

Linearna regresija

PREDAVANJA 13

Univerzitet u Novom Sadu, FTN

- Posmatramo dva obeležja X i Y .
- Metodom momenata dobijamo ocenu R koeficijenta korelacije:

$$R = \frac{\bar{s}_{XY}}{\bar{s}_X \bar{s}_Y},$$

čiju realizovanu vrednost označavamo sa r

$$r = \frac{\bar{s}_{xy}}{\bar{s}_x \bar{s}_y}$$

$$\bar{s}_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2}; \quad \bar{s}_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_n)^2};$$

$$\bar{s}_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n); \quad \bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \text{ i } \bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i.$$

- Posmatramo dva obeležja X i Y .
- Metodom momenata dobijamo ocenu R koeficijenta korelacije:

$$R = \frac{\bar{s}_{XY}}{\bar{s}_X \bar{s}_Y},$$

čiju realizovanu vrednost označavamo sa r

$$r = \frac{\bar{s}_{xy}}{\bar{s}_x \bar{s}_y}$$

$$\bar{s}_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2}; \quad \bar{s}_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_n)^2};$$

$$\bar{s}_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n); \quad \bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \text{ i } \bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i.$$

- Posmatramo dva obeležja X i Y .
- Metodom momenata dobijamo ocenu R koeficijenta korelacije:

$$R = \frac{\bar{s}_{XY}}{\bar{s}_X \bar{s}_Y},$$

čiju realizovanu vrednost označavamo sa r

$$r = \frac{\bar{s}_{xy}}{\bar{s}_x \bar{s}_y}$$

$$\bar{s}_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2}; \quad \bar{s}_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_n)^2};$$

$$\bar{s}_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n); \quad \bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \text{ i } \bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i.$$

Ako je $|r| < 0.3$ obeležja X i Y nisu u korelacionoj vezi.

Ako je $0.3 \leq |r| \leq 0.7$ postoji neznatna korelacija između posmatranih obeležja X i Y .

Ako je $|r| > 0.7$ postoji znatna korelacija između posmatranih obeležja X i Y . Pri tome:

- Ako je $0.7 < |r| < 0.9$, postoji znatna linearna korelacija.
- Ako je $|r| \geq 0.9$, linearna korelacija je veoma jaka.

Ako je r **pozitivno**, tada je linearna regresiona funkcija **rastuća**, a ako je r **negativno**, tada je linearna regresiona funkcija **opadajuća**.

Ako je vrednost $|r|$ bliska jedinici, možemo da zaključimo da je korelacija između X i Y jaka, a ako je vrednost $|r|$ bliska nuli, možemo da zaključimo da su X i Y veoma slabo ili nikako povezane.

Ako je $|r| < 0.3$ obeležja X i Y nisu u korelacionoj vezi.

Ako je $0.3 \leq |r| \leq 0.7$ postoji neznatna korelacija između posmatranih obeležja X i Y .

Ako je $|r| > 0.7$ postoji znatna korelacija između posmatranih obeležja X i Y . Pri tome:

- Ako je $0.7 < |r| < 0.9$, postoji znatna linearna korelacija.
- Ako je $|r| \geq 0.9$, linearna korelacija je veoma jaka.

Ako je r pozitivno, tada je linearna regresiona funkcija rastuća, a ako je r negativno, tada je linearna regresiona funkcija opadajuća.

Ako je vrednost $|r|$ bliska jedinici, možemo da zaključimo da je korelacija između X i Y jaka, a ako je vrednost $|r|$ bliska nuli, možemo da zaključimo da su X i Y veoma slabo ili nikako povezane.

Ako je $|r| < 0.3$ obeležja X i Y nisu u korelacionoj vezi.

Ako je $0.3 \leq |r| \leq 0.7$ postoji neznatna korelacija između posmatranih obeležja X i Y .

Ako je $|r| > 0.7$ postoji znatna korelacija između posmatranih obeležja X i Y . Pri tome:

- Ako je $0.7 < |r| < 0.9$, postoji znatna linearna korelacija.
- Ako je $|r| \geq 0.9$, linearna korelacija je veoma jaka.

Ako je r pozitivno, tada je linearna regresiona funkcija rastuća, a ako je r negativno, tada je linearna regresiona funkcija opadajuća. Ako je vrednost $|r|$ bliska jedinici, možemo da zaključimo da je korelacija između X i Y jaka, a ako je vrednost $|r|$ bliska nuli, možemo da zaključimo da su X i Y veoma slabo ili nikako povezane.

Ako je $|r| < 0.3$ obeležja X i Y nisu u korelacionoj vezi.

Ako je $0.3 \leq |r| \leq 0.7$ postoji neznatna korelacija između posmatranih obeležja X i Y .

Ako je $|r| > 0.7$ postoji znatna korelacija između posmatranih obeležja X i Y . Pri tome:

- Ako je $0.7 < |r| < 0.9$, postoji znatna linearna korelacija.
- Ako je $|r| \geq 0.9$, linearna korelacija je veoma jaka.

Ako je r **pozitivno**, tada je linearna regresiona funkcija **rastuća**, a ako je r **negativno**, tada je linearna regresiona funkcija **opadajuća**.

Ako je vrednost $|r|$ bliska jedinici, možemo da zaključimo da je korelacija između X i Y jaka, a ako je vrednost $|r|$ bliska nuli, možemo da zaključimo da su X i Y veoma slabo ili nikako povezane.

Ako je $|r| < 0.3$ obeležja X i Y nisu u korelacionoj vezi.

Ako je $0.3 \leq |r| \leq 0.7$ postoji neznatna korelacija između posmatranih obeležja X i Y .

Ako je $|r| > 0.7$ postoji znatna korelacija između posmatranih obeležja X i Y . Pri tome:

- Ako je $0.7 < |r| < 0.9$, postoji znatna linearna korelacija.
- Ako je $|r| \geq 0.9$, linearna korelacija je veoma jaka.

Ako je r **pozitivno**, tada je linearna regresiona funkcija **rastuća**, a ako je r **negativno**, tada je linearna regresiona funkcija **opadajuća**.

Ako je vrednost $|r|$ bliska jedinici, možemo da zaključimo da je korelacija između X i Y jaka, a ako je vrednost $|r|$ bliska nuli, možemo da zaključimo da su X i Y veoma slabo ili nikako povezane.

Jednačina linearne regresije (Y u zavisnosti od X):

Regresiona prava

$$\hat{y} = a_1x + a_0,$$

gde je $a_1 = r \frac{\bar{s}_y}{\bar{s}_x} = \frac{\bar{s}_{xy}}{\bar{s}_x^2}$ koeficijent pravca prave, $a_0 = \bar{y}_n - a_1\bar{x}_n$.

Jednačina linearne regresije (X u zavisnosti od Y):

Regresiona prava

$$\hat{x} = a_1y + a_0,$$

gde je $a_1 = r \frac{\bar{s}_x}{\bar{s}_y} = \frac{\bar{s}_{xy}}{\bar{s}_y^2}$ koeficijent pravca prave, $a_0 = \bar{x}_n - a_1\bar{y}_n$.

Jednačina linearne regresije (Y u zavisnosti od X):

Regresiona prava

$$\hat{y} = a_1x + a_0,$$

gde je $a_1 = r \frac{\bar{s}_y}{\bar{s}_x} = \frac{\bar{s}_{xy}}{\bar{s}_x^2}$ koeficijent pravca prave, $a_0 = \bar{y}_n - a_1\bar{x}_n$.

Jednačina linearne regresije (X u zavisnosti od Y):

Regresiona prava

$$\hat{x} = a_1y + a_0,$$

gde je $a_1 = r \frac{\bar{s}_x}{\bar{s}_y} = \frac{\bar{s}_{xy}}{\bar{s}_y^2}$ koeficijent pravca prave, $a_0 = \bar{x}_n - a_1\bar{y}_n$.

Jednačina linearne regresije (Y u zavisnosti od X):

Regresiona prava

$$\hat{y} = a_1x + a_0,$$

gde je $a_1 = r \frac{\bar{s}_y}{\bar{s}_x} = \frac{\bar{s}_{xy}}{\bar{s}_x^2}$ koeficijent pravca prave, $a_0 = \bar{y}_n - a_1\bar{x}_n$.

Jednačina linearne regresije (X u zavisnosti od Y):

Regresiona prava

$$\hat{x} = a_1y + a_0,$$

gde je $a_1 = r \frac{\bar{s}_x}{\bar{s}_y} = \frac{\bar{s}_{xy}}{\bar{s}_y^2}$ koeficijent pravca prave, $a_0 = \bar{x}_n - a_1\bar{y}_n$.

Primer

Dati su parovi podataka o temperaturi ($[^{\circ}\text{C}]$) eksperimentalne pločice od ispitivanog materijala i veličini pukotine ($[\text{mm}]$) na njoj.

<i>temperatura $[^{\circ}\text{C}]$</i>	<i>-1.5</i>	<i>-1.0</i>	<i>-0.5</i>	<i>0.0</i>	<i>1.0</i>	<i>2.0</i>
<i>veličina pukotine $[\text{mm}]$</i>	<i>0.2</i>	<i>0.4</i>	<i>0.5</i>	<i>0.8</i>	<i>1.0</i>	<i>1.3</i>

Podatke predstaviti dijagramom rasipanja (scatter). Naći pravu linearne regresije veličine pukotine u zavisnosti od temperature i ucrtati je na dijagram. Izračunati koeficijent linearne korelacije. Prognozirati širinu pukotine na 2.5°C .

Proglasimo temperaturu za obeležje X , a odgovarajuću širinu pukotine za obeležje Y . Veličina uzorka je $n = 6$.

Izračunavamo aritmetičke sredine:

$$\bar{x}_6 = \frac{1}{6} \cdot (-1.5 - 1.0 - 0.5 + 0.0 + 1.0 + 2.0) = 0.0$$

$$\bar{y}_6 = \frac{1}{6} \cdot (0.2 + 0.4 + 0.5 + 0.8 + 1.0 + 1.3) = 0.7$$

x_i	y_i	$x_i - \bar{x}_6$	$y_i - \bar{y}_6$	$(x_i - \bar{x}_6)^2$	$(y_i - \bar{y}_6)^2$	$A \cdot B$
-1.50	0.20	-1.5	-0.5	2.25	0.25	0.75
-1.00	0.40	-1.0	-0.3	1.00	0.09	0.30
-0.50	0.50	-0.5	-0.2	0.25	0.04	0.10
0.00	0.80	0.0	0.1	0.00	0.01	0.00
1.00	1.00	1.0	0.3	1.00	0.09	0.30
2.00	1.30	2.0	0.6	4.00	0.36	1.20
				8.50	0.84	2.65

gde je $A \cdot B = (x_i - \bar{x}_6) \cdot (y_i - \bar{y}_6)$.

Standardne greške su:

$$\bar{s}_x = \sqrt{\frac{1}{n} \sum_{i=1}^6 (x_i - \bar{x}_6)^2} = \sqrt{\frac{1}{6} \cdot 8.50} = 1.19024;$$

$$\bar{s}_y = \sqrt{\frac{1}{n} \sum_{i=1}^6 (y_i - \bar{y}_6)^2} = \sqrt{\frac{1}{6} \cdot 0.84} = 0.37417;$$

$$\bar{s}_{xy} = \frac{1}{6} \sum_{i=1}^6 (x_i - \bar{x}_6)(y_i - \bar{y}_6) = \frac{1}{6} \cdot 2.65 = 0.44166.$$

$$\text{Odatle je } r = \frac{0.44166}{1.19024 \cdot 0.37417} = 0.9917;$$

$$a_1 = \frac{\bar{s}_{xy}}{\bar{s}_x^2} = \frac{0.44166}{1.19024^2} = 0.31176;$$

$$a_0 = \bar{y}_6 - a_1 \bar{x}_6 = 0.7 - 0.31176 \cdot 0 = 0.7.$$

Vidimo da jednačina prave linearne regresije glasi

$$\hat{y} = 0.31176x + 0.7 ,$$

tako da prognoziramo da će širina pukotine na 2.5°C biti

$$\hat{y}(2.5^{\circ}) = 0.31176 \cdot 2.5 + 0.7 = 1.4794 \text{ mm.}$$

Kako je r u ovom primeru po apsolutnoj vrednosti približno 1, linearna korelacija između posmatranih obeležja je veoma jaka. Dijagram rasipanja dobijamo ucrtavanjem tačaka $(-1.5, 0.2)$, $(-1.0, 0.4)$, $(-0.5, 0.5)$, $(0.0, 0.8)$, $(1.0, 1.0)$, $(2.0, 1.3)$ na xy -ravan. Jednačina regresije je jednačina prave. Uzmimo dve proizvoljne vrednosti za x , na primer $x_1 = 0$ i $x_2 = 1$.

Uvrštavanjem u jednačinu regresije dobijamo vrednosti $\hat{y}_1 = 0.7$ i $\hat{y}_2 = 1.0117$. Spajanjem tačaka $(0, 0.7)$ i $(1, 1.0117)$ dobijamo traženu regresionu pravu (Slika 1).

Vidimo da jednačina prave linearne regresije glasi

$$\hat{y} = 0.31176x + 0.7 ,$$

tako da prognoziramo da će širina pukotine na 2.5°C biti

$$\hat{y}(2.5^{\circ}) = 0.31176 \cdot 2.5 + 0.7 = 1.4794 \text{ mm.}$$

Kako je r u ovom primeru po apsolutnoj vrednosti približno 1, linearna korelacija između posmatranih obeležja je veoma jaka.

Dijagram rasipanja dobijamo ucrtavanjem tačaka $(-1.5, 0.2)$, $(-1.0, 0.4)$, $(-0.5, 0.5)$, $(0.0, 0.8)$, $(1.0, 1.0)$, $(2.0, 1.3)$ na xy -ravan. Jednačina regresije je jednačina prave. Uzmimo dve proizvoljne vrednosti za x , na primer $x_1 = 0$ i $x_2 = 1$.

Uvrštavanjem u jednačinu regresije dobijamo vrednosti $\hat{y}_1 = 0.7$ i $\hat{y}_2 = 1.0117$. Spajanjem tačaka $(0, 0.7)$ i $(1, 1.0117)$ dobijamo traženu regresionu pravu (Slika 1).

Vidimo da jednačina prave linearne regresije glasi

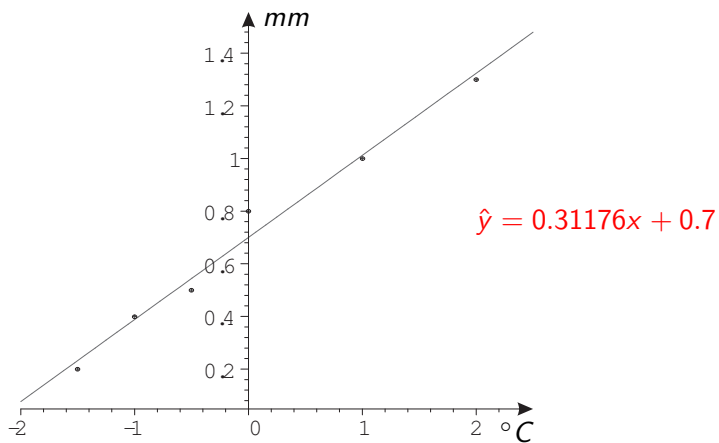
$$\hat{y} = 0.31176x + 0.7 ,$$

tako da prognoziramo da će širina pukotine na 2.5°C biti

$$\hat{y}(2.5^{\circ}) = 0.31176 \cdot 2.5 + 0.7 = 1.4794 \text{ mm}.$$

Kako je r u ovom primeru po apsolutnoj vrednosti približno 1, linearna korelacija između posmatranih obeležja je veoma jaka. Dijagram rasipanja dobijamo ucrtavanjem tačaka $(-1.5, 0.2)$, $(-1.0, 0.4)$, $(-0.5, 0.5)$, $(0.0, 0.8)$, $(1.0, 1.0)$, $(2.0, 1.3)$ na xy -ravan. Jednačina regresije je jednačina prave. Uzmimo dve proizvoljne vrednosti za x , na primer $x_1 = 0$ i $x_2 = 1$.

Uvrštavanjem u jednačinu regresije dobijamo vrednosti $\hat{y}_1 = 0.7$ i $\hat{y}_2 = 1.0117$. Spajanjem tačaka $(0, 0.7)$ i $(1, 1.0117)$ dobijamo traženu regresionu pravu (Slika 1).



Slika 1.