

Capitolul 2. Modelul de regresie liniară simplă

Aplicații

I. Se consideră datele cu privire la *nivelul studiilor* (ani) și *venitul* (lei) pentru un eșantion de 5 angajați. Datele sunt prezentate în tabelul următor.

x_i	y_i
10	800
12	1000
12	1200
14	1600
16	1800
64	6400

Pe baza datelor prezentate, se cere:

- 1) Să se specifice variabila dependentă și variabila independentă analizate;
- 2) Să se reprezinte grafic legătura dintre cele două variabile;
- 3) Să se explice legătura de dependență dintre cele două variabile;
- 4) Să se aleagă forma modelului de regresie;
- 5) Să se estimeze punctual parametrii modelului de regresie;
- 6) Să se interpreteze estimațiile parametrilor modelului de regresie;
- 7) Să se scrie ecuația estimată a modelului de regresie;

i	x_i	y_i	$x_i y_i$	x_i^2	$n = 5$
1	10	800	8000	100	$\sum_i x_i = 64$
2	12	1000	12000	144	$\sum_i y_i = 6400$
3	12	1200	14400	144	$\sum_i x_i y_i = 85600$
4	14	1600	22400	196	$\sum_i x_i^2 = 840$
5	16	1800	28800	256	$\bar{y} = 1280$
5	64	6400	85600	840	$\bar{x} = 12,8$

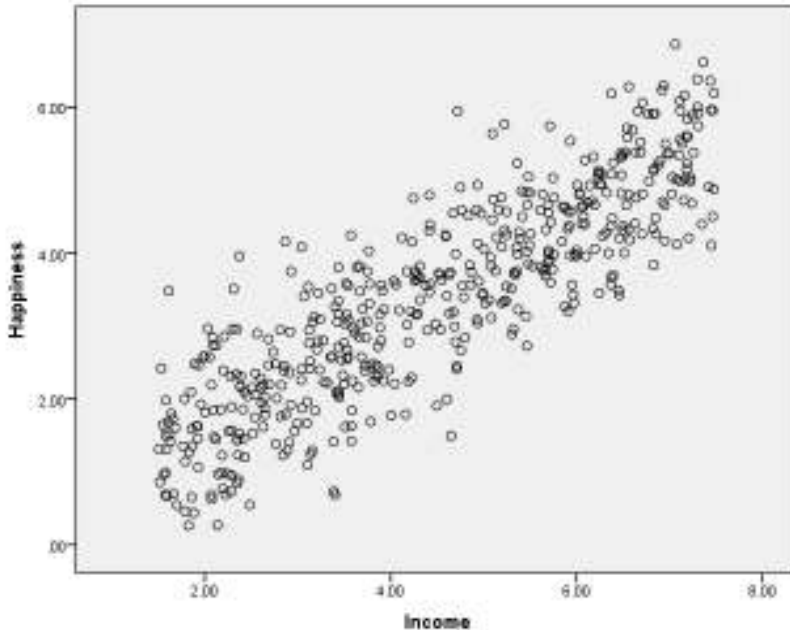
$$b_1 = \frac{n \sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n \sum_i x_i^2 - (\sum_i x_i)^2}$$

$$b_0 = \frac{\sum_i y_i \sum_i x_i^2 - \sum_i x_i \sum_i x_i y_i}{n \sum_i x_i^2 - (\sum_i x_i)^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

II. Se consideră datele unei anchete de sondaj privind scorul fericirii (cu valori de la 0 la 10 puncte) și venitul (exprimat în 10 mii de dolari) înregistrate pentru 498 de persoane.

1. Reprezentarea grafică a legăturii dintre cele două variabile analizate



Pe baza reprezentării grafice a legăturii dintre cele două variabile analizate, se cere:

- 1) Pe baza datelor prezentate, să se specifice variabila dependentă și variabila independentă analizate.
- 2) Să se explice legătura de dependență din cele două variabile.
- 3) Să se identifice forma și sensul legăturii dintre cele două variabile.
- 4) Să se aleagă forma modelului de regresie.

2. Estimarea și testarea parametrilor modelului de regresie

Model	1		2		3	4	5	
	Unstandardized Coefficients		Standardized Coefficients		t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta				Lower Bound	Upper Bound
1 (Constant)	.204	.089			2.299	.022	.030	.379
Income	.714	.019	.866		38.505	.000	.677	.750

a. Dependent Variable: Happiness

1.1 1.2 5.1 5.2

Legenda tabelului <i>Coefficients</i>	Notații și rezultate obținute pe baza datelor de la nivelul eșantionului	
	Constant (b_0)	PIB (b_1)
1. Coeficienți nestandardizați		
1.1. Estimațiile parametrilor modelului de regresie	$b_0 =$	$b_1 =$
1.2. Estimațiile erorilor standard ale estimatorilor parametrilor modelului de regresie	$s_{\hat{\beta}_0} =$	$s_{\hat{\beta}_1} =$
2. Coeficienți standardizați		$\tilde{b}_1 =$
3. Valoarea calculată a statisticii test Student	$t_{calc} = \frac{b_0}{s_{\hat{\beta}_0}} =$	$t_{calc} = \frac{b_1}{s_{\hat{\beta}_1}} =$
4. Probabilitatea asociată statisticii test Student (semnificația testului)	$Sig\ t =$	$Sig\ t =$
5. Intervalul de încredere al parametrilor modelului de regresie	$IC(\beta_0): [\quad ; \quad]$	$IC(\beta_1): [\quad ; \quad]$
5.1. Limita inferioară a intervalului	$b_0 - t_{\alpha/2; n-2} \cdot s_{\hat{\beta}_0}$	$b_1 - t_{\alpha/2; n-2} \cdot s_{\hat{\beta}_1}$
5.2. Limita superioară a intervalului	$b_0 + t_{\alpha/2; n-2} \cdot s_{\hat{\beta}_0}$	$b_1 + t_{\alpha/2; n-2} \cdot s_{\hat{\beta}_1}$

Pe baza rezultatelor privind estimarea punctuală și prin interval de încredere a parametrilor modelului de regresie construit, se cere:

- 5) Să se scrie ecuația estimată a modelului de regresie.
- 6) Să se interpreteze estimațiile parametrilor modelului de regresie.
- 7) Să se interpreteze intervalele de încredere ale coeficienților de regresie, considerând o probabilitate de 95%.
- 8) Să se estimeze prin interval de încredere ordonata la origine și panta dreptei de regresie, pentru o probabilitate de 0,99.

Pe baza modelului estimat, se cere:

- 9) Să se estimeze scorul fericirii pentru un individ pentru care venitul este de 3 unități (30 de mii de dolari).
- 10) Să se precizeze cât ar trebui să fie nivelul venitului unui individ pentru a obține un scor al fericirii de 4 puncte.
- 11) Să se estimeze cu cât scade scorul fericirii dacă are loc o scădere a venitului cu 0,5 unități (5000 de dolari).
- 12) Să se determine valoarea cu care ar trebui să crească nivelul venitului pentru a avea o creștere medie a scorului fericirii cu 1,5 puncte.

Pe baza rezultatelor privind testarea parametrilor modelului de regresie, se cere:

13) Să se testeze semnificația coeficienților de regresie, aplicând toți pașii demersului testării.

Etapele testării	Testarea parametrului β_0	Testarea parametrului β_1
1. Formularea ipotezelor	$H_0: \beta_0 = 0$ (parametrul β_0 nu diferă <u>semnificativ</u> de 0 SAU constanta modelului nu este <u>semnificativă</u> statistic) $H_1: \beta_0 \neq 0$ (parametrul β_0 diferă <u>semnificativ</u> de 0 SAU constanta modelului este <u>semnificativă</u> statistic)	$H_0: \beta_1 = 0$ (parametrul β_1 nu diferă semnificativ de 0, ceea ce înseamnă că între cele două variabile nu există o legătură liniară <u>semnificativă</u> SAU variabila independentă X nu are o influență <u>semnificativă</u> asupra variabilei dependente Y) $H_1: \beta_1 \neq 0$ (parametrul β_1 diferă semnificativ de 0 ceea ce înseamnă că între cele două variabile există o legătură liniară <u>semnificativă</u> SAU variabila independentă X explică <u>semnificativ</u> variația variabilei dependente Y)
2. Alegerea pragului de semnificație	$\alpha = 0,05$	$\alpha = 0,05$
3. Alegerea statisticii test	$t = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \sim t(n-2)$	$t = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t(n-2)$
4. Determinarea valorii teoretice a statisticii test	$t_{teoretic} = t_{\alpha/2; n-2} =$	$t_{teoretic} = t_{\alpha/2; n-2} =$
5. Determinarea valorii calculate a statisticii test (în condițiile acceptării ipotezei nule)	$t_{calc} = \frac{b_0}{S_{\hat{\beta}_0}} =$	$t_{calc} = \frac{b_1}{S_{\hat{\beta}_1}} =$
6. Regula de decizie	Dacă se ține cont de valoarea calculată a testului, regula de decizie este următoarea: <ul style="list-style-type: none"> - dacă $t_{calc} \leq t_{\alpha/2; n-2}$, nu se respinge ipoteza nulă (H_0); - dacă $t_{calc} > t_{\alpha/2; n-2}$, se respinge ipoteza nulă (H_0), cu probabilitatea $(1 - \alpha)$. Dacă se ține cont de semnificația testului, regula de decizie este următoarea: <ul style="list-style-type: none"> - dacă $Sigt \geq \alpha$, nu se respinge ipoteza nulă (H_0); - dacă $Sigt < \alpha$, se respinge H_0, cu probabilitatea $(1 - \alpha)$. 	
7. Luarea deciziei		

8. Interpretarea rezultatului		
-------------------------------	--	--

3. Estimarea și testarea indicatorilor de corelație

Observație: Pentru această etapă, ne folosim fie de rezultatele:

- din tabelul **Coefficients** (coloana Standardized Coefficients) pentru a estima coeficientul de corelație, fie de rezultatele din tabelul **Correlations** pentru a estima și testa coeficientul de corelație
- din tabelul **Model Summary** sau tabelul **Anova** pentru a estima și testa raportul de determinație și raportul de corelație

3.1. Estimarea indicatorilor de corelație

Indicatori de corelație	Coeficientul de corelație	Raportul (coeficientul) de determinație	Raportul de corelație
Definiție	măsoară intensitatea și indică sensul legăturii dintre două variabile.	măsoară cât din variația totală a variabilei dependente este explicat de modelul de regresie.	măsoară intensitatea legăturii dintre două variabile.
Parametru	$\rho = \beta_1 \sqrt{\frac{V(X)}{V(Y)}}$	$\eta^2 = \frac{V_E}{V_T} = 1 - \frac{V_R}{V_T}$ $V_T = V_E + V_R$	$\eta = \sqrt{\eta^2}$
Condiție	$-1 \leq \rho \leq 1$	$0 \leq \eta^2 \leq 1$	$0 \leq \eta \leq 1$
Estimator	$\hat{\rho} = \hat{\beta}_1 \sqrt{\frac{V(X)}{\hat{V}(Y)}}$	$\hat{\eta}^2 = \frac{\hat{V}_E}{\hat{V}_T} = 1 - \frac{\hat{V}_R}{\hat{V}_T}$ $\hat{V}_T \sim \chi^2(n-1),$ $\hat{V}_E \sim \chi^2(k-1),$ $\hat{V}_R \sim \chi^2(n-k),$	$\hat{\eta} = \sqrt{\frac{\hat{V}_E}{\hat{V}_T}} = \sqrt{1 - \frac{\hat{V}_R}{\hat{V}_T}}$
Estimație	$r = b_1 \sqrt{\frac{s_x^2}{s_y^2}}$	$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$ $TSS = ESS + RSS$	$R = \sqrt{R^2}$
Condiție	$-1 \leq r \leq 1$	$0 < R^2 < 1$	$0 \leq R \leq 1$
Observații	$r = \tilde{b}_1$	$r^2 = R^2$	$ r = R$

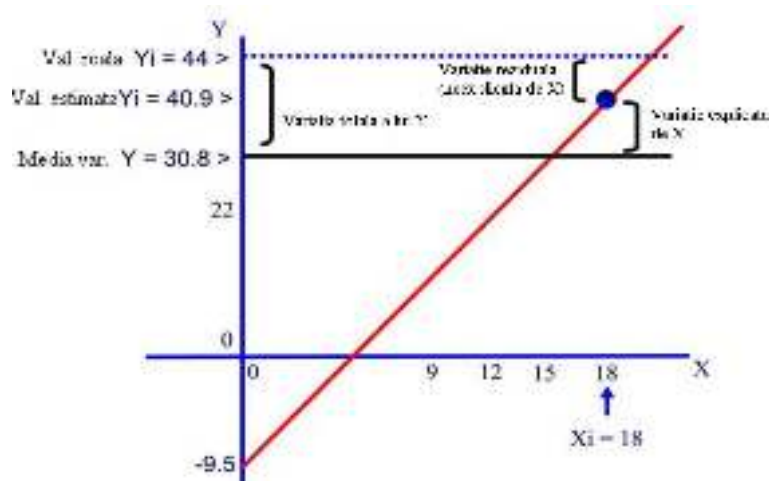


Figura 2. Componentele variației

3.2. Estimarea și testarea coeficientului de corelație

Correlations

		Happiness	Income	
Happiness	Pearson Correlation	1	.866**	1
	Sig. (2-tailed)		.000	2
	N	498	498	3
Income	Pearson Correlation	.866**	1	
	Sig. (2-tailed)	.000		
	N	498	498	

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

Legenda tabelului <i>Correlations</i>	Notății și rezultate obținute la nivelul eșantionului
1. Estimația coeficientului de corelație Pearson	$r =$
2. Probabilitatea asociată statisticii test Student (semnificația testului)	$Sig\ t =$
3. Volumul eșantionului	$n =$
Observație (doar în cazul regresiei liniare simple)	$r = \tilde{b}_1 =$

Pe baza rezultatelor obținute privind corelația dintre scorul fericirii și venit, se cere:

- 1) Să se interpreteze estimația coeficientului de corelație.
- 2) Să se testeze legătura dintre cele două variabile.

- 1) Să se interpreteze estimația coeficientului de corelație.

$$r = -0,866$$

Interpretare:

- în funcție de **semnul** coeficientului de corelație:

- în funcție de **valoarea în modul** a coeficientului de corelație ($|r| = 0,866$):

$ r = 0$	$0 \leftarrow r $	$ r \rightarrow 0,5 \leftarrow r $	$ r \rightarrow 1$	$ r = 1$
nu există o leg. liniară între Y și X	leg. liniară de intensitate slabă între Y și X	leg. liniară de intensitate moderată între Y și X	leg. liniară de intensitate puternică între Y și X	leg. liniară perfectă între Y și X

2) Să se testeze legătura dintre cele două variabile.

Etapete testării	Testarea coeficientului de corelație ρ
1. Formularea ipotezelor	<p>$H_0: \rho = 0$ (coeficientul de corelație ρ nu diferă semnificativ de 0, ceea ce înseamnă între cele două variabile nu există o legătură liniară semnificativă SAU cele două variabile nu sunt corelate semnificativ)</p> <p>$H_1: \rho \neq 0$ (coeficientul de corelație ρ diferă semnificativ de 0, ceea ce înseamnă între cele două variabile există o legătură liniară semnificativă SAU cele două variabile sunt corelate semnificativ)</p>
2. Alegerea pragului de semnificație	$\alpha = 0,05$
3. Alegerea statisticii test	$t = \frac{\hat{\rho}}{\sqrt{\frac{1 - \hat{\rho}^2}{n - 2}}} \sim t(n - 2)$
4. Determinarea valorii teoretice a statisticii test	$t_{teoretic} = t_{\alpha/2; n-2}$
5. Determinarea valorii calculate a statisticii test	$t_{calc} = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}} =$
6. Regula de decizie	<p>Dacă se ține cont de valoarea calculată a testului, regula de decizie este următoarea:</p> <ul style="list-style-type: none"> - dacă $t_{calc} \leq t_{\alpha/2; n-2}$, nu se respinge ipoteza nulă (H_0); - dacă $t_{calc} > t_{\alpha/2; n-2}$, se respinge ipoteza nulă (H_0), cu probabilitatea $(1 - \alpha)$. <p>Dacă se ține cont de semnificația testului, regula de decizie este următoarea:</p> <ul style="list-style-type: none"> - dacă $Sigt \geq \alpha$, nu se respinge ipoteza nulă (H_0); - dacă $Sigt < \alpha$, se respinge H_0, cu probabilitatea $(1 - \alpha)$.
7. Luarea deciziei	
8. Interpretarea deciziei luate	

3.3. Estimarea și testarea raportului de determinație și a raportului de corelație

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.866 ^a	.749	.749	.7181004959

a. Predictors: (Constant), Income

Legenda tabelului <i>Model Summary</i>	Notații și rezultate obținute la nivelul eșantionului
1. Estimația raportului de corelație	$R = \sqrt{R^2} =$
2. Estimația raportului de determinație	$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} =$
Observația 1 (doar în cazul regresiei liniare simple)	$ r = R =$ și $r^2 = R^2 =$
Observația 2: indicatorii de corelație se pot calcula și pe baza componentelor variației (ale căror valori estimate le găsim în Tabelul Anova)	

Pe baza rezultatelor obținute privind indicatorii de corelație, se cere:

- 1) Să se interpreteze valoarea estimată a raportului de determinație.
- 2) Să se interpreteze valoarea estimată a raportului de corelație.
- 3) Să se testeze semnificația raportului de corelație și de determinație, urmând toți pașii demersului testării.

- 1) Să se interpreteze valoarea estimată a raportului de determinație.

$$R^2 =$$

Interpretare:

- 2) Să se interpreteze valoarea estimată a raportului de corelație.

$$R = \sqrt{R^2} =$$

Interpretare:

$R = 0$	$0 \leftarrow R$	$R \rightarrow 0,5 \leftarrow R$	$R \rightarrow 1$	$R = 1$
<i>nu</i> există o leg. liniară	leg. liniară de intensitate <i>slabă</i> între Y și X	leg. liniară de intensitate <i>moderată</i> între Y și X	leg. liniară de intensitate <i>puternică</i> între Y și X	leg. liniară perfectă

3.5. Testarea modelului de regresie

			1	2	3	4	5	
				ANOVA ^a				
			Model	Sum of Squares	df	Mean Square	F	Sig.
1.1; 2.1; 3.1	1	Regression	764.546	1	764.546	1482.632		.000 ^b
		Residual	255.771	496	.516			
1.3; 2.3		Total	1020.318	497				
								1.2; 2.2; 3.2

a. Dependent Variable: Happiness

b. Predictors: (Constant), Income

Legenda tabelului ANOVA	Notații și rezultate obținute la nivelul eșantionului
1. Estimațiile componentelor variației 1.1. Estimația variației explicate a modelului 1.2. Estimația variației reziduale a modelului 1.3. Estimația variației totale a modelului	$TSS = ESS + RSS$ $ESS =$ $RSS =$ $TSS =$
2. Gradele de libertate corespunzătoare fiecărei componente a variației 2.1. Variația explicată a modelului 2.2. Variația reziduală a modelului 2.3. Variația totală a modelului	$(k - 1) = \quad = 1$ $(n - k) = \quad = 40$ $(n - 1) = \quad = 41$
3. Estimația dispersiei explicate și reziduale (raportul dintre estimația variației și numărul corespunzător de grade de libertate) 3.1. Pentru variația explicată a modelului 3.2. Pentru variația reziduală a modelului	$ESS/(k - 1) =$ $RSS/(n - k) =$
4. Valoarea calculată a statisticii test Fisher (pe baza componentelor variației)	$F_{calc} = \frac{ESS}{RSS} \cdot \frac{(n - k)}{(k - 1)} =$
4. Valoarea calculată a statisticii test Fisher (pe baza raportului de determinație)	$F_{calc} = \frac{R^2}{(1 - R^2)} \cdot \frac{(n - k)}{(k - 1)} =$
5. Probabilitatea asociată valorii calculate a statisticii test Fisher	$Sig F =$

Pe baza rezultatelor din tabelul Anova, se cere:

- 1) Să se estimeze punctual indicatorii de corelație (coeficientul de corelație; raportul de corelație, raportul de determinație).

$$R^2 = \frac{ESS}{TSS} =$$

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

$$R = \sqrt{R^2} =$$

- 2) Să se testeze dacă modelul de regresie construit este corect specificat, parcurgând toate etapele demersului testării.

Etapele testării	Testarea modelului de regresie	Testarea raportului de determinație η^2 (sau raportului de corelație η)
1. Formularea ipotezelor	$H_0: \beta_0 = 0, \beta_1 = 0$ (modelul de regresie nu explică semnificativ dependența liniară dintre cele două variabile SAU între cele două variabile nu există o legătură liniară semnificativă SAU modelul de regresie construit nu este corect specificat) $H_1: \beta_1 \neq 0$ (modelul de regresie explică semnificativ dependența liniară dintre cele două variabile SAU între cele două variabile există o legătură liniară semnificativă)	$H_0: \eta = 0$ (raportul de determinație η^2 sau raportul de corelație η nu diferă semnificativ de 0, ceea ce înseamnă că între cele două variabile nu există o legătură liniară semnificativă) $H_1: \eta > 0$ (raportul de determinație η^2 sau raportul de corelație η este semnificativ mai mare decât 0, ceea ce înseamnă că între cele două variabile există o legătură liniară semnificativă)
2. Alegerea pragului de semnificație	$\alpha = 0,05$	$\alpha = 0,05$
3. Alegerea statisticii test	$F = \frac{\frac{\hat{V}_E}{k-1}}{\frac{\hat{V}_R}{n-k}} \sim F(k-1; n-k)$	$F = \frac{\hat{\eta}^2}{(1-\hat{\eta}^2)} \cdot \frac{n-k}{k-1} \sim F(k-1; n-k)$
4. Determinarea valorii teoretice a statisticii test	$F_{teoretic} = F_{\alpha; k-1; n-k}$	$F_{teoretic} = F_{\alpha; k-1; n-k}$
5. Determinarea valorii calculate a statisticii test	$F_{calc} = \frac{\frac{ESS}{k-1}}{\frac{RSS}{n-k}} = \frac{ESS}{RSS} \cdot \frac{n-k}{k-1}$ $TSS = ESS + RSS$	$F_{calc} = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1}$ $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$

6. Regula de decizie	<p>Dacă se ține cont de valoarea calculată a testului, regula de decizie este următoarea:</p> <ul style="list-style-type: none"> - dacă $F_{calc} \leq F_{\alpha; k-1; n-k}$, nu se respinge ipoteza nulă (H_0); - dacă $F_{calc} > F_{\alpha; k-1; n-k}$, se respinge ipoteza nulă (H_0), cu probabilitatea $(1 - \alpha)$. <p>Dacă se ține cont de semnificația testului, regula de decizie este următoarea:</p> <ul style="list-style-type: none"> - dacă $SigF \geq \alpha$, nu se respinge ipoteza nulă (H_0); - dacă $SigF < \alpha$, se respinge H_0, cu probabilitatea $(1 - \alpha)$.
7. Luarea deciziei	
8. Interpretarea deciziei luate	