ECONOMETRIE

CURS 3

- note de curs -

IAŞI

- 2023-

C 3 - REGRESIA LINIARĂ SIMPLĂ

TEMATICA CURS 03

Indicatorii de corelație

1. Estimarea indicatorilor de corelaţie (coeficient, raport)
2. Relația coeficient corelație - coeficient de regresie
3. Testarea indicatorilor de corelaţie
4. Testarea modelului
5. Probleme specifice utilizând SPSS si Excel

1. Estimarea indicatorilor de corelaţie

Coeficientul de corelaţie - *se foloseşte doar pentru modelul liniar*

Parametrul coeficientul de corelație (ρ)

$$\rho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{N\sigma_x \sigma_y}$$ sau

$$\rho(X,Y) = \frac{N\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{[N\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2][N\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2]}}$$, **-1≤ρ≤+1**

Interpretarea lui ρ

Coeficientul de corelație ρ caracterizează **intensitatea legăturii liniare** dintre două variabile X și Y.

Astfel:

Pentru **$-1 \leq \rho < 0$** – > legătura între X și Y este **negativă**;

Pentru **$0 < \rho \leq +1$** ->legătura între X și Y este **pozitivă**;

Pentru **$|\rho| \to 0$** => legătura este **foarte slabă**

$|\rho| \to 1$=> legătura este **foarte puternică**.

Estimația coeficientului de corelație (r)

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n s_x s_y} =$$

$$= \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{[n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2][n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2]}}$$

r se interpretează la fel ca **ρ**.

Raportul de determinaţie

Parametrul raportului de determinație (η^2)(desen)

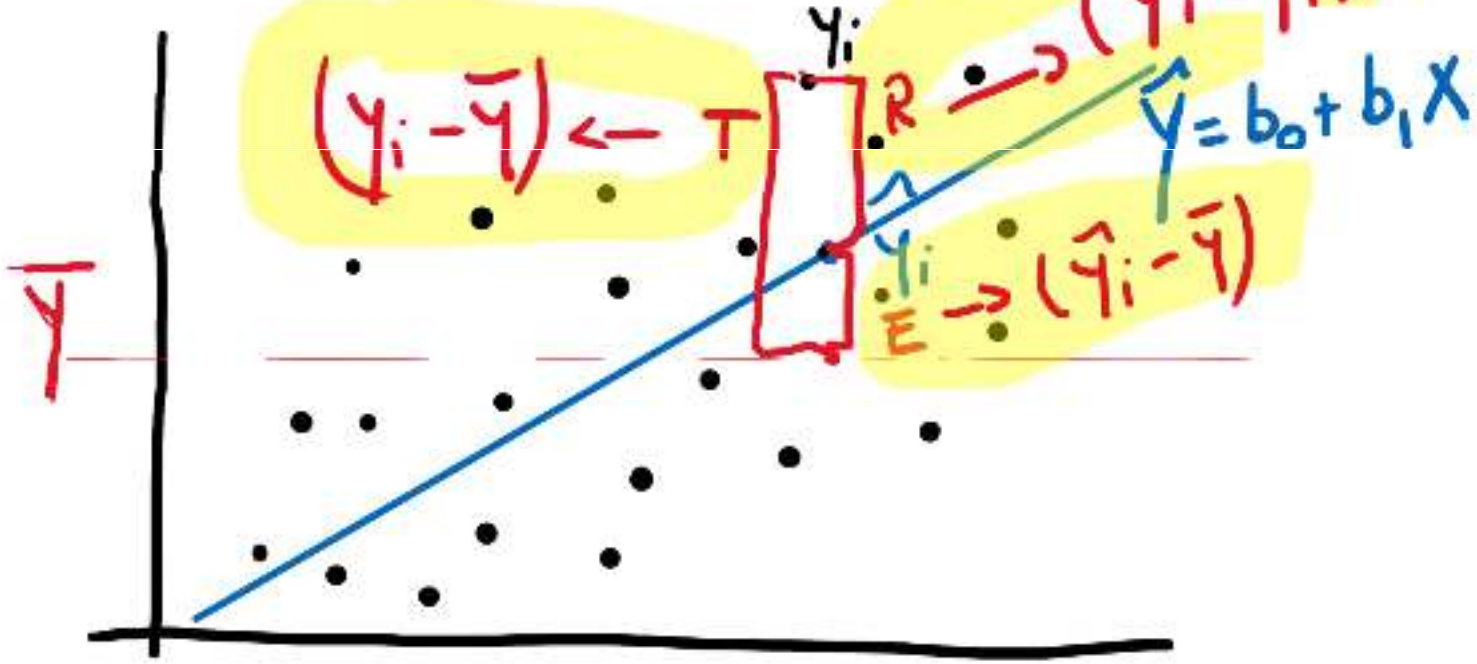
$$\eta^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = \frac{V_E}{V_T} = 1 - \frac{V_R}{V_T}$$, cu $0<\eta^2<1$

V_E, V_T și V_R reprezintă **parametrii variației explicată**, **variația totale și variației reziduale**.

$V_T = V_E + V_R$

$$TSS = \sum_i (y_i - \bar{y})^2; \quad ESS = \sum_i (\hat{y}_i - \bar{y})^2; \quad RSS = \sum_i (y_i - \hat{y}_i)^2;$$

$$df_T = n - 1; \qquad df_Б = k - 1; \qquad df_R = n - k$$



Interpretarea raportului de determinație

η^2 aparține intervalului **[0; 1]**.

Cu cât **η^2 -> 0** cu atât **legătura este mai slabă** între variabila dependentă și variabila independentă.

Cu cât η^2 -> 1 cu atât **legătura este mai puternică** între variabila dependentă și variabila independentă.

OBS: Raportul de determinație poate fi calculat pentru toate tipurile de legaturi: **liniare** sau **neliniare**, **simple** sau **multiple**.

Interpretarea lui η^2 în modelul de regresie

De obicei η^2 este exprimat în valoare procentuală, ia valori de la **0%** la **100%**, și arată **cât la %** din **variația variabilei dependente** (**Y**) este **explicată** de **variația variabilei independente** (**X**) printr-un model specificat (de ex prin modelul liniar $y_{xi} = \beta_0 + \beta_1 X$)

OBS: Pentru **modelul liniar simplu** (**MLS**) se stabilește următoarea relație între η^2 și ρ: $\eta^2 = \rho^2 \Leftrightarrow |\rho| = \eta$ relație care se extinde și asupra etimațiilor acestora: $R^2 = r^2 \Leftrightarrow |r| = R$

Estimația raportului de determinație (R^2)

$$R^2 = \frac{\sum_i (\hat{y} - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

ESS, **TSS** și **RSS** reprezintă **estimațiile variației explicate, variației totale respectiv variației reziduale**.

SS-este **Suma pătratelor [abaterilor]** (eng. **Sum of Sqare**)

TSS=ESS+RSS

Interpretarea **estimației** R^2 este aceeași cu cea a **parametrului** η^2.

Calculul TSS, ESS și RSS și determinarea gradelor de libertate corespunzătoare acestora

$$TSS = \sum_i (y_i - \bar{y})^2; \quad ESS = \sum_i (\hat{y}_i - \bar{y})^2; \quad RSS = \sum_i (y_i - \hat{y}_i)^2;$$

$$\downarrow \downarrow \qquad \downarrow \downarrow \qquad \downarrow \downarrow$$

$$df_T = n - 1; \qquad df_E = k - 1; \qquad df_R = n - k$$

$$df_T = df_E + df_R$$

2. Relația coeficientul de corelație (r) - coeficientul de regresie liniara simpla (b_1)

Legătura dintre *estimația* **coeficientului de corelație (r)** și *estimația* **coeficientului de regresie liniară (b_1)** se realizează prin relația:

$$\left.\begin{aligned} r &= \frac{\operatorname{cov}(y,x)}{s_x s_y} \\ b_1 &= \frac{\operatorname{cov}(y,x)}{s_x^2} \end{aligned}\right\} r = b_1 \frac{s_x}{s_y}$$

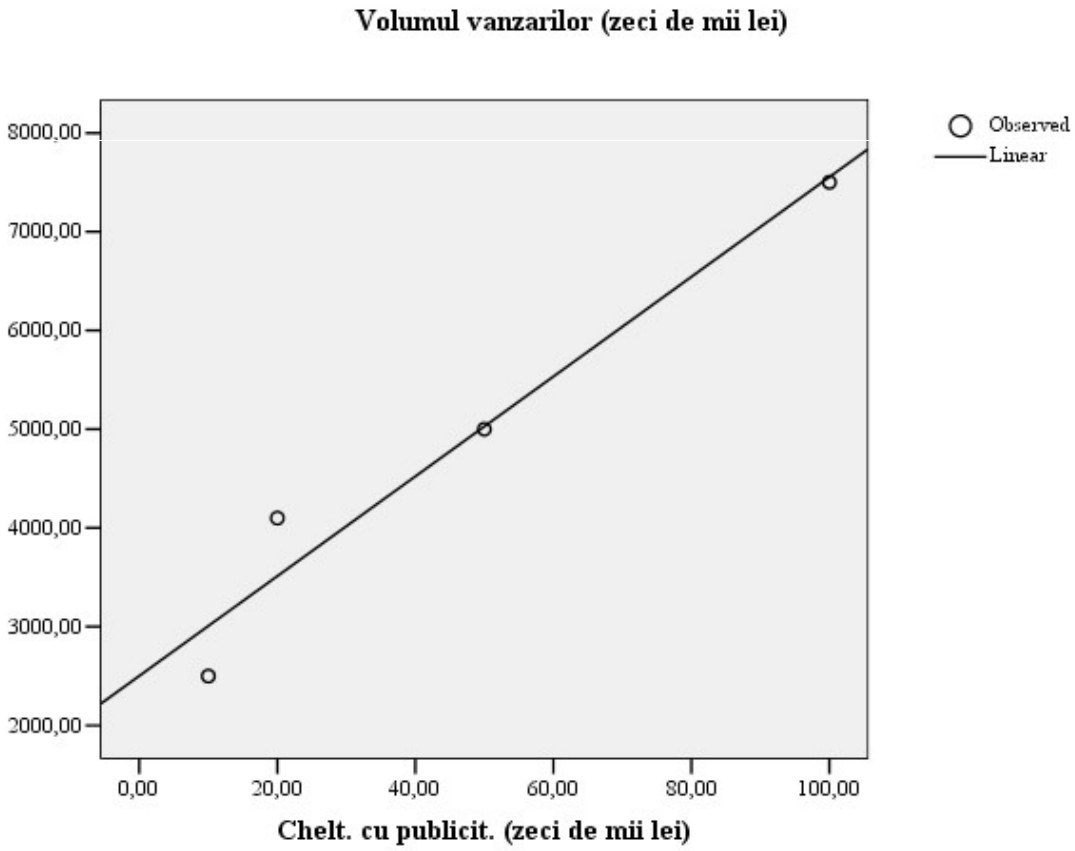
Unde s_x^2, s_y^2 si s_x, s_y reprezintă estimațiile variațelor respectiv estimațiile abaterilor standard ale variabilelor X și Y.

OBS: În cazul **modelului de regresie standardizat (s^2_X=1, s^2_Y=1)** **$r=b_1$**.

EXEMPLU

Se consideră datele cu privire la *Valoarea vânzărilor* şi *Cheltuielile cu publicitatea* pentru un eşantion de 4 firme. Datele sunt prezentate în tabelul următor.

x_i	y_i
10	2500
20	4100
50	5000
100	7500
180	**19100**



Rezultatele analizei in SPSS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.977[a]	.954	.931	550.55630

a. Predictors: (Constant), X

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12501275.51	1	12501275.51	41.243	.023[b]
	Residual	606224.490	2	303112.245		
	Total	13107500.00	3			

a. Dependent Variable: Y

b. Predictors: (Constant), X

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	2502.041	448.379		5.580	.031	572.821	4431.260
	X	50.510	7.865	.977	6.422	.023	16.669	84.351

a. Dependent Variable: Y

3. Testarea indicatorilor de corelaţie

3.I. Testarea coeficientul de corelaţie

1. Formularea ipotezelor H_0: $\rho=0(->\rho_0=0)$

H_1: $\rho\neq0$ (->Test Bilateral->$t_{\alpha/2,\ n-k}=t_{th}$)

2. Fixarea pragului de semnificaţie α (ex. $\alpha=0,05$)

3. Alegerea statisticii test si aflarea valorii critice a acesteia pentru un α dat : $\rho \approx t_{\alpha/2,\,n-k}$

4. Calcularea statisticii test $t_{calc} = \dfrac{\hat{\rho}-\rho_0}{\sigma_{\hat{\rho}}} = \dfrac{\hat{\rho}}{\sqrt{\dfrac{1-\hat{\rho}^2}{n-2}}} \approx \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

5. Criterii de decizie:

$|t_{calc}| \leq t_{teoretic} = t_{\alpha/2,\ n-2} \Leftrightarrow$ sig.$\geq\alpha$ =>se **acceptă/ nu se respinge** H_0 cu o **probabilitate de 1-α**

$|t_{calc}| > t_{teoretic} = t_{\alpha/2,\ n-2} \Leftrightarrow$ sig.$<\alpha$ =>se **respinge** H_0 cu un **risc asumat α**.

3.II. Testarea raportului de corelaţie

1. Formularea ipotezelor : $\mathbf{H_0}$: $\eta=0$ $\mathbf{H_1}$: $\eta\neq0$ (->Test Unilateral ($\eta>0$)->α)

2. Fixarea pragului de semnificaţie α (ex. $\alpha=0{,}05$)

3. Alegerea statisticii test si aflarea valorii critice a acesteia pentru un α dat ($F_{\alpha, k-1, n-k}$).

NB: $v_1=k-1\leq v_2=n-k$

4. Calcularea statisticii test $F_{calc}=\frac{\hat{\eta}}{1-\hat{\eta}}\frac{n-k}{k-1}\cong\frac{R^2}{1-R^2}\frac{n-k}{k-1}$

5. Criterii de decizie:

$F_{calc}\leq F_{\alpha, k-1, n-k}$ $\Leftrightarrow$ sig.$\geq\alpha$. => se **acceptă** $\mathbf{H_0}$ cu **o probabilitate** de **1-α**

$F_{calc}> F_{\alpha, k-1, n-k}$ $\Leftrightarrow$ sig.$<\alpha$ => se **respinge** $\mathbf{H_0}$ cu un **risc asumat α**.

4. Testarea modelului de regresie - testul F omnibus

1. **Formularea ipotezelor**

   H_0: $\beta_0 = 0$ și $\beta_1 = 0$

   H_1: $\beta_0 \neq 0$ sau/si $\beta_1 \neq 0$

2. **Fixarea pragului de semnificaţie** $\alpha = 0{,}05$

3. **Alegerea statisticii test si aflarea valorii critice a acesteia pentru un** α **dat** ($F_{\alpha,\, k-1,\, n-k}$)

4. **Calcularea statisticii test** $$F_{calc} = \frac{V_E}{V_R}\frac{n-k}{k-1} = \frac{\frac{ESS}{k-1}}{\frac{RSS}{n-k}} \cong \frac{ESS}{RSS}\frac{n-k}{k-1}$$

5. **Criterii de decizie**:

$F_{calc} \leq F_{\alpha,\, k-1,\, n-k} \Leftrightarrow sig. \geq \alpha$ => se **acceptă** H_0 cu **o probabilitate** de **1-α**.

$F_{calc} > F_{\alpha,\, k-1,\, n-k} \Leftrightarrow sig. \geq \alpha$ => se **respinge** H_0 cu un **risc asumat** **α**.

Legatura dintre testarea lui R^2 și testarea modelului

Dacă îl exprimăm pe **R^2** în funcție de **SS** vom obține:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

și înlocuind în expresia lui F obținem o expresie a lui F în funcție de SS:

$$F = \frac{ESS}{RSS}\frac{n-k}{k-1} = \frac{R^2}{1-R^2}\frac{n-k}{k-1}$$

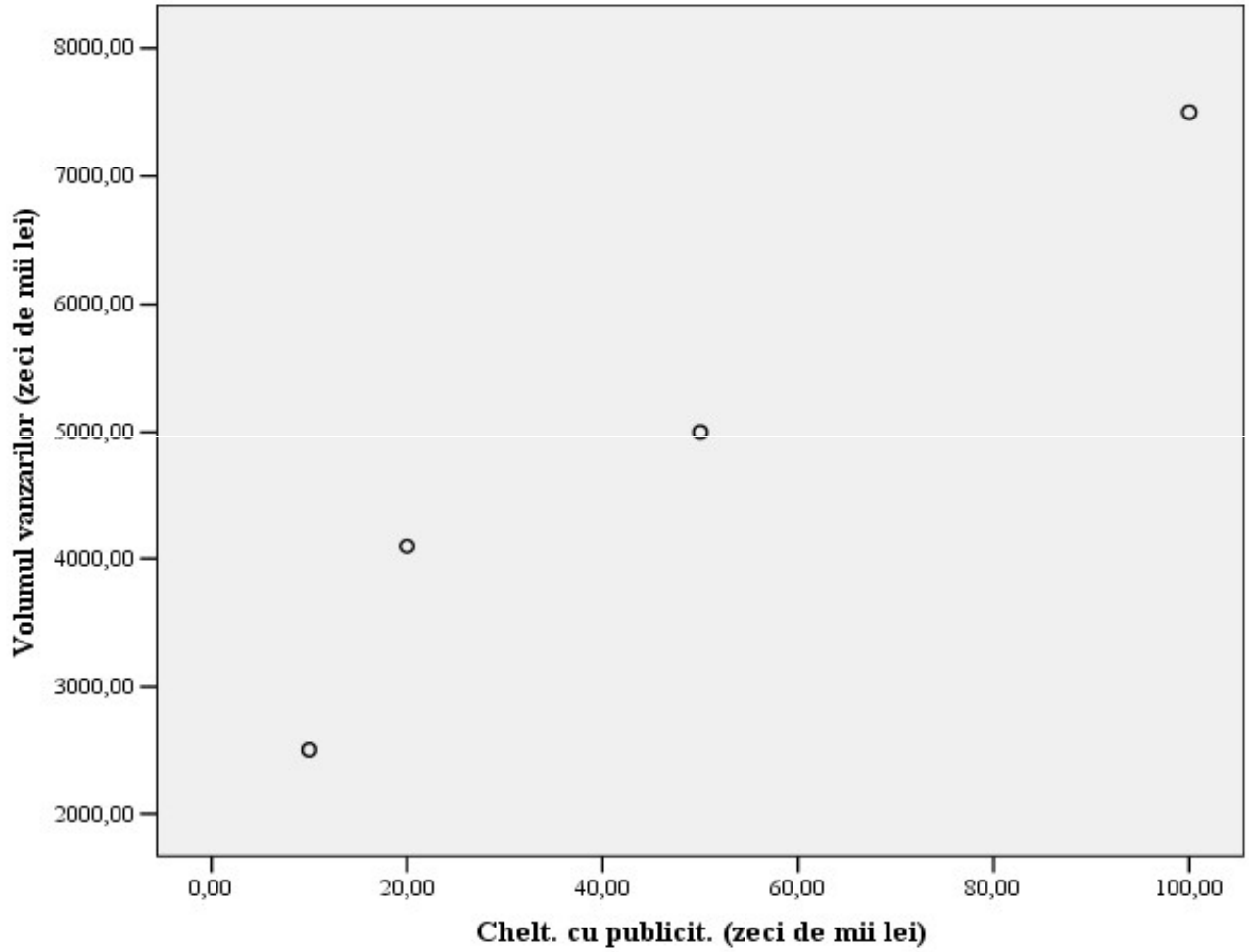
5. Probleme specifice analizei de corelație și regresie

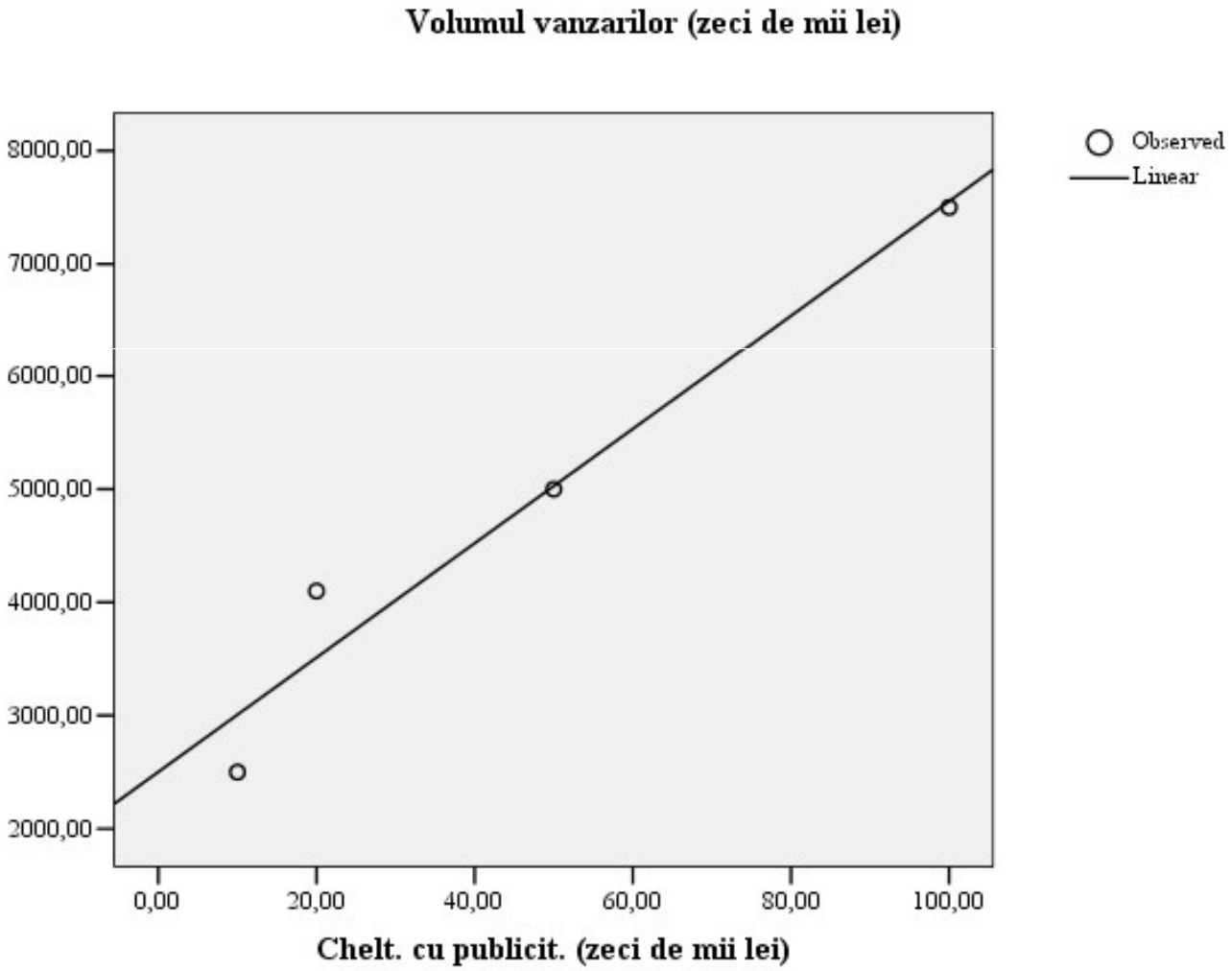
Analiza de regresie si corelație în SPSS

Se consideră datele cu privire la *Valoarea vânzărilor* şi *Cheltuielile cu publicitatea* pentru un eşantion de 4 firme.

Datele sunt prezentate în tabelul alăturat.

x_i	y_i
10	2500
20	4100
50	5000
100	7500
180	**19100**


8000,00
7000,00
6000,00
5000,00
4000,00
3000,00
2000,00
Volumul vanzarilor (zeci de mii lei)
0,00
20,00
40,00
60,00
80,00
100,00
Chelt. cu publicit. (zeci de mii lei)


Volumul vanzarilor (zeci de mii lei)
8000,00
7000,00
6000,00
5000,00
4000,00
3000,00
2000,00
0,00
20,00
40,00
60,00
80,00
100,00
Chelt. cu publicit. (zeci de mii lei)
Observed
Linear

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.977[a]	.954	.931	550.55630

a. Predictors: (Constant), X

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12501275.51	1	12501275.51	41.243	.023[b]
	Residual	606224.490	2	303112.245		
	Total	13107500.00	3			

a. Dependent Variable: Y

b. Predictors: (Constant), X

Correlations

		X	Y
X	Pearson Correlation	1	.977*
	Sig. (2-tailed)		.023
	N	4	4
Y	Pearson Correlation	.977*	1
	Sig. (2-tailed)	.023	
	N	4	4

*. Correlation is significant at the 0.05 level (2-tailed).

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	2502.041	448.379		5.580	.031	572.821	4431.260
	X	50.510	7.865	.977	6.422	.023	16.669	84.351

a. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	b_0 2502,041	$S_{\hat{\beta}_0}$ 448,379		5,580	,031	572,821	4431,260
	X	b_1 50,510	$S_{\hat{\beta}_1}$ 7,865	,877	6,422	,023	16,669	84,351

a. Dependent Variable: Y

$P(\beta_0 \in [572,821;\ 4431,260]) = 1-\alpha = 95\%$

$P(\beta_1 \in [16,669;\ 84,351]) = 1-\alpha = 95\%$

L_i L_s

$n = 4 \Rightarrow n-2 = 2$

$IC_{\beta_0} = b_0 \pm \Delta\hat{\beta}_0 = b_0 \pm t_{\alpha/2;2} \cdot S_{\hat{\beta}_0} = 2502,041 \pm 4,303 \cdot 448,379 \cong \begin{matrix} L_i(\beta_0) \\ L_s(\beta_0) \end{matrix}$ TABEL

$\alpha = 0,05 : t_{\alpha/2;2} = 4,303$

$IC_{\beta_1} = b_1 \pm \Delta\hat{\beta}_1 = b_1 \pm t_{\alpha/2;2} \cdot S_{\hat{\beta}_1} = 50,510 \pm 4,303 \cdot 7,865 \cong \begin{matrix} L_i(\beta_1) \\ L_s(\beta_0) \end{matrix}$