



Cosmina Ioana Bondor – autor

Probabilități – partea 2



ALWAYS



SEEK



KNOWLEDGE

$P(A | B)$ = probabilitatea conditionată

- Probabilitatea ca evenimentul A să apară dacă evenimentul B s-a realizat deja

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

Exemplu

- B – evenimentul de a prezenta o boala (de exemplu diabet, HIV etc.),
- T - evenimentul obținerii unui test nou pozitiv în cazul aplicării unui test de diagnostic T pentru detectarea bolii B (de ex. test rapid)
- non (B) (persoană fără condiția B) și non (T) (test negativ) - evenimente complementare evenimentelor B, respectiv T

- Fie:
 - B - evenimentul ca o persoană să aibă o anumită **afecțiune B** (de exemplu, TBC , HIV etc.),
 - T - evenimentul de obținere a unui **test pozitiv** în cazul aplicării unui test diagnostic T pentru detectarea afecțiunii B
 - Prin non(B) (persoană fără afecțiunea B) și non(T) (test negativ) - evenimentele complementare evenimentelor B și respectiv T.

- Test applied to n persons:

Boala / Test	B Cu boala	non(B) Fara boala	Total
T test pozitiv	a (AP)	b (FP)	a+b
non (T) test negativ	c (FN)	d (AN)	c+d
Total	a+c	b+d	n

Valoarea predictiva pozitiva PPV

- Probabilitatea ca un test nou pozitiv sa indice boala (sa fie corect):

$$VPP = Pr(B|T) = \frac{Pr(B \cap T)}{Pr(T)} = \frac{\frac{AP}{n}}{\frac{AP + FP}{n}} = \frac{AP}{AP + FP} = \frac{a}{a + b}$$

Boala / Test	B Cu boala	non(B) Fara boala	Total
T Test pozitiv	a (AP)	b (FP)	a+b
non (T) test negativ	c (FN)	d (AN)	c+d
Total	a+c	b+d	n

Negative predictive value NPV

- Probabilitatea ca un test nou negativ sa nu indice boala (sa fie corect):

$$VPN = \Pr(nonB / nonT) = \frac{\Pr(nonB \cap nonT)}{\Pr(nonT)} = \frac{AN}{FN + AN} = \frac{d}{c + d}$$

Boala / Test	B Cu boala	non(B) Fara boala	Total
T Test pozitiv	a (AP)	b (FP)	a+b
non (T) test negativ	c (FN)	d (AN)	c+d
Total	a+c	b+d	n

Sensitivitatea testului Se

- Probabilitatea ca un bolnav sa aiba test nou pozitiv:

$$Se = \Pr(T / B) = \frac{AP}{AP + FN} = \frac{a}{a + c} = \frac{\frac{a}{n}}{\frac{a + c}{n}} = \frac{\Pr(T \cap B)}{\Pr(B)}$$

Boala / Test	B Cu boala	non(B) Fara boala	Total
T Test pozitiv	a (AP)	b (FP)	a+b
non (T) test negativ	c (FN)	d (AN)	c+d
Total	a+c	b+d	n

Specificitatea testului Sp

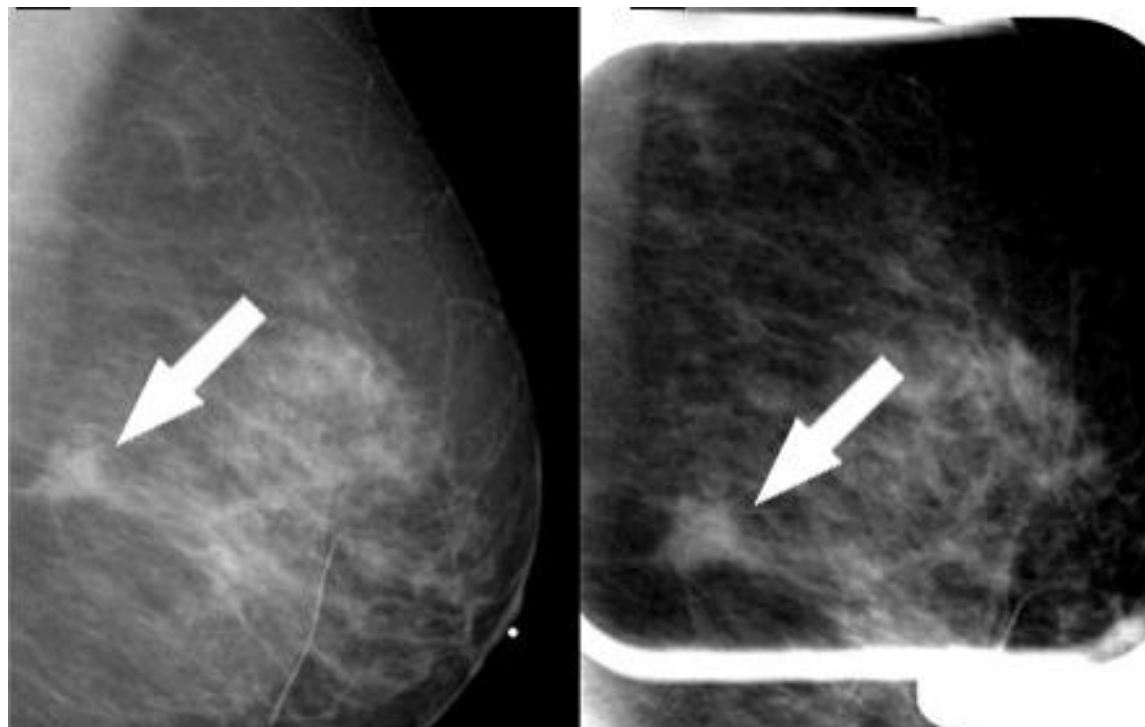
- Probabilitatea ca un indemn de boala sa aiba test nou negativ:

$$Sp = \Pr(\text{non}(T) / \text{non}(B)) = \frac{AN}{FP + AN} = \frac{d}{b + d} = \frac{\frac{d}{n}}{\frac{b + d}{n}} = \frac{\Pr(\text{non}T \cap \text{non}B)}{\Pr(\text{non}B)}$$

Boala / Test	B Cu boala	non(B) Fara boala	Total
T Test pozitiv	a (AP)	b (FP)	a+b
non (T) test negativ	c (FN)	d (AN)	c+d
Total	a+c	b+d	n

EXEMPLU

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



EXEMPLU

- Din 100.000 de femei cu mamografii negative 20 au cancer de sân
- Din 10 femei cu mamografii pozitive 1 are cancer de sân
- Calculați Se, Sp, VPP, VPN
- Pragul de interpretare 0,80

Exemplu: Se, Sp, VPP, VPN

Prevalenta = $21/100.010 = 0,21$ la mie

Afecțiunea / Testul	Subiecți care au dezvoltat afecțiunea	Subiecți sănătoși	Total
Mamografie pozitivă	1	9	10
Mamografie negativă	20	99.980	100000
Total	21	99.989	100.010

$$VPP = 1/10 = 0,1$$

$$VPN = 99.980/100.000 = 0,99$$

$$Se = 1/21 = 0,047$$

$$Sp = 99.980/99.989 = 0,99$$

HIV Test

	Cu HIV (B ⁺)	Fără HIV (B ⁻)	Total
Test pozitiv (T ⁺)	99	1	100
Test negativ (T ⁻)	1	99.999	100.000
Total	100	100.000	100.100

Sensitivitatea testului nou HIV = $\Pr(T^+ | B^+) = \Pr(T^+ \cap B^+) / \Pr(B^+) = (99/100) \sim 99\%$

Specificitatea testului nou HIV = $\Pr(T^- | B^-) = \Pr(T^- \cap B^-) / \Pr(B^-) = (99.999/100.000) \sim 99.999\%$

Valoarea predictivă pozitivă a testului nou HIV = $\Pr(B^+ | T^+) = \Pr(T^+ \cap B^+) / \Pr(T^+) = (99/100)$

Valoarea predictivă negativă a testului nou HIV = $\Pr(B^- | T^-) = \Pr(T^- \cap B^-) / \Pr(T^-) = (99.999/100.000)$

Măsurarea riscului: RR

Riscul relativ RR

= raportul dintre cazurile de imbolnavire la cei cu expunere la factorul de risc si cazurile de imbolnavire la cei fara expunere la factorul de risc

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

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

Non A

RR=1

Nu exista risk de imbolnavire la persoanele expuse la factorul de risc

RR>1

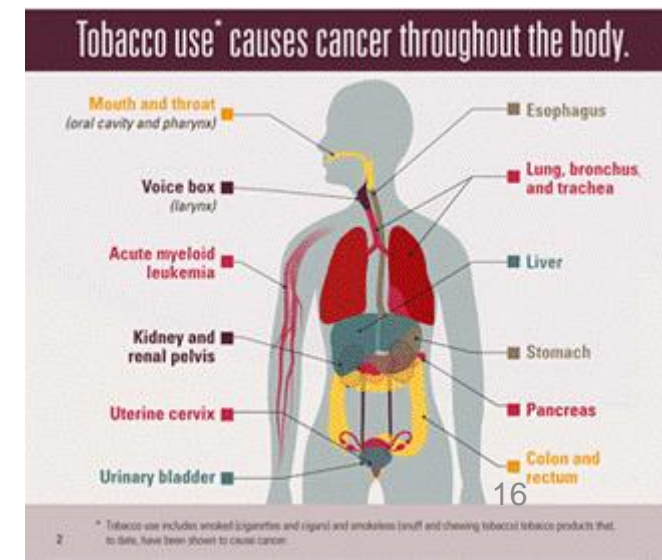
Exista risk de a face boala daca exista expunere la factorul de risc

RR<1

Expunerea este factor de protective pentru boala

Exemplu

- 2 nefumatori din 1000 face cancer de plamani
- 46 fumatori din 1000 fac cancer de plamani
- Probabilitatea de cancer pulmonar la nefumatori= 0.002
- Probabilitatea de cancer pulmonar la fumatori = 0.046
- $RR = \Pr(B | A) / \Pr(B | \text{non } A)$
- $RR = 0.046 / 0.002 = 23$ (de 23 de ori mai mare)
- Risk difference $0.046 - 0.002 = 0.044$
- With 29 individuals more from 10 000



Cauzalitatea!!!

A depend de B echivalent cu dacă B atunci A ($B \Rightarrow A$)

! Non B ~~\Rightarrow~~ non A

Non A \Rightarrow non B

Eroare logică !!!

În 1970 - 1980 au fost realizate studii:

hipercolesterolemie => mortalitate ridicată din cauze cardiovasculare

– s-a presupus că dieta bogată în grăsimi ~~duce~~ la hipercolesterolemie

- Dieta săracă în grăsimi ~~=>~~ colesterol scăzut ~~=>~~ **mortalitate redusă din cauze cardiovasculare**

Reduceți grăsimile din dietă

- A fost sfatul oficial **eronat** în SUA din 1977 și în UK din 1983

!

- A crescut obezitatea, a crescut mortalitatea din cauze cardiovasculare
- Până și tinerii sunt afectați

2017

Studiile au arătat altceva:

Sfatul 2017 → prezent:

- Se poate mânca odată / săptămână carne de porc
- Carnea de porc nu crește colesterolul

2017

- ! Reduceți zahărul din dietă

Alte exemple de erori logice

- Carbohidrații sunt calorii, mulți carbohidrați îngrașă, pâinea are mulți carbohidrați, deci îngrașă (silogism)
 - pâinea este transformată de bacteriile din intestine în grăsimi
- Persoanele cu deficiențe de vitamine au sistem imunitar slăbit, deci dacă suplimentăm cu vitamine impulsionăm sistemul imunitar (negația nu e valabilă automat fără demonstrație)

Teorema lui Bayes

Evenimente dependente (cauză și efect)

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

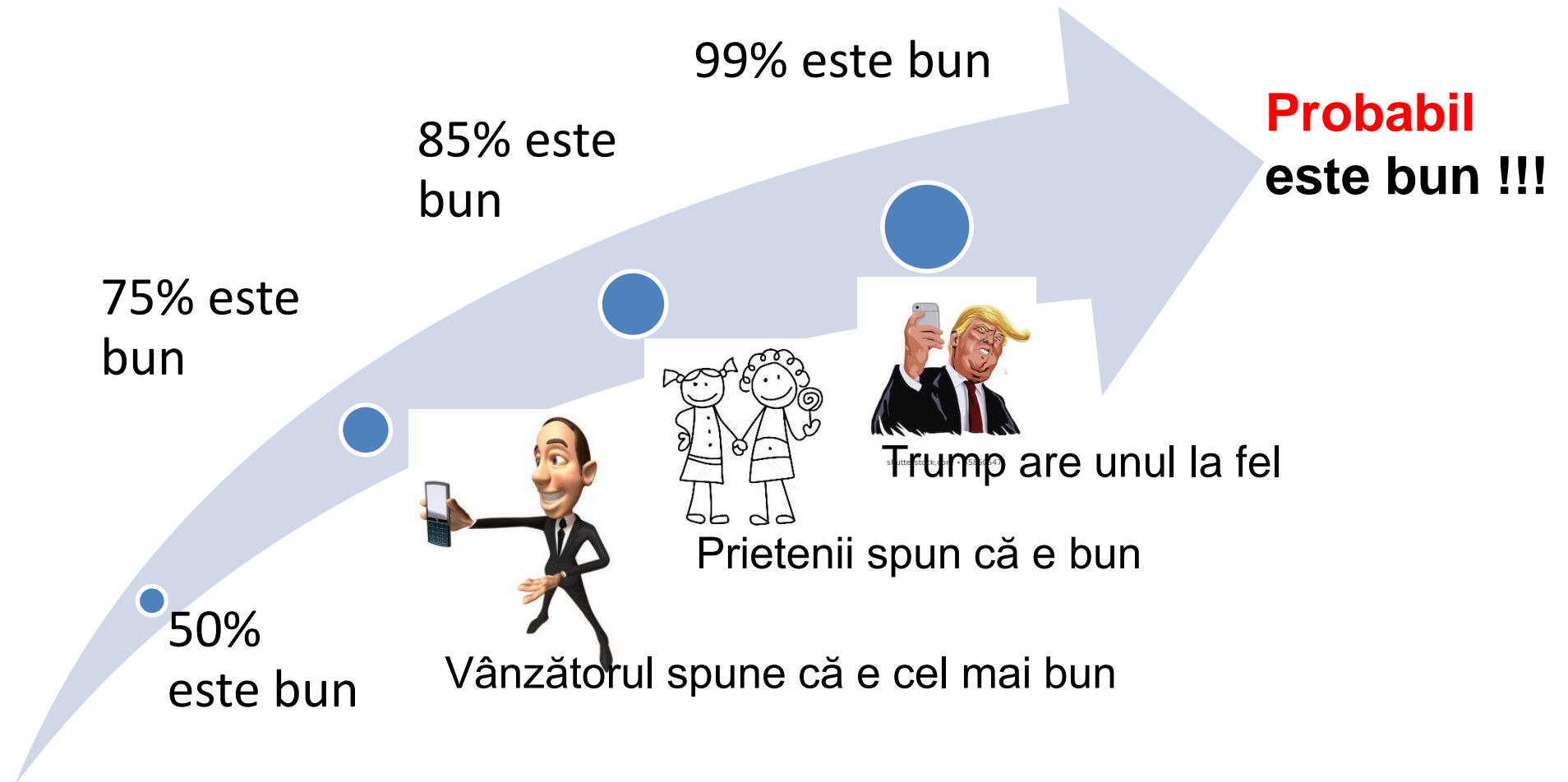
$$Pr(A | B) = \frac{Pr(B | A) * Pr(A)}{Pr(B)}.$$

Implicată în luarea deciziilor medicale - teste diagnostice.

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

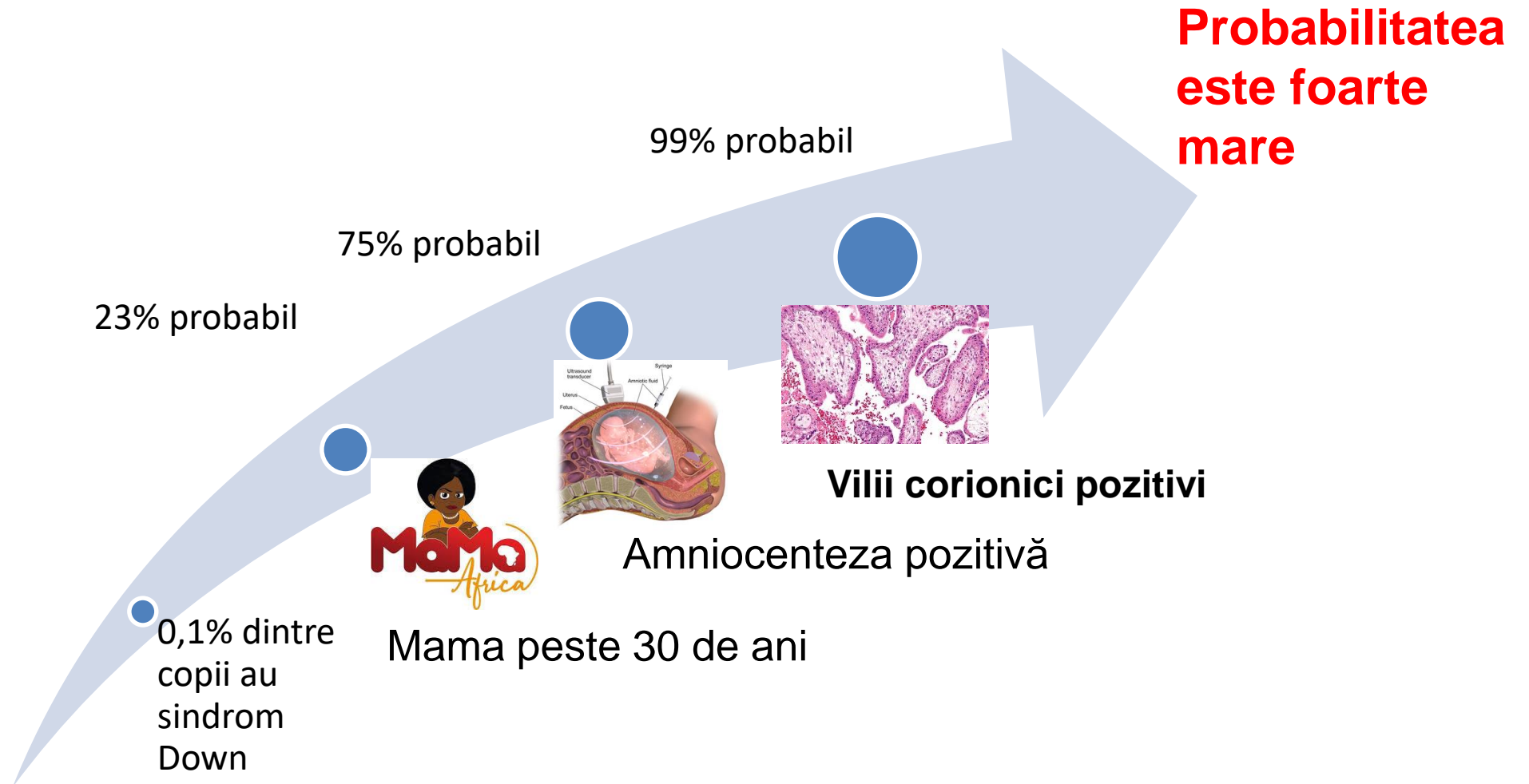
- Așa învață calculatoarele
- Așa funcționează filtrele de spam
- Așa pun medicii diagnosticele
- Așa știm ce le place prietenilor noștri

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



Sindrom Down la făt în timpul sarcinii?

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





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

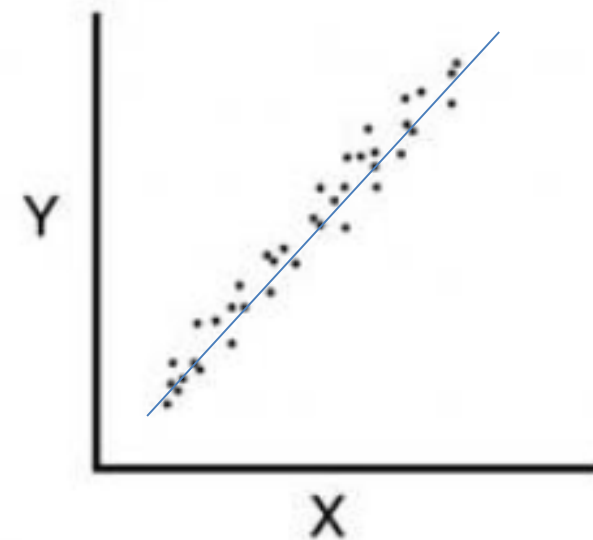
- Relația dintre două caracteristici
- Cum este relația?
- Putem face predicții?
- Ce eroare putem să acceptăm?

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

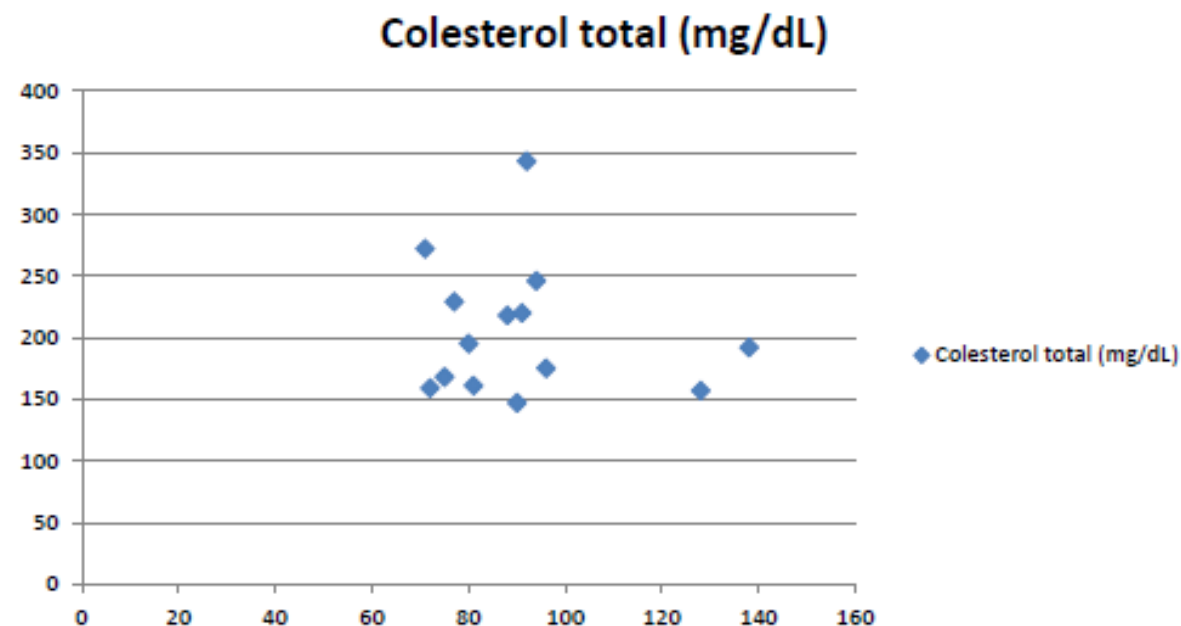
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		

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

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

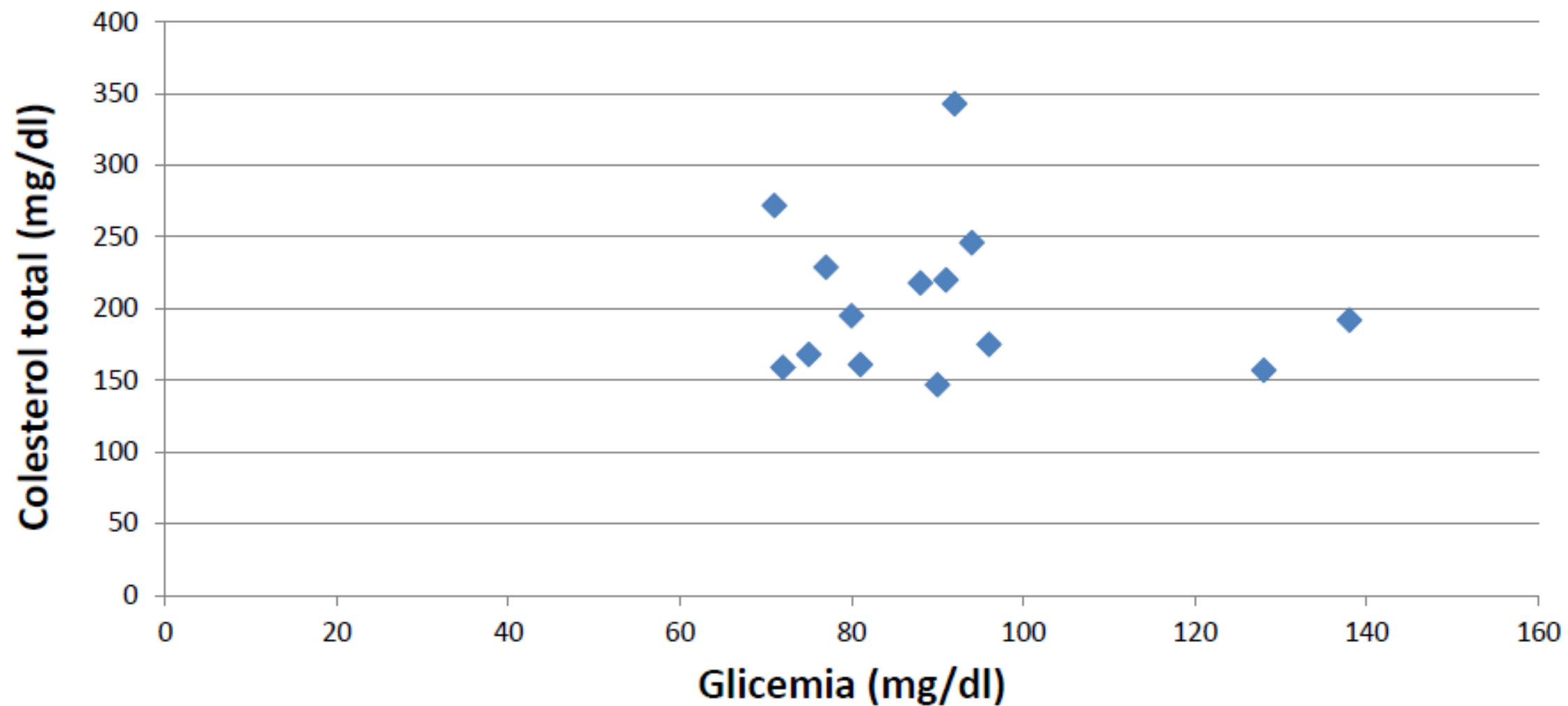
Glicemie (mg/dL)	Colesterol total (mg/dL)
75	168
92	343
77	229
128	157
81	161
138	192
88	218
72	159
71	272
80	195
91	220
94	246
90	147
96	175

Selectam cele doua coloane cu date



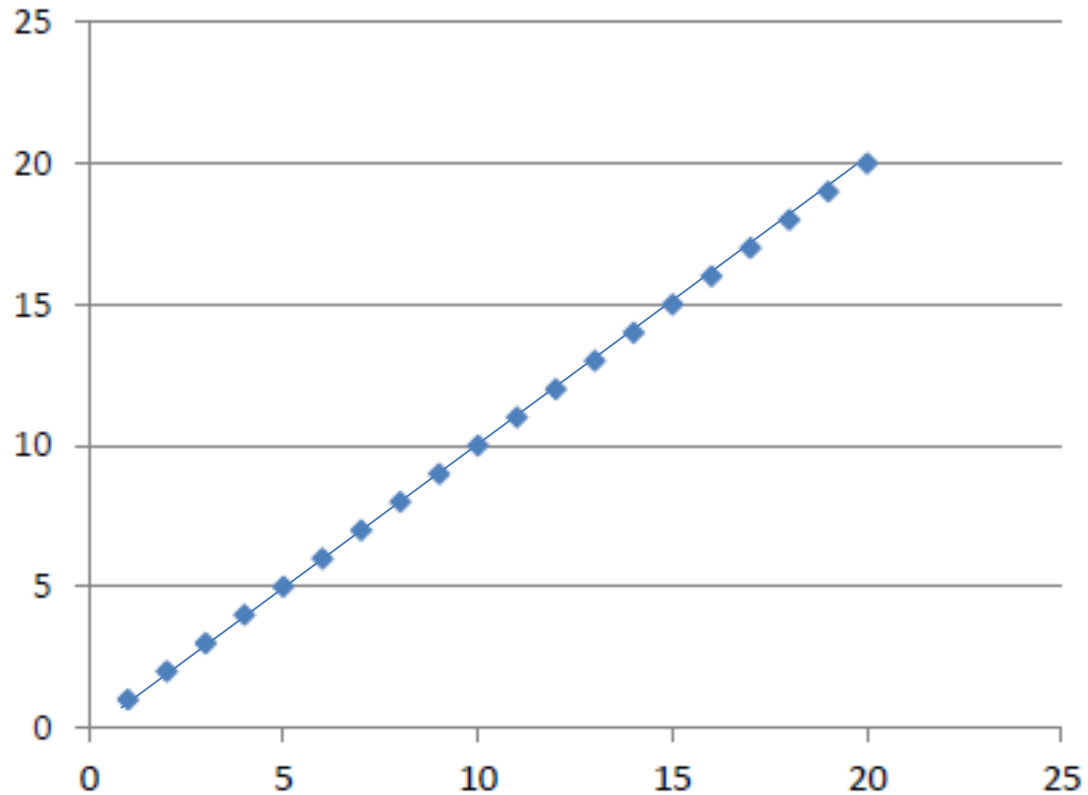
Inserăm grafic Scatter

Corelatia dintre colesterol si glicemie



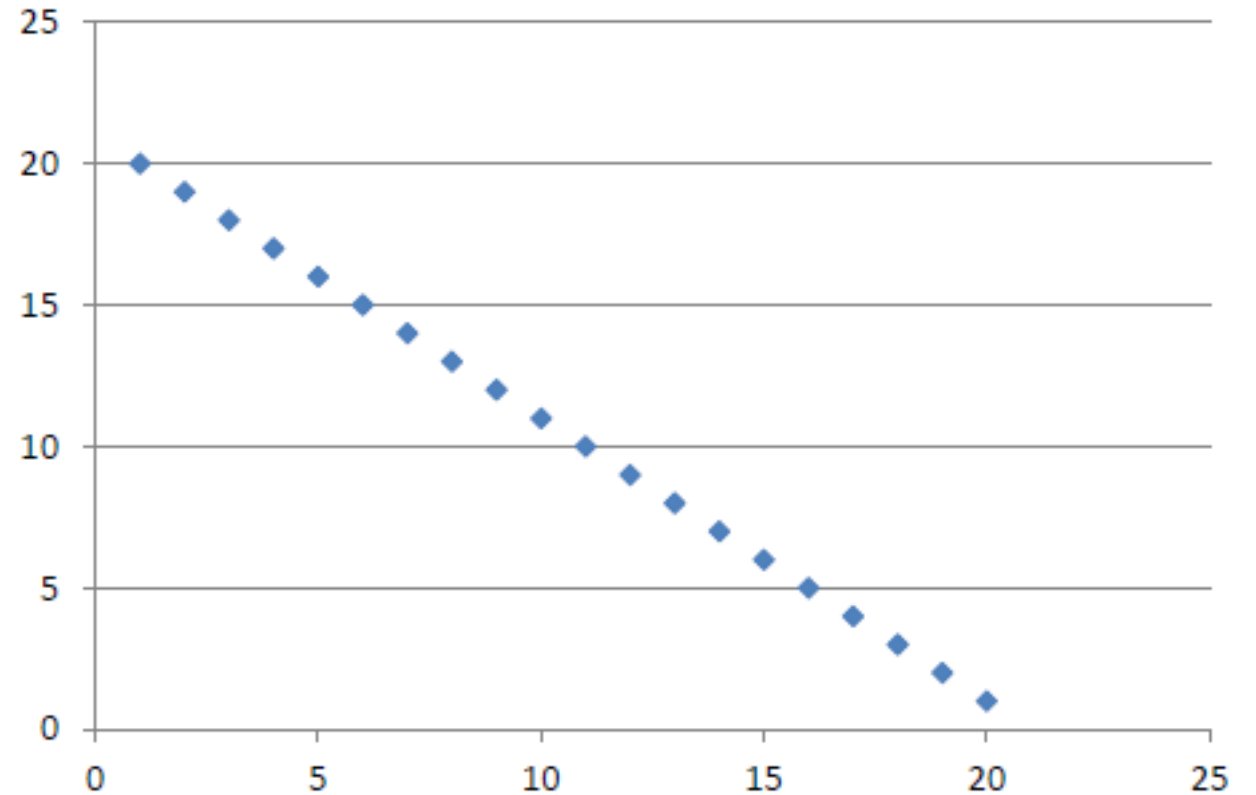
Corelația perfectă

- $r=1$



Pozitiva (direct proporțională)

$r=-1$



Negativa (invers proporțională)

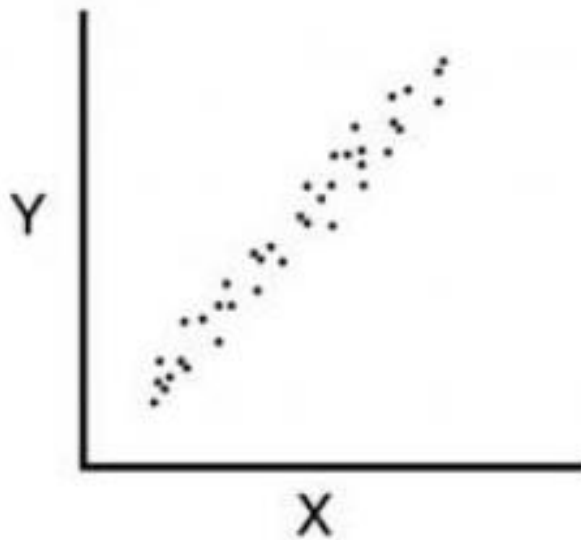
Coeficient de corelație Pearson

- 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

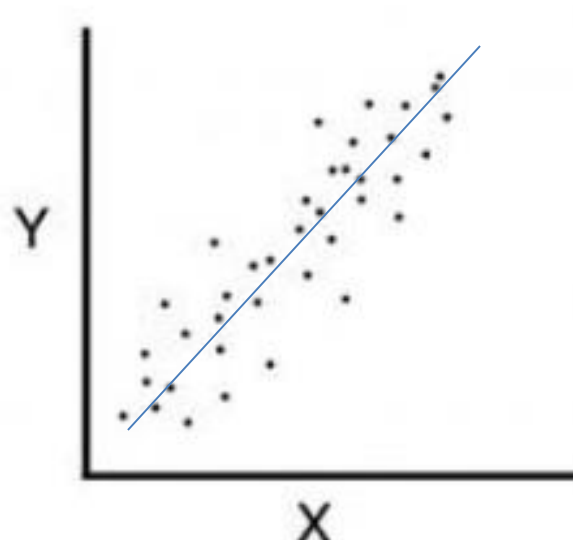


Graficul de corelație – XY Scatter

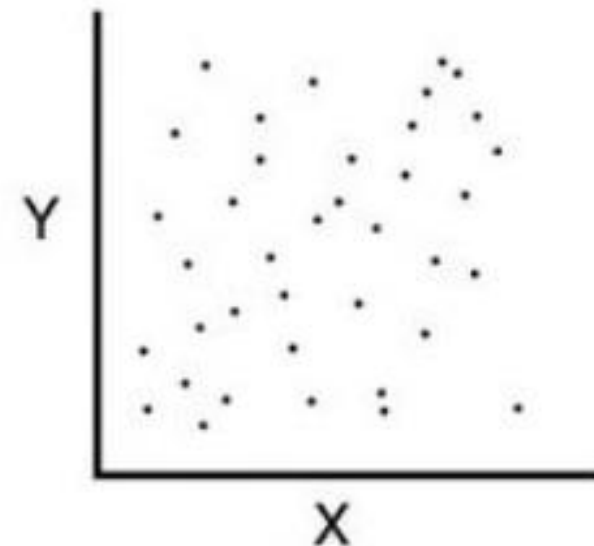
- Relație directă, pozitivă



Corelație puternică
 $r \approx 1$



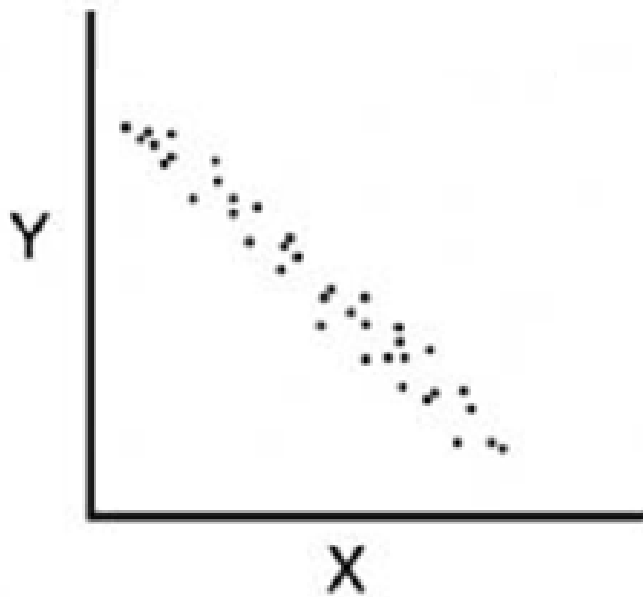
Corelație moderată
 $r \approx 0.5$



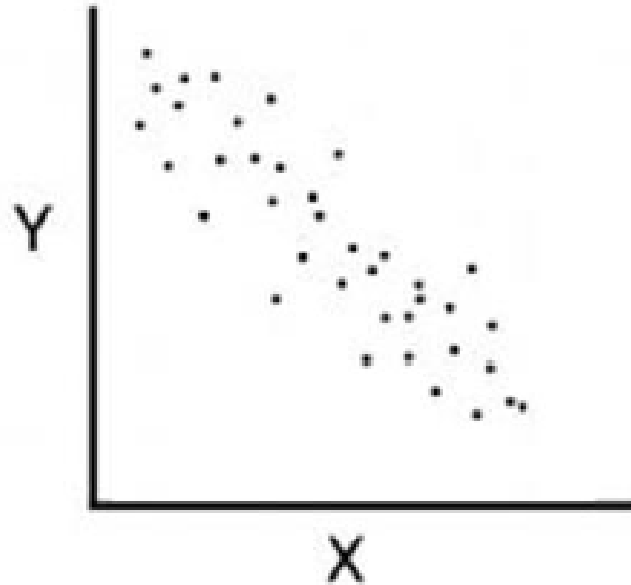
Corelație nulă
 $r \approx 0$

Graficul de corelație – XY Scatter

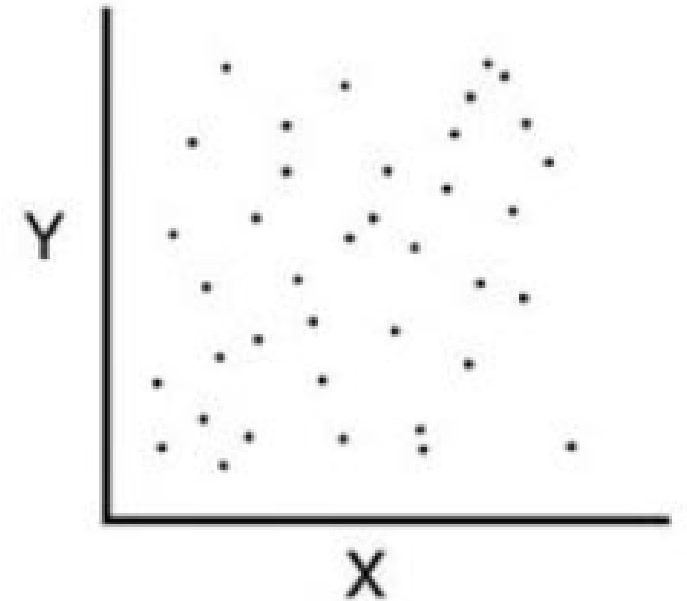
- Relație inversă, negativă



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!