



**Cosmina Ioana Bondor – autor**

# Asociere și predicție



ALWAYS



SEEK



KNOWLEDGE

# Obiective

- Corelație
- Regresie liniară
- Exerciții

# Corelație

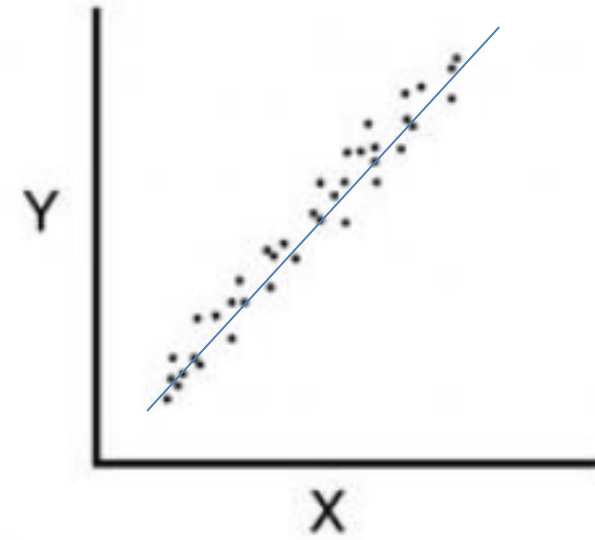
- Relația dintre două caracteristici cantitative
- Cum este relația?
- Putem face predicții?
- Ce eroare putem să acceptăm?

# 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

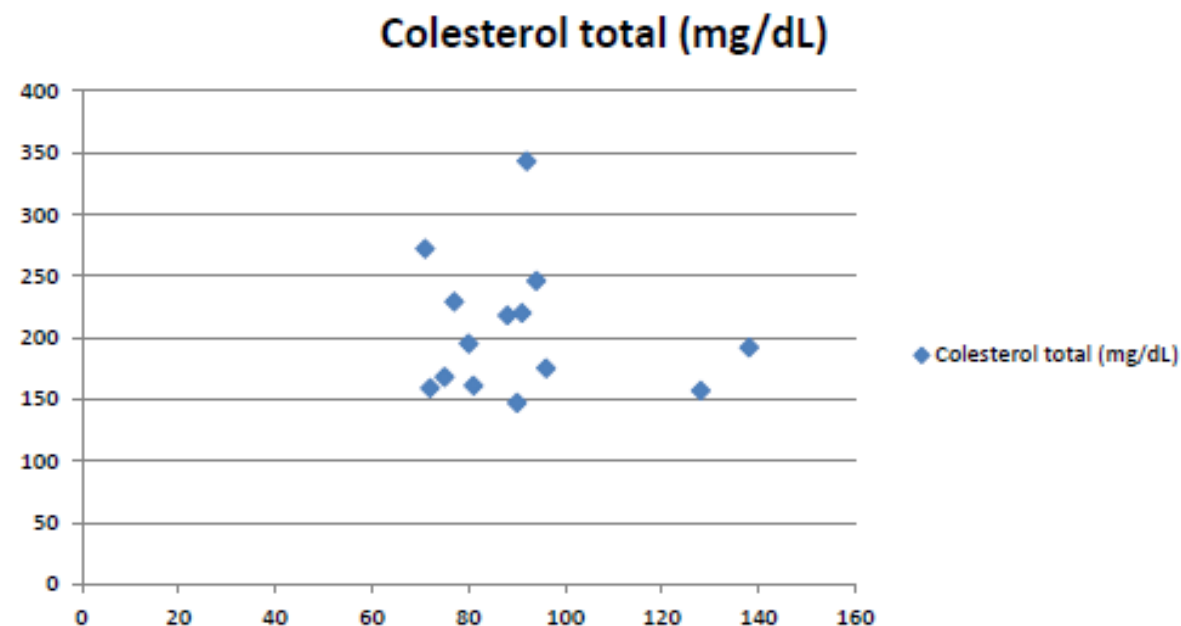
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		

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



# 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



# 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!