

# OSNOVI STATISTIČKE ANALIZE

**Predavanje I dvočas**

***Višestruka regresija***

*(Uvod u statistiku, Poglavlje 14)*

## **VAŽNE NAPOMENE ZA POGLAVLJE 14:**

1. Pitanja iz 14. poglavlja neće biti zastupljena na usmenom ispitu zaključno sa januarsko-februarskim rokom školske 2020/2021.

2. Zadaci iz 14. poglavlja neće biti zastupljeni na pismenom ispitu zaključno sa januarsko-februarskim rokom školske 2020/2021.

# Višestruka linearna regresija

- Obraditi iz knjige odeljke 14.1-14.5 (str. 678-691)

➤ **Posebno obratiti pažnju na:**

- 1) definiciju višestrukog linearnog regresionog modela i interpretaciju parametara (str.679)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

- Deterministički deo modela je  $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$
- Slučajna greška je  $\varepsilon$
- Y je zavisna promenljiva, a  $X_1, X_2, \dots, X_k$  su objašnjavajuće promenljive
- $\beta_0$  -pokazuje **prosečnu** vrednost zavisne promenljive Y kada su sve objašnjavajuće promenljive jednake 0
- $\beta_1$  - pokazuje **prosečnu** promenu Y kada se  $X_1$  poveća za jednu svoju jedinicu, a sve ostale objašnjavajuće promenljive ostanu nepromenjene.
- $\beta_2$  - pokazuje **prosečnu** promenu Y kada se  $X_2$  poveća za jednu svoju jedinicu, a sve ostale objašnjavajuće promenljive ostanu nepromenjene.
- Jednostavno je zaključiti šta pokazuju i ostali regresioni koeficijenti  $\beta_3, \dots, \beta_k$ .

2) metod najmanjih kvadrata (str.680)

3) parametre, ocene, ocenjene vrednosti (str.679-680)

regresioni parametar	ocena	ocenjena vrednost
$\beta_0$	$b_0$	$b_0$
$\beta_1$	$b_1$	$b_1$
$\beta_2$	$b_2$	$b_2$
...	...	...
$\beta_k$	$b_k$	$b_k$

4) regresioni model uzorka (str. 615 i 618)

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

( $\hat{Y}$  predstavlja ocenjenu ili prilagođenu vrednost promenljive Y. Važi da je  $n > k + 1$ )

5) pretpostavke višestrukog linearnog regresionog modela (str. 680-681)

1. linearnost ( $E(Y) = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_kX_k$ ); 2. Y je slučajna promenljiva, jer predstavlja funkciju od  $\varepsilon$ , a  $X_1, X_2, \dots, X_k$  nisu slučajne promenljive; 3. očekivana vrednost slučajne greške je 0; 4. nema autokorelacije (slučajne greške su međusobno nezavisne); 5. homoskedastičnost (jednakost varijansi slučajnih grešaka); 6. normalnost raspodele slučajnih grešaka; 7. objašnjavajuće promenljive nisu međusobno linearno zavisne odnosno ne postoji problem multikolinearnosti.

6) standardnu grešku regresije kao apsolutnu meru reprezentativnosti (str. 681)

$$S = \sqrt{\frac{\text{SKR}}{n-k-1}} \quad \text{gde je} \quad \text{SKR} = \sum(Y - \hat{Y})^2$$

7) koeficijent višestruke determinacije kao relativnu meru reprezentativnosti (str. 682)

(pokazuje koliko je učešće objašnjenog varijabiliteta u ukupnom, odnosno koliki je deo varijacija zavisne promenljive objašnjen višestrukim regresionim modelom)

8) korigovani koeficijent višestruke determinacije (str. 682)

$$\bar{R}^2 = 1 - (1 - R^2) \left( \frac{n-1}{n-k-1} \right) \quad \text{ili} \quad \bar{R}^2 = 1 - \frac{\text{SKN} / (n-k-1)}{\text{SKU} / (n-1)}$$

9) broj stepeni slobode u višestrukoj regresiji (str. 679)

( $df=n-k-1$ )

10) intervale poverenja za regresione parametre (str. 688)

**Interval poverenja za parametar  $\beta_i$**   $(1 - \alpha) \times 100\%$  interval poverenja za parametar  $\beta_i$  glasi:

$$b_i \pm ts_{b_i}$$

$t$  vrednost u prethodnom izrazu dobija se iz Tablice Studentove  $t$  raspodele za površinu od  $\alpha/2$  na svakom kraju ispod krive  $t$  raspodele i  $(n - k - 1)$  stepena slobode. Vrednosti  $b_i$  i  $s_{b_i}$  se dobijaju korišćenjem statističkog softvera.

### 11) ocene $b_i$ imaju normalnu raspodelu (str. 689)

Zbog pretpostavke o normalnoj raspodeli slučajnih grešaka, uzoračka raspodela svake ocene  $b_i$  je takođe normalna sa srednjom vrednošću  $\beta_i$  i standardnom greškom,  $\sigma_{b_i}$ . Kao što je ranije objašnjeno, standardna greška  $\sigma_{b_i}$  nije poznata, pa zato koristimo njenu ocenu iz uzorka  $S_{b_i}$ , a testiranje hipoteze o parametru  $\beta_i$  zasnivamo na  $t$  raspodeli.

### 12) testiranje hipoteza o regresionim parametrima (str. 689)

**Statistika testa za regresioni parametar  $\beta_i$**  Za testiranje hipoteze o regresionom parametru  $\beta_i$  koristi se sledeća statistika testa:

$$t = \frac{b_i - \beta_i}{S_{b_i}}$$

Vrednost regresionog parametra  $\beta_i$  u prethodnom izrazu zamenjuje se hipotetičkom vrednošću, koja je definisana nultom hipotezom. Nultu hipotezu najčešće formulišemo u sledećem obliku:  $H_0: \beta_i = 0$ . Softver MINITAB sadrži i proceduru testiranja korišćenjem statistike  $t$  testa.

### 13) Formulisanje alternativne hipoteze

$H_1$  se formuliše na jedan od sledeća tri načina:

- (1)  $H_1: \beta_i \neq 0$  ( $X_i$  utiče na  $Y$  u osnovnom skupu)
- (2)  $H_1: \beta_i > 0$  ( $X_i$  pozitivno utiče na  $Y$ ) i
- (3)  $H_1: \beta_i < 0$  ( $X_i$  negativno utiče na  $Y$ ).

## Višestruki linearni regresioni model sa dve objašnjavajuće promenljive (dodatna objašnjenja)

- **regresioni model skupa je**  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
- **regresiona ravan osnovnog skupa je**  $\mu_{Y|X} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ , gde je  $\mu_{Y|X}$ , odnosno  $E(Y)$ , prosečna (očekivana) vrednost promenljive  $Y$
- **$\beta_0$  je odsečak**-pokazuje **prosečnu** vrednost zavisne promenljive  $Y$  kada su i  $X_1$  i  $X_2$  jednaki 0;
- **$\beta_1$**  - pokazuje prosečnu promenu promenljive  $Y$  nastalu usled jedinične promene  $X_1$  pod uslovom da  $X_2$  ostane nepromenjena.
- **$\beta_2$**  -pokazuje prosečnu promenu promenljive  $Y$  nastalu usled jedinične promene  $X_2$  pod uslovom da  $X_1$  ostane nepromenjena.
- **regresioni model uzorka** (regresiona ravan uzorka)  $\hat{Y} = \mathbf{b}_0 + \mathbf{b}_1 X_1 + \mathbf{b}_2 X_2$

# Ilustracija

U jednom istraživanju analiziran je uticaj izvoznih cena i mesečnog fizičkog obima proizvodnje proizvoda A na mesečnu vrednost izvoza tog proizvoda (u mil.din.) na uzorku od 7 preduzeća na teritoriji Srbije. Osnovni rezultati analize prikazani su u sledećoj tabeli:

	Koeficijent	Standardna greška	t-vredn.
Odsečak	4,50	1,5	3,0
Izvozna cena (u 000 din)	1,01	0,1	10,1
Proizvodnja (u tonama)	0,55	0,19	2,9
Koeficijent determinacije=0,65;		Standardna greška regresije = 8,43	

Uočavamo da je zavisna promenljiva Y- vrednost izvoza proizvoda (u mil.din.), a da su objašnjavajuće promenljive  $X_1$ - izvozna cena (u 000 din.) i  $X_2$ - proizvodnja (u tonama). Ocenjene vrednosti regresionih parametara su  $b_0 = 4,5$ ,  $b_1 = 1,01$  i  $b_2 = 0,55$  a regresiona jednačina glasi:

$$\hat{y} = 4,5 + 1,01x_1 + 0,55x_2.$$

**$b_0 = 4,5$**  → ako bi izvozna cena iznosila 0 din i ako ne bi bilo proizvodnje prosečna mesečna vrednost izvoza bi iznosila 4,5 mil.din. Treba napomenuti da je ovo samo ilustracija tumačenja ocenjene vrednosti  $b_0$ , jer ovaj rezultat nema ekonomski smisao.



$b_1 = 1,01$  → ako se izvozna cena poveća za 1000 din, a proizvodnja ostane nepromenjena, vrednost izvoza proizvoda će se u proseku povećati za 1,01 mil.din.

$b_2 = 0,55$  → ako se fizički obim proizvodnje poveća za 1t, a izvozna cena ostane nepromenjena, vrednost izvoza će se u proseku povećati za 0,55 mil.din.

**Standardna greška regresije  $S=8,43$**  → apsolutna mera varijacija empirijskih podataka od regresione ravni uzorka (vrednost izvoza proizvoda pojedinih preduzeća u uzorku u proseku odstupa od ocenjenih vrednosti izvoza za 8,43 mil.din).

**Koeficijent višestruke determinacije  $R^2 = 0,65$**  → 65% varijacija vrednosti izvoza proizvoda je objašnjeno uticajem izvoznih cena i mesečnog fizičkog obima proizvodnje. **Nedostatak koeficijenta determinacije** je što zavisi od broja promenljivih i veličine uzorka. Ako je  $n$  malo a posmatra se veliki broj objašnjavajućih promenljivih, vrednost  $R^2$  je nerealan visok. Takođe uključivanjem nove objašnjavajuće promenljivih u model vrednost  $R^2$  se povećava bez obzira na uticaj te promenljive.

**Korigovani koeficijent višestruke determinacije  $\bar{R}^2 = 0,475$**  → vrednost koeficijenta višestruke determinacije korigovanog za broj stepeni slobode. Nakon korekcije proizilazi da pomenute dve objašnjavajuće promenljive objašnjavaju 47,5% varijacija zavisne promenljive.

U koloni **Standardna greška** u tabeli brojevi pokazuju sledeće:

**1,5** → ocenjena vrednost standardne greške ocene  $b_0$

**0,1** → ocenjena vrednost standardne greške ocene  $b_1$

**0,19** → ocenjena vrednost standardne greške ocene  $b_2$

## Testiranje hipoteza o regresionom parametru $\beta_1$

$H_0: \beta_1 = 0$  (promenljiva  $X_1$  ne utiče na promenljivu  $Y$ ).

$H_1$  se formuliše na jedan od sledeća tri načina:

(1)  $H_1: \beta_1 \neq 0$  (  $X_1$  utiče na  $Y$  u osnovnom skupu)

(2)  $H_1: \beta_1 > 0$  (  $X_1$  pozitivno utiče na  $Y$ ) i

(3)  $H_1: \beta_1 < 0$  (  $X_1$  negativno utiče na  $Y$ ).

Realizovana vrednost statistike  $t$  testa u prethodnoj ilustraciji je  $t = \frac{b_1}{s_{b_1}} = 10,1$

## Testiranje hipoteza o regresionom parametru $\beta_2$

$H_0: \beta_2 = 0$  (promenljiva  $X_2$  ne utiče na promenljivu  $Y$ ).

$H_1$  se formuliše na jedan od sledeća tri načina:

(1)  $H_1: \beta_2 \neq 0$  (  $X_2$  utiče na  $Y$  u osnovnom skupu)

(2)  $H_1: \beta_2 > 0$  (  $X_2$  pozitivno utiče na  $Y$ ) i

(3)  $H_1: \beta_2 < 0$  (  $X_2$  negativno utiče na  $Y$ ).

- Realizovana vrednost statistike  $t$  testa u prethodnoj ilustraciji je  $t = \frac{b_2}{s_{b_2}} = 2,9$