

## 13.6 LINEARNA KORELACIJA

---

- Koeficijent linearne korelacije
- Testiranje hipoteze o koeficijentu proste linearne korelacije

# Koeficijent linearne korelacije

---

Vrijednost koeficijenta proste linearne korelacije

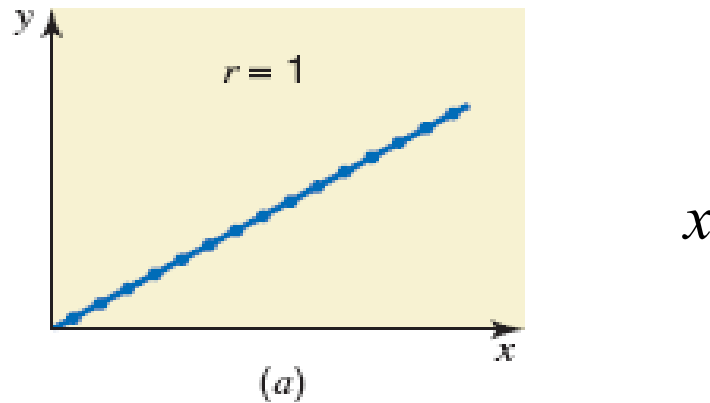
**Vrijednost koeficijenta proste linearne korelacije** je uvijek u intervalu od  $-1$  do  $1$ ; odnosno,

$$-1 \leq \rho \leq 1 \quad \text{i} \quad -1 \leq r \leq 1$$

# Slika 13.18 Linearna korelacija između dvije promjenljive.

---

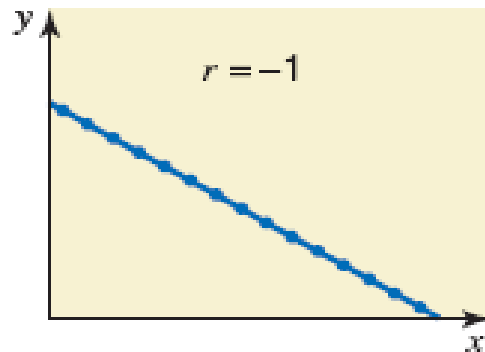
(a) Perfektna pozitivna linearna korelacija,  $r = 1$



# Slika 13.18 Linearna korelacija između dvije promjenljive.

---

(b) Perfektna negativna linearna korelacija,  $r = -1$



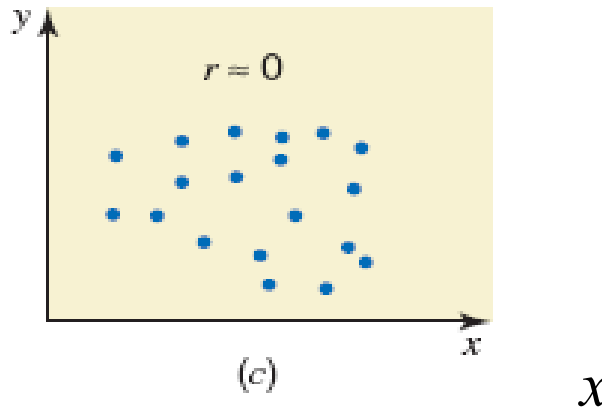
(b)

$x$

# Slika 13.18 Linearna korelacija između dvije promjenljive.

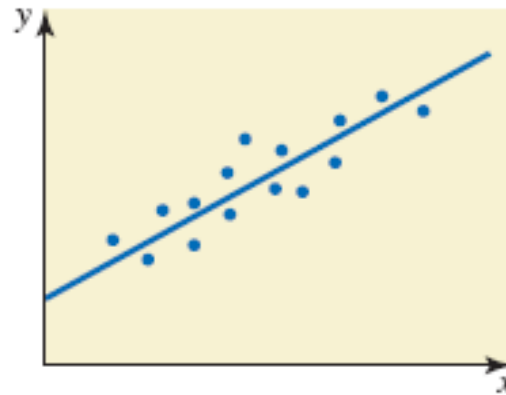
---

(c) Nepostojanje linearne korelacije,  $r \approx 0$



# Slika 13.19 Linearna korelacija između dvije promjenljive.

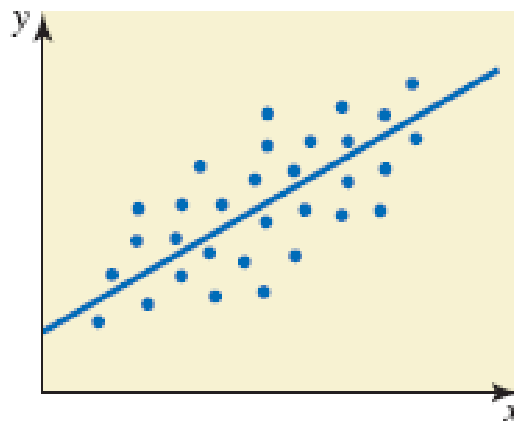
---



(a) Strong positive linear correlation  
( $r$  is close to 1)

# Slika 13.19 Linearna korelacija između dvije promjenljive.

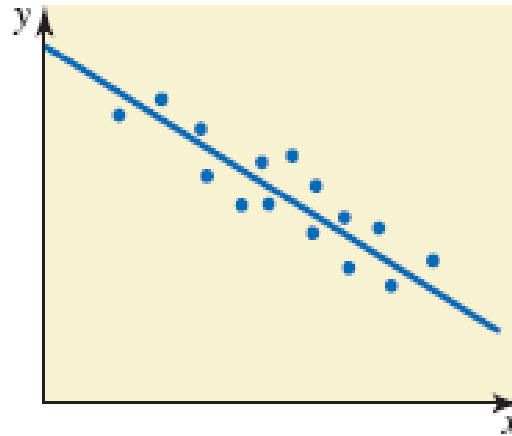
---



(b) Weak positive linear correlation  
( $r$  is positive but close to zero)

# Slika 13.19 Linearna korelacija između dvije promjenljive.

---

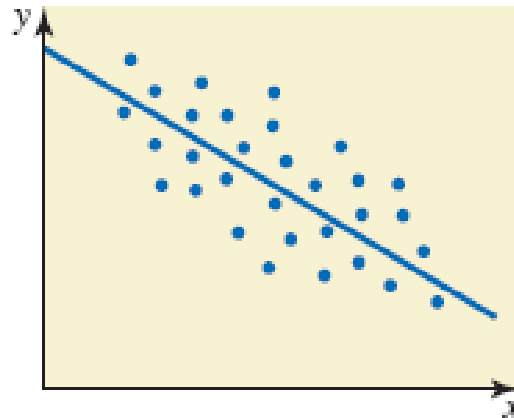


(c) Strong negative linear correlation  
( $r$  is close to  $-1$ )



# Slika 13.19 Linearna korelacija između dvije promjenljive.

---



(d) Weak negative linear correlation  
( $r$  is negative and close to zero)

## Koeficijent proste linearne korelacije

---

Koeficijent proste linearne korelacije

**Koeficijent proste linearne korelacije**,

označen sa  $r$ , mjeri jačinu linearne veze između dvije promjenljive u uzorku i glasi:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}}$$

## Primjer 13-6

---

Izračunati koeficijent korelacije za primjer o dohotku i izdacima za hranu sedam domaćinstava.

## Primjer 13-6: Rješenje

---

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}}$$
$$= \frac{447.5714}{\sqrt{(1772.8571)(125.7143)}} = .95$$

# Testiranje hipoteze o koeficijentu proste linearne korelacije

---

## Statistika testa za $r$

Pod pretpostavkom da je zajednički raspored dvije promjenljive normalan i za  $H_0: \rho = 0$ , statistika  $t$  testa glasi

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$

Ovdje oznaka  $n - 2$  predstavlja stepene slobode.

## Primjer 13-7

---

Koristeći nivo značajnosti 1% i podatke iz primjera 13-1, testirati da li je koeficijent linearne korelacije između dohotka i izdataka za hranu statistički značajan. Pretpostavka je da je zajednička raspodjela dvije promjenljive normalna.

## Primjer 13-7: Rješenje

---

- Etapa 1:
- $H_0: \rho = 0$  (koeficijent linearne korelacije je jednak nuli)
- $H_1: \rho \neq 0$  (koeficijent linearne korelacije je različit od nule)
  
- Etapa 2: Pod pretpostavkom da je zajednička raspodjela dvije promjenljive normalna, za testiranje hipoteze o koeficijentu proste linearne korelacije koristimo  $t$  raspodjelu.

## Primjer 13-7: Rješenje

---

- Etapa 3:
- Površina na oba kraja raspodjele =  $0.01/2$   
=  $0.005$
- $df = n - 2 = 7 - 2 = 5$
- Kritična vrijednost  $t = 4.032$



## Primjer 13-7: Rješenje

---

Etapa 4:

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$
$$= .95 \sqrt{\frac{7-2}{1-(.95)^2}} = 6.803$$

## Primjer 13-7: Rješenje

---

- Etapa 5:
- Vrijednost statistike testa  $t = 6.803$ 
  - Veća je od kritične vrijednosti  $t=4.032$
  - Nalazi se u oblasti odbacivanja
- Dakle, odbacujemo nultu hipotezu
- Zaključujemo da postoji pozitivna veza između dohotka i izdataka za hranu.

## 13.7 REGRESIONA ANALIZA: KOMPLETAN PRIMJER

---

### Primjer 13-8

Izabran je slučajnan uzorak od osam vozača koji posjeduju polise auto osiguranja. U sledećoj tabeli nalaze se podaci o vozačkom iskustvu (u godinama) i iznosu mjesečne premije auto osiguranja (u dolarima).

## Primjer 13-8

---

<b>Driving Experience (years)</b>	<b>Monthly Auto Insurance Premium (\$)</b>
5	64
2	87
12	50
9	71
15	44
6	56
25	42
16	60

## Primjer 13-8

---

- a) Da li premija osiguranja zavisi od vozačkog iskustva ili vozačko iskustvo zavisi od premije osiguranja? Da li očekujete pozitivnu ili negativnu vezu između ove dvije promjenljive?
- b) Izračunajte  $SS_{xx}$ ,  $SS_{yy}$  i  $SS_{xy}$ .
- c) Odrediti regresionu pravu po metodu najmanjih kvadrata na osnovu identifikovane objašnjavajuće i zavisne promjenljive u dijelu zadatka pod a.
- d) Objasnite značenje dobijenih ocijenjenih vrijednosti a i b izračunatih u dijelu pod c.

## Primjer 13-8

---

- e) Nacrtajte dijagram raspršenosti i ucrtajte ocijenjenu regresionu pravu.
- f) Izračunajte koeficijente  $r$  i  $r^2$  i objasnite njihovo značenje.
- g) Ocijenite mjesečnu premiju auto osiguranja vozača sa 10 godina iskustva u vožnji.
- h) Izračunajte standardnu grešku regresije.
- i) Formirajte 90% interval povjerenja za parametar  $B$ .
- j) Testirati sa nivoom značajnosti od 5% da li je  $B$  statistički značajno.
- k) Uz  $\alpha = 0.05$ , testirati da li se koeficijent  $\rho$  značajno razlikuje od nule.

## Primjer 13-8: Rješenje

---

- a) Na osnovu teorije i intuitivno, očekujemo da će premija osiguranja zavisi od vozačkog iskustva
- Premija osiguranja je zavisna promjenljiva
  - Vozačko iskustvo je objašnjavajuća promjenljiva

# Tabela 13.5

Experience	Premium				
$x$	$y$	$xy$	$x^2$	$y^2$	
5	64	320	25	4096	
2	87	174	4	7569	
12	50	600	144	2500	
9	71	639	81	5041	
15	44	660	225	1936	
6	56	336	36	3136	
25	42	1050	625	1764	
16	60	960	256	3600	
$\Sigma x = 90$	$\Sigma y = 474$	$\Sigma xy = 4739$	$\Sigma x^2 = 1396$	$\Sigma y^2 = 29,642$	



## Primjer 13-8: Rješenje

---

$$\text{b) } \bar{x} = \sum x / n = 90 / 8 = 11.25$$

$$\bar{y} = \sum y / n = 474 / 8 = 59.25$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 4739 - \frac{(90)(474)}{8} = -593.5000$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 1396 - \frac{(90)^2}{8} = 383.5000$$

$$SS_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} = 29,642 - \frac{(474)^2}{8} = 1557.5000$$

## Primjer 13-8: Rješenje

---

c)

$$b = \frac{SS_{xy}}{SS_{xx}} = \frac{-593.5000}{383.5000} = -1.5476$$

$$a = \bar{y} - b\bar{x} = 59.25 - (-1.5476)(11.25) = 76.6605$$

$$\hat{y} = 76.6605 - 1.547x$$

## Primjer 13-8: Rješenje

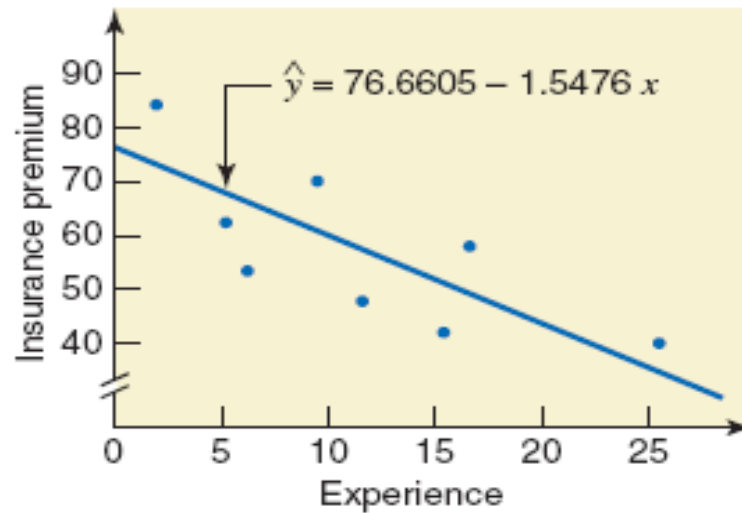
---

d) Ocijenjena vrijednost  $a = 76.6605$  predstavlja vrijednost  $\hat{y}$  za  $x = 0$ ; odnosno, prosječni mjesečni nivo premije osiguranja za vozača bez vozačkog iskustva.

Ocijenjena vrijednost  $b = -1.5476$  pokazuje da sa porastom vozačkog iskustva za 1 godinu, mjesečna premija osiguranja u prosjeku opada za \$1.55.

# Slika 13.21 Dijagram raspršenosti i regresiona prava.

e) Regresiona prava je opadajuća.



## Primjer 13-8: Rješenje

---

f)

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{-593.5000}{\sqrt{(383.5000)(1557.5000)}} = -.77$$

$$r^2 = \frac{bSS_{xy}}{SS_{yy}} = \frac{(-1.5476)(-593.5000)}{1557.5000} = .59$$

## Primjer 13-8: Rješenje

---

f) Vrijednost  $r = -0.77$  pokazuje da između vozačkog iskustva i mjesečne premije auto osiguranja postoji jaka negativna korelaciona veza.

Vrijednost  $r^2 = 0.59$  pokazuje da je 59% ukupnih varijacija mjesečne premije objašnjeno vozačkim iskustvom, dok je ostatak od 41% rezultat uticaja drugih faktora.

## Primjer 13-8: Rješenje

---

- g) Na osnovu regresione prave uzorka, ocijenjena vrijednost  $y$  za  $x = 10$  iznosi

$$\hat{y} = 76.6605 - 1.5476(10) = \$61.18$$

Dakle, očekujemo da će mjesečni iznos premije auto osiguranja vozača sa 10 godina vozačkog iskustva biti \$61.18.

## Primjer 13-8: Rješenje

---

h)

$$\begin{aligned} s_e &= \sqrt{\frac{SS_{yy} - bSS_{xy}}{n - 2}} \\ &= \sqrt{\frac{1557.5000 - (-1.5476)(-593.5000)}{8 - 2}} \\ &= 10.3199 \end{aligned}$$



## Primjer 13-8: Rješenje

---

$$i) \quad s_b = \frac{s_e}{\sqrt{SS_{xx}}} = \frac{10.3199}{\sqrt{383.5000}} = .5270$$

$$\alpha / 2 = .5 - (.90 / 2) = .05$$

$$df = n - 2 = 8 - 2 = 6$$

$$t = 1.943$$

$$b \pm ts_b = -1.5476 \pm 1.943(.5270)$$

$$= -1.5476 \pm 1.0240 = -2.57 \quad \text{to} \quad -.52$$

## Primjer 13-8: Rješenje

---

j)

- Etapa 1:
- $H_0: B = 0$  ( $B$  je jednak nuli)
- $H_1: B \neq 0$  ( $B$  je različito od nule)

## Primjer 13-8: Rješenje

---

- Etapa 2: Budući da standardna devijacija slučajne greške nije poznata, koristimo  $t$  raspodjelu za testiranje
- Etapa 3:
- Površina na oba kraja raspodjele =  $\alpha/2 = 0.05/2 = 0.025$
- $df = n - 2 = 8 - 2 = 6$
- Kritične vrijednosti  $t$  su  $-2.447$  i  $2.447$

## Primjer 13-8: Rješenje

---

Etapa 4:

$$t = \frac{b - B}{s_b} = \frac{-1.5476 - 0}{.5270} = -2.937$$

Iz  $H_0$

## Primjer 13-8: Rješenje

---

- Etapa 5:
- Vrijednost statistike testa  $t = -2.937$ 
  - Nalazi se u oblasti odbacivanja
- Dakle, odbacujemo nultu hipotezu i zaključujemo da je parametar  $B$  statistički značajan
- Mjesečni iznos premije auto osiguranja u osnovnom skupu vozača u prosjeku opada sa povećanjem vozačkog iskustva.

## Primjer 13-8: Rješenje

---

- k)
- Etapa 1:
  - $H_0: \rho = 0$  (koeficijent linearne korelacije je jednak nuli)
  - $H_1: \rho \neq 0$  (koeficijent linearne korelacije je različit od nule)

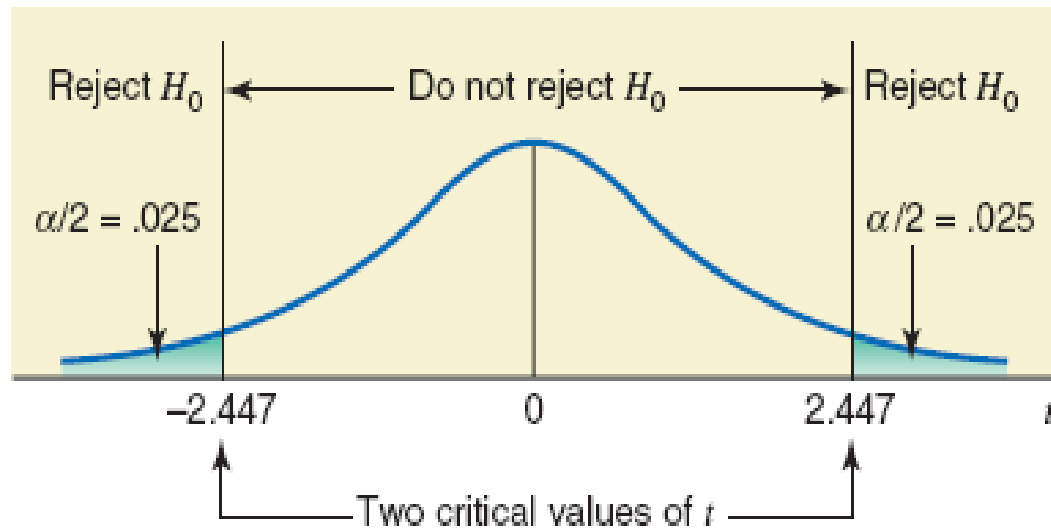
## Primjer 13-8: Rješenje

---

- Etapa 2: Pod pretpostavkom da je zajednička raspodjela dvije promjenljive  $x$  i  $y$  normalna, za testiranje hipoteze o koeficijentu linearne korelacije koristimo  $t$  raspodjelu.
- Etapa 3:
- Površina na svakom kraju =  $0.05/2 = 0.025$
- $df = n - 2 = 8 - 2 = 6$
- Kritične vrijednosti  $t$  su  $-2.447$  i  $2.447$

## Slika 13.23

---





## Primjer 13-8: Rješenje

---

Etapa 4:

$$t = r \sqrt{\frac{n-2}{1-r^2}}$$
$$= -.77 \sqrt{\frac{8-2}{1-(-.77)^2}} = -2.956$$

## Primjer 13-8: Rješenje

---

- Etapa 5:
- Vrijednost statistike testa  $t = -2.956$ 
  - Nalazi se u oblasti odbacivanja
- Dakle, odbacujemo nultu hipotezu
- Zaključujemo da se koeficijent proste linearne korelacije između godina vozačkog iskustva i visine premije osiguranja u osnovnom skupu vozača statistički značajno razlikuje od nule.

## 13.9 PROBLEMI U KORIŠĆENJU REGRESIONOG MODELA

---

- ❑ Ekstrapolacija: Regresiona prava ocijenjena na osnovu uzorka odnosi se samo na opseg vrijednosti  $x$  koje su obuhvaćene uzorkom.
- ❑ Uzročnost: Regresiona prava ne pokazuje uzročnost između dvije promjenljive: odnosno, ne predviđa da je promjena  $y$  *uzrokovana* promjenom  $x$ .