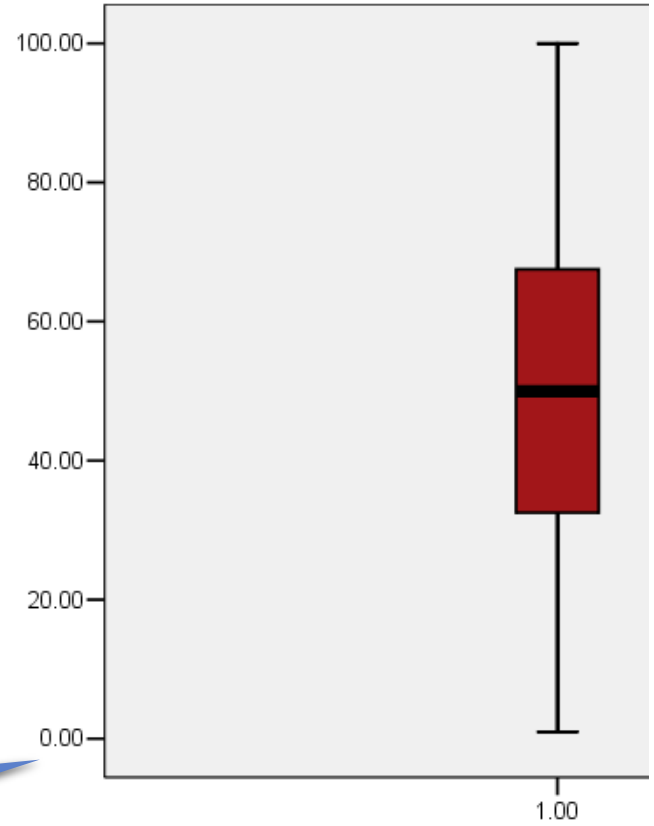
Media ± dev.st. = [23,28; 76,72] sunt 87,5% din date

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

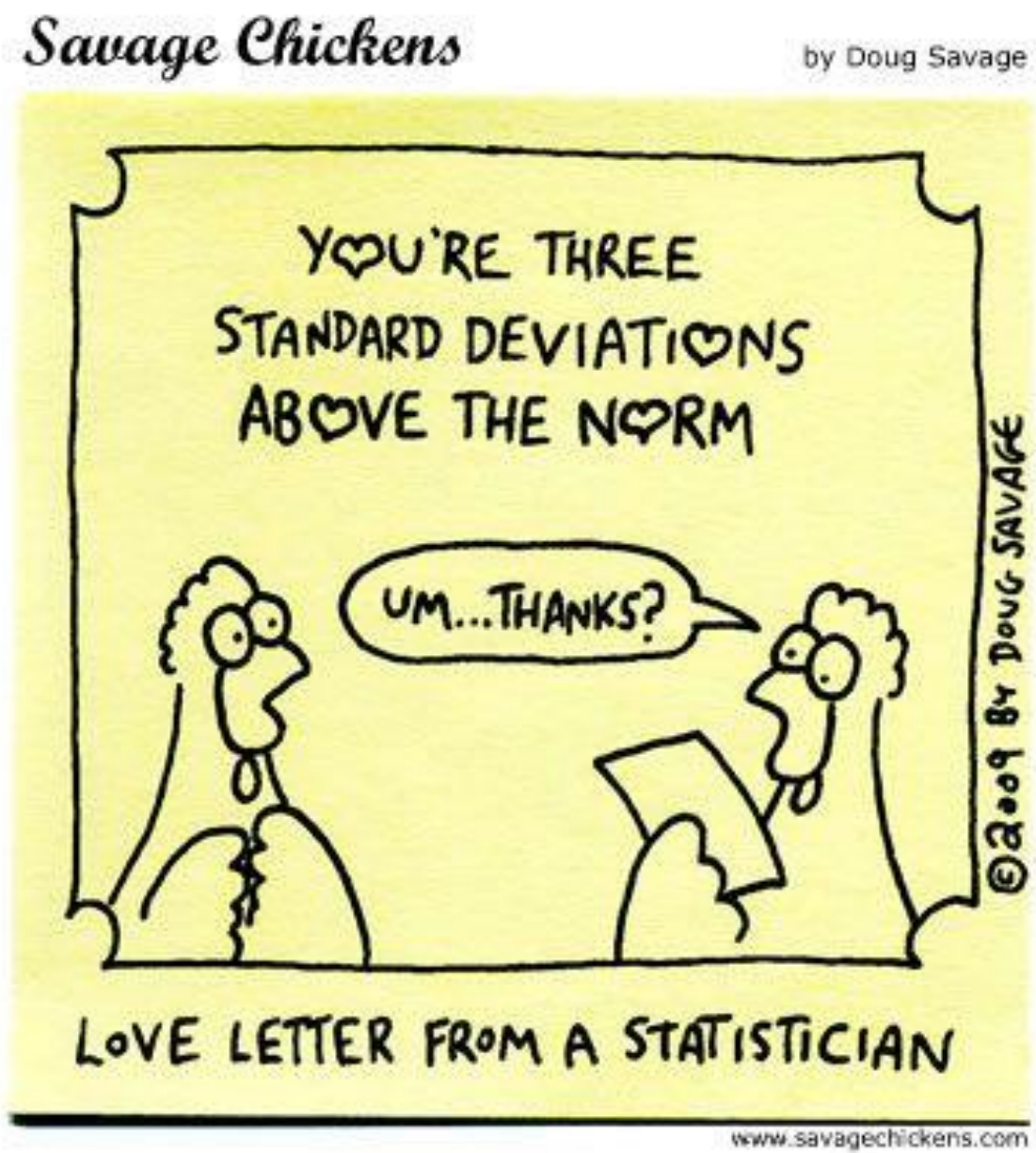
Media ± 3*dev. st. = [-30,15; 130,15] sunt 100% din date

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

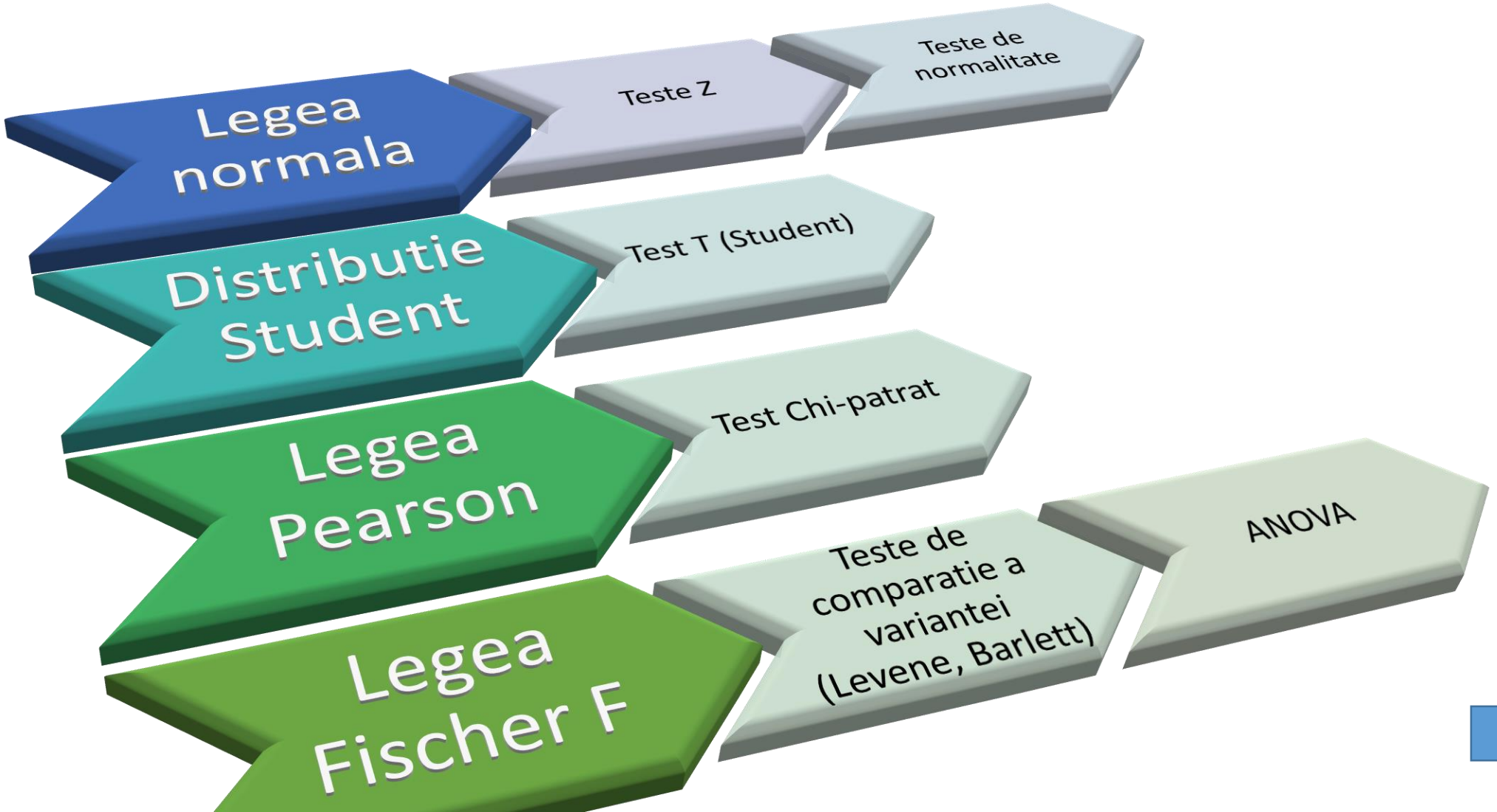




Distribuția este apropiată de cea normală


Savage Chickens
by Doug Savage
YOU'RE THREE STANDARD DEVIATIONS ABOVE THE NORM
UM...THANKS?
©2009 BY DOUG SAVAGE
LOVE LETTER FROM A STATISTICIAN
www.savagechickens.com

Distributie – Test statistic



Distribuția paranormală

- Mulțumesc!!!