Capitolul 2. Modelul de regresie liniară simplă
Aplicaţii

I. Se consideră datele cu privire la *nivelul studiilor* (ani) şi *venitul* (lei) pentru un eşantion de 5 angajaţi. 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;
   Variabila dependentă (Y) este *venitul*, iar variabila independentă (X) este *nivelul studiilor*.

2) Să se reprezinte grafic legătura dintre cele două variabile;

3) Să se explice legătura de dependență dintre cele două variabile;
   Se consideră că *venitul* depinde într-o oarecare măsură de *nivelul studiilor*, adică dacă *nivelul studiilor* crește, ne așteptăm ca și *venitul* să fie mai ridicat.

4) Să se aleagă forma modelului de regresie;
   Pentru analiza acestei legături se poate utiliza un model de regresie liniară simplă, după expresia:

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

5) Să se estimeze punctual parametrii modelului de regresie;

i	x_i	y_i	$\boldsymbol{x_i y_i}$	$\boldsymbol{x_i^2}$	$n = 5$
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**	$\bar{x} = 12{,}8$

Estimarea punctuală:

- *a parametrului* β_1 *(a pantei dreptei de regresie)*

$$b_1 = \frac{n\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n\sum_i x_i^2 - (\sum_i x_i)^2} = \frac{5 \cdot 85600 - 64 \cdot 6400}{5 \cdot 840 - (64)^2} = 176{,}923$$

- *a parametrului β_0 (a constantei sau a ordonatei la origine)*

$$b_0 = \frac{\sum_i y_i \sum_i x_i^2 - \sum_i x_i \sum_i x_i y_i}{n\sum_i x_i^2 - (\sum_i x_i)^2} = \frac{6400 \cdot 840 - 64 \cdot 85600}{5 \cdot 840 - (64)^2} = -984{,}614$$

$$b_0 = \bar{y} - b_1\bar{x} = 1280 - 176{,}923 \cdot 12{,}8 = -984{,}614$$

6) Să se interpreteze estimaţiile parametrilor modelului de regresie;

Interpretare:

- *estimația parametrului β_0 (a constantei sau a ordonatei la origine)*
  $b_0 = -984{,}614$ lei: *venitul mediu estimat* atunci când nivelul studiilor este 0.
- *estimația parametrului β_1 (a pantei dreptei de regresie)*
  $b_1 = 176{,}923$ lei: arată că *venitul va crește, în medie, cu 176,923 de lei*, dacă nivelul studiilor va crește cu 1 an.

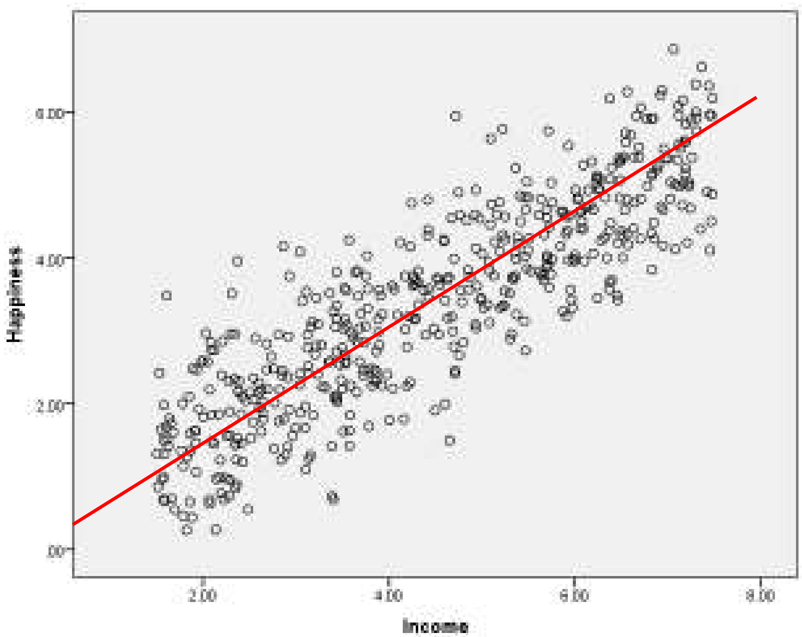
7) Să se scrie ecuaţia estimată a modelului de regresie.

$$y_{x_i} = b_0 + b_1 x_i = -984{,}614 + 176{,}923 x_i$$

$$Y_X = b_0 + b_1 X = -984{,}614 + 176{,}923 X$$

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.

   Variabila dependentă (Y) este *scorul fericirii*, iar variabila independentă (X) este *venitul*. Identificarea celor două variabile se bazează pe analiza logică a legăturii funcționale dintre ele, dar are și un corespondent teoretic/empiric în studiile din literatura de specialitate.

2) Să se explice legătura de dependență din cele două variabile.

   Se consideră că scorul fericirii depinde într-o oarecare măsură de nivelul venitului, adică dacă nivelul venitului crește, ne așteptăm ca și nivelul fericirii să fie mai ridicat.

3) Să se identifice forma și sensul legăturii dintre cele două variabile.

   În funcție de reprezentarea grafică, se poate aprecia că norul de puncte poate fi ajustat printr-o dreaptă, ceea ce înseamnă că legătura dintre *scorul fericirii* și *venit* este de tip *liniar*, după natura dependenței, și *directă*, după sensul acesteia.

4) Să se aleagă forma modelului de regresie.

   Pentru analiza acestei legături se poate utiliza un model de regresie liniară simplă, după expresia:

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

2. Estimarea şi testarea parametrilor modelului de regresie

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)	.204	.089		2.299	.022	.030	.379
	Income	.714	.019	.866	38.505	.000	.677	.750

a. Dependent Variable: Happiness

Legenda tabelului *Coefficients*	Notaţii şi rezultate obţinute pe baza datelor de la nivelul eşantionului	
	Constant ($\boldsymbol{\beta_0}$)	**Income ($\boldsymbol{\beta_1}$)**
1. Coeficienţi nestandardizaţi (coloana *Unstandardized Coefficients*) 1.1. Estimaţiile parametrilor modelului de regresie (coloana *B*) 1.2. Estimaţiile erorilor standard ale estimatorilor parametrilor modelului de regresie (coloana *Std. Error*)	$b_0 = 0{,}204$ $s_{\hat{\beta}_0} = 0{,}089$	$b_1 = 0{,}714$ $s_{\hat{\beta}_1} = 0{,}019$
2. Coeficienţi standardizaţi (coloana *Standardized Coefficients - Beta*)		$\tilde{b}_1 = 0{,}866$
3. Valoarea calculată a statisticii test Student (coloana t)	$t_{calc} = \frac{b_0}{s_{\hat{\beta}_0}} = 2{,}299$	$t_{calc} = \frac{b_1}{s_{\hat{\beta}_1}} = 38{,}505$
4. Probabilitatea asociată statisticii test Student (semnificaţia testului)	$Sig\ t = 0{,}022$	$Sig\ t = 0{,}000$
5. Intervalul de încredere al parametrilor modelului de regresie (coloana *95% Confidence Interval for B*) 5.1. Limita inferioară a intervalului (coloana *Lower Bound*) 5.2. Limita superioară a intervalului (coloana *Upper Bound*)	$IC(\beta_0)$: $[0{,}030; 0{,}379]$ $b_0 - t_{\alpha/2;n-2} \cdot s_{\hat{\beta}_0}$ $b_0 + t_{\alpha/2;n-2} \cdot s_{\hat{\beta}_0}$	$IC(\beta_1)$: $[0{,}677; 0{,}750]$ $b_1 - t_{\alpha/2;n-2} \cdot s_{\hat{\beta}_1}$ $b_1 + t_{\alpha/2;n-2} \cdot s_{\hat{\beta}_1}$

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

5) Să se scrie ecuaţia estimată a modelului de regresie.

Regresia estimată, scrisă pentru fiecare valoare, respectiv pentru toate valorile variabilelor:
$y_{x_i} = b_0 + b_1 x_i = 0{,}204 + 0{,}714 x_i$
$Y_X = b_0 + b_1 X = 0{,}204 + 0{,}714 X$

6) Să se interpreteze estimaţiile parametrilor modelului de regresie.

$b_0 = 0{,}204$ puncte: *scorul mediu estimat al fericirii* atunci când venitul este 0.

$b_1 = 0{,}714$ puncte: la o creștere a venitului cu o unitate (10 mii de dolari), scorul fericirii *crește, în medie, cu 0,714 puncte*.
Observație: Deoarece $b_1 > 0$ se confirmă că legătura dintre cele două variabile este una directă.

7) Să se interpreteze intervalele de încredere ale coeficienților de regresie, considerând o probabilitate de 95%.

Interpretarea intervalului de încredere pentru cei doi parametri:

$IC(\beta_0): \left[b_0 - t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_0}; b_0 + t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_0}\right]$
$IC(\beta_0): [0{,}030; 0{,}379]$
Interpretare: Cu o probabilitate de 95%, se poate garanta că parametrul β_0 (sau ordonata la origine) este acoperit(ă) de intervalul $[0{,}030; 0{,}379]$.

$IC(\beta_1): \left[b_1 - t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_1}; b_1 + t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_1}\right]$
$IC(\beta_1): [0{,}677; 0{,}750]$
Interpretare: Cu o probabilitate de 95%, se poate garanta că intervalul $[0{,}677; 0{,}750]$ acoperă parametrul β_1 (sau panta dreptei de regresie).

8) Să se estimeze prin interval de încredere ordonata la origine și panta dreptei de regresie, pentru o probabilitate de 0,99.

$IC(\beta_0): \left[b_0 \pm t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_0}\right]$ $(1-\alpha) = 99\%$
$t_{\alpha/2;n-2} = t_{0{,}01/2;498-2} = t_{0{,}005;496} = 2{,}576$
$IC(\beta_0): \left[b_0 - t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_0}; b_0 + t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_0}\right]$
$IC(\beta_0): [0{,}204 - 2{,}576 \cdot 0{,}089; 0{,}204 + 2{,}576 \cdot 0{,}089]$
$IC(\beta_0): [0{,}025; 0{,}433]$
Interpretare: Cu o probabilitate de 99%, se poate garanta că ordonata la origine este acoperită de intervalul $[0{,}025; 0{,}433]$.

$IC(\beta_1): \left[b_1 \pm t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_1}\right]$ $(1-\alpha) = 99\%$
$t_{\alpha/2;n-2} = t_{0{,}01/2;498-2} = t_{0{,}005;496} = 2{,}576$
$IC(\beta_1): \left[b_1 - t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_1}; b_1 + t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_1}\right]$
$IC(\beta_1): [0{,}714 - 2{,}576 \cdot 0{,}019; 0{,}714 + 2{,}576 \cdot 0{,}019]$
$IC(\beta_1): [0{,}665; 0{,}763]$
Interpretare: Cu o probabilitate de 99%, se poate garanta că panta dreptei de regresie este acoperită de intervalul $[0{,}665; 0{,}763]$.

Pe baza modelului estimat, se cere:

10) Să se estimeze scorul fericirii (y_{x_i}) pentru un individ pentru care venitul (x_i) este de 3 unități (30 de mii de dolari).

$$y_{x_i} = 0{,}204 + 0{,}714x_i = 0{,}204 + 0{,}714 \cdot 3 = 2{,}346 \text{ puncte}$$

Interpretare: Pentru un venit de 30 de mii de dolari, ne așteptăm ca scorul fericirii unui individ să fie de 2,346 puncte.

11) Să se precizeze cât ar trebui să fie nivelul venitului (x_i) unui individ pentru a obține un scor al fericirii (y_{x_i}) de 4 puncte.

$$y_{x_i} = 0{,}204 + 0{,}714x_i \Rightarrow 4 = 0{,}204 + 0{,}714x_i \Rightarrow$$
$$x_i = \frac{4 - 0{,}204}{0{,}714} = 5{,}316$$

Interpretare: Pentru a obține un scor al fericirii de 4 puncte, nivelul venitului unui individ ar trebui să fie de 53,16 mii de dolari.

12) Să se estimeze cu cât scade scorul fericirii ($\Delta(Y)$) dacă are loc o scădere a venitului ($\Delta(X)$) cu 0,5 unități (5000 de dolari).

$$b_1 = \frac{\Delta Y}{\Delta X} \Rightarrow 0{,}714 = \frac{\Delta Y}{-0{,}5} \Rightarrow \Delta Y = -0{,}357$$

Interpretare: La o scădere a venitului cu 5000 de dolari, scorul fericirii scade, în medie, cu 0,357 puncte.

13) Să se determine valoarea cu care ar trebui să crească nivelul venitului ($\Delta(X)$) pentru a avea o creștere medie a scorului fericirii ($\Delta(Y)$) cu 1,5 puncte.

$$b_1 = \frac{\Delta Y}{\Delta X} \Rightarrow 0{,}714 = \frac{1{,}5}{\Delta X} \Rightarrow \Delta X = \frac{1{,}5}{0{,}714} = 2{,}1$$

Interpretare: Pentru o creștere, în medie, a scorului fericirii cu 1,5 puncte, nivelul venitului ar trebui să crească cu 21 de mii de dolari.

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

14) Să se testeze semnificaţia coeficienţilor de regresie, aplicând toţi paşii demersului testării.

Etapele testării	**Testarea parametrului $\boldsymbol{\beta_0}$**	**Testarea parametrului $\boldsymbol{\beta_1}$**
1. Formularea ipotezelor	$H_0: \beta_0 = 0$ (parametrul β_0 nu diferă semnificativ de 0 **SAU** constanta modelului nu este semnificativă statistic) $H_1: \beta_0 \neq 0$ (parametrul β_0 diferă semnificativ de 0 **SAU** constanta modelului este semnificativă 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 semnificativă **SAU** variabila independentă *X* nu are o influență semnificativă 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 semnificativă **SAU** variabila independentă *X* explică semnificativ variația variabilei dependente *Y*)
2. Alegerea pragului de semnificaţie	$\alpha = 0{,}05$	$\alpha = 0{,}05$
3. Alegerea statisticii test	$t = \dfrac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \sim t(n-2)$	$t = \dfrac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t(n-2)$
4. Determinarea valorii teoretice a statisticii test	$t_{teoretic} = t_{\alpha/2;\,n-2} =$ $t_{0{,}05/2;\,498-2} = t_{0{,}025;\,496} = 1{,}96$	$t_{teoretic} = t_{\alpha/2;\,n-2} =$ $t_{0{,}05/2;\,498-2} = t_{0{,}025;\,496} = 1{,}96$
5. Determinarea valorii calculate a statisticii test (în condițiile acceptării ipotezei nule)	$t_{calc} = \dfrac{b_0}{s_{\hat{\beta}_0}} = \dfrac{0{,}204}{0{,}089} = 2{,}299$	$t_{calc} = \dfrac{b_1}{s_{\hat{\beta}_1}} = \dfrac{0{,}714}{0{,}019} = 38{,}505$
6. Regula de decizie	Dacă se ţine cont de valoarea calculată a testului, regula de decizie este următoarea: - dacă $\lvert t_{calc}\rvert \leq t_{\alpha/2;\,n-2}$, nu se respinge ipoteza nulă (H_0); - dacă $\lvert t_{calc}\rvert > t_{\alpha/2;\,n-2}$, se respinge ipoteza nulă (H_0), cu probabilitatea $(1-\alpha)$. Dacă se ţine cont de semnificaţia testului, regula de decizie este următoarea: - 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	$\lvert t_{calc}\rvert = 2{,}299 > t_{\alpha/2;\,n-2} = 1{,}96$ **SAU**	$\lvert t_{calc}\rvert = 38{,}505 > t_{\alpha/2;\,n-2} = 1{,}96$ **SAU**

	$Sigt = 0{,}000 < \alpha = 0{,}05$ $\Rightarrow$ că se respinge ipoteza H_0 (95%)	$Sigt = 0{,}000 < \alpha = 0{,}05$ $\Rightarrow$ că se respinge ipoteza H_0 (95%)
8. Interpretarea rezultatului	Pentru o probabilitate de 95%, se poate afirma că parametrul β_0 diferă semnificativ de 0 **SAU** este semnificativ statistic.	Pentru o probabilitate de 95%, se poate afirma că parametrul β_1 diferă semnificativ de 0, **ceea ce înseamnă că** legătura liniară dintre scorul fericirii (Y) și venit (X) este semnificativă statistic **SAU** că venitul (X) are o influență semnificativă asupra scorului fericirii (Y) **SAU** că venitul (X) explică semnificativ variația scorului fericirii (Y)

3. Estimarea şi testarea indicatorilor de corelaţie

Observaţie: Pentru această etapă, ne folosim fie de rezultatele:
- din tabelul ***Coefficients*** (coloana Standardized Coefficients) pentru a estima coeficientul de corelaţie, fie de rezultatele din tabelul ***Correlations*** pentru a estima şi testa coeficientul de corelaţie
- din tabelul ***Model Summary*** sau tabelul ***Anova*** pentru a estima şi testa raportul de determinaţie şi raportul de corelaţie

3.1. Estimarea indicatorilor de corelaţie

Indicatori de corelaţie	**Coeficientul de corelaţie**	**Raportul (coeficientul) de determinaţie**	**Raportul de corelaţie**
Definiţie	măsoară intensitatea şi indică sensul legăturii dintre două variabile.	măsoară cât din variaţia totală a variabilei dependente este explicat de modelul de regresie.	măsoară intensitatea legăturii dintre două variabile.
Parametru	$\rho = \beta_1 \sqrt{\frac{V(X)}{V(Y)}}$	$\eta^2 = \frac{V_E}{V_T} = 1 - \frac{V_R}{V_T}$ $V_T = V_E + V_R$	$\eta = \sqrt{\eta^2}$
Condiţie	$-1 \le \rho \le 1$	$0 \le \eta^2 \le 1$	$0 \le \eta \le 1$
Estimator	$\hat{\rho} = \hat{\beta}_1 \sqrt{\frac{\hat{V}(X)}{\hat{V}(Y)}}$	$\hat{\eta}^2 = \frac{\hat{V}_E}{\hat{V}_T} = 1 - \frac{\hat{V}_R}{\hat{V}_T}$ $\hat{V}_T \sim \chi^2(n-1)$, $\hat{V}_E \sim \chi^2(k-1)$, $\hat{V}_R \sim \chi^2(n-k)$,	$\hat{\eta} = \sqrt{\frac{\hat{V}_E}{\hat{V}_T}} = \sqrt{1 - \frac{\hat{V}_R}{\hat{V}_T}}$
Estimaţie	$r = b_1 \sqrt{\frac{s_x^2}{s_y^2}}$	$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$ $TSS = ESS + RSS$	$R = \sqrt{R^2}$
Condiţie	$-1 \le r \le 1$	$0 < R^2 < 1$	$0 \le R \le 1$
Observaţii	$r = \tilde{b}_1$	$r^2 = R^2$	$\lvert r \rvert = R$

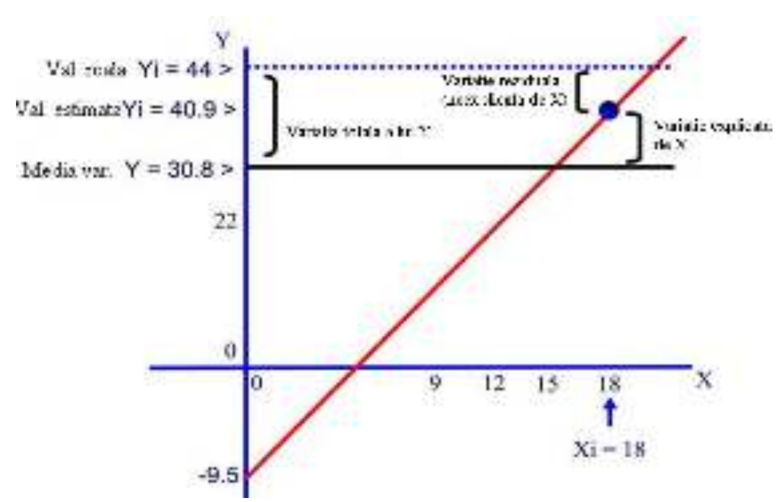


Figura 2. *Componentele variaţiei*

3.2. Estimarea şi testarea coeficientului de corelaţie

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 *Correlations*	**Notaţii şi rezultate obţinute la nivelul eşantionului**
1. Estimaţia coeficientului de corelaţie Pearson (linia *Pearson Correlation*)	$r = 0{,}866$
2. Probabilitatea asociată statisticii test Student - semnificaţia testului (linia *Sig. (2-tailed)*)	$Sig\ t = 0{,}000$
3. Volumul eşantionului (linia *N*)	$n = 498$
Observaţie *(doar în cazul regresiei liniare simple)*	$r = \tilde{b}_1 = 0{,}866$

Pe baza rezultatelor obţinute privind corelaţia dintre scorul fericirii şi venit, se cere:

1) Să se interpreteze estimaţia coeficientului de corelaţie.
2) Să se testeze legătura dintre cele două variabile.

1) Să se interpreteze estimaţia coeficientului de corelaţie.

$r = -0{,}866$

Interpretare:

- în funcție de ***semnul*** coeficientului de corelație: legătura liniară dintre scorul fericirii (Y) și venit (X) este inversă
- în funcție de ***valoarea în modul*** a coeficientului de corelație ($|r| = 0{,}866$), această legătură este de intensitate puternică

Interpretarea integrală:

- între scorul fericirii (Y) și venit (X) există o legătură liniară inversă și de intensitate puternică.

$\mathbf{|r| = 0}$	$\mathbf{0 \leftarrow |r|}$	$\mathbf{|r| \rightarrow\ 0,5\ \leftarrow |r|}$	$\mathbf{|r| \rightarrow 1}$	$\mathbf{|r| = 1}$
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.

Etapele testării	**Testarea coeficientului de corelație ρ**
1. Formularea ipotezelor	$H_0: \rho = 0$ (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) $H_1: \rho \neq 0$ (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)
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	$t_{teoretic} = t_{\alpha/2;\, n-2} =$ $t_{0{,}05/2;\, 498-2} = t_{0{,}025;\, 496} = 1{,}96$
5. Determinarea valorii calculate a statisticii test	$t_{calc} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0{,}866}{\sqrt{\frac{1-0{,}866^2}{498-2}}} = \frac{0{,}866}{\sqrt{\frac{1-0{,}749}{496}}} = \frac{0{,}866}{0{,}022} = 39{,}363$
6. Regula de decizie	Dacă se ţine cont de valoarea calculată a testului, regula de decizie este următoarea: - dacă $\lvert t_{calc}\rvert \leq t_{\alpha/2;\, n-2}$, nu se respinge ipoteza nulă (H_0); - dacă $\lvert t_{calc}\rvert > t_{\alpha/2;\, n-2}$, se respinge ipoteza nulă (H_0), cu probabilitatea $(1-\alpha)$. Dacă se ţine cont de semnificaţia testului, regula de decizie este următoarea: - 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	$\lvert t_{calc}\rvert = 39{,}363 > t_{\alpha/2;\, n-2} = 1{,}96 \Rightarrow$ că se respinge ipoteza H_0 (95%) **SAU** $Sigt = 0{,}000 < \alpha = 0{,}05 \Rightarrow$ că se respinge ipoteza H_0 (95%)
8. Interpretarea deciziei luate	Cu o probabilitate de 95%, se poate garanta că între scorul fericirii și venit există o legătură liniară semnificativă **SAU** scorul fericirii și venitul sunt corelate semnificativ.

3.3. Estimarea şi testarea raportului de determinaţie şi a raportului de corelaţie

Model Summary

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

a. Predictors: (Constant), Income

Legenda tabelului *Model Summary*	**Notaţii şi rezultate obţinute la nivelul eşantionului**
1. Estimaţia raportului de corelaţie (coloana R)	$R = \sqrt{R^2} = 0{,}866$
2. Estimaţia raportului de determinaţie (coloana $R\ Square$)	$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 0{,}749$
Observaţia 1 *(doar în cazul regresiei liniare simple)*	$\lvert r \rvert = R = 0{,}866$ şi $r^2 = R^2 = 0{,}749$
Observaţia 2: indicatorii de corelaţie se pot calcula şi pe baza componentelor variaţiei (ale căror valori estimate le găsim în Tabelul Anova)	

Pe baza rezultatelor obţinute privind indicatorii de corelaţie, se cere:
1) Să se interpreteze valoarea estimată a raportului de determinaţie.
2) Să se interpreteze valoarea estimată a raportului de corelaţie.
3) Să se testeze semnificaţia raportului de corelaţie şi de determinaţie, urmând toţi paşii demersului testării.

1) Să se interpreteze valoarea estimată a raportului de determinaţie.

$R^2 = 0{,}749$

Interpretare:
- 74,9% din variația totală a scorului fericirii este explicată de variația venitului. Iar restul de 25,1%% (diferența până la 100%) din variația totală a scorului fericirii este explicată de influența factorilor aleatori sau neincluși în model.

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

$R = \sqrt{R^2} = 0{,}866$

Interpretare:
- între scorul fericirii (Y) și venit (X) există o legătură liniară de intensitate puternică.

$\boldsymbol{R = 0}$	$\boldsymbol{0 \leftarrow R}$	$\boldsymbol{R \rightarrow\ 0{,}5\ \leftarrow R}$	$\boldsymbol{R \rightarrow 1}$	$\boldsymbol{R = 1}$
nu există o leg. liniară	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

3.5. Testarea modelului de regresie

ANOVA[a]

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

a. Dependent Variable: Happiness

b. Predictors: (Constant), Income

Legenda tabelului *ANOVA*	**Notaţii şi rezultate obţinute la nivelul eşantionului**
1. Estimaţiile componentelor variaţiei 1.1. Estimaţia variaţiei explicate a modelului (linia *Regression* şi coloana *Sum of Square*) 1.2. Estimaţia variaţiei reziduale a modelului (linia *Residual* şi coloana *Sum of Square*) 1.3. Estimaţia variaţiei totale a modelului (linia *Total* şi coloana *Sum of Square*)	$TSS = ESS + RSS$ $ESS = 764{,}546$ $RSS = 255{,}771$ $TSS = 1020{,}318$
2. Gradele de libertate corespunzătoare fiecărei componente a variaţiei 2.1. Variaţia explicată a modelului (linia *Regression* şi coloana *df*) 2.2. Variaţia reziduală a modelului (linia *Residual* şi coloana *df*) 2.3. Variaţia totală a modelului (linia *Total* şi coloana *df*)	$(k-1) = 2 - 1 = 1$ $(n-k) = 498 - 2 = 496$ $(n-1) = 498 - 1 = 497$
3. Estimaţia dispersiei explicate şi reziduale (raportul dintre estimaţia variaţiei şi numărul corespunzător de grade de libertate) 3.1. Pentru variaţia explicată a modelului (linia *Regression* şi coloana *Mean Square*) 3.2. Pentru variaţia reziduală a modelului (linia *Residual* şi coloana *Mean Square*)	$ESS/(k-1) = 764{,}546$ $RSS/(n-k) = 0{,}516$
4. Valoarea calculată a statisticii test Fisher - pe baza componentelor variaţiei (coloana *F*)	$F_{calc} = \frac{ESS}{RSS} \cdot \frac{(n-k)}{(k-1)} = 1482{,}632$
4. Valoarea calculată a statisticii test Fisher - pe baza raportului de determinaţie)	$F_{calc} = \frac{R^2}{(1-R^2)} \cdot \frac{(n-k)}{(k-1)} = 1482{,}632$
5. Probabilitatea asociată valorii calculate a statisticii test Fisher (coloana *Sig.*)	$Sig\ F = 0{,}000$

Pe baza rezultatelor din tabelul Anova, se cere:

1) Să se estimeze punctual indicatorii de corelaţie (coeficientul de corelaţie; raportul de corelaţie, raportul de determinaţie).

Estimarea raportului de determinație

$$R^2 = \frac{ESS}{TSS} = \frac{764,546}{1020,318} = 0,749$$

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{255,771}{1020,318} = 1 - 0,251 = 0,749$$

- în cazul regresiei liniare simple, raportul de determinație se poate determina și pe baza coeficientului de corelație astfel:

  $$R^2 = r^2 = 0,866^2 = 0,749$$

Estimarea raportului de corelație

$$R = \sqrt{R^2} = \sqrt{0,749} = 0,866$$

- în cazul regresiei liniare simple, raportul de corelație se poate determina și pe baza coeficientului de corelație astfel:

  $$R = |r| = |0,866| = 0,866$$

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 corelația 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 corelația η este semnificativ mai mare decât 0, **ceea ce înseamnă** că între cele două variabile există o legătură liniară semnificativă)
2. Alegerea pragului de semnificaţie	$\alpha = 0{,}05$	$\alpha = 0{,}05$
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	$F_{teoretic} = F_{\alpha;\, k-1;\, n-k} = F_{0,05;\, 1;\, 496} = 3{,}842$	$F_{teoretic} = F_{\alpha;\, k-1;\, n-k} = F_{0,05;\, 1;\, 496} = 3{,}842$
5. Determinarea valorii calculate a statisticii test	$F_{calc} = \frac{ESS}{RSS} \cdot \frac{n-k}{k-1} = \frac{764{,}546}{255{,}771} \cdot \frac{496}{1}$ $F_{calc} = \frac{\frac{ESS}{k-1}}{\frac{RSS}{n-k}} = \frac{764{,}546}{0{,}516}$ $F_{calc} = 1482{,}632$ $TSS = ESS + RSS$	$F_{calc} = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1}$ $F_{calc} = \frac{0{,}749}{1-0{,}749} \cdot \frac{496}{1}$ $F_{calc} = 1482{,}632$ $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$
6. Regula de decizie	Dacă se ţine cont de valoarea calculată a testului, regula de decizie este următoarea: - 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 $(1-\alpha)$.	

	Dacă se ţine cont de semnificaţia testului, regula de decizie este următoarea: - 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	$F_{calc} = 1482{,}632 > F_{\alpha;\, k-1;\, n-k} = 3{,}842$ ⇒ că se respinge ipoteza H_0 (95%) **SAU** $SigF = 0{,}000 < \alpha = 0{,}05$ ⇒ că se respinge ipoteza H_0 (95%)
8. Interpretarea deciziei luate	Cu o probabilitate de 95%, se poate garanta că între scorul fericirii și venit există o legătură liniară semnificativă **SAU** modelul de regresie explică semnificativ dependența liniară dintre scorul fericirii și venit **SAU** modelul de regresie liniară simplă ales este corect specificat.