

REGRESIONA I KORELACIONA ANALIZA

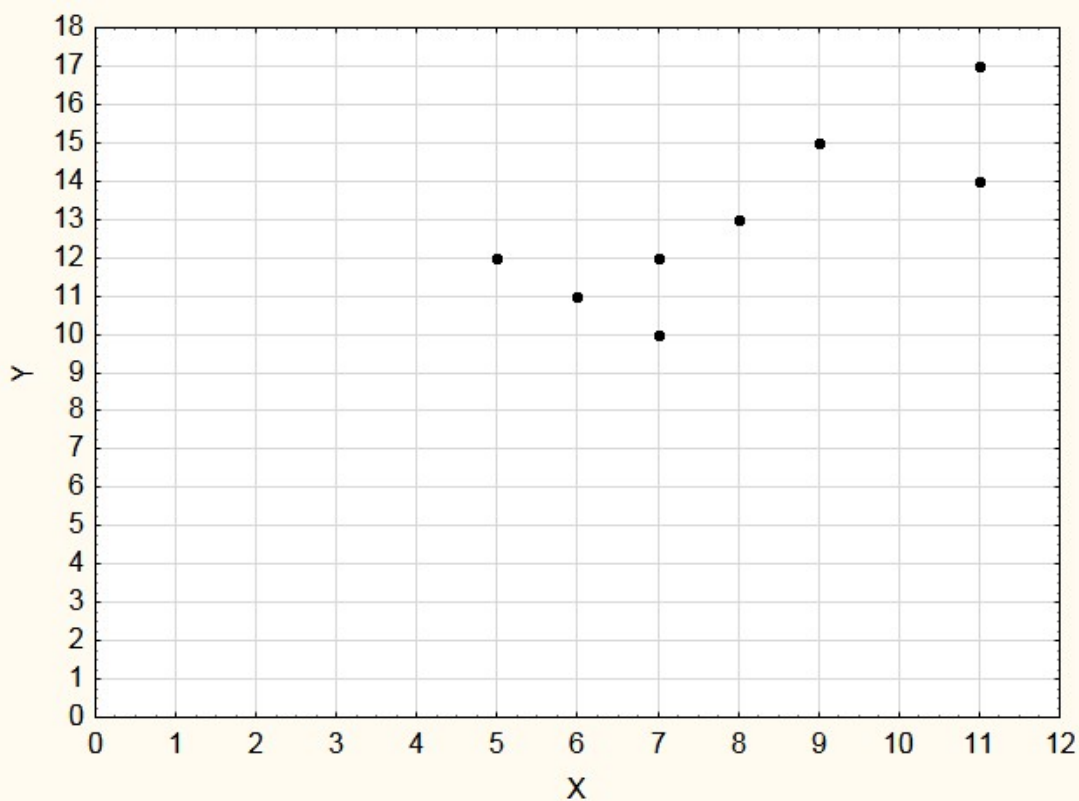
Na osnovu 8 parova vrednosti numeričkih obeležja X i Y:

Y	12	14	10	13	17	12	11	15
X	5	11	7	8	11	7	6	9

- a) Formirati dijagram rasturanja.
- b) Oceniti i grafički predstaviti linearni regresioni model: $\hat{Y} = a + bX$.
- c) Izračunati ocenjene vrednosti regresije i standardnu grešku regresije.
- d) Izračunati koeficijente korelacije i determinacije. Testirati statističku značajnost koeficijenta korelacije.

Rešenje:

a)



b) Opšti oblik linearnog regresionog modela sa jednom nezavisnom promenljivom je:

$$Y = \alpha + \beta X + \varepsilon, \quad \alpha \text{ i } \beta \text{ su nepoznati parametri, } \varepsilon \text{ je slučajna greška.}$$

Ocene nepoznatih parametara α i β izračunate na osnovu uzorka za X i Y veličine n su a i b, tako da je ocenjeni regresioni model oblika $\hat{Y} = a + bX$.

Jedan od metoda ocene nepoznatih parametara je **metod najmanjih kvadrata**, koji polazi od kriterijuma da je zbir kvadrata razlika izmerenih i ocenjenih vrednosti zavisne promenljive najmanji tj. $\min \sum (Y - \hat{Y})^2$. Ocene a i b su:

$$b = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} \quad \text{i} \quad a = \bar{Y} - b\bar{X}.$$

Ocena b se može izraziti i u obliku:

$$b = \frac{\sum XY - \frac{(\sum X \sum Y)}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}}.$$

Ocena a može da se izrazi i kao:

$$a = \frac{\sum Y(\sum X^2) - \sum X \sum XY}{n(\sum X^2) - (\sum X)^2}$$

I način

$$\bar{X} = \frac{\sum X}{n} = \frac{64}{8} = 8 \quad \bar{Y} = \frac{\sum Y}{n} = \frac{104}{8} = 13$$

Formira se tabela:

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$	$(X - \bar{X})^2$	$(Y - \bar{Y})^2$
5	12	-3	-1	3	9	1
11	14	3	1	3	9	1
7	10	-1	-3	3	1	9
8	13	0	0	0	0	0
11	17	3	4	12	9	16
7	12	-1	-1	1	1	1
6	11	-2	-2	4	4	4
9	15	1	2	2	1	4
64	104	0	0	28	34	36

$$b = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} = \frac{28}{34} = 0,8235 \quad a = \bar{Y} - b\bar{X} = 13 - 0,8235 \cdot 8 = 6,412$$

II način

Formira se tabela:

Y	X	X^2	Y^2	XY
12	5	25	144	60
14	11	121	196	154
10	7	49	100	70
13	8	64	169	104
17	11	121	289	187
12	7	49	144	84
11	6	36	121	66
15	9	81	225	135
104	64	546	1388	860

$$b = \frac{\sum XY - \frac{(\sum X \sum Y)}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}} = \frac{860 - \frac{64 \cdot 104}{8}}{546 - \frac{64^2}{8}} = \frac{28}{34} = 0,8235$$

$$a = \bar{Y} - b\bar{X} = 13 - 0,8235 \cdot 8 = 6,412$$

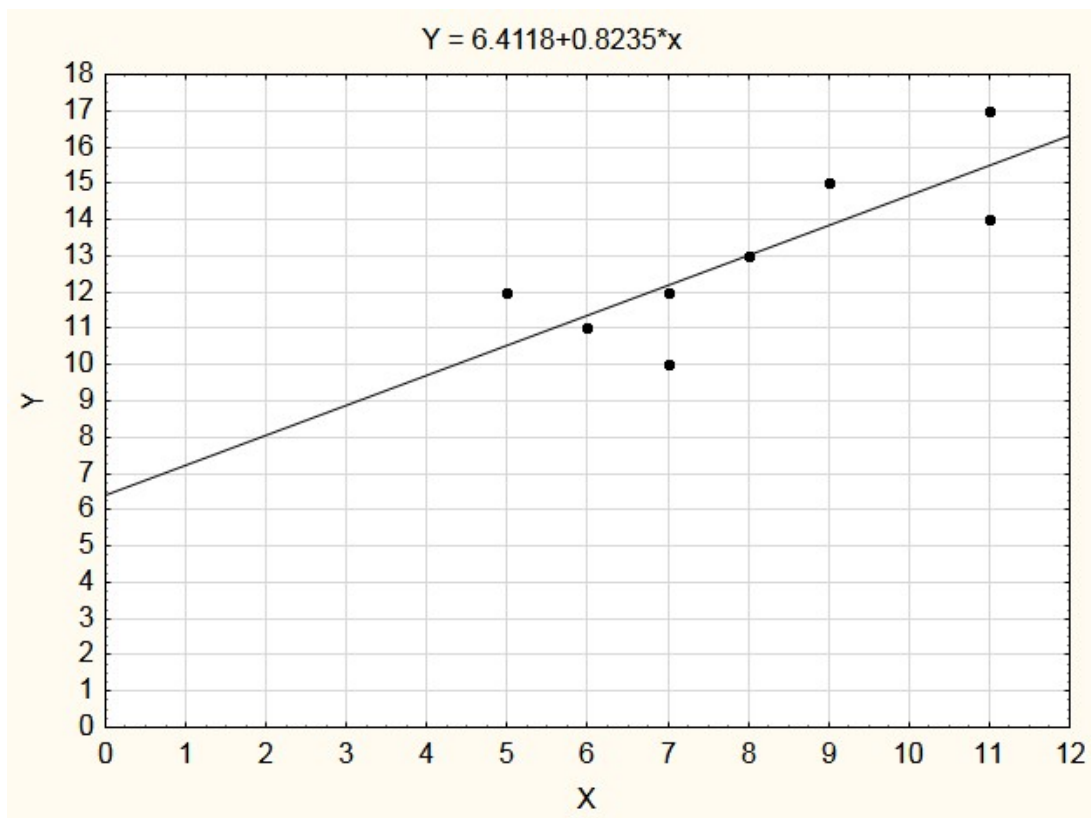
Ili:

$$a = \frac{\sum Y(\sum X^2) - \sum X \sum XY}{n(\sum X^2) - (\sum X)^2} = \frac{104 \cdot 546 - 64 \cdot 860}{8 \cdot 546 - 64^2} = \frac{1744}{272} = 6,4118$$

Ocenjeni regresioni model je:

$$\hat{Y} = 6,412 + 0,8235X.$$

Da bi se grafički predstavila regresiona prava dovoljno je odrediti dve tačke koje pripadaju pravoj npr. tačke $X=0$, $\hat{Y} = a$ i $X = \bar{X}$, $Y = \bar{Y}$ pripadaju pravoj. U slučaju ocenjenog regresionog modela to su tačke $X=0$, $\hat{Y} = 6,412$ i $X = 8$, $\hat{Y} = 13$.



c) Ocenjene vrednosti regresije su:

$$\hat{Y}_1 = 6,412 + 0,8235 \cdot 5 = 10,5295$$

$$\hat{Y}_2 = 6,412 + 0,8235 \cdot 11 = 15,4705$$

$$\hat{Y}_3 = 6,412 + 0,8235 \cdot 7 = 12,1765$$

$$\hat{Y}_4 = 6,412 + 0,8235 \cdot 8 = 13,0000$$

$$\hat{Y}_5 = 6,412 + 0,8235 \cdot 11 = 15,4705$$

$$\hat{Y}_6 = 6,412 + 0,8235 \cdot 7 = 12,1765$$

$$\hat{Y}_7 = 6,412 + 0,8235 \cdot 6 = 11,3530$$

$$\hat{Y}_8 = 6,412 + 0,8235 \cdot 9 = 13,8235$$

Napomena:

Kod linearnog regresionog modela čiji su parametri ocenjeni metodom najmanjih kvadrata je uvek $\sum Y = \sum \hat{Y}$.

Da bi se izračunala standardna greška regresije formira se tabela:

X	Y	\hat{Y}	$Y - \hat{Y}$	$(Y - \hat{Y})^2$
5	12	10,5295	1,4705	2,16237
11	14	15,4705	-1,4705	2,16237
7	10	12,1765	-2,1765	4,73715
8	13	13,0000	0,0000	0,00000
11	17	15,4705	1,5295	2,33937
7	12	12,1765	-0,1765	0,03115
6	11	11,3530	-0,3530	0,12461
9	15	13,8235	1,1765	1,38415
64	104	104,0000	0,0000	12,94117

$$s_e = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{12,94117}{6}} = 1,4686$$

$$s_e = \sqrt{\frac{\sum Y^2 - a \sum Y - b \sum XY}{n-2}} = \sqrt{\frac{1388 - 6,412 \cdot 104 - 0,8235 \cdot 860}{6}} = \sqrt{\frac{12,9420}{6}} = 1,4687$$

Napomena:

Varijansa regresije je: $s_e^2 = \frac{\sum (Y - \hat{Y})^2}{n-2}$.

Standardna greška regresije je: $s_e = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}}$

Reziduali regresije su: $e_i = Y_i - \hat{Y}_i$

d)
način

$$r = \frac{\sum XY - \frac{(\sum X \sum Y)}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right] \cdot \left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}} = \frac{860 - \frac{64 \cdot 104}{8}}{\sqrt{\left(546 - \frac{64^2}{8}\right) \cdot \left(1388 - \frac{104^2}{8}\right)}} = \frac{28}{\sqrt{34 \cdot 36}} = 0,8003$$

II način

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} = \frac{28}{\sqrt{34 \cdot 36}} = 0,8003$$

Napomena:

- Prost (Pirsonov) koeficijent korelacije r je ocena na osnovu uzorka populacijskog parametra ρ koji pokazuje jačinu **linearne** međuzavisnosti promenljivih X i Y .
- $r \in [-1,1]$

Koeficijent determinacije:

$$r^2 = 64,05\%$$

Koeficijent determinacije pokazuje koliki je deo varijacije zavisne promenljive objašnjen regresionim modelom.

Testiranje značajnosti koeficijenta korelacije

$$\begin{array}{|l} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{array}$$

$$t = \frac{r}{s_r} \sim t_{n-2} \quad s_r = \sqrt{\frac{1-r^2}{n-2}}$$

$$t = \frac{0,8003}{\sqrt{1-0,6405}} \cdot \sqrt{6} = 3,27^* \quad t_{6;0,05} = 2,447 \quad t_{6;0,05} = 3,707$$

Kako je $2,447 < t < 3,707$ sledi da se nulta hipoteza odbacuje na pragu značajnosti 5%, dok se se odbacuje na pragu značajnosti 1%. Prema tome linearna međuzavisnost promenljivih X i Y je statistički značajna.