ECONOMETRIE

Planul cursului

1. Introducere
2. Modelul de regresie liniară simplă
3. Modelul de regresie liniară multiplă
4. Modele de regresie neliniară
5. Ipoteze statistice: normalitatea erorilor, homoscedasticitatea, necorelarea erorilor, multicoliniaritatea.

2. Modelul de regresie liniară simplă

2.1. Noțiuni introductive
2.2. Forma generală a modelului de regresie liniară simplă
2.3. Ipoteze clasice formulate
2.4. Estimarea parametrilor modelului
2.5. Estimarea indicatorilor de corelație
2.6. Testarea parametrilor modelului
2.7. Testarea modelului de regresie
2.8. Testarea indicatorilor de corelație
2.9. Regresia prin origine
2.10. Aplicație

2.1. Noțiuni introductive

- Natura datelor: variabile numerice.
- Obiectivele analizei de regresie: studiul legăturilor dintre fenomene şi folosirea modelului în scop predictiv.
- Corelație: intensitatea legăturii (legătura dintre variabile este puternică sau slabă).

2.2. Forma generală a modelului de regresie liniară simplă

A. *Identificarea pe cale grafică* a formei legăturii dintre variabile:

2.2. Forma generală a modelului de regresie liniară simplă

B. *Modelul econometric de regresie liniară simplă*

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon \quad \text{(scris cu variabile)}$$

$$y_i = \beta_o + \beta_1 \cdot x_i + \varepsilon_i \quad \text{(scris cu valori)}$$

Dacă se consideră că regresia este o medie condiționată, componenta deterministă este:

$$M(Y / X = x_i) = \beta_o + \beta_1 \cdot x_i$$

iar modelul este: $y_i = M(Y|X = x_i) + \varepsilon_i$

Parametrii modelului:

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon$$

$$y_i = b_o + b_1 \cdot x_i + e_i$$

$$y_{x_i} = b_o + b_1 \cdot x_i$$

Parametrii modelului sunt:

β0 – este constanta modelului (*intercept*). Arată nivelul mediu a lui *Y* atunci când *X=0*. Este ordonata la origine (atunci când dreapta de regresie trece prin origine, β_0=0)

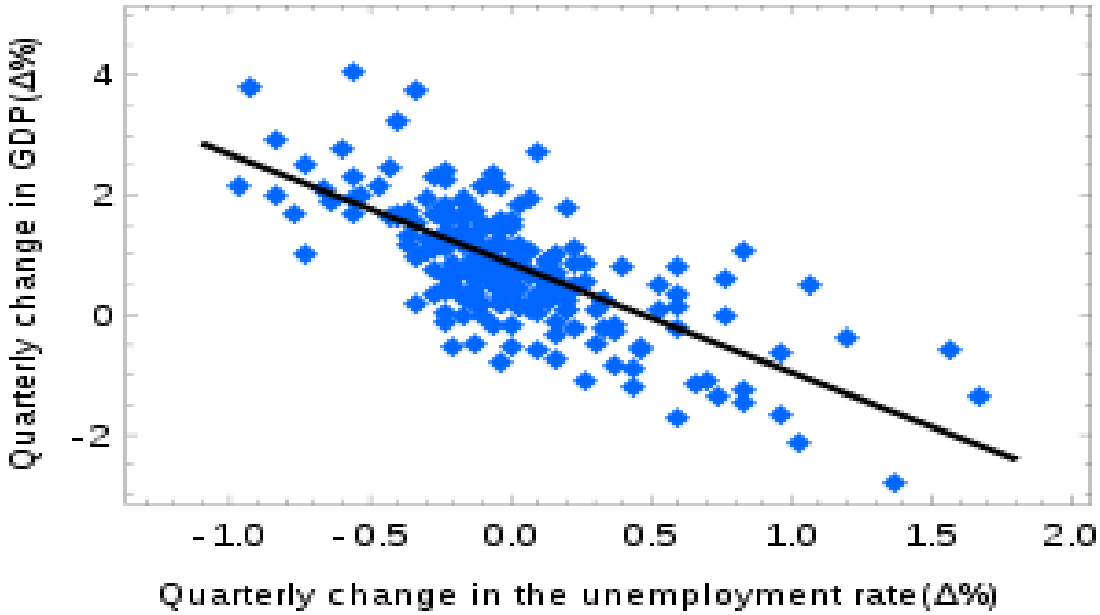
Parametrii modelului:

β1 – este panta dreptei de regresie. *Semnul* pantei arată sensul legăturii dintre variabile:

-O valoare pozitivă arată o legătură directă dintre *X* și *Y*
-O valoare negativă arată o legătură inversă dintre *X* și *Y*

Valoarea pantei arată cu cât variază, în medie, nivelul lui Y la o creștere cu o unitate a lui X.

Exemplu



Conform legii lui Okun (1962), există o legătură inversă între rata de creştere a economiei şi rata şomajului. Pentru a menţine anumite niveluri de ocupare, o economie trebuie să crească în fiecare an, cu o rată cuprinsă între 2,6% şi 3%. Orice creştere mai mică, pentru economist, a însemnat o creştere a şomajului, datorită îmbunătăţirii productivităţii.

2.3. Ipoteze clasice

- Normalitatea, homoscedasticitatea și autocorelarea erorilor;
- Multicoliniaritatea.

(tratate în ultimul capitol)

Dacă se admite ipoteza $\varepsilon \sim N(0,\sigma^2)$, atunci variabila dependentă este o variabilă aleatoare normal distribuită de forma: $Y \sim N(\beta_0 + \beta_1 X; \sigma^2)$

2.4. Estimarea parametrilor modelului

a. Noțiuni teoretice
b. Metoda de estimare
c. Estimarea punctuală a parametrilor modelului
d. Estimarea prin interval de încredere
e. Valorile estimate ale parametrilor modelului de regresie în SPSS

a. Noțiuni teoretice

- Estimarea reprezintă procedeul de determinare a unui parametru al unei populații (β_0, β_1) pe baza datelor înregistrate la nivelul unui eşantion.
- Se poate realiza prin:

1. *estimare punctuală*.
2. *estimare prin interval de încredere (IC)*.

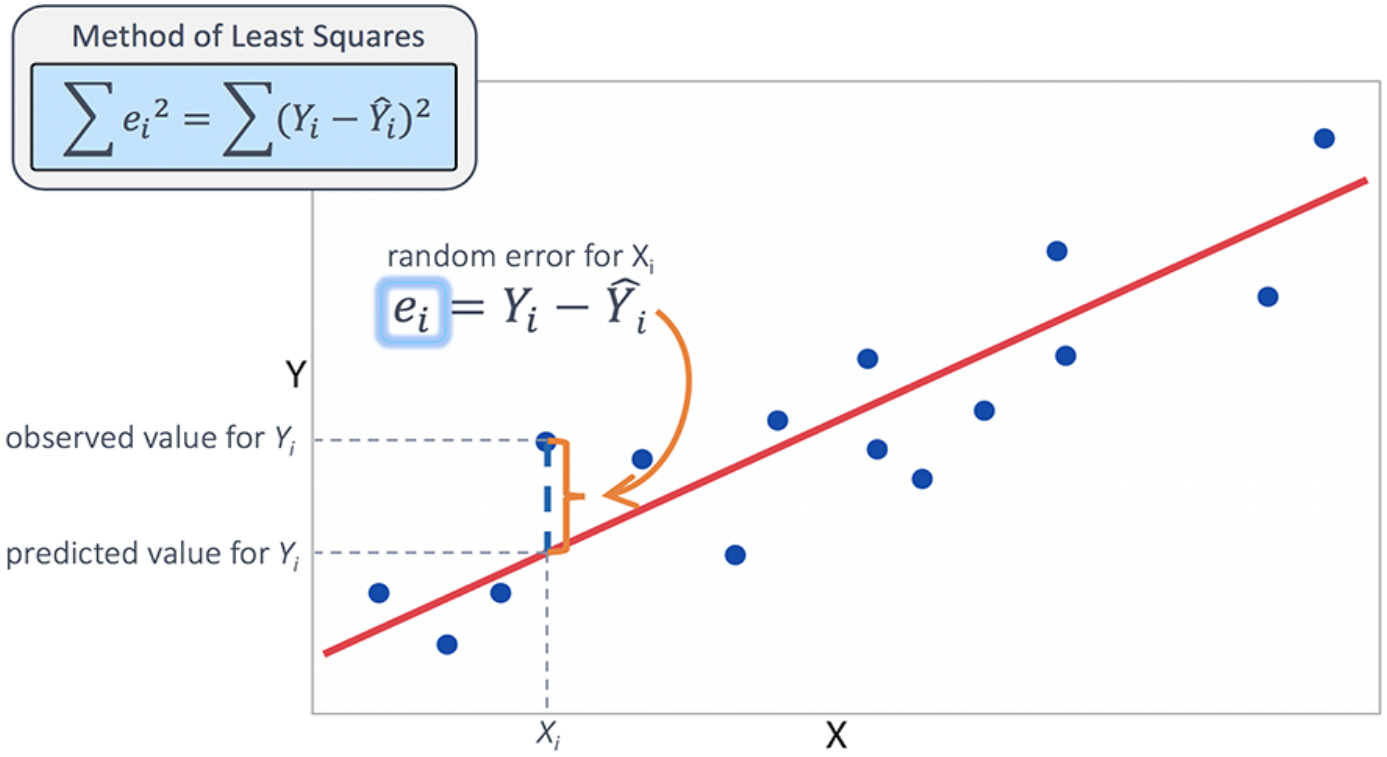
b. Metoda de estimare

$$\sum_i (e_i^2) \min$$

b. Metoda de estimare

Metoda celor mai mici pătrate (MCMMP):

$$S = \sum_i e_i^2 = \sum_i (y_i - y_{x_i})^2 = \sum_i (y_i - b_o - b_1 \cdot x_i)^2 = \min$$


Method of Least Squares
$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$
random error for X_i
$e_i = Y_i - \hat{Y}_i$
Y
observed value for Y_i
predicted value for Y_i
X_i
X

c. Estimarea punctuală a parametrilor modelului

$$S = \sum_i e_i^2 = \min$$

$$\frac{\partial S}{\partial b_0} = 2 \cdot \Sigma (y_i - b_0 - b_1 \cdot x_i) \cdot (- 1) = 0$$

$$\frac{\partial S}{\partial b_1} = 2 \cdot \Sigma (y_i - b_0 - b_1 \cdot x_i) \cdot (- x_i) = 0$$

Relații de calcul:

- n este numărul de observații

$$b_0 = \frac{\sum y_i \cdot \sum x_i^2 - \sum x_i \cdot \sum x_i \cdot y_i}{n \cdot \sum x_i^2 - (\sum x_i)^2} \quad sau \quad b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{n \cdot \sum x_i y_i - \sum x_i \cdot \sum y_i}{n \cdot \sum x_i^2 - (\sum x_i)^2}$$

Derivatele parțiale de ordinul doi:

$$\frac{\partial^2 S}{\partial {b_0}^2} = 2n; \frac{\partial^2 S}{\partial b_0 \partial b_1} = 2\sum_i x_i; \frac{\partial^2 S}{\partial {b_1}^2} = 2\sum_i x_i^2.$$

Matricea derivatelor parțiale de ordinul doi:

$$\begin{pmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}$$

Derivatele parțiale de ordinul doi – pozitiv definite:

$$n\sum_i x_i^2 - \left(\sum_i x_i\right)^2 = n \cdot \sigma^2 > 0$$

d. Estimarea prin interval de încredere (IC) a parametrilor modelului

- Estimatorii parametrilor β_i urmează o lege normală şi sunt:
- *nedeplasați* (în condițiile respectării ipotezei că variabila *X* este nestochastică şi pe baza proprietății că variabilele aleatoare y_i urmează aceeaşi lege de repartiție);
- *convergenți* (pentru un volum al eşantionului suficient de mare) (n mai mare decât 30);
- *eficienți*.

d. Estimarea prin interval de încredere (IC) a parametrilor modelului

A.- Parametrul β_0 $\quad \hat{\beta}_0 \sim N(\beta_0, \sigma^2_{\hat{\beta}_0})$

$$M(\hat{\beta}_0) = \beta_0 \; ; \; V(\hat{\beta}_0) = \sigma^2_{\hat{\beta}_0} = \frac{\sum_i x_i^2}{n \cdot \sum_i (x_i - \bar{x})^2} \cdot \sigma^2_\varepsilon$$

- I.C. este definit de limitele de încredere care acoperă valoarea unui parametru, pentru un coeficient de încredere.

$$\hat{\beta}_0 \sim N(\beta_0, \sigma^2_{\hat{\beta}_0}) \Rightarrow Z \sim N(0,1)$$

după relația:

$$Z = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}}$$

sau când nu se cunoaşte varianța:

$$t = \frac{\hat{\beta}_0 - \beta_0}{s_{\hat{\beta}_0}}$$

- Această variabilă t Student (Z) permite să se construiască un interval de încredere, astfel:

$$P(-t_{\alpha/2} \leq \frac{\hat{\beta}_0 - \beta_0}{s_{\hat{\beta}_0}} \leq +t_{\alpha/2}) = 1 - \alpha$$

unde: α este un nivel al probabilității cuprins între zero şi unu (numit şi risc asumat; de regulă, egal cu 0,05 sau 5%).

Prin prelucrarea datelor la nivelul unui eşantion, se obține o estimație punctuală a parametrului β_0, respectiv valoarea b_0.

$$P(-t_{\alpha/2} \leq \frac{b_0 - \beta_0}{s_{\hat{\beta}_0}} \leq +t_{\alpha/2}) = 1 - \alpha$$

I.C. pentru parametrul β_0:

$$\left[b_0 \pm t_{\alpha/2;n-2} \cdot s_{\hat{\beta}_0} \right]$$

unde:

- b_0 este o estimație punctuală a parametrului β_0;

- $t_{\alpha/2,n-2}$ este o valoare a statisticii *t Student* care se citeşte pentru un risc α (de regulă, egal cu 0,05) şi (n-2) grade de libertate (*df*).

Obs.: $df=n-k$, unde k este numărul de parametri ai modelului.

s este o estimație a abaterii standard a estimatorului acestui parametru (Std. Error)

$$s_{\hat{\beta}_0} = \sqrt{\frac{\sum_i x_i^2}{n \cdot \sum_i (x_i - \bar{x})^2} \cdot s_e^2}$$

$$s_e^2 = \frac{\sum_i e_i^2}{n-2}; e_i = y_i - y_{x_i}; y_{x_i} = b_0 + b_1 \cdot x_i$$

B. Parametrul β_1:

$$\hat{\beta}_1 \sim \mathrm{N}(\beta_1, \sigma^2_{\hat{\beta}_1})$$

$$M(\hat{\beta}_1) = \beta_1;\ V(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \frac{\sigma^2_\varepsilon}{\sum_i (x_i - \bar{x})^2}$$

- I.C. pentru parametrul β_1:

$$\left[b_1 \pm t_{\alpha|2;n-2} \cdot s_{\hat{\beta}_1} \right]$$

unde: b_1 este o estimație punctuală a parametrului β_1;

$$s_{\hat{\beta}_1} = \sqrt{\frac{s_e^2}{\sum_i (x_i - \bar{x})^2}}$$

$$s_e^2 = \frac{\sum_i e_i^2}{n - 2}; e_i = y_i - y_{x_i}.$$

$$y_{x_i} = b_0 + b_1 \cdot x_i$$

e. Valorile estimate ale parametrilor modelului de regresie în SPSS

X-Pret (lei/kg)	Y-Consum (kg)
5	15
7	12
9	13
11	10
14	8

e. Valorile estimate ale parametrilor modelului de regresie în SPSS

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18,311	1,443		12,689	,001
	X	-,730	,149	-,943	-4,912	,016

a. Dependent Variable: Y

Exemplu:

Pe baza datelor din output-ul de mai sus, cunoscând n=5, se cere:

1. Să se scrie modelul legăturii dintre cele două variabile;
2. Să se precizeze sensul legăturii dintre aceste variabile;
3. Să se calculeze limitele intervalului de încredere pentru parametrul β_1, considerând un risc de 0.05.

2.5. Estimarea indicatorilor de corelație

- *Obiectiv:* studiul intensității legăturii dintre variabile.

2.5.1 Indicatori de corelație:

1. Coeficientul de corelație
2. Raportul de determinație
3. Raportul de corelație

1. Coeficientul de corelație

$$\rho = \frac{\text{cov}(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\sum_i (x_i - \mu_X) \cdot (y_i - \mu_Y)}{N \cdot \sigma_X \cdot \sigma_Y}$$

$$\rho = \beta_1 \cdot \frac{\sigma_X}{\sigma_Y}$$

- Domeniul de variație
- Interpretare

Observație:

- Dacă se realizează o standardizare a variabilelor X şi Y, atunci estimatorul coeficientului de corelație pentru aceste variabile este identic cu cel al coeficientului de regresie β_1.

Exemplu

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18,311	1,443		12,689	,001
	X	-,730	,149	-,943	-4,912	,016

a. Dependent Variable: Y

2. *Raportul de determinație*

- măsoară gradul de corelație dintre variabile şi calitatea ajustării norului de puncte prin dreapta de regresie (*goodness of fit*).

- ecuația de analiză a variației:

  Variația totală = Variația explicată + Variația reziduală sau

  $V_T = V_E + V_R$

$$\eta^2 = \frac{\sum_i (y_x - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = \frac{V_E}{V_T} = 1 - \frac{V_R}{V_T}$$

- Domeniul de variație
- Interpretare
- Estimarea raportului de determinație

$$R^2 = \frac{\sum_i (y_{x_i} - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = \frac{\sum_i (b_0 + b_1 \cdot x_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

- *ESS* este estimația variației explicate;
- *RSS* este estimația variației reziduale;
- *TSS* este estimația variației totale.

Exemplu

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25.970	1	25.970	24.125	.016[a]
	Residual	3.230	3	1.077		
	Total	29.200	4			

a. Predictors: (Constant), Pret

b. Dependent Variable: Consum

Observații:

- raportul de determinație poate fi folosit pentru compararea calității modelelor care au aceeaşi variabilă dependentă Y.

- pentru modelul de regresie liniară simplă:

  $\rho^2=\eta^2$ *şi* $r^2=R^2$.

3. Raportul de corelație

$$\eta = \sqrt{\eta^2} = \sqrt{\frac{\sum_i (y_x - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}} = \sqrt{\frac{V_E}{V_T}} = \sqrt{1 - \frac{V_R}{V_T}}$$

- domeniul de variație.

Estimarea raportului de corelație

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

2.5.2 Indicatori de corelație în SPSS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.943[a]	.889	.853	1.03755

a. Predictors: (Constant), Pret

Exemplu

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25.970	1	25.970	24.125	.016[a]
	Residual	3.230	3	1.077		
	Total	29.200	4			

a. Predictors: (Constant), Pret

b. Dependent Variable: Consum

2.6 Testarea parametrilor modelului

- Ipoteze statistice
- Calculul statisticii test
- Regula de decizie
- Interpretare

Exemplu

Coefficients[a]

Model		Unstandardized Coefficients	
		B	Std. Error
1	(Constant)	18,311	1,443
	X	-,730	,149

a. Dependent Variable: Y

Exemplu

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18,311	1,443		12,689	,001
	X	-,730	,149	-,943	-4,912	,016

a. Dependent Variable: Y

Pe baza datelor din output-ul de mai sus, se cere:

1. Să se testeze semnificația parametrului β_0
2. Să se testeze semnificația parametrului β_1 , pentru un risc de 5%.

2.7 Testarea modelului de regresie

- Ipoteze

H_o: $\beta_0 = 0$; $\beta_1 = 0$

H_1: $\beta_0 \# 0$; $\beta_1 \# 0$

- Calculul statisticii test

$$F_{calc} = \frac{\frac{ESS}{k-1}}{\frac{RSS}{n-k}}$$

- Regula de decizie

Dacă $F_{calc} > F_{\alpha,k-1,n-k}$, atunci se

respinge ipoteza Ho, cu o probabilitate de $1-\alpha$

Exemplu

ANOVA[a]

Model		Sum of Squares	df
1	Regression	25,970	1
	Residual	3,230	3
	Total	29,200	4

a. Dependent Variable: Y

Exemplu

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25,970	1	25,970	24,125	,016[b]
	Residual	3,230	3	1,077		
	Total	29,200	4			

a. Dependent Variable: Y

b. Predictors: (Constant), X

2.8. Testarea indicatorilor de corelație

a) Testarea coeficientului de corelație

- Ipoteze statistice
- Statistica test
- Calculul statisticii test
- Decizie

Exemplu

Correlations

		X	Y
X	Pearson Correlation	1	-,943*
	Sig. (2-tailed)		,016
	N	5	5
Y	Pearson Correlation	-,943*	1
	Sig. (2-tailed)	,016	
	N	5	5

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

Exemplu

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25.970	1	25.970	24.125	.016[a]
	Residual	3.230	3	1.077		
	Total	29.200	4			

a. Predictors: (Constant), Pret

b. Dependent Variable: Consum

b) Testarea raportului de corelație

- Ipoteze statistice
- Statistica test
- Calculul statisticii test
- Decizie

c) Testarea raportului de determinație

Exemplu

ANOVA[a]

Model		Sum of Squares	df
1	Regression	25,970	1
	Residual	3,230	3
	Total	29,200	4

a. Dependent Variable: Y

Exemplu

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,943[a]	,889	,853	1,03755

a. Predictors: (Constant), X

2.9 Regresia prin origine

- Parametrul β_0 poate să nu fie semnificativ statistic.
- Modelul de regresie este:

$$Y = \beta_1 \cdot X + \varepsilon$$

- Metoda celor mai mici pătrate:

$$S = \sum_i e_i^2 = \sum_i (y_i - y_{x_i})^2 = \sum_i (y_i - b_1 \cdot x_i)^2 = min$$

$$\frac{\partial S}{\partial b_1} = 2 \cdot \Sigma (y_i - b_1 \cdot x_i) \cdot (- x_i) = 0$$

$$b_1 \cdot \Sigma x_i^2 = \Sigma x_i y_i$$

Probleme:

1. Nerespectarea condiției $\sum_i e_i = 0$
2. Pentru acest model de regresie prin origine, R^2 poate fi negativ. În literatură există un alt indicator:

$$r^2 = \frac{\sum_i (x_i y_i)^2}{\sum_i x_i^2 \sum_i y_i^2}$$

Observații:

- Se recomandă evitarea eliminării parametrului β_0 din ecuația de regresie.
- În cazul variabilelor standardizate, problemele menționate mai sus dispar (panta dreptei de regresie este egală cu valoarea coeficientului de corelație Pearson şi nu este necesară calcularea unui alt coeficient).

2.10. Aplicație modelul de regresie liniară simplă

Vânzări (mil lei)	Chelt. public. (mii lei)
7	8
14	10
12	12
27	20
19	15
25	17
15	10
18	15
25	17
8	9

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-5.092	2.649		-1.922	.091
	X	1.661	.191	.951	8.680	.000

a. Dependent Variable: Y

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	408	1	408.616	75.348	.000[b]
	idual	.384	8	5.423		
	Total	452.000	9			

a. Dependent Variable: Y

b. Predictors: (Constant), X

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.951[a]	.904	.892	2.32874

a. Predictors: (Constant), X

Correlations

		Y	X
Y	Pearson Correlation	1	.951**
	Sig. (2-tailed)		.000
	N	10	10
X	Pearson Correlation	.951**	1
	Sig. (2-tailed)	.000	
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

Se cere:

1. Să se scrie ecuația estimată a legăturii dintre *Vânzări* și *Cheltuieli de publicitate* și să se interpreteze panta dreptei de regresie.
2. Pe baza rezultatelor din tabelul *Coefficients,* să se aprecieze intensitatea legăturii dintre variabile.
3. Să se precizeze cu cât cresc, în medie, vânzările dacă se măresc cheltuielile de publicitate cu 10 mii lei.
4. Să se estimeze cât ar trebui să se cheltuiască pentru publicitate pentru a avea vânzări de 20 mil. lei.
5. Să se testeze dacă influența cheltuielilor de publicitate asupra vânzărilor este semnificativă statistic (pentru un risc de 0.05).

Se cere:

6. Pe baza rezultatelor din tabelul *Model summary,* să se estimeze valoarea coeficientului de corelație.

7. Să se testeze semnificația statistică a ordonatei la origine (pentru un risc de 10%).

8. Să se testeze semnificația coeficientului de corelație, folosind statistica test t Student (pentru un risc de 0.05).

9. Să se testeze semnificația raportului de corelație (pentru un risc de 0.05).

10. Să se testeze dacă modelul de regresie este corect specificat (pentru un risc de 0.05).