



**Cosmina Ioana Bondor – autor**

---

## Probabilități – partea 2



ALWAYS



SEEK

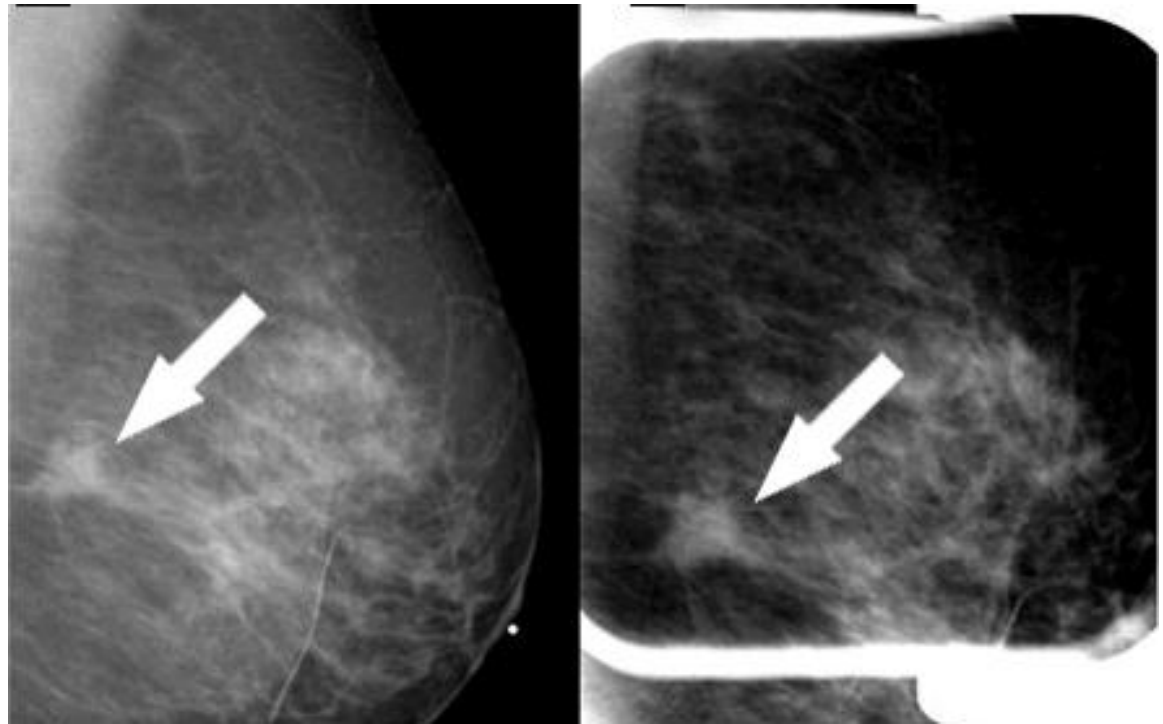


KNOWLEDGE

Obiectiv: recapitulare test diagnostic

# EXEMPLU

- Se recomandă femeilor peste 50 de ani să își facă o mamografie odată la 1-2 ani;
- testul diagnostic (de “aur”) pentru cancerul de san este biopsia dar este prea invaziv



- Din 82 de femei cu mamografii pozitive 77 au cancer de sân
- Din 9 femei cu mamografii negative 1 are cancer de sân

Ferranti FR, Vasselli F, Barba M, Sperati F, Terrenato I, Graziano F, Vici P, Botti C, Vidiri A. Diagnostic Accuracy of Contrast-Enhanced, Spectral Mammography (CESM) and 3T Magnetic Resonance Compared to Full-Field Digital Mammography plus Ultrasound in Breast Lesions: Results of a (Pilot) Open-Label, Single-Centre Prospective Study. *Cancers (Basel)*. 2022 Mar 7;14(5):1351.

Afecțiunea / Testul	Subiecți care au dezvoltat afecțiunea	Subiecți sănătoși	Total
Mamografie pozitivă	77	5	82
Mamografie negativă	1	8	9
Total	78	13	91

$$VPP = 77/82 = 0,93$$

$$VPN = 8/9 = 0,89$$

$$Se = 77/78 = 0,99$$

$$Sp = 8/13 = 0,62$$

$Se=0,99$

- performanța mamografiei de a detecta pozitivi

$Sp=0,62$

- performanța mamografiei de a detecta negativi

$VPP=0,93$

- corectitudinea unui test pozitiv

$VPN = 0,89$

- corectitudinea unui test negativ

# ! Confuzie între Se, Sp și VPP, VPN

Se performanța mamografiei de a detecta pozitivi

Sp performanța mamografiei de a detecta negativi

} din literatura sau prospect

VPP corectitudinea unui test pozitiv

VPN corectitudinea unui test negativ

} vedem în cabinet

Se performanța mamografiei de a detecta pozitivi

Sp performanța mamografiei de a detecta negativi

nu depind de prevalența  
bolii

VPP corectitudinea unui test pozitiv

VPN corectitudinea unui test negativ

depind de prevalența bolii  
crește prevalența bolii → crește VPP  
scade prevalența bolii → crește VPN

**! VPP, VPN – depind de prevalența bolii**



# Exemplu de fluctuație a VPP cu fluctuația prevalenței

- grup de femei selectate din populație
  - prevalența cancerului de sân: 0,0003%
  - VPP = 0,01
    - 1% dintre testele pozitive sunt corecte
- grup de femei selectate dintre cele suspecte
  - frecvența cancerului de sân cancerul de sân: 25%
  - VPP = 0,95
    - 95% dintre testele pozitive sunt corecte

în condițiile în care aplicăm același test pt. mamografie

# Exemple de fluctuație a VPN cu fluctuația prevalenței

- prevalența gripei **vara** 0,001
  - VPN=0,99
    - 99% dintre testele negative sunt corecte
- prevalența gripei **iarna** 0,01
  - VPN=0,45
    - 45% dintre testele negative sunt corecte

în condițiile în care aplicăm același test pt. gripa

# Măsurarea riscului: RR

# Riscul relativ RR

= raportul dintre cazurile de îmbolnăvire la cei cu expunere la factorul de risc și cazurile de îmbolnăvire la cei fără expunere la factorul de risc

$$RR = \frac{P(B|A)}{P(B|\bar{A})}$$

Non A

$$RR = \frac{P(Boala|Factor\ de\ risc)}{P(Boala|\overline{Fără\ factor\ de\ risc})}$$

# Exemplu

- Este pesimismul factor de risc pentru depresie?
- 2000 pesimiști din 10.000 fac depresie

$$P(\text{depresie} | \text{pesimiști}) = \frac{2000}{10.000} = 0,2$$

- 2000 optimiști din 10.000 fac depresie

$$P(\text{depresie} | \text{optimiști}) = \frac{2000}{10.000} = 0,2$$

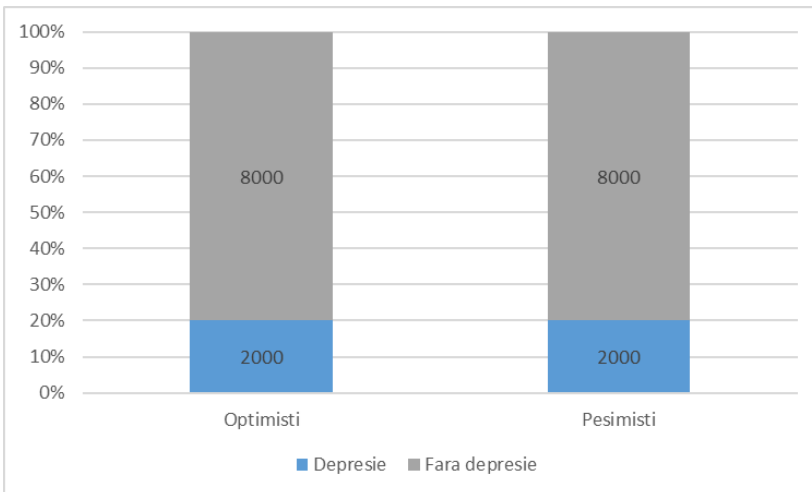
— riscul relativ

$$RR = \frac{P(\text{depresie} | \text{pesimiști})}{P(\text{depresie} | \text{optimiști})} = \frac{0,2}{0,2} = 1$$

# Interpretare RR

RR=1

Nu există risc de îmbolnăvire la persoanele expuse



riscul relativ

$$RR = \frac{0,2}{0,2} = 1$$

# Interpretare RR

$$RR=1$$

Nu există risc de îmbolnăvire la persoanele expuse

2000 pesimiști din 10.000 fac depresie

$$P(\text{depresie}|\text{pesimiști}) = 0,2$$

2000 optimiști din 10.000 fac depresie

$$P(\text{depresie}|\text{optimiști}) = 0,2$$

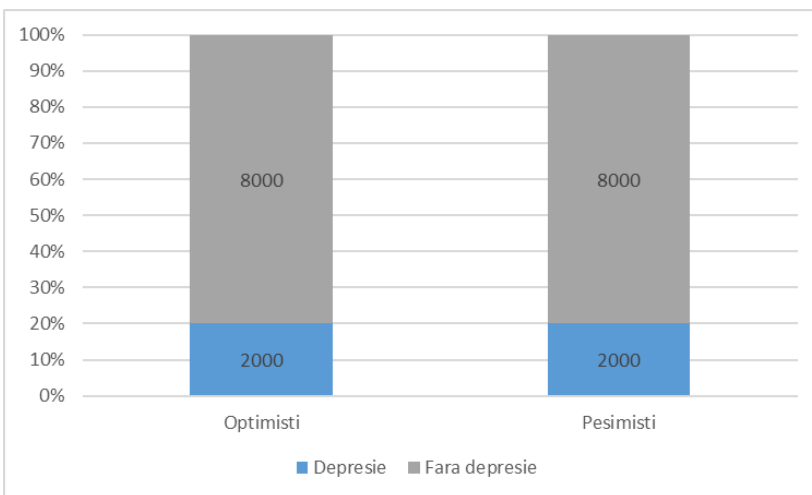
riscul relativ

$$RR = \frac{0,2}{0,2} = 1$$

# Interpretare RR

$$RR=1$$

Nu există risc de îmbolnăvire la persoanele expuse

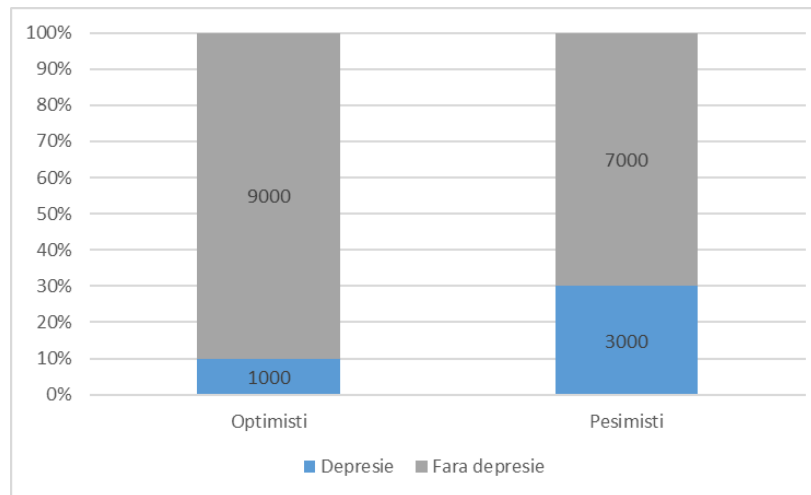


riscul relativ

$$RR = \frac{0,2}{0,2} = 1$$

$$RR>1$$

Există risc de a face boala dacă există expunere la factorul de risc



riscul relativ

$$RR = \frac{0,3}{0,1} = 3$$



# Interpretare RR

$$RR=1$$

Nu există risc de îmbolnăvire la persoanele expuse

2000 pesimiști din 10.000 fac depresie

$$P(\text{depresie}|\text{pesimiști}) = 0,2$$

2000 optimiști din 10.000 fac depresie

$$P(\text{depresie}|\text{optimiști}) = 0,2$$

riscul relativ

$$RR = \frac{0,2}{0,2} = 1$$

$$RR>1$$

Există risc de a face boala dacă există expunere la factorul de risc

3000 pesimiști din 10.000 fac depresie

$$P(\text{depresie}|\text{pesimiști}) = 0,2$$

1000 optimiști din 10.000 fac depresie

$$P(\text{depresie}|\text{optimiști}) = 0,1$$

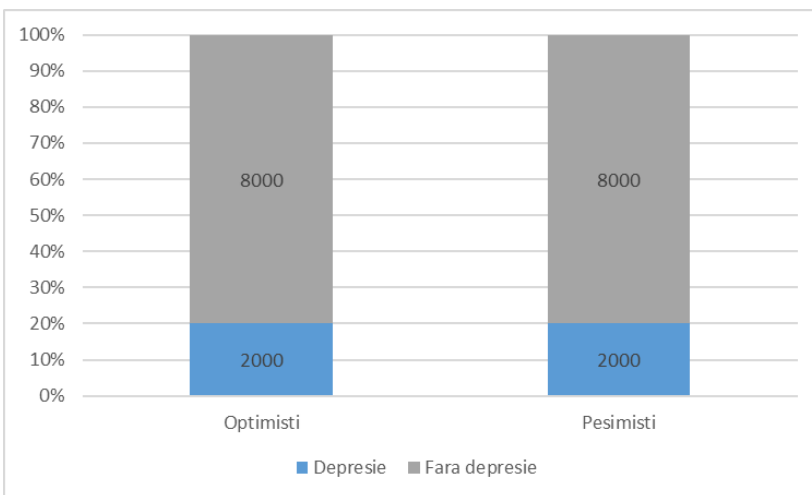
riscul relativ

$$RR = \frac{0,2}{0,1} = 2$$

# Interpretare RR

## RR=1

Nu există risc de îmbolnăvire la persoanele expuse

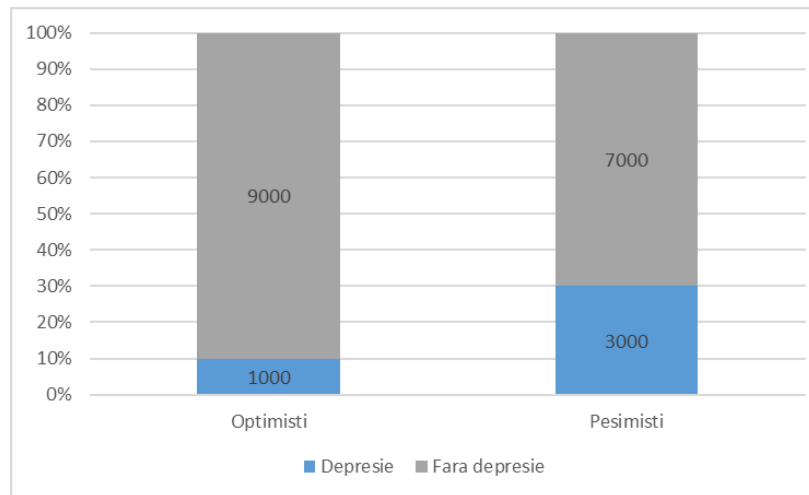


riscul relativ

$$RR = \frac{0,2}{0,2} = 1$$

## RR>1

Există risc de a face boala dacă există expunere la factorul de risc

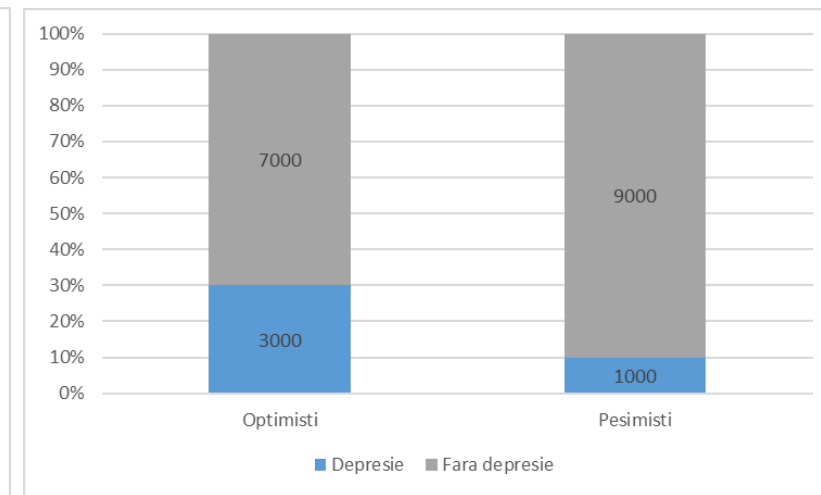


riscul relativ

$$RR = \frac{0,3}{0,1} = 3$$

## RR<1

Expunerea este factor de protecție pentru boală



riscul relativ

$$RR = \frac{0,1}{0,3} = 0,33$$

# Interpretare RR

$$RR=1$$

Nu există risc de îmbolnăvire la persoanele expuse

2000 pesimiști din 10.000 fac depresie

$$P(\text{depresie}|\text{pesimiști}) = 0,2$$

2000 optimiști din 10.000 fac depresie

$$P(\text{depresie}|\text{optimiști}) = 0,2$$

riscul relativ

$$RR = \frac{0,2}{0,2} = 1$$

$$RR>1$$

Există risc de a face boala dacă există expunere la factorul de risc

3000 pesimiști din 10.000 fac depresie

$$P(\text{depresie}|\text{pesimiști}) = 0,2$$

1000 optimiști din 10.000 fac depresie

$$P(\text{depresie}|\text{optimiști}) = 0,2$$

riscul relativ

$$RR = \frac{0,3}{0,1} = 3$$

$$RR<1$$

Expunerea este factor de protecție pentru boală

1000 pesimiști din 10.000 fac depresie

$$P(\text{depresie}|\text{pesimiști}) = 0,2$$

3000 optimiști din 10.000 fac depresie

$$P(\text{depresie}|\text{optimiști}) = 0,2$$

riscul relativ

$$RR = \frac{0,1}{0,3} = 0,33$$

# Interpretare RR

$$RR=1$$

Nu există risc de îmbolnăvire la persoanele expuse

$$RR=20$$

Există risc de 20 de ori mai mare de a face boala dacă există expunere la factorul de risc

$$RR=0,05$$

Există risc de 20 de ori mai mic de a face boala dacă există expunere la factorul de risc

# Teorema lui Bayes

Evenimente dependente (cauză și efect)

Probabilitatea condiționată - probabilitatea unui rezultat care depinde de un rezultat anterior.

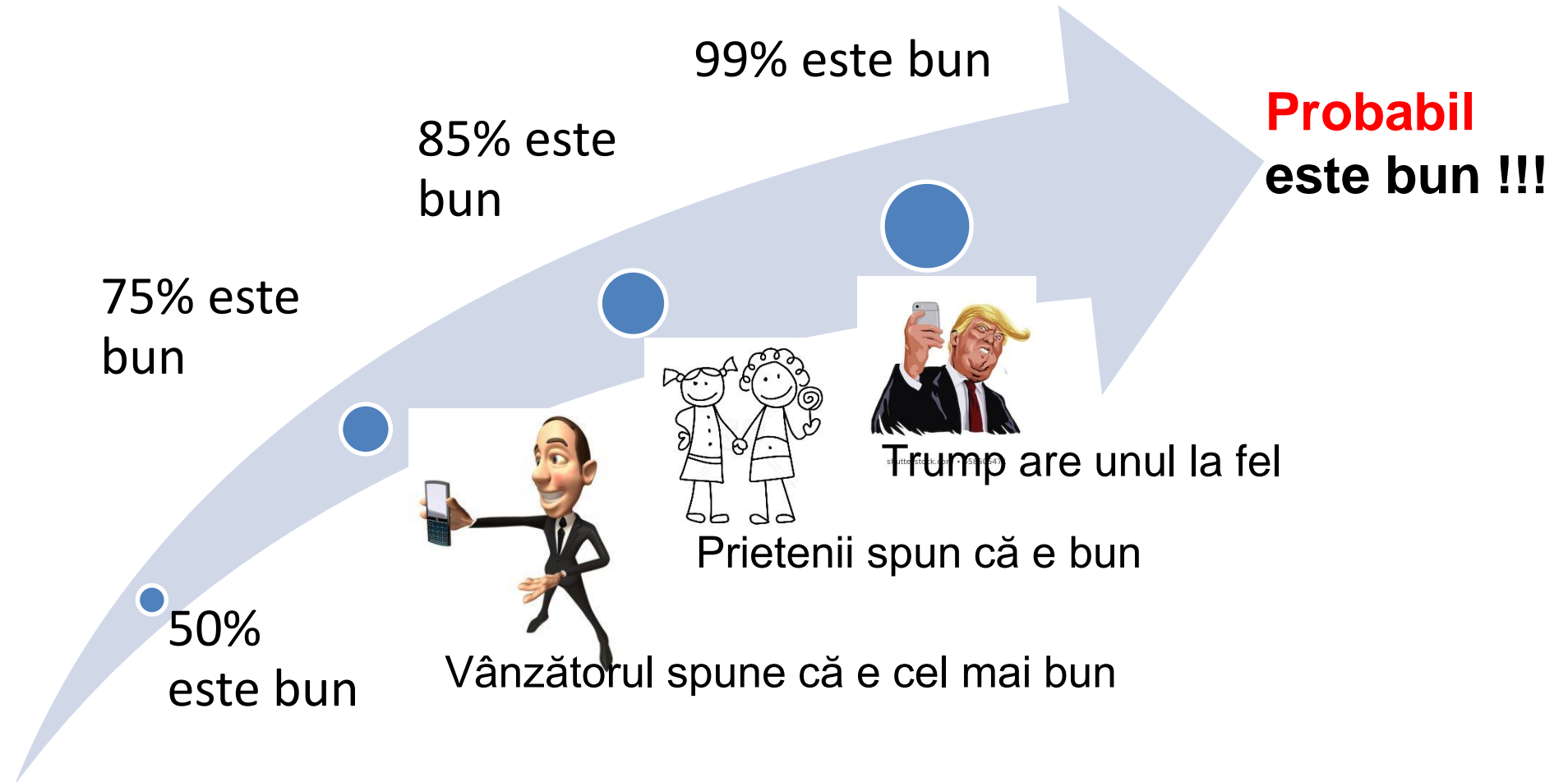
$$Pr(A | B) = \frac{Pr(B | A) * Pr(A)}{Pr(B)}.$$

Implicată în luarea deciziilor medicale - teste diagnostice.

# Teorema lui Bayes - după producerea unui nou eveniment îmbunătățim ceea ce știm deja

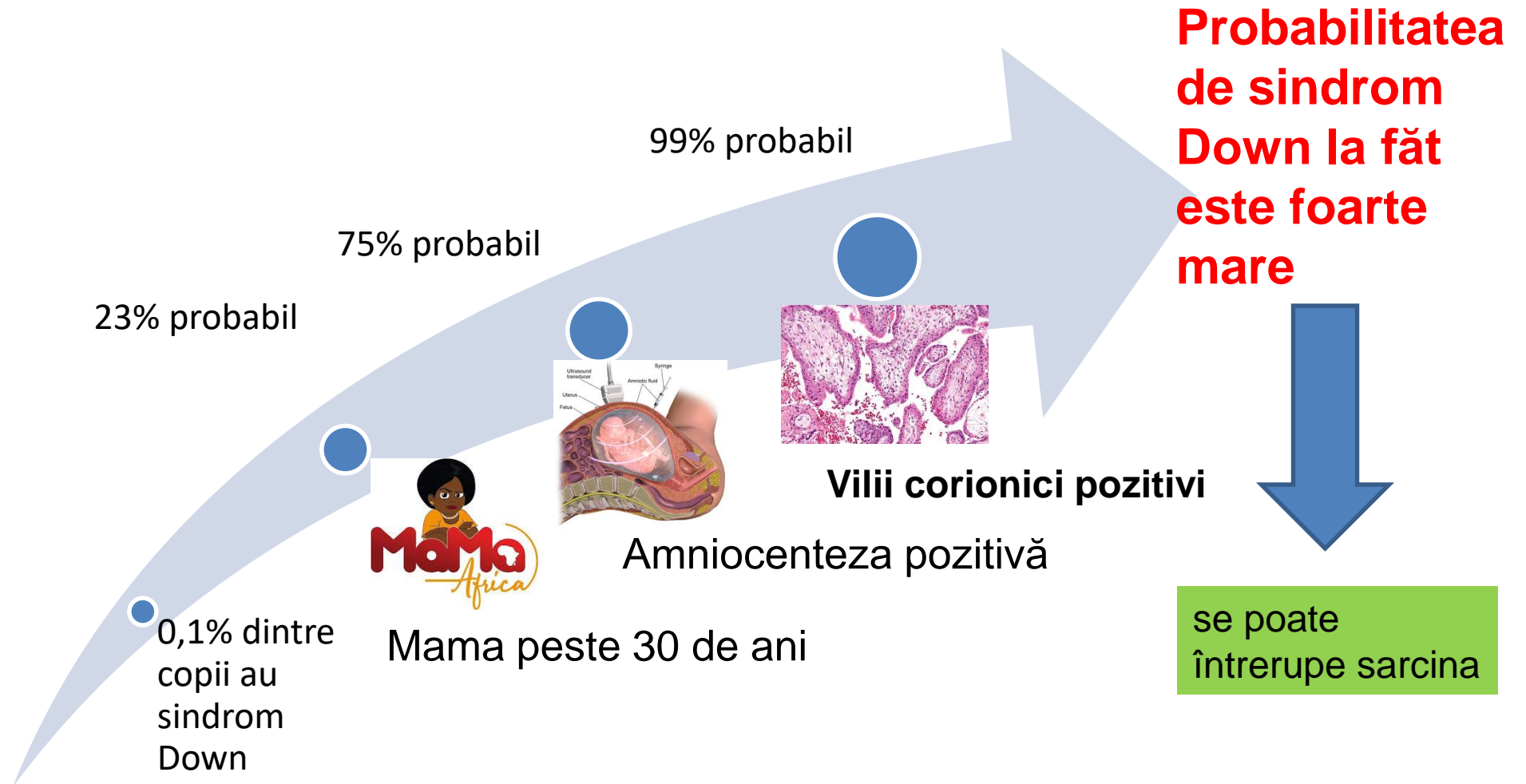
- Așa învață calculatoarele
- Așa funcționează filtrele de spam
- Așa pun medicii diagnosticele
- Așa știm ce le place prietenilor noștri

# Vreau să cumpăr un telefon smart Apple. Este bun sau nu este bun?

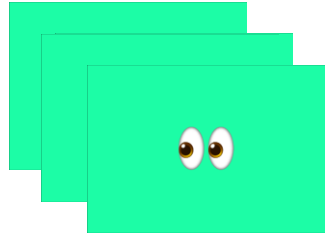


# Sindrom Down la făt în timpul sarcinii?

Prezintă fătul afecțiunea? Putem să fim destul de siguri ca să facem o  
întrerupere de sarcină?







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



**Cosmina Ioana Bondor – autor**

# Asociere și predicție



**ALWAYS**



**SEEK**



**KNOWLEDGE**

# Obiective

- Corelație
- Regresie liniară
- Exerciții

36 de pacienți cu artroplastie bilaterală totală a genunchiului

- A fost măsurată:
  - Greutatea
  - Densitatea minerală osoasă
- Scop: Există asociere între greutate și densitatea minerală osoasă

Ishii Y, Noguchi H, Sato J, Ishii H, Todoroki K, Toyabe SI. Association between body weight and proximal tibial bone mineral density after bilateral total knee arthroplasty. Knee. 2017 Oct;24(5):1153-1159.

# Corelație

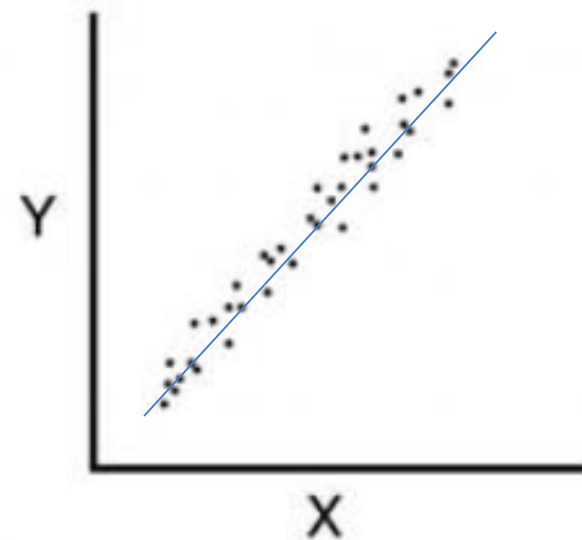
- Relația dintre două caracteristici
- Cum este relația?
- Putem face predicții?
- Ce eroare putem să acceptăm?

# Asociere

- Termeni similari: relație, dependență.
- Termenul **corelație** se utilizează numai în cazul a două variabile numerice sau ordinale



- Relație neliniară



Relație liniară

# Coeficient de corelație Pearson

Dacă X și Y sunt două variabile cantitative sau ordinale.  
Coeficientul de corelație Pearson:

$$r_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

unde n - numărul de pacienți (dimensiunea eșantionului),  $\bar{X}, \bar{Y}$  - media aritmetică a variabilelor X respectiv Y

Nu se cere să știți să-l calculați, pentru pasionați: așa arată formula



# Coeficient de corelație Pearson

- Indică asocierea dintre variabilele X și Y
- $r=1$  sau  $r=-1$  corelație perfectă
- la valori între -1 și 1:

$$r \in [-1, 1]$$

Cu cât  $|r|$  în modul se apropie de 1 cu atât asocierea este mai puternică

Cu cât  $|r|$  în modul se apropie de 0 cu atât asocierea este mai slabă

! Vorbim despre relație liniară



# Graficul de corelație – XY Scatter

- Două variabile – pot fi cu două unități de măsură diferite
- Variabila dependentă pe axa OY și variabila independentă pe axa OX

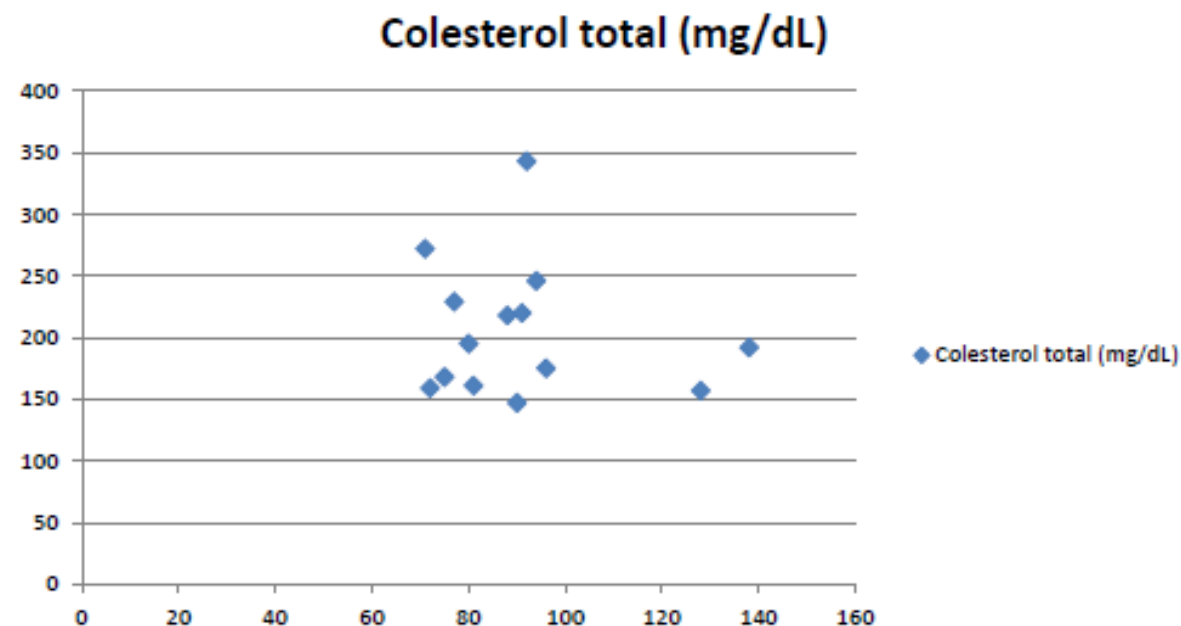
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		

Glicemie (mg/dL)	Colesterol total (mg/dL)
75	168
92	343
77	229
128	157
81	161
138	192
88	218
72	159
71	272
80	195
91	220
94	246
90	147
96	175

Selectam cele doua coloane cu date

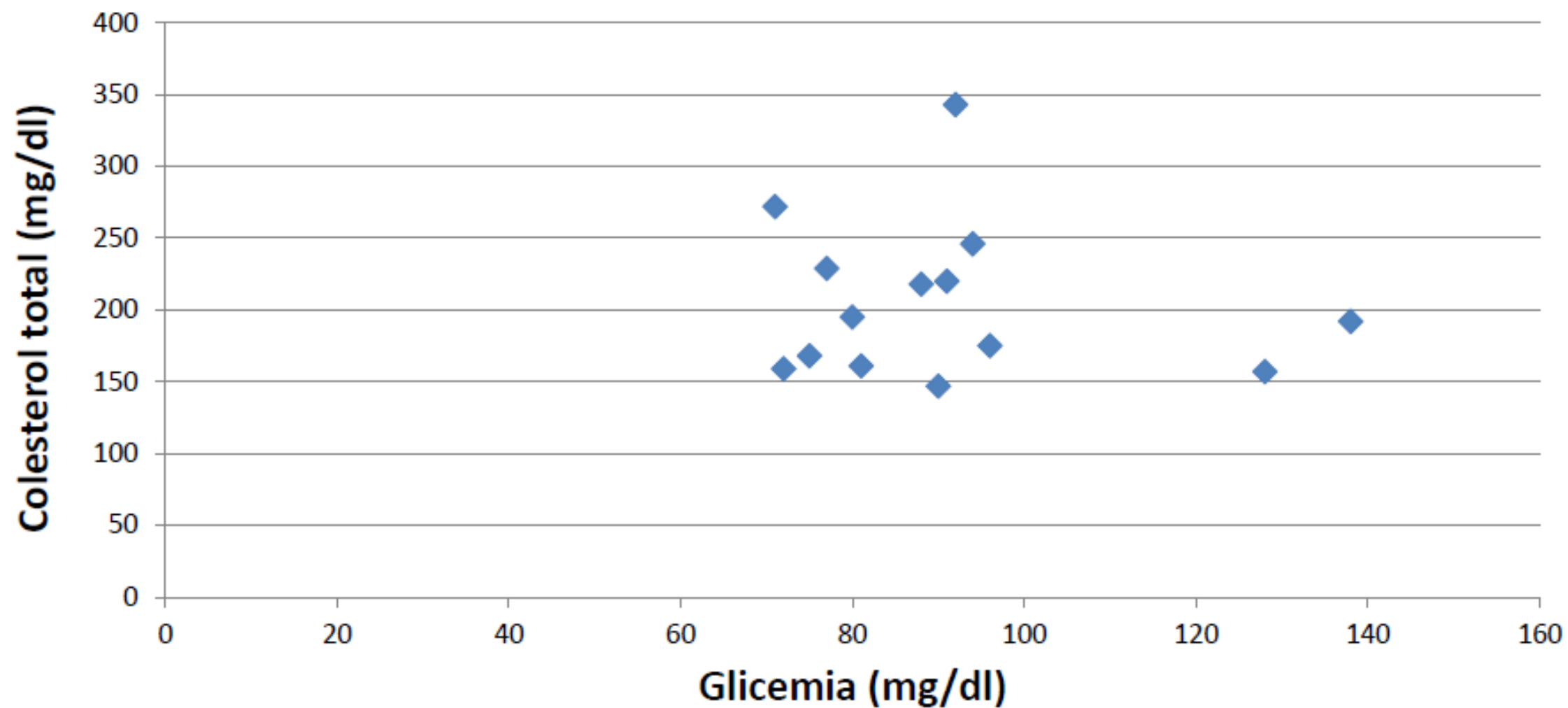
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		

Glicemie (mg/dL)	Colesterol total (mg/dL)
75	168
92	343
77	229
128	157
81	161
138	192
88	218
72	159
71	272
80	195
91	220
94	246
90	147
96	175



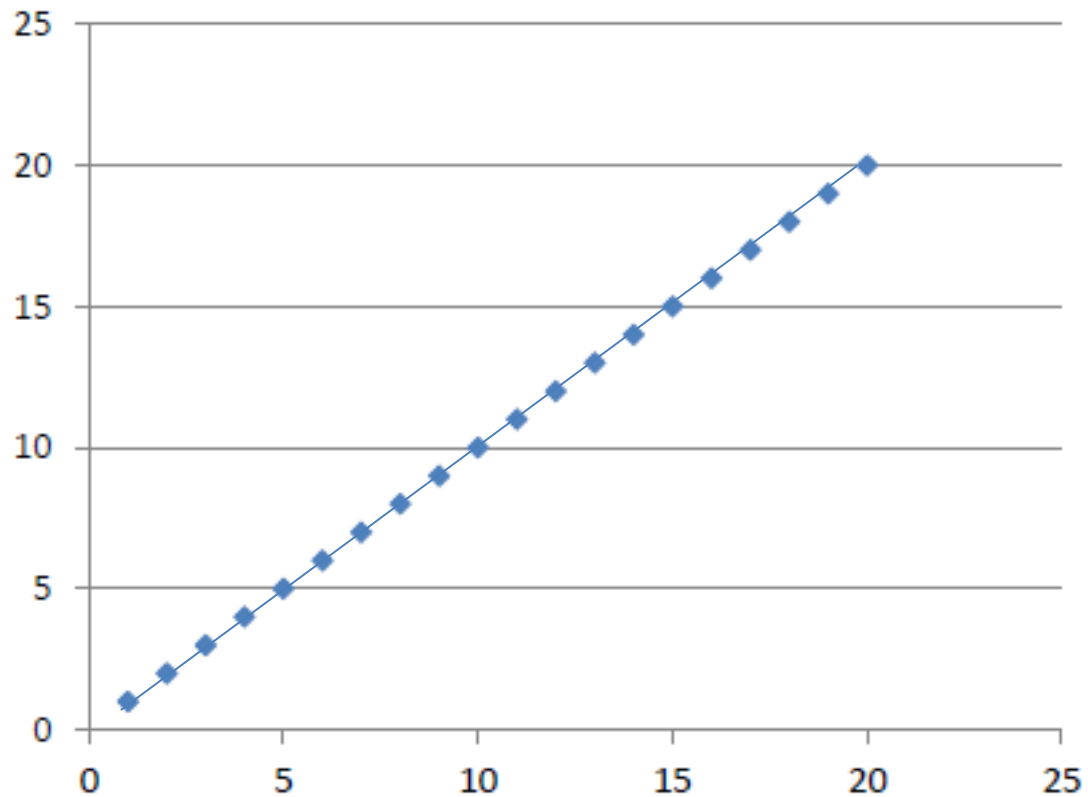
Inserăm grafic Scatter

## Corelatia dintre colesterol si glicemie



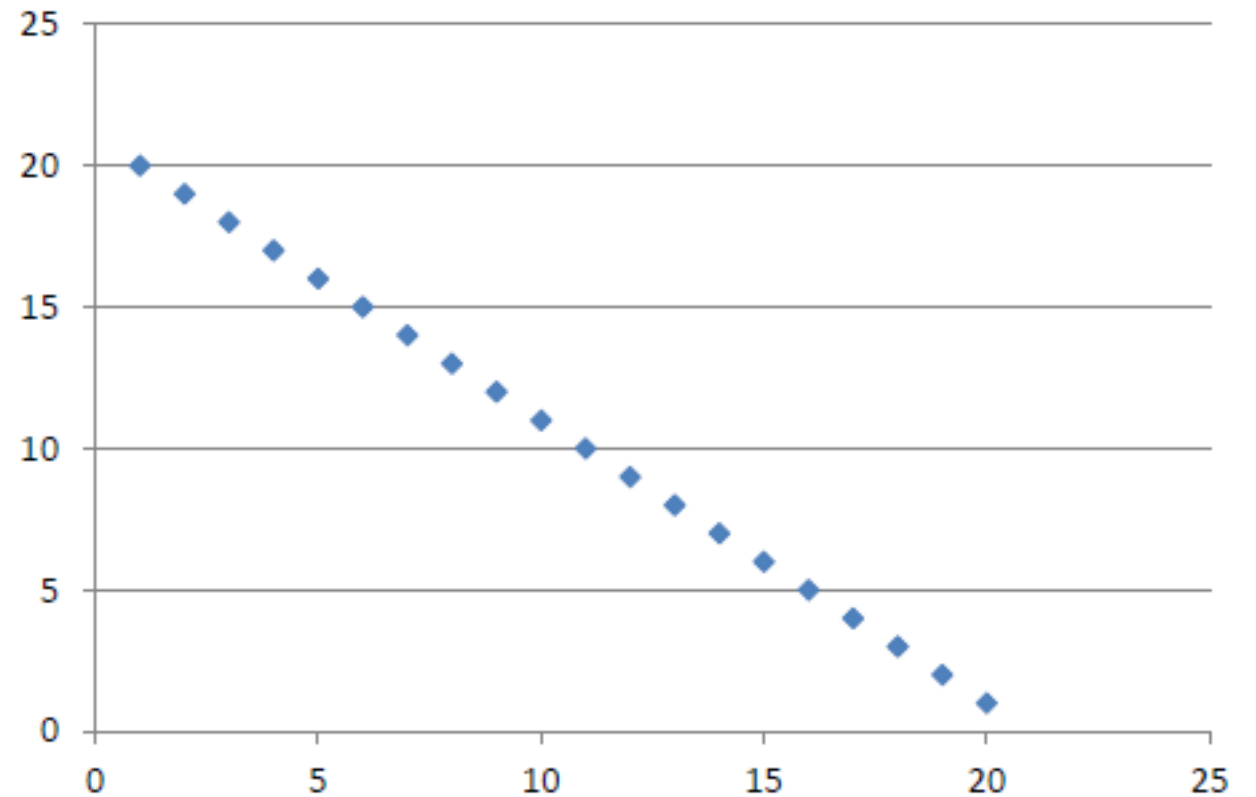
# Corelația perfectă

- $r=1$



Pozitiva (direct proporțională)

$r=-1$



Negativa (invers proporțională)

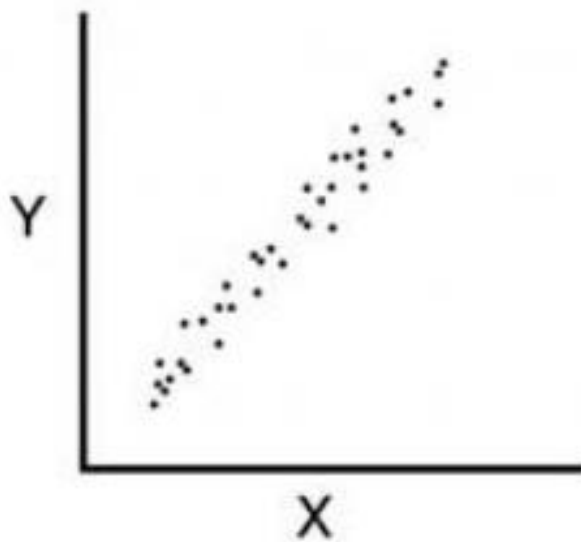
# Coeficient de corelație Pearson

- Dacă  $r > 0$  atunci relația dintre X și Y este **pozitivă** (directă)
  - valorilor mici ale lui X – le corespund valori mici ale lui Y,
  - valorilor mari ale lui X – le corespund valori mari ale lui Y
- Dacă  $r < 0$  atunci relația dintre X și Y este **negativă** (inversă)
  - valorilor mici ale lui X – le corespund valori mari ale lui Y,
  - valorilor mari ale lui X – le corespund valori mici ale lui Y

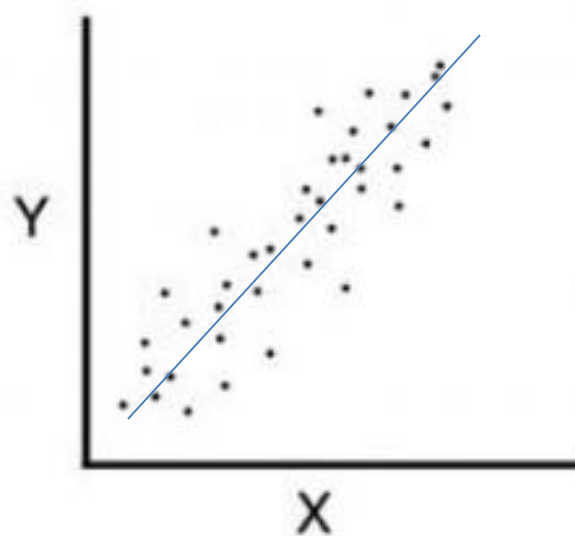


# Graficul de corelație – XY Scatter

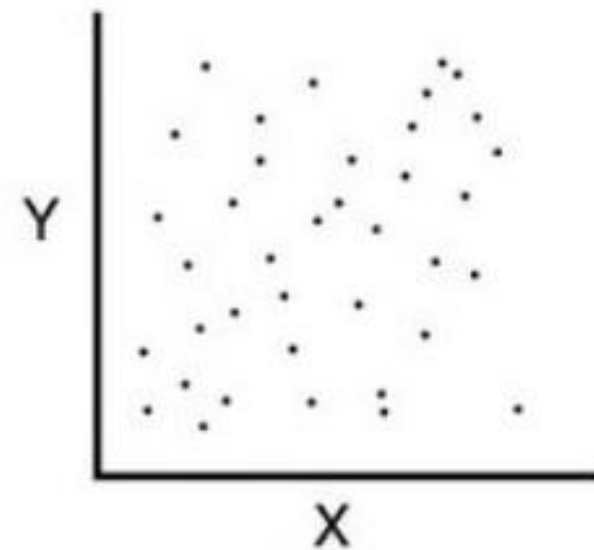
- Relație directă, pozitivă



Corelație puternică  
 $r \approx 1$



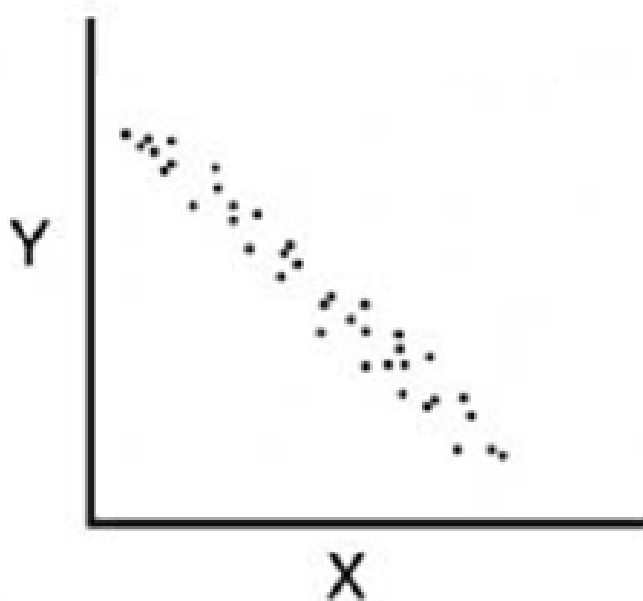
Corelație moderată  
 $r \approx 0.5$



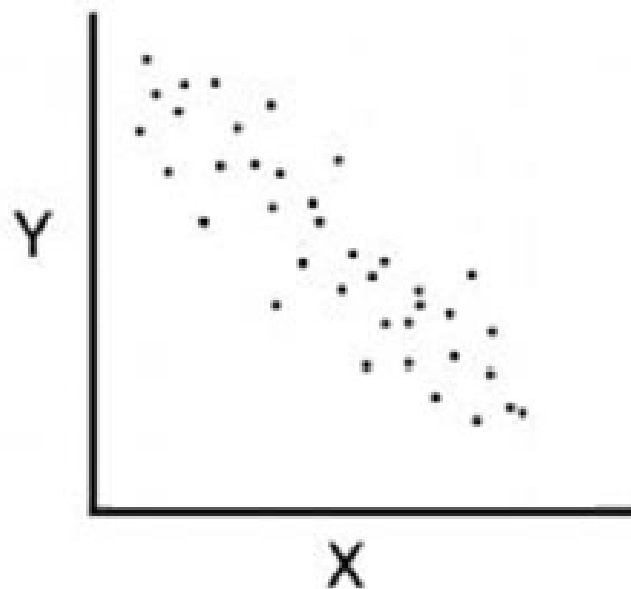
Corelație nulă  
 $r \approx 0$

# Graficul de corelație – XY Scatter

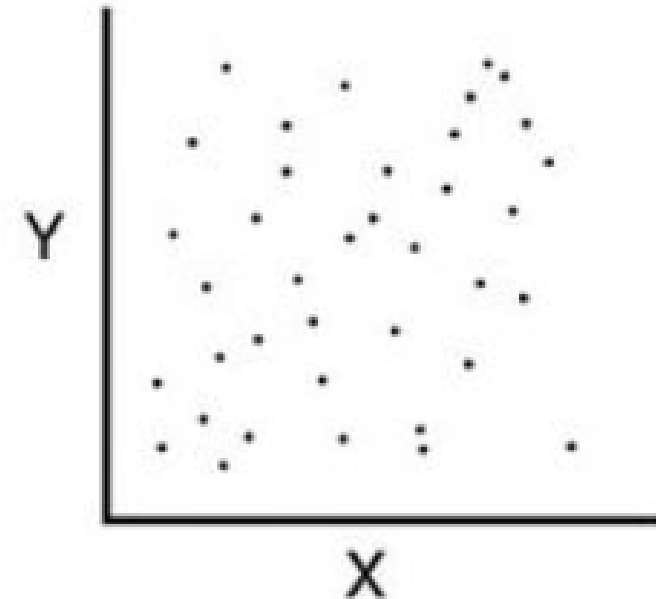
- Relație inversă, negativă



Corelație puternică  
 $r \approx -1$



Corelație moderată  
 $r \approx -0.5$



Corelație nulă  
 $r \approx 0$

# Regulile lui Colton de interpretare a coeficientului de corelație

- [Colton T. Statistics in Medicine. Little Brown and Company, New York, NY 1974]:
- $r \in [-0,25 ; +0,25] \rightarrow$  Nu există asocieri spre asocieri slabă
- $r \in (0,25 ; +0,50]$  sau  $r \in (-0,50 ; -0,25] \rightarrow$  Asocieri acceptabilă
- $r \in (0,50 ; +0,75]$  sau  $r \in (-0,75 ; -0,50] \rightarrow$  Asocieri moderată sau bună
- $r \in (0,75 ; +1]$  sau  $r \in (-0,75 ; -1] \rightarrow$  Asocieri foarte bună





# Exemplu

- Ex. Corelația dintre greutate și densitatea minerală osoasă ( $r=0.43$ ) la pacienții cu artroplastie bilaterală totală a genunchiului
- **Pozitivă** – Greutatea mare – corespunde cu valori mari de densitate osoasă (DMO crește cu greutatea subiectului)
- Regulile lui Colton --> **relație acceptabilă**

Ishii Y, Noguchi H, Sato J, Ishii H, Todoroki K, Toyabe SI. Association between body weight and proximal tibial bone mineral density after bilateral total knee arthroplasty. Knee. 2017 Oct;24(5):1153-1159.

# Multumesc!