

INSTRUCȚIUNI CORELAȚII ȘI REGRESII

Contents

Calcularea coeficientului de corelație Pearson dintre două variabile cantitative cu ajutorul funcției CORREL:.....1

Realizarea unui grafic de tip nor de puncte – diagrama de dispersie (XY Scatter).....1

 Adăugarea dreptei de regresie pe grafic.....4

 Modificarea valorii minime și maxime afișate pe axele norului de puncte.....5

 Schimbarea schimbarea axelor între ele (XYScatter):.....6

Interpretări.....6

 Coeficientul de corelație.....6

 Coeficientul de determinare.....6

 Diagrama de dispersie (Norul de puncte).....7

Calcularea coeficientului de corelație Pearson dintre două variabile cantitative cu ajutorul funcției CORREL:

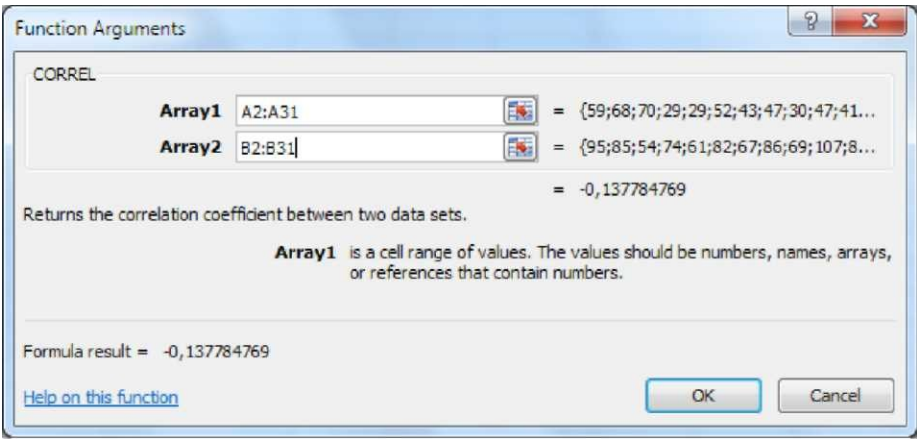
Coeficientul de corelație Pearson este un indice numeric ce dă o măsură a relației dintre două **variabile cantitative continue sau discrete (!!! Nu se calculează pentru altfel de variabile).**

Realizați un tabel de felul următor

Selectați celula unde vom calcula coeficientul de

		Coeficientul de corelație Pearson
Varsta si Greutate		

Meniul **Formulas** - Butonul **Insert Function** - Selectați funcția **CORREL**



- Array 1 - se selecteaza celulele primei variabile (ex. Varsta)
- Array 2 - se selecteaza celulele celei de a doua variabile (ex. Greutate)

Apăsați **OK**

Realizarea unui grafic de tip nor de puncte - diagrama de dispersie (XY Scatter)

Copiați cele două coloane cu date, mai întâi coloana care va apărea pe axa Ox, apoi coloan care va apărea pe Oy, exemplu:

Microsoft Excel interface showing the 'Insert' tab. The ribbon includes 'Clipboard', 'Font', and 'Align' groups. The active cell is B25. The spreadsheet data is as follows:

	A	B	C	D	E	F
	Greutate	Vârsta gestațională (saptamani)				
1						
2	5500	40				
3	4100	39				
4	3700	38				
5	4900	41				
6	4200	41				
7	2800	30				
8	4300	36				
9	4200	35				
10	4900	38				
11	4400	37				

Excel spreadsheet showing a selection of data from the previous table, specifically rows 1 through 20, columns A and B. The selection is highlighted in blue.

	A	B	C
	Greutate	Vârsta gestațională (saptamani)	
1			
2	5500	40	
3	4100	39	
4	3700	38	
5	4900	41	
6	4200	41	
7	2800	30	
8	4300	36	
9	4200	35	
10	4900	38	
11	4400	37	
12	5200	39	
13	4800	39	
14	4100	36	
15	5200	38	
16	4000	37	
17	3800	39	
18	4500	40	
19	4200	37	
20	4900	38	
21			
22			

Selectați cele două coloane: domeniul A1:B20

Meniul **Insert**

Secțiunea

Chart

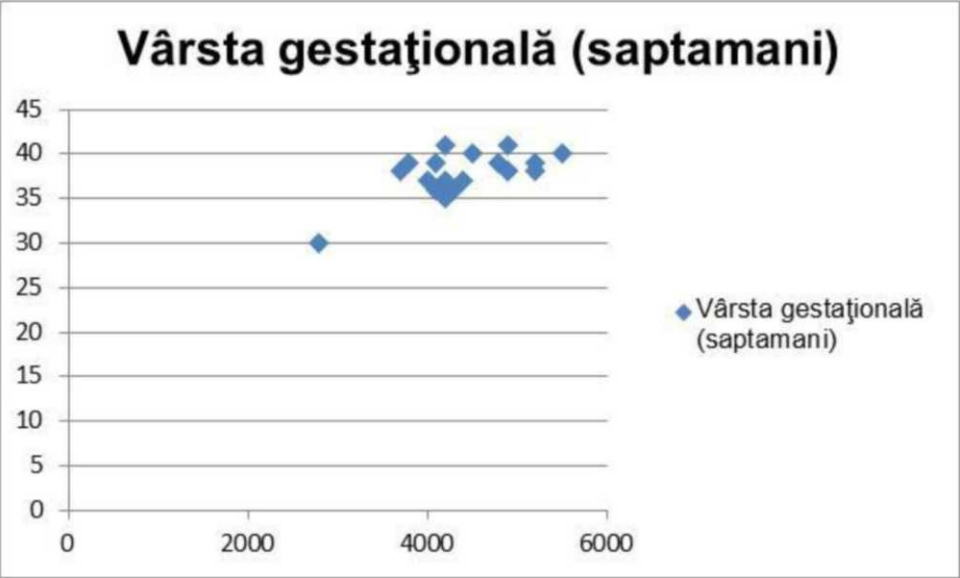
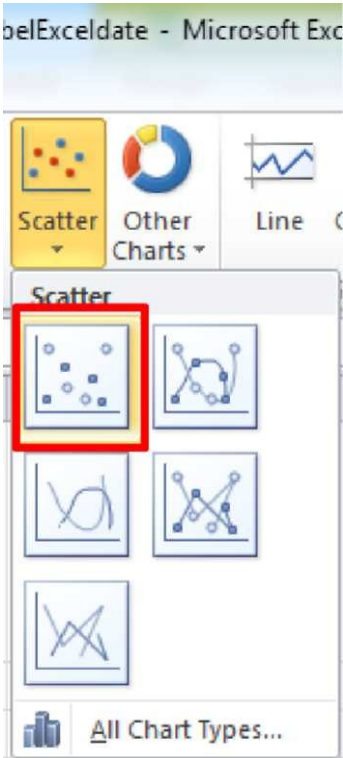
Butonul

Column



Alegeți subtipul

Rezultat
exemplu:

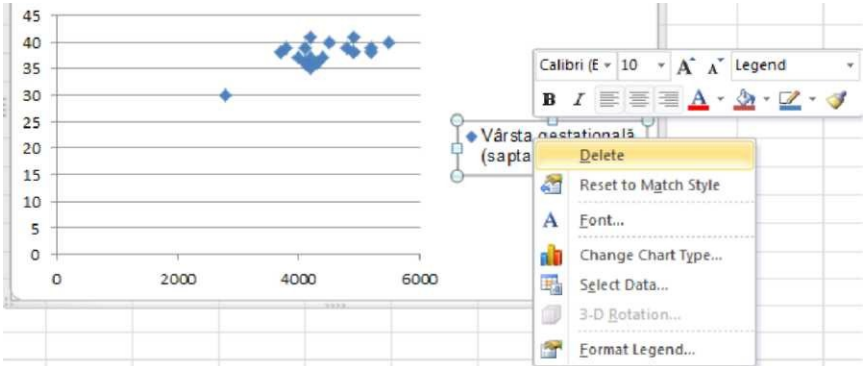


Introduceți un titlu potrivit

Dezactivarea legendei:

Click dreapta pe
Legendă

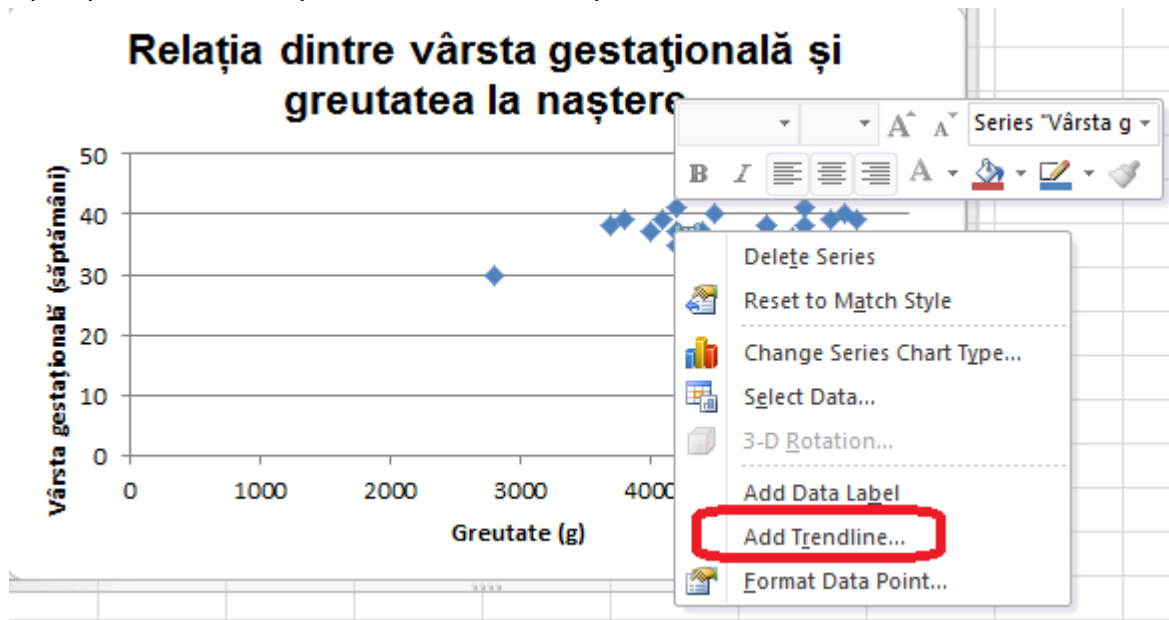
Opțiunea **Delete**



Introduceți titlurile adecvate pentru axe

Adăugarea dreptei de regresie pe grafic

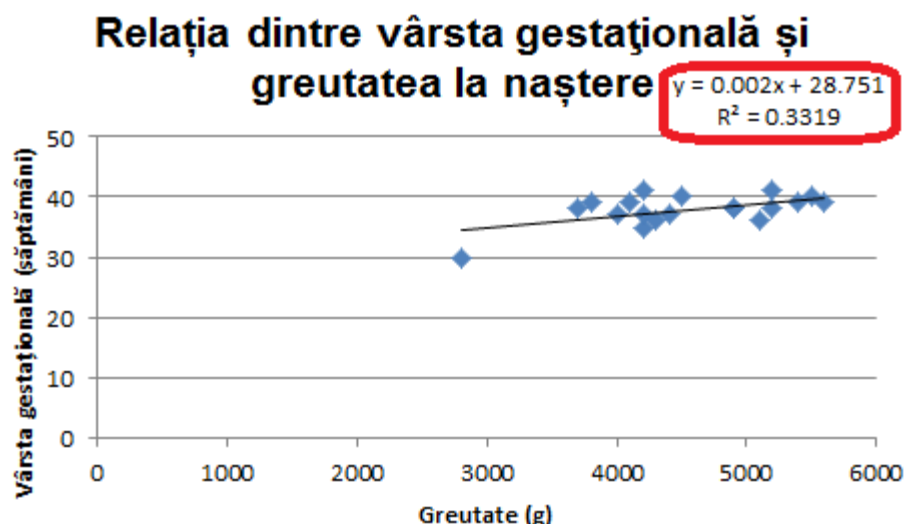
Click dreapta pe oricare din punctele norului de puncte, selectare **Add Trendline...**



În fereastra apărută se selectează **Linear** – pentru afișarea dreptei de regresie lineare, respectiv se selectează **Display Equation on chart** – pentru afișarea ecuației dreptei de regresie și **Display R-squared value on chart** – pentru afișarea coeficientului de determinare

The "Format Trendline" dialog box is shown with the "Trendline Options" tab selected. Under "Trend/Regression Type", the "Linear" option is selected and highlighted with a red box. The "Trendline Name" section shows "Automatic" selected, with the name "Linear (Vârsta gestațională (săptămâni))". The "Forecast" section shows "Forward" and "Backward" both set to 0.0 periods. At the bottom, the "Set Intercept" is set to 0.0. The options "Display Equation on chart" and "Display R-squared value on chart" are both checked and highlighted with a red box. The "Close" button is at the bottom right.

Apăsând apoi butonul **Close** se obține (după re poziționarea ecuației și coeficientului de determinare):



Modificarea valorii minime și maxime afișate pe axele norului de puncte

Pentru a modifica aceste valori pentru axa verticală se dă dublu-click pe valorile axei verticale și se obține următoarea fereastră:

Format Axis

Axis Options

Minimum: ☐ Auto ☒ Fixed 28.0

Maximum: ☒ Auto ☐ Fixed 42.0

Major unit: ☒ Auto ☐ Fixed 2.0

Minor unit: ☒ Auto ☐ Fixed 0.4

☐ Values in reverse order

☐ Logarithmic scale Base: 10

Display units: None

☐ Show display units label on chart

Major tick mark type: Outside

Minor tick mark type: None

Axis labels: Next to Axis

Horizontal axis crosses: ☒ Automatic

☐ Axis value: 28.0

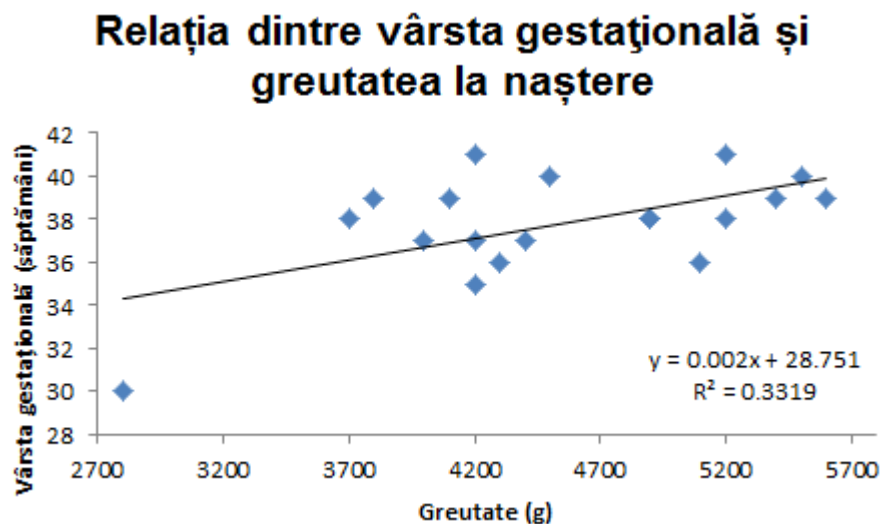
☐ Maximum axis value

Close

Pentru a modifica valoarea minimă, la Minimum se alege Fixed și se trece manual o valoare apropiată de cel mai jos punct din grafic. Asemănător se face și pentru valoarea maximă.

În mod similar se modifică și axa orizontală.

Se poate astfel obține un grafic similar cu următoarea imagine:

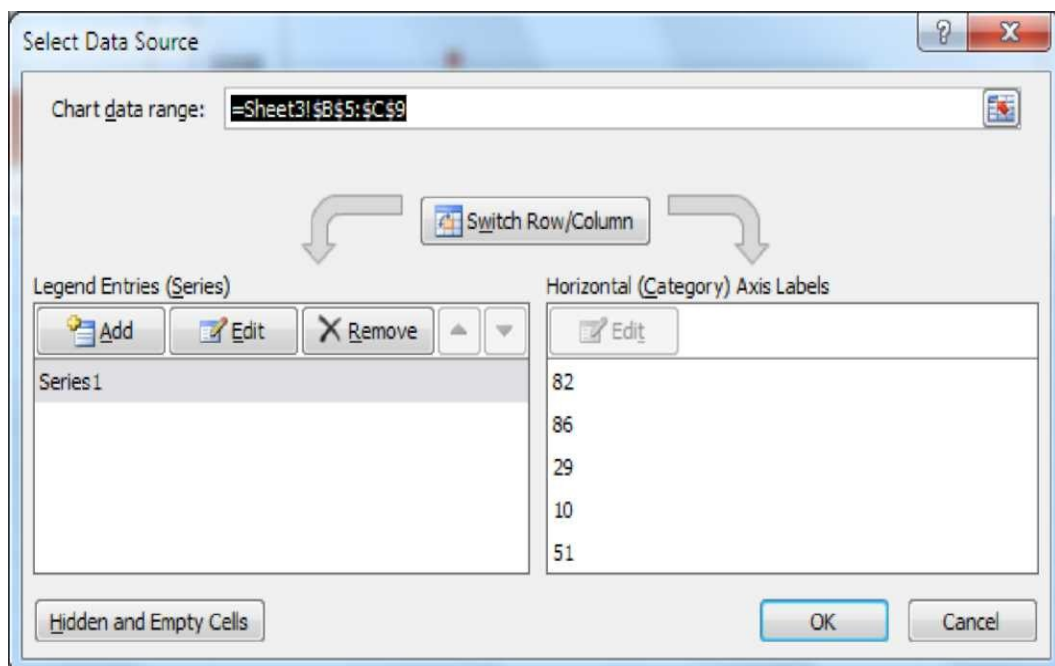


Schimbarea schimbarea axelor între ele (XYScatter):

Click dreapta pe o coloană sau pe un punct

Opțiunea **Select Data**

Row/Column



Observație: Din această fereastră se pot schimba titlurile de pe linii și coloane, domeniul datelor, se pot adăuga serii de date etc.

Interpretări

Coeficientul de corelație

Interpretarea Coeficientului de corelație conform regulilor empirice ale lui Colton (1974):

1. un coeficient de corelație de la -0,25 la 0,25 indică o corelație slabă sau nulă,
2. un coeficient de corelație de la 0,25 la 0,50 (sau de la -0,25 la -0,50) indică un grad de asociere acceptabil
3. un coeficient de corelație de la 0,5 la 0,75 (sau de la -0,5 la -0,75) indică o corelație moderată spre bună
4. un coeficient de corelație mai mare decât 0,75 (sau mai mic decât -0,75) indică o foarte bună asociere sau corelație

Coeficientul de determinare

Este pătratul coeficientului de corelație r , adică $d = r^2$.

Valoarea coeficientului de determinare exprimă o intensitate a relației liniare între cele două variabile sau răspunde la întrebarea: cât la sută din variația lui Y se poate explica prin relația liniară cu X . Exemplu: un coeficient de determinare pentru relația între tensiunea arterială sistolică și vârstă este egal cu 0,51. În cazul acesta $r^2 = 0,51$,

deci 51% din variația tensiunii arteriale sistolice se poate explica prin relația liniară cu Vârsta.

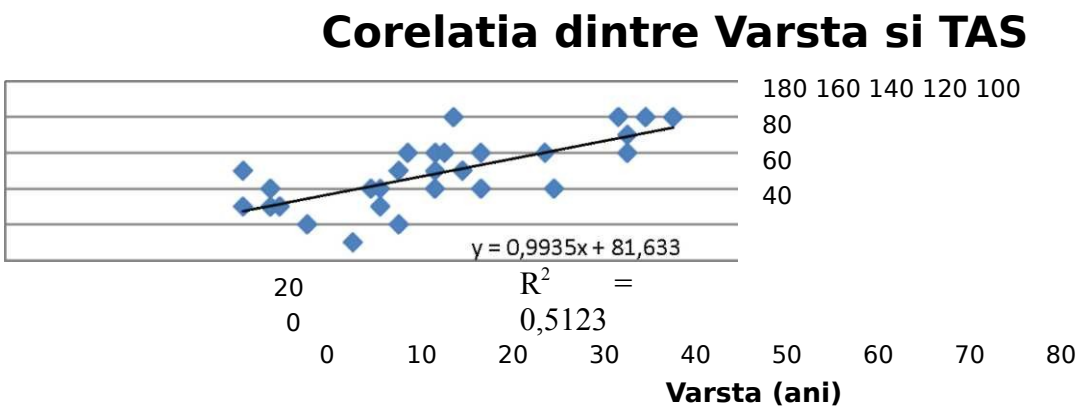
Diagrama de dispersie (Norul de puncte)

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 (\bar{X}, \bar{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 TAS și Vârsta este pozitivă: o creștere a Vârstei implică o creștere a TAS. Dreapta de regresie are un trend ascendent. Diagrama de dispersie indică o tendință crescătoare. Interpretarea coeficientului de determinare (R^2): 51% din variația TAS se explica prin relația liniară cu vârsta. Interpretarea coeficientului variabilei x din ecuația de regresie: pentru fiecare creștere a vârstei (x) cu o unitate de măsură (cu un an), TAS (y) crește cu 0,9935 unități de măsură (mmHg). Interpretarea coeficientului liber – dacă vârsta ar fi egală cu 0, TAS ar fi egală cu 81,633 mmHg.