

PROSTA LINEARNA REGRESIJA

- Ispitivanje zavisnosti između dve vrednosti obeležja, pri čemu je jedna zavisna promenljiva, a druga nezavisna promenljiva
- Nezavisna promenljiva (x_i) je promenljiva koja utiče na kretanje zavisne promenljive (y_i)
- Npr. na tražnju potrošača utiče cena nekog proizvoda ili usluge, što znači da je cena nezavisna promenljiva koja utiče na tražnju (zavisnu promenljivu)

ZADATAK 1

Dati su podaci o proizvodnji i prihodu jednog preduzeća:

| | | | | | | | |
|-----------------------|---|---|---|---|---|----|----|
| Proizvodnja (000) t | 2 | 3 | 5 | 6 | 8 | 10 | 13 |
| Prihodu (mil.) dinara | 5 | 7 | 8 | 9 | 9 | 11 | 15 |

- Oceniti linearnu regresiju i predstaviti je grafički na dijagramu raspršenosti.
- Ispitati stepen zavisnosti između proizvodnje i prihoda.
- Koliko je prosečno odstupanje podataka od linije regresije?
- Predvideti sa 95% pouzdanosti koliki će biti prihod pri proizvodnji od 9.000 tona.
- Predvideti sa 99% pouzdanosti koliki će biti prihod pri proizvodnji od 7.000 tona.
- Sa koliko % je ukupni varijabilitet objašnjen vezom između proizvodnje i prihoda preduzeća? (koeficijent determinacije)

Napomena: Student treba samostalno da odredi u postavljenom zadatku šta je nezavisna, a šta zavisna promenljiva!!!

| Proizvodnja (000 t) | Prihod (10 ⁶ din) | | | |
|---------------------|------------------------------|-----------------|------------|------------|
| x_i | y_i | $x_i \cdot y_i$ | x_i^2 | y_i^2 |
| 2 | 5 | 10 | 4 | 25 |
| 3 | 7 | 21 | 9 | 49 |
| 5 | 8 | 40 | 25 | 64 |
| 6 | 9 | 54 | 36 | 81 |
| 8 | 9 | 72 | 64 | 81 |
| 10 | 11 | 110 | 100 | 121 |
| 13 | 15 | 195 | 169 | 225 |
| 47 | 64 | 502 | 407 | 646 |

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{47}{7} = 6,714$$

Odgovor: Prosečna proizvodnja iznosi 6.714 tona.

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{64}{7} = 9,143$$

Odgovor: Prosečan prihod iznosi 9.143.000 dinara

$$b = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum x_i^2 - n \bar{x}^2} = \frac{502 - 7 * 6,714 * 9,143}{407 - 7 * 6,714^2} = \frac{72,297286}{91,455428} = 0,791$$

Odgovor: U slučaju povećanja proizvodnje za 1.000 tona prihod će se u proseku povećati za 791.000 dinara.

Napomena: Ako je vrednost koeficijenta b negativan broj, to znači da povećanje nezavisne promenljive za 1 jedinicu izaziva smanjenje zavisne promenljive za dobijeni iznos!!!

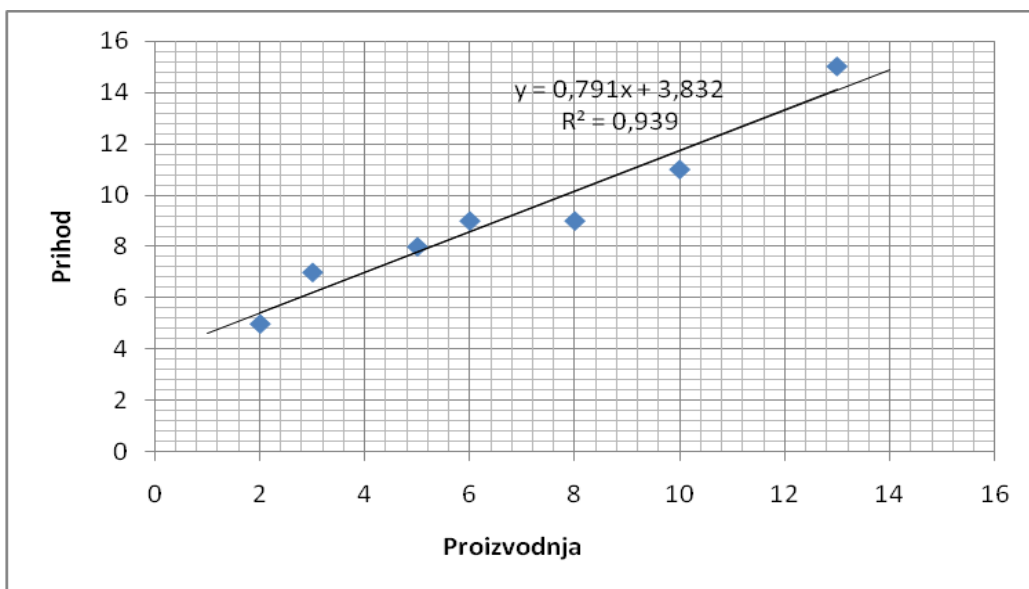
$$a = \bar{y} - b\bar{x} = 9,143 - 0,791 * 6,714 = 3,832$$

Odgovor: U slučaju kada je proizvodnja $x_i=0$ prihod će iznositi 3.832.000 dinara.

Ocenjena jednačina linearne regresije ima sledeći oblik:

$$\hat{y}_i = a + b \cdot x_i = 3,832 + 0,791 \cdot x_i$$

Dijagram rasturanja (raspršenosti)



b)

$$r = \frac{\sum x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{(\sum x_i^2 - n \cdot \bar{x}^2) \cdot (\sum y_i^2 - n \cdot \bar{y}^2)}} = \frac{502 - 7 \cdot 6,714 \cdot 9,143}{\sqrt{(407 - 7 \cdot 6,714^2) \cdot (646 - 7 \cdot 9,143^2)}} = \frac{72,297286}{\sqrt{91,455428 \cdot 60,838857}}$$
$$= \frac{72,297286}{\sqrt{5564,043706}} = \frac{72,297286}{74,59251776} = 0,969$$

Odgovor: Na osnovu izračunatog koeficijenta korelacije između posmatranih promenljivih postoji jaka korelaciona veza.

Napomena za tumačenje korelacione veze:

c)

$$s_e = \sqrt{\frac{\sum y_i^2 - a \cdot \sum y_i - b \cdot \sum x_i y_i}{n - 2}} = \sqrt{\frac{646 - 3,832 \cdot 64 - 0,791 \cdot 502}{7 - 2}} = \sqrt{\frac{3,67}{5}} = \sqrt{0,734} = 0,857$$

Odgovor: Prosečno odstupanje zavisne promenljive (prihoda) od linije regresije iznosi 0,857.

d)

$$\hat{y}_0 = a + b \cdot x_0 = 3,832 + 0,791 \cdot 9 = 10,951$$

Odgovor: Na osnovu jednačine linearne regresije a pri proizvodnji od 9.000 tona možemo očekivati da će prihod biti 10.591.000 dinara.

$$s_p = s_e \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum x_i^2 - n \cdot \bar{x}^2}} = 0,857 \cdot \sqrt{1 + \frac{1}{7} + \frac{(6,714 - 9)^2}{407 - 7 \cdot 6,714^2}} = 0,857 \cdot \sqrt{1,199997502} = 0,939$$

$$t_{\alpha/2; n-2} = t_{0,025; 5} = 2,57$$

$$\alpha - \text{nivo greške} = 5\% = 0,05$$

Napomena: Tablična vrednost (t) se očitava iz studentove t-distribucije koja se nalazi u formulama!

$$\hat{y}_0 - t_{\alpha/2; n-2} \cdot s_p < y_0 < \hat{y}_0 + t_{\alpha/2; n-2} \cdot s_p$$

$$10,951 - 2,57 \cdot 0,939 < y_0 < 10,951 + 2,57 \cdot 0,939$$

$$8,538 < y_0 < 13,364$$

Odgovor: Može se očekivati da će se sa nivoom pouzdanosti od 95% i pri proizvodnji od 9.000 tona prihod kretati u intervalu od 8.538.000 do 13.364.000 dinara.

e)

$$\hat{y}_0 = a + b \cdot x_0 = 3,832 + 0,791 \cdot 7 = 9,369$$

Odgovor: Na osnovu jednačine linearne regresije a pri proizvodnji od 7.000 tona možemo očekivati da će prihod biti 9.369.000 dinara.

$$s_p = s_e \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum x_i^2 - n \cdot \bar{x}^2}} = 0,857 \cdot \sqrt{1 + \frac{1}{7} + \frac{(6,714 - 7)^2}{407 - 7 \cdot 6,714^2}} = 0,857 \cdot \sqrt{1,143751524} = 0,917$$

$$t_{\alpha/2; n-2} = t_{0,005; 5} = 4,03$$

$$\alpha = 1\% = 0,01$$

$$\hat{y}_0 - t_{\alpha/2; n-2} \cdot s_p < y_0 < \hat{y}_0 + t_{\alpha/2; n-2} \cdot s_p$$

$$9,369 - 4,03 \cdot 0,917 < y_0 < 9,369 + 4,03 \cdot 0,917$$

$$5,673 < y_0 < 13,065$$

Odgovor: Može se očekivati da će se sa nivoom pouzdanosti od 99% i pri proizvodnji od 7.000 tona prihod kretati u intervalu od 5.673.000 do 13.065.000 dinara.

f)

$$r^2 = \frac{a \cdot \sum y_i + b \cdot \sum x_i y_i - n \cdot \bar{y}^2}{\sum y_i^2 - n \cdot \bar{y}^2} = \frac{3,832 \cdot 64 + 0,791 \cdot 502 - 7 \cdot 9,143^2}{646 - 7 \cdot 9,143^2} = \frac{57,1689}{60,8389} = 0,9397 = 93,97\%$$

Odgovor: Na osnovu izračunatog koeficijenta determinacije zaključujemo da je 93,97% varijabiliteta prihoda određeno količinom proizvodnje.