

# Statistika

2016/17

Regresiona i korelaciona analiza

# Regresiona i korelaciona analiza



**Oblik povezanosti**  
između zavisne i  
nezavisne(ih) promenljive



**Jačina povezanosti**  
između promenljivih  
[kod više od dve promenljive neophodno je  
definisati zavisnu promenljivu]

Veza između promenljivih može biti

- deterministička (funkcionalna)  $y = f(x)$   
[podrazumeva perfektno slaganje podataka]
- stohastička (statistička)  
[slaganje podataka u proseku]

# Regresiona i korelaciona analiza

Prosta regresiona i korelaciona analiza:

Linearna      Krivolinijska

Višestruka regresiona i korelaciona analiza  
(prostorna povezanost):

Linearna      Nelinearna

# Prosta linearna regresija i korelacija

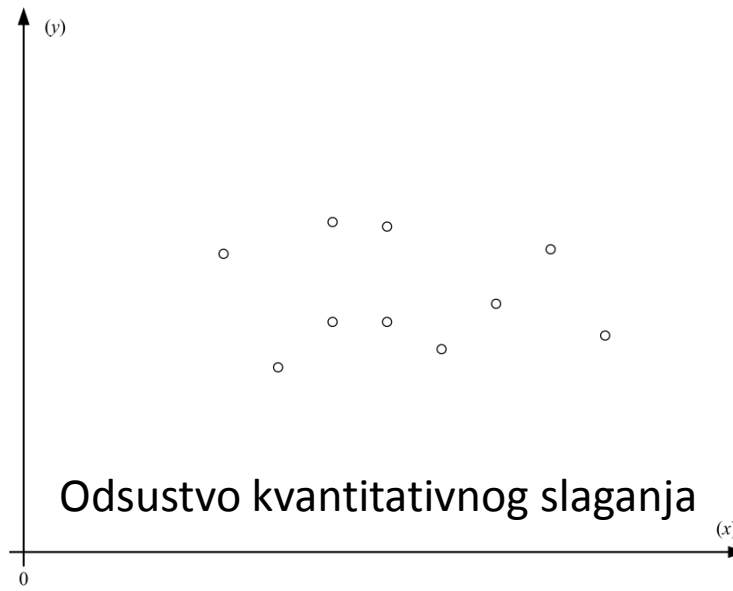
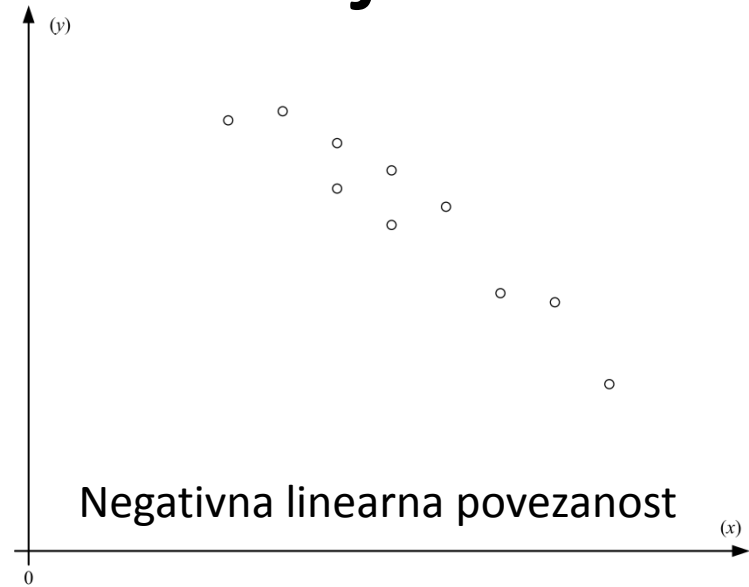
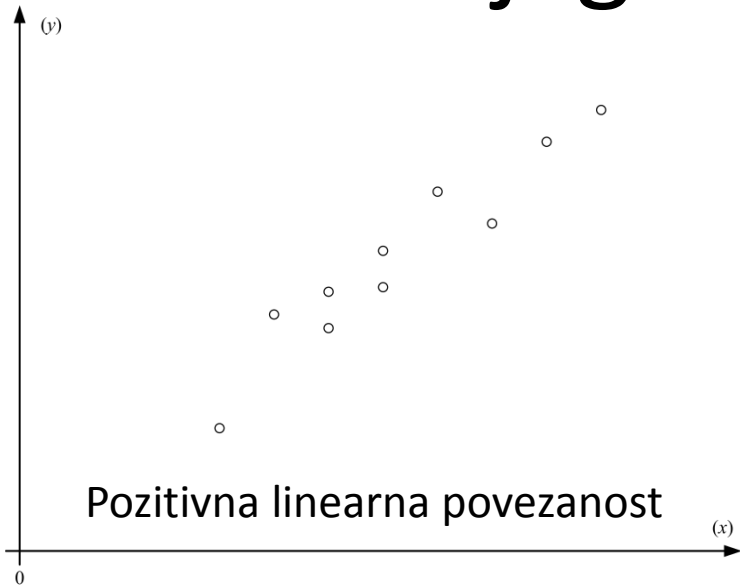
"Parametri" proste linearne regresije

- Dijagram rasturanja
- Jednačina regresije
- Standardna greška

# Dijagram rasturanja

- Da li postoji kvantitativno slaganje?
- Linearni ili krivolinijski oblik?
- Smer (samo za linearnu vezu)?

# Dijagram rasturanja



# Jednačina regresije

Za osnovni skup (cilj):  $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$        $x_{i1.2} = A_{12} + B_{12} x_{i2} + \varepsilon_i$

$y$  – zavisna promenljiva

$x$  – nezavisna promenljiva

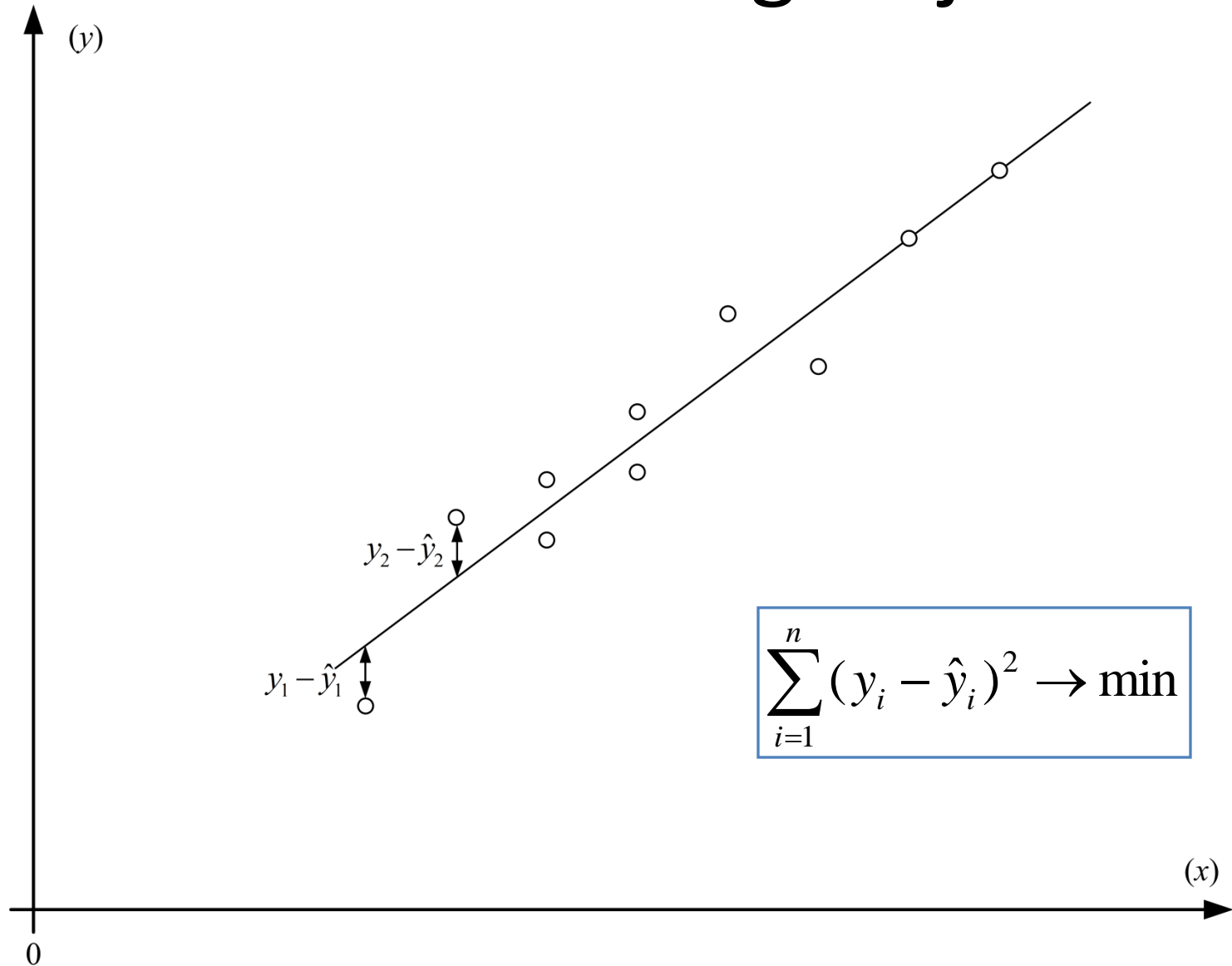
$\beta_0$  – presek linije regresije sa ordinatnom osom

$\beta_1$  – pokazuje za koliko se vrednost zavisne promenljive menja u proseku, ako se nezavisna promenljiva poveća za jednu jedinicu

$\varepsilon$  – slučajna greška

Za uzorak:  $\hat{y}_i = b_0 + b_1 x_i$        $\hat{x}_{i1.2} = a_{12} + b_{12} x_{i2}$

# Jednačina regresije





# Jednačina regresije

$$\sum_{i=1}^n y_i = nb_0 + b_1 \sum_{i=1}^n x_i$$

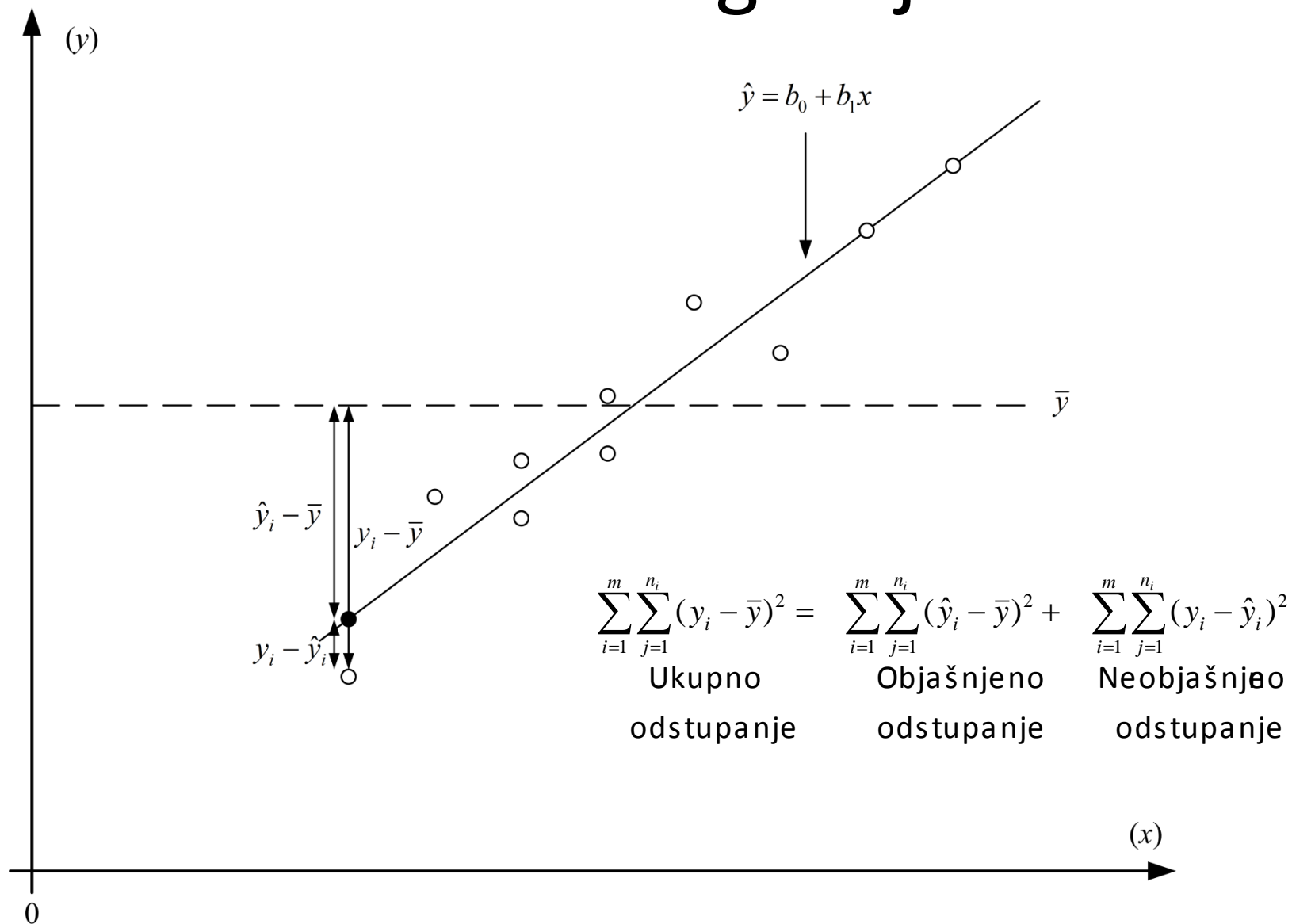
$$\sum_{i=1}^n x_i y_i = b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2$$

ili

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i}$$

# Jednačina regresije



# Standardna greška

Prosečno odstupanje originalnih vrednosti zavisne promenljive od linije regresije.

$$\sigma_{\hat{y}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}}$$

# Prosta linearna regresija i korelacija

## "Parametri" proste linearne korelacije

- Koeficijent determinacije

[Procenat objašnjenih varijacija  
zavisne promenljive preko linearne veze]

$$r^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{\sum_{i=1}^n (y_i - \bar{y})}$$

$$0 \leq r^2 \leq 1$$

- Koeficijent nedeterminacije  $k^2 = 1 - r^2$

[Procenat neobjašnjenih varijacija zavisne promenljive preko linearne veze]

- Koeficijent korelacije

[Stepen linearne povezanosti između zavisne i nezavisne promenljive]

$$r = \pm \sqrt{r^2}$$

$$-1 \leq r \leq +1$$

- Koeficijent alineacije

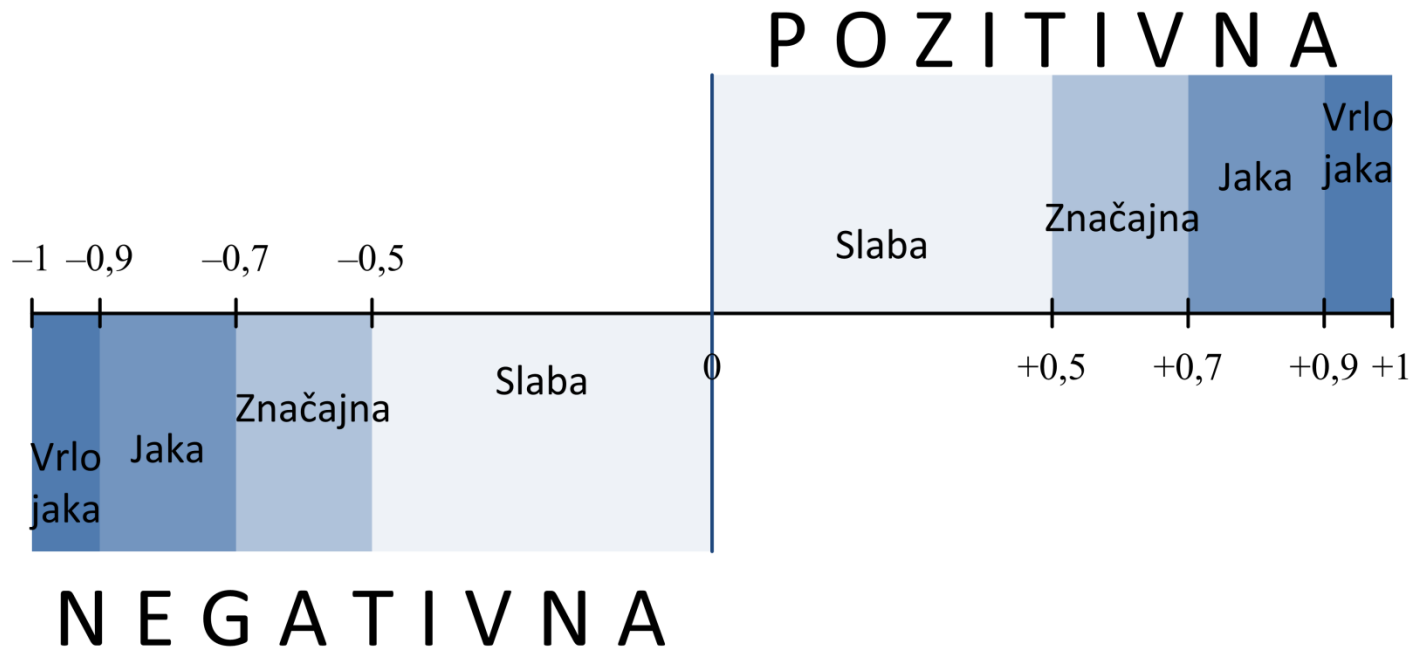
[Stepen nepostojanja linearne povezanosti između  
zavisne i nezavisne promenljive]

$$k = \sqrt{k^2}$$

# Koeficijent korelacije

$r = \pm \sqrt{r^2}$   
↑  
Znak  
regresionog  
koeficijenta

$0 <  r  \leq 0,5$	Slaba korelaciona veza
$0,5 <  r  \leq 0,7$	Značajna korelaciona veza
$0,7 <  r  \leq 0,9$	Jaka korelaciona veza
$0,9 <  r  \leq 1$	Vrlo jaka korelaciona veza



# Zadatak 1

Vreme pripreme kolokvijuma u  $h$  i broj osvojenih bodova kod 10 na slučajan način izabranih studenata:

Vreme pripreme:	11	14	17	17	20	20	23	26	29	32
Broj bodova:	27	52	49	57	66	58	79	72	89	98

Na osnovu ovih podataka:

- Formirati dijagram rasturanja.
- Odrediti jednačinu linearne regresije za zavisnost broja bodova od vremena pripreme ispita.
- Izračunati standardnu grešku.
- Odrediti parametre proste linearne korelacije.

# Zadatak 2

Starost automobila u godinama i prodajna cena u stotinama novčanih jedinica kod 8 na slučajan način izabranih vozila:

Starost:	5	6	7	7	8	8	9	10
Cena:	74	82	69	75	71	65	70	54

Na osnovu ovih podataka:

- Formirati dijagram rasturanja.
- Odrediti jednačinu linearne regresije za zavisnost cene od starosti vozila.
- Izračunati standardnu grešku.
- Odrediti parametre proste linearne korelacije.

# Spirmanov koeficijent korelacije ranga

- Mera povezanosti između rangova podataka
- Veza između originalnih podataka ne mora da bude linearna
- Najčešće se koristi na podacima sa ordinalne merne skale



# Kendalov koeficijent konkordacije

- Mera slaganja između rangova podataka  
(najčešće slaganja između mišljenja ispitanika)