

# CORELAȚII LINIARE

## SOLUȚII

### Testarea normalității datelor folosind DataAnalysis

Se copiază variabilele pentru care dorim să testăm normalitatea într-o foaie de calcul nouă. Coloanele (variabilele) pe care le vom folosi trebuie să fie alăturate, astfel încât acestea să poată să fie selectate folosind o referință continuă:

Descriptive Statistics

Input

Input Range:

Grouped By: ☒ Columns ☐ Rows

☒ Labels in first row

Output options

☒ Output Range:

☐ New Worksheet Ply:

☐ New Workbook

☒ Summary statistics

☐ Confidence Level for Mean:  %

☐ Kth Largest:

☐ Kth Smallest:

OK Cancel Help

	Circumferinta abdominala(cm)	IMC(kg/m2)	TAS(mmHg)	HDL-colesterol (mg/dl)	cGIM(mm)
Mean	95.78070417	26.68485941	140.8130664	45.64727041	0.939088276
Standard Error	1.259622837	0.499492536	1.312476482	1.102029599	0.016127249
Median	98.73700652	26.67873009	141.412813	44.73237426	0.948169436
Mode	110	#N/A	#N/A	#N/A	#N/A
Standard Deviation	12.59622837	4.994925364	13.12476482	11.02029599	0.16127249
Sample Variance	158.6649692	24.94927939	172.2594516	121.4469238	0.026008816
Kurtosis	-0.114082845	-0.845722736	-0.54861901	0.146971868	-0.41044788
Skewness	-0.579437696	0.160920087	-0.091308423	0.120906738	-0.127977938
Range	56.01240213	18.94075994	61.15683391	57.44179361	0.718308752
Minimum	60.98759787	18.01140438	108.3221102	18.29444785	0.575931682
Maximum	117	36.95216431	169.4789441	75.73624146	1.294240434
Sum	9578.070417	2668.485941	14081.30664	4564.727041	93.90882757
Count	100	100	100	100	100

Cele cinci variabile urmează o distribuție normală, deci putem să calculăm coeficientul de corelație Pearson pentru orice pereche formată (**Atenție**, coeficientul de corelație Pearson are sens să îl calculăm doar dacă avem o distribuție normală a valorilor!!)

## Calcularea coeficientului de corelație Pearson dintre două variabile cantitative folosind funcția CORREL

B	C	D	E	F	G	H	I	J	K	L
31.73	157	55	1.13							
22.66	145	18	0.91							
33.39	142	42	0.95							
20.56	138	29	0.85							
20.22	146	55	0.67							
18.86	149	44	1.08							
27.00	140	60	0.99							
21.80	142	56	0.59							
19.41	134	46	0.81							
35.37	158	33	1.08							
29.50	143	54	1.00							
24.13	130	45	0.84							
21.51	152	53	0.94							
26.08	169	57	0.99							
24.77	153	66	0.94							
28.92	164	40	1.22							
23.71	154	43	1.29							
27.94	163	46	1.07							
25.00	147	51	1.01							

Array1: valorile circumferinței abdominale pe eșantion

Array2: valorile IMC pe eșantion

### Interpretarea coeficientului de corelație Pearson (r)

Pentru a **interpreta intensitatea relației liniare** dintre două variabile (cantitative cu distribuție normală) folosim regulile empirice ale coeficientului de corelație ale lui Colton [1]

- Dacă  $r \in [-0,25 ; +0,25]$  → corelație liniară slabă sau inexistentă
- Dacă  $r \in (0,25 ; 0,50] \cup (-0,25 ; -0,50]$  → corelație liniară acceptabilă
- Dacă  $r \in (0,50 ; 0,75] \cup (-0,50 ; -0,75]$  → corelație liniară moderată spre bună
- Dacă  $r \in (0,75 ; 1] \cup (-0,75 ; -1]$  → corelație liniară puternică (foarte bună)

**Interpretarea semnului relației liniare** dintre două variabile (cantitative cu distribuție normală)

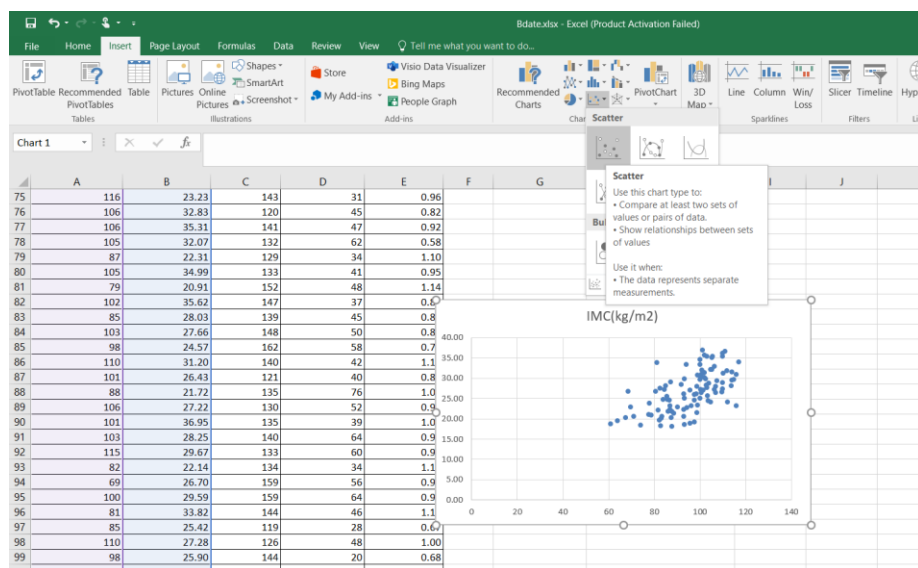
Dacă  $r > 0$  → între cele două variabile există o relație de direct proporționalitate (când valoarea uneia dintre ele crește, crește și valoarea celeilalte)

Dacă  $r < 0$  → între cele două variabile există o relație de invers proporționalitate (când valoarea uneia dintre ele crește, valoarea celeilalte scade)

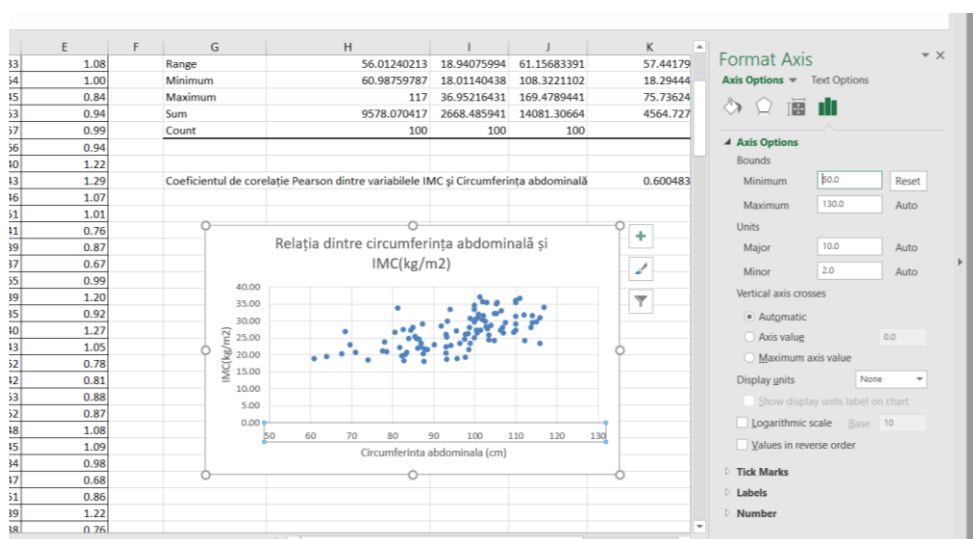
<sup>1</sup> Colton T. Statistics in Medicine. Little Brown and Company, New York, NY, 1974

## Realizarea graficului nor de puncte (scatter)

Se selectează valorile celor doua variabile (circumferinta abdominala și IMC) din tabelul cu date și se alege tipul de grafic **Scatter**

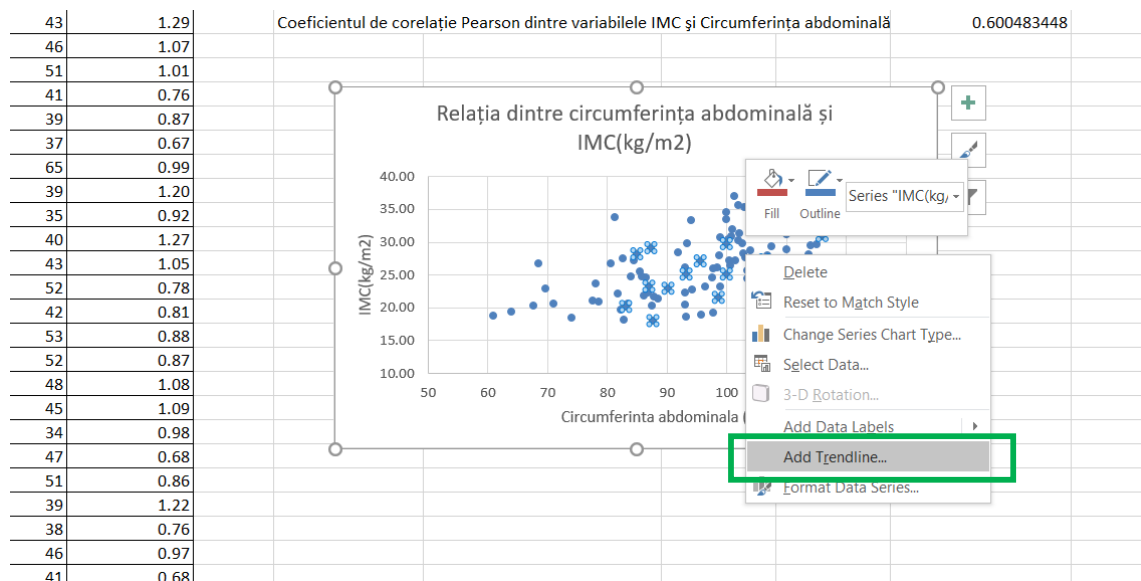


Graficul obținut va trebui formatat în așa fel încât să aibă denumirile pe axe (cu unitățile de măsură) (coloana din stânga va fi pe axa OX, iar coloana din dreapta va fi pe axa OY), titlul corespunzător, dezactivare legendă și schimbarea originii pe axe, acolo unde există spații mari goale inestetice.

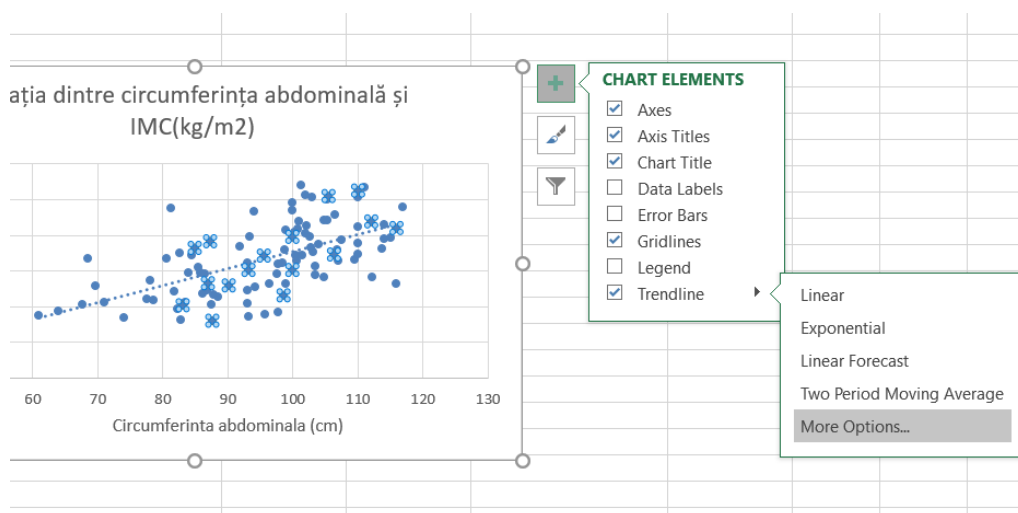


## Adăugarea pe grafic a dreptei de regresie, a ecuației dreptei de regresie și a coeficientului de determinare ( $r^2$ )

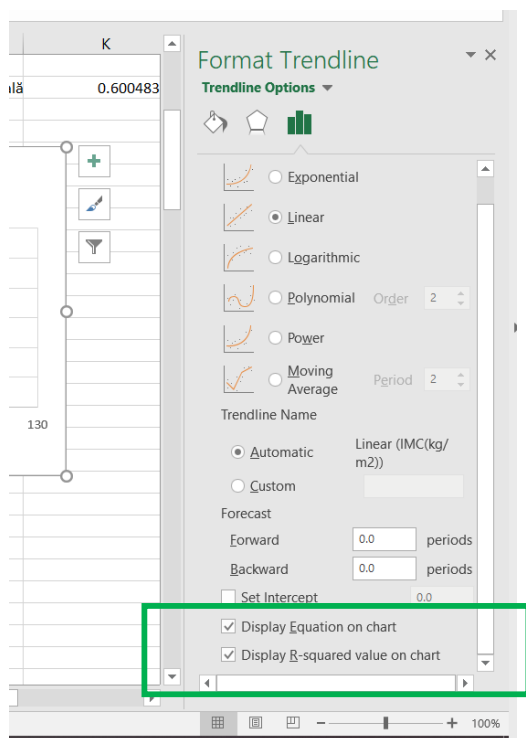
Se selectează un punct oarecare de pe grafic și cu click dreapta de mouse se bifează opțiunea **Add Trendline** și apoi se alege **More Options**



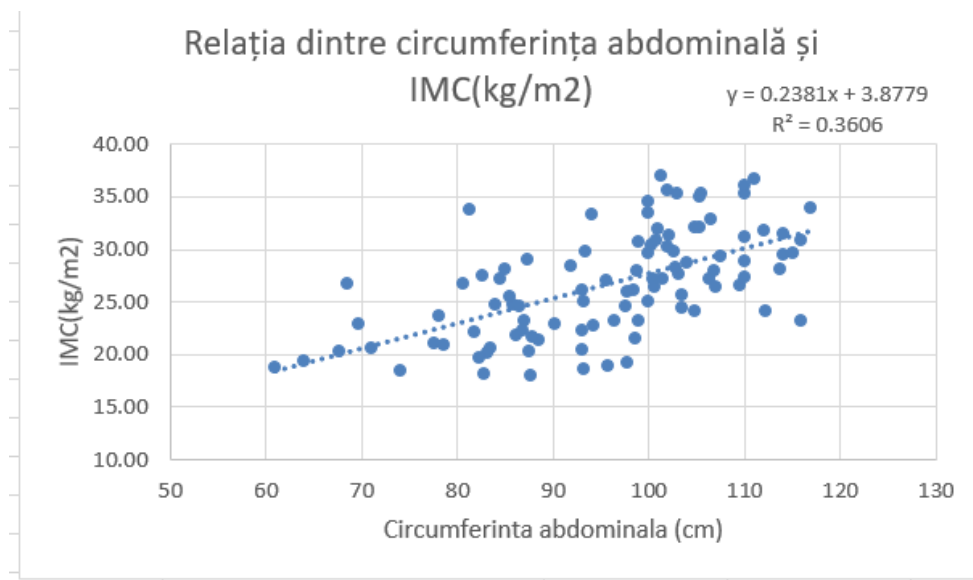
sau



Se bifează apoi opțiunile:



La final, graficul va arăta așa:



### Interpretarea coeficientului de determinare $r^2$

Coeficientul de determinare ( $d$ ) este pătratul coeficientului de corelație  $r$ .

Valoarea lui (în procente) exprimă cât la sută din variația variabilei dependente (Y) se poate explica prin relația liniară a acesteia cu variabila independentă (X)

**$r^2 = 0,36$ , deci, 36% din variația IMC-ului se poate explica prin relația liniară a acestuia cu circumferința abdominală.**

### Interpretarea dreptei de regresie

Se va interpreta coeficientul variabilei x din ecuația de regresie: **pentru fiecare creștere a variabilei independente (X=Circumferința abdominală) cu o unitate de măsură (cu 1 cm), variabila dependentă (Y=IMC (kg/m<sup>2</sup>)) crește în medie cu 0,238 unități (kg/m<sup>2</sup>)**

### Interpretarea diagramei de dispersie (norul de puncte)

O idee privind relația între cele două caracteristici se obține împărțind diagrama de dispersie în patru cadrane prin două drepte perpendiculare care trec prin punctul (X, Y ), având coordonatele egale cu mediile celor două variabile.

II	I
III	IV

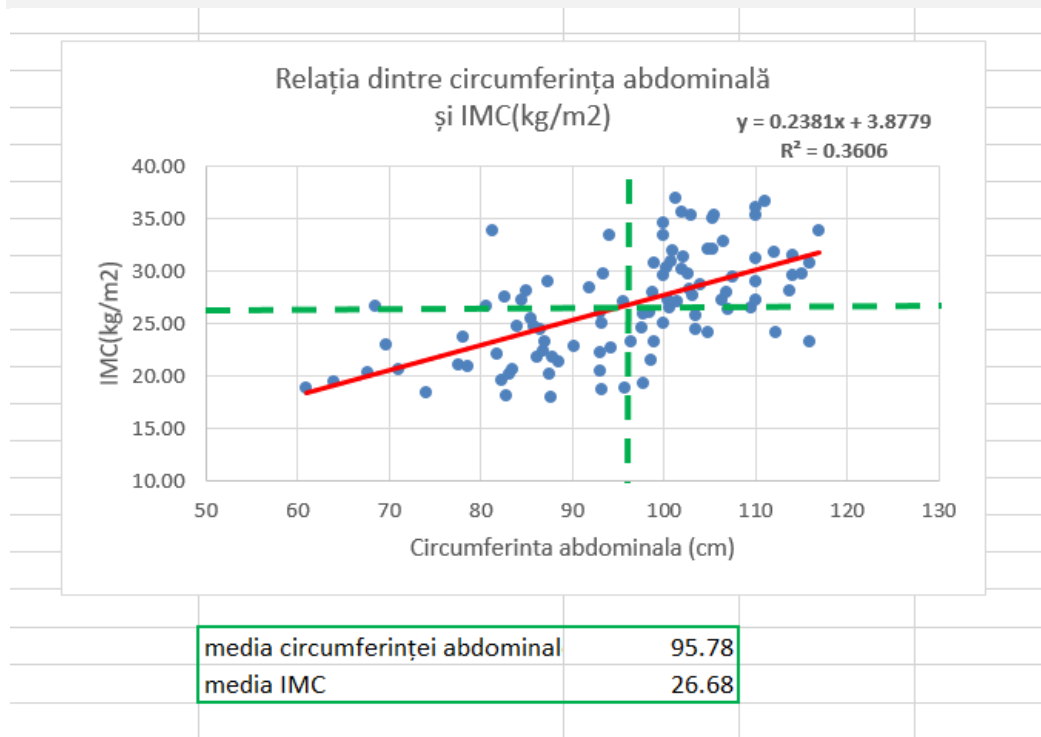
Dacă există o relație liniară între cele două variabile, atunci punctele diagramei se vor repartiza preferențial în anumite cadrane (II și IV sau I și III).

Dacă punctele sunt repartizate în cadranele **I și III**, atunci diagrama de dispersie are o tendință crescătoare (**dreapta de regresie va avea un trend ascendent**).

Dacă punctele sunt repartizate în cadranele **II și IV**, atunci diagrama de dispersie are o tendință descrescătoare (**dreapta de regresie va avea un trend descendent**).

Dacă punctele sunt repartizate în toate cele patru cadrane în mod egal, atunci diagrama de dispersie ne indică o corelație nulă.

## Exemplu de interpretare



*Punctele diagramei sunt repartizate preferențial în cadranele I și III. Dependența dintre Circumferința abdominală și IMC este pozitivă: o creștere a circumferinței abdominale implică o creștere a IMC-ului. Dreapta de regresie are un trend ascendent (tendință crescătoare).*

*Conform valorii estimate a coeficientului de determinare ( $R^2$ ), putem afirma că aproximativ 36% din variația IMC-ului se explică prin relația liniară a acestuia cu circumferința abdominală. În plus, pentru fiecare creștere a circumferinței abdominale ( $x$ ) cu o unitate de măsură (cu 1 cm), IMC-ul ( $y$ ) crește cu 0,238 unități de măsură (kg/m<sup>2</sup>) (rezultă din coeficientul lui  $x$  din ecuația dreptei de regresie).*