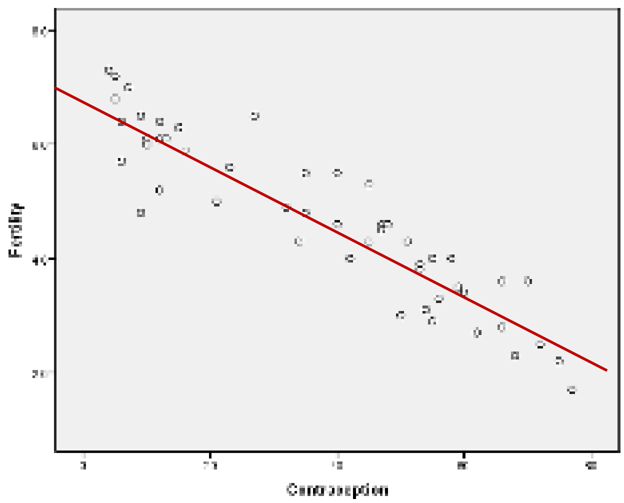
Capitolul 2. Modelul de regresie liniară simplă

Se consideră rata totală a fertilității (copii născuți de o femeie de-a lungul vieții) și procentul femeilor care folosesc metode contraceptive înregistrate pentru 50 de țări.



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.920[a]	.847	.844	.5745

a. Predictors: (Constant), Contraceptors

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	87.672	1	87.672	265.668	.000[b]
	Residual	15.840	48	.330		
	Total	103.513	49			

a. Dependent Variable: Fertility
b. Predictors: (Constant), Contraceptors

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.875	.157		43.829	.000
	Contraceptors	-.058	.004	-.920	-16.299	.000

a. Dependent Variable: Fertility

Correlations

		Fertility	Contraceptors
Pearson Correlation	Fertility	1.000	-.920
	Contraceptors	-.920	1.000
Sig. (1-tailed)	Fertility	.	.000
	Contraceptors	.000	.
N	Fertility	50	50
	Contraceptors	50	50

Pe baza rezultatelor modelării econometrice, se cere:

1. Considerând reprezentarea grafică de mai sus pentru cele două variabile, analizați ce se poate spune despre o posibilă legătură dintre cele două variabile.

În funcție de reprezentarea grafică, se poate aprecia că forma norului de puncte poate fi aproximată printr-o dreaptă, ceea ce înseamnă că legătura dintre cele două variabile este liniară, iar după sens, inversă.

2. Să se scrie ecuaţia estimată a modelului de regresie (atât pentru toate valorile variabilelor, cât şi pentru fiecare valoare a acestora).

$Y_X = b_0 + b_1 X = 6{,}875 - 0{,}058X$

$y_{x_i} = b_0 + b_1 x_i = 6{,}875 - 0{,}058x_i$

3. Să se interpreteze estimaţiile coeficienţilor de regresie.

$b_0 = 6{,}875$ copii născuți de o femeie de-a lungul vieții: *nivelul mediu estimat* al ratei totale a fertilității (Y) atunci când procentul femeilor care folosesc metode contraceptive (X) ia valoarea zero.

$b_1 = -0{,}058$ copii născuți de o femeie de-a lungul vieții: la o creștere a procentului femeilor care folosesc metode contraceptive (X) cu 1 procent, rata totală a fertilității (Y) *scade, în medie, cu 0,058 copii născuți de o femeie de-a lungul vieții.*

4. Pentru o probabilitate de 90%, estimaţi prin interval de încredere ordonata la origine a modelului şi interpretaţi rezultatul.

$IC(\beta_0): \left[b_0 \pm t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_0}\right]$ $(1-\alpha) = 90\%$

$t_{\alpha/2;n-2} = t_{0,10/2;50-2} = t_{0,05;48} = 1{,}645$

$IC(\beta_0): \left[b_0 - t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_0}; b_0 + t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_0}\right]$

$IC(\beta_0): [6{,}875 - 1{,}645 \cdot 0{,}157; 6{,}875 + 1{,}645 \cdot 0{,}157]$

$IC(\beta_0): [6{,}617; 7{,}133]$

Interpretare: Cu o probabilitate de 90%, se poate garanta că ordonata la origine este acoperită de intervalul $[6{,}617; 7{,}133]$.

5. Considerând o probabilitate de 95%, estimaţi prin interval de încredere panta dreptei de regresie şi interpretaţi rezultatul.

$IC(\beta_1): \left[b_1 \pm t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_1}\right]$ $(1-\alpha) = 95\%$

$t_{\alpha/2;n-2} = t_{0,05/2;50-2} = t_{0,025;48} = 1{,}96$

$IC(\beta_1): \left[b_1 - t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_1}; b_1 + t_{\alpha/2;n-2} \cdot s_{\widehat{\beta}_1}\right]$

$IC(\beta_1): [-0{,}058 - 1{,}96 \cdot 0{,}004; -0{,}058 - 1{,}96 \cdot 0{,}004]$

$IC(\beta_1): [-0{,}066; -0{,}050]$

Interpretare: Cu o probabilitate de 99%, se poate garanta că panta dreptei de regresie este acoperită de intervalul $[-0{,}066; -0{,}050]$.

6. Formulați o cerință care presupune o predicție pentru variabila dependentă/ independentă, cunoscând o anumită valoare a variabilei independente/dependente.

Să se specifice procentului femeilor care folosesc metode contraceptive pentru a obține o rată totală a fertilității de 2,5 copii născuți de o femeie de-a lungul vieții.

$$y_{x_i} = 6{,}875 - 0{,}058x_i \Longrightarrow 2{,}5 = 6{,}875 - 0{,}058x_i \Longrightarrow$$
$$x_i = \frac{2{,}5 + 0{,}058}{6{,}875} = 0{,}372$$

Interpretare: Pentru a obține o rată totală a fertilității de 2,5 copii născuți de o femeie, procentul femeilor care folosesc metode contraceptive ar trebui să fie de 0,372 procente.

7. Formulați o cerință care presupune o predicție pentru variația variabilei dependente/ independente, pentru o modificare dată a variabilei independente/dependente.

Să se estimeze cu cât va scădea rata totală a fertilității pentru o creștere a procentului femeilor care folosesc metode contraceptive cu 1,5 procente.

$$b_1 = \frac{\Delta Y}{\Delta X} \Longrightarrow -0{,}058 = \frac{\Delta Y}{1{,}5} \Longrightarrow \Delta Y = -0{,}087$$

Interpretare: La o scădere a procentului femeilor care folosesc metode contraceptive cu 1,5 procente, rata totală a fertilității scade, în medie, cu 0,087 puncte.

8. Verificați dacă procentul femeilor care folosesc metode contraceptive explică semnificativ variația ratei totale a fertilității.

Etapele testării	**Testarea parametrului $\boldsymbol{\beta_1}$**
1. Formularea ipotezelor	$H_0: \beta_1 = 0$ (parametrul β_1 nu diferă semnificativ de 0, **ceea ce înseamnă că** între rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) nu există o legătură liniară semnificativă **SAU** procentul femeilor care folosesc metode contraceptive (X) nu are o influență semnificativă asupra rata totală a fertilității (Y)) $H_1: \beta_1 \neq 0$ (parametrul β_1 diferă semnificativ de 0 **ceea ce înseamnă că** între rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) există o legătură liniară semnificativă **SAU** procentul femeilor care folosesc metode contraceptive (X) are o influență semnificativă asupra rata totală a fertilității (Y))
2. Alegerea pragului de semnificație	$\alpha = 0{,}05$
3. Alegerea statisticii test	$t = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t(n-2)$

4. Determinarea valorii teoretice a statisticii test	$t_{teoretic} = t_{\alpha/2;\,n-2} =$ $t_{0{,}05/2;50-2} = t_{0{,}025;48} = 1{,}96$
5. Determinarea valorii calculate a statisticii test	$t_{calc} = \frac{b_1}{s_{\hat{\beta}_1}} = \frac{-0{,}058}{0{,}004} = -16{,}299$
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), în condițiile unui risc α. 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, în condițiile unui risc α.
7. Luarea deciziei	$\lvert t_{calc}\rvert = 20 > t_{\alpha/2;\,n-2} = 1{,}96$ **SAU** $Sigt = 0{,}000 < \alpha = 0{,}05 \Rightarrow$ că se respinge ipoteza H_0 (5%)
8. Interpretarea rezultatului	În condițiile unui risc 5%, se consideră că parametrul β_1 diferă semnificativ de 0, **ceea ce înseamnă că** legătura liniară dintre rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) este semnificativă statistic **SAU** că procentul femeilor care folosesc metode contraceptive (X) are o influență semnificativă asupra ratei totale a fertilității (Y) **SAU** procentul femeilor care folosesc metode contraceptive (X) explică semnificativ variația ratei totale a fertilității (Y).

9. Calculați valoarea testului Student pentru parametrul β_0 și luați decizia, considerând un risc de 10%, privind semnificația parametrilor modelului de regresie.

Etapele testării	Testarea parametrului β_0
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)
2. Alegerea pragului de semnificaţie	$\alpha = 0{,}10$
3. Alegerea statisticii test	$t = \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \sim t(n-2)$
4. Determinarea valorii teoretice a statisticii test	$t_{teoretic} = t_{\alpha/2;\,n-2} =$ $t_{0{,}10/2;\,50-2} = t_{0{,}05;\,48} = 1{,}645$
5. Determinarea valorii calculate a statisticii test	$t_{calc} = \frac{b_0}{s_{\hat{\beta}_0}} = \frac{6{,}875}{0{,}157} = 43{,}829$
6. Regula de decizie	Dacă se ţine cont de valoarea calculată a testului, regula de decizie este următoarea:

	- 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), în condițiile unui risc α. 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, în condițiile unui risc α.
7. Luarea deciziei	$\|t_{calc}\| = 43{,}829 > t_{\alpha/2;\,n-2} = 1{,}645$ **SAU** $Sigt = 0{,}000 < \alpha = 0{,}10 \Rightarrow$ că se respinge ipoteza H_0 (10%)
8. Interpretarea rezultatului	În condițiile unui risc de 10%, se consideră că parametrul sau constanta modelului diferă semnificativ de 0.

10. Interpretaţi valoarea estimată a coeficientului de corelaţie.

$r = -0{,}920$

Interpretare:

- în funcție de ***semnul*** coeficientului de corelație: legătura liniară dintre rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) este inversă
- în funcție de ***valoarea în modul*** a coeficientului de corelație $(|r| = 0{,}920)$, această legătură este de intensitate puternică

Interpretarea integrală:

- între rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) există o legătură liniară inversă și de intensitate puternică.

11. Verificați dacă cele două variabile sunt corelate semnificativ.

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 rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) 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 rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) există o legătură liniară semnificativă **SAU** cele două variabile sunt corelate semnificativ)
2. Alegerea pragului de semnificație	$\alpha = 0{,}05$
3. Alegerea statisticii test	$t = \dfrac{\hat{\rho}}{\sqrt{\dfrac{1-\hat{\rho}^2}{n-2}}} \sim t(n-2)$
4. Determinarea valorii teoretice a statisticii test	$t_{teoretic} = t_{\alpha/2;\,n-2} =$ $t_{0{,}05/2;\,50-2} = t_{0{,}025;\,48} = 1{,}96$

5. Determinarea valorii calculate a statisticii test	$t_{calc} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{-0{,}920}{\sqrt{\frac{1-(-0{,}959)^2}{50-2}}} = \frac{-0{,}921}{0{,}056} = -16{,}446$
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), în condițiile unui risc α. 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, în condiţiile unui risc α.
7. Luarea deciziei	$\lvert t_{calc}\rvert = 16{,}446 > t_{\alpha/2;\,n-2} = 1{,}96 \Rightarrow$ că se respinge ipoteza H_0 (5%) **SAU** $Sigt = 0{,}000 < \alpha = 0{,}05 \Rightarrow$ că se respinge ipoteza H_0 (5%)
8. Interpretarea deciziei luate	În condițiile unui risc de 5%, se poate garanta că între rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) există o legătură liniară semnificativă **SAU** între rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) sunt corelate semnificativ.

12. Estimaţi raportul de determinaţie şi interpretaţi rezultatul.

Estimarea raportului de determinație

$$R^2 = \frac{ESS}{TSS} = \frac{87.672}{103.513} = 0{,}847$$

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{15.840}{103.513} = 1 - 0{,}153 = 0{,}847$$

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

$$R^2 = r^2 = (-0{,}920)^2 = 0{,}847$$

Interpretare:

- 84,7% din variația totală a ratei totale a fertilității este explicată de variația procentului femeilor care folosesc metode contraceptive. Iar restul de 15,3% (diferența până la 100%) din variația totală a ratei totale a fertilității este explicată de influența factorilor aleatori sau nespecificați în model.

13. Estimaţi raportul de corelație şi interpretaţi rezultatul.

Estimarea raportului de determinație

$$R = \sqrt{R^2} = \sqrt{0{,}847} = 0{,}920$$

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

$R = |r| = |-0{,}920| = 0{,}920$

Interpretare:

- între rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) există o legătură liniară de intensitate puternică.

14. Testaţi semnificaţia raportului de corelaţie.

Etapele testării	**Testarea raportului de corelație η**
1. Formularea ipotezelor	$H_0: \eta = 0$ (raportul de determinație η^2 sau raportul de corelația η nu diferă semnificativ de 0, **ceea ce înseamnă** că între rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) nu există o legătură liniară semnificativă) $H_1: \eta > 0$ (raportul de determinație η^2 sau raportul de corelația η este semnificativ statistic, **ceea ce înseamnă** că între rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) există o legătură liniară semnificativă)
2. Alegerea pragului de semnificaţie	$\alpha = 0{,}05$
3. Alegerea statisticii test	$F = \frac{\hat{\eta}^2}{1-\hat{\eta}^2} \cdot \frac{n-k}{k-1} \sim F(k-1; n-k)$
4. Determinarea valorii teoretice a statisticii test	$F_{teoretic} = F_{\alpha;\, k-1;\, n-k} =$ $F_{0,05;\, 1;\, 48} = 4{,}085$
5. Determinarea valorii calculate a statisticii test	$F_{calc} = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1}$ $F_{calc} = \frac{0{,}847}{1-0{,}847} \cdot \frac{48}{1} = 265{,}658$
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), în condițiile unui risc α. 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, în condițiile unui risc α.
7. Luarea deciziei	$F_{calc} = 265{,}658 > F_{\alpha;\, k-1;\, n-k} = 4{,}085 \Rightarrow$ că se respinge ipoteza H_0 (5%) **SAU** $SigF = 0{,}000 < \alpha = 0{,}05 \Rightarrow$ că se respinge ipoteza H_0 (5%)
8. Interpretarea deciziei luate	În condițiile unui risc de 5%, se consideră că între rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) există o legătură liniară semnificativă **SAU** modelul de regresie explică semnificativ dependența liniară dintre rata totală a fertilității (Y) și procentul femeilor care folosesc

	metode contraceptive (X) **SAU** modelul de regresie liniară simplă ales este corect specificat.

15. Interpretaţi probabilitatea asociată statisticii test Student în vederea testării modelului de regresie.

$SigF = 0{,}000 < \alpha = 0{,}05 \Rightarrow$ că se respinge ipoteza H_0 (5%)

Interpretare:
În condițiile unui risc de 5%, se consideră că între rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) există o legătură liniară semnificativă **SAU** modelul de regresie explică semnificativ dependența liniară dintre rata totală a fertilității (Y) și procentul femeilor care folosesc metode contraceptive (X) **SAU** modelul de regresie liniară este corect specificat.