Klasični višestruki linearni regresioni model

Zorica Mladenović

Struktura

- Osnove višestrukog linearnog regresionog modela
- Dvostruki linearni regresioni model
- Ocene parametara primenom metoda ONK (na vežbama)
- Koeficijent determinacije u višestrukom KLRM
- Pretpostavke višestrukog KLRM
- Svojstva ocena dobijenih primenom metoda ONK u višestrukom KLRM (bez dokaza)
- Statističko zaključivanje u višestrukom KLRM

- Osnove

Klasični višestruki linearni regresioni model: osnove

- Na jednom od prethodnih primera ispitivali smo zavisnost tražnje datog proizvoda od njegove cene
$$Y_i = \beta_0 + \beta X_i + \varepsilon_i, i = 1,2,\ldots, n.$$
- Ovakva postavka može biti suviše restriktivna.
- Tražnja za datim proizvodom može zavisiti i od:
  1. dohotka potrošača,
  2. cena konkurentnih proizvoda i
  3. opšteg indeksa cena.
- Neophodno je napraviti uopštenje linearnog regresionog modela: izabrana zavisna promenljiva zavisi od većeg broja nezavisnih promenljivih.
- Time dolazimo do višestrukog linearnog regresionog modela.

Postavka modela

- Analitički oblik višestrukog linearnog regresionog modela (ukupan broj parametara je k):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_{k-1} X_{k-1} + \varepsilon$$

- Parametri modela: $\beta_1, \beta_2, ..., \beta_{k-1}$ su parcijalni koeficijenti nagiba
- Na primer: ako se X_1 poveća za jednu jedinicu, očekivana promena Y je β_1 jedinica, pod pretpostavkom da se ne menja uticaj ostalih objašnjavajućih promenljivih
- Parametar slobodnog člana: β_0

- Dvostruki linearni regresioni model

Dvostruki linearni regresioni model

- Najjednostavniji je model sa dve objašnjavajuće promenljive:

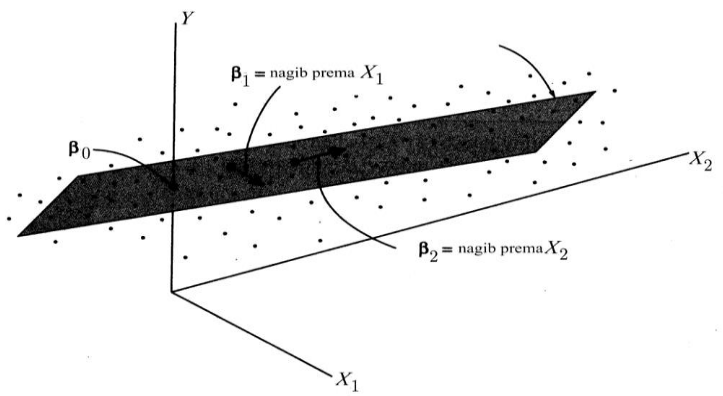
$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\varepsilon$$

- **Populaciona regresiona jednačina** (za $E(\varepsilon)=0$) je:

$$E(Y)=\beta_0+\beta_1X_1+\beta_2X_2$$

- Parametri β_0, β_1 i β_2 su populacioni parametri ili regresioni koeficijenti.
- Populaciona **regresiona jednačina** ne opisuje pravu, nego **ravan.**
- Parametar β_0 je odsečak (presek ravni sa y-osom).
- Parametri β_1 i β_2 **parcijalni koeficijenti nagiba.**

Grafički prikaz modela



Dvostruki linearni regresioni model II

- Uključivanjem konkretnih podataka model postaje:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, \quad za\ i = 1,2,\ldots,n.$$

- **Uzoračka regresiona funkcija je**:

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i}$$

gde su b_0, b_1 i b_2 ocene parametara.

- Ocene nepoznatih parametara dobijaju se primenom metoda ONK.

- Metod ONK

Izvođenje ONK ocena

- Potrebno je minimizirati zbir (tzv. rezidualnu sumu kvadrata):

$$e_1^2 + e_2^2 + \ldots + e_n^2 = \sum_{i=1}^{n} e_i^2$$

- Šta je e_i? To je razlika $e_i = Y_i - \hat{Y}_i$
- Naći minimum $\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$ je ekvivalentno određivanju minimuma $\sum_{i=1}^{n} e_i^2$

11

Izvođenje ocena metoda ONK (II)

- Kako je $\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i}$, možemo definisati funkciju

$$S(b_0, b_1, b_2) = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i})^2$$

- Potrebno je minimizirati funkciju S u odnosu na b_0 , b_1 i b_2 :

$$\frac{\partial S(b_0, b_1, b_2)}{\partial b_0} = -2 \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i}) = 0 \qquad (1)$$

$$\frac{\partial S(b_0, b_1, b_2)}{\partial b_1} = -2 \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i}) X_{1i} = 0 \qquad (2)$$

$$\frac{\partial S(b_0, b_1, b_2)}{\partial b_2} = -2 \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i}) X_{2i} = 0 \qquad (3)$$

12

Izvođenje ocena metoda ONK (III)

- Iz (1): $\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i}) = 0$

$$\Rightarrow$$

$$\sum_{i=1}^{n} Y_i - nb_0 - b_1 \sum_{i=1}^{n} X_{1i} - b_2 \sum_{i=1}^{n} X_{2i} = 0$$

$$b_0 = \overline{Y} - b_1 \overline{X}_1 - b_2 \overline{X}_2 \qquad (4)$$

- Zamenom (4) u (2) i (3)

$$\sum_{i=1}^{n}\left(Y_i - \left(\overline{Y} - b_1 \overline{X}_1 - b_2 \overline{X}_2\right) - b_1 X_{1i} - b_2 X_{2i}\right) X_{1i} = 0 \qquad (5)$$

$$\sum_{i=1}^{n}\left(Y_i - \left(\overline{Y} - b_1 \overline{X}_1 - b_2 \overline{X}_2\right) - b_1 X_{1i} - b_2 X_{2i}\right) X_{2i} = 0 \qquad (6)$$

13

Izvođenje ocena metoda ONK (IV)

$$\sum_{i=1}^{n}\left(\left(Y_i - \overline{Y}\right) - b_1\left(X_{1i} - \overline{X}_1\right) - b_2\left(X_{2i} - \overline{X}_2\right)\right) X_{1i} = 0 \qquad (7)$$

$$\sum_{i=1}^{n}\left(\left(Y_i - \overline{Y}\right) - b_1\left(X_{1i} - \overline{X}_1\right) - b_2\left(X_{2i} - \overline{X}_2\right)\right) X_{2i} = 0 \qquad (8)$$

Znamo da je:

$$\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right) X_{1i} = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(X_{1i} - \overline{X}_1\right)$$

$$\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right) X_{2i} = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(X_{2i} - \overline{X}_1\right)$$

$$\sum_{i=1}^{n}\left(X_{1i} - \overline{X}_1\right) X_{1i} = \sum_{i=1}^{n}\left(X_{1i} - \overline{X}_1\right)^2, \sum_{i=1}^{n}\left(X_{2i} - \overline{X}_2\right) X_{2i} = \sum_{i=1}^{n}\left(X_{2i} - \overline{X}_2\right)^2$$

$$\sum_{i=1}^{n}\left(X_{1i} - \overline{X}_1\right) X_{2i} = \sum_{i=1}^{n}\left(X_{1i} - \overline{X}_1\right)\left(X_{2i} - \overline{X}_2\right),$$

$$\sum_{i=1}^{n}\left(X_{2i} - \overline{X}_2\right) X_{1i} = \sum_{i=1}^{n}\left(X_{1i} - \overline{X}_1\right)\left(X_{2i} - \overline{X}_2\right).$$

14

Izvođenje ocena metoda ONK: sistem normalnih jednačina

$$b_1 \sum_{i=1}^{n} \left(X_{1i} - \overline{X}_1\right)^2 + b_2 \sum_{i=1}^{n} \left(X_{1i} - \overline{X}_1\right)\left(X_{2i} - \overline{X}_2\right) = \sum_{i=1}^{n} \left(X_{1i} - \overline{X}_1\right)\left(Y_i - \overline{Y}\right)$$

$$b_1 \sum_{i=1}^{n} \left(X_{1i} - \overline{X}_1\right)\left(X_{2i} - \overline{X}_2\right) + b_2 \sum_{i=1}^{n} \left(X_{2i} - \overline{X}_2\right)^2 = \sum_{i=1}^{n} \left(X_{2i} - \overline{X}_2\right)\left(Y_i - \overline{Y}\right)$$

Notacija za centrirane podatke :

$$\sum_{i=1}^{n} x_{1i}^2 = \sum_{i=1}^{n} \left(X_{1i} - \overline{X}_1\right)^2, \sum_{i=1}^{n} x_{2i}^2 = \sum_{i=1}^{n} \left(X_{2i} - \overline{X}_2\right)^2,$$

$$\sum_{i=1}^{n} x_{1i} x_{2i} = \sum_{i=1}^{n} \left(X_{1i} - \overline{X}_1\right)\left(X_{2i} - \overline{X}_2\right)$$

$$\sum_{i=1}^{n} x_{1i} y_i = \sum_{i=1}^{n} \left(X_{1i} - \overline{X}_1\right)\left(Y_i - \overline{Y}\right), \sum_{i=1}^{n} x_{2i} y_i = \sum_{i=1}^{n} \left(X_{2i} - \overline{X}_2\right)\left(Y_i - \overline{Y}\right)$$

15

Izvođenje ocena metoda ONK: najčešća forma

$$b_1 \sum_{i=1}^{n} x_{1i}^2 + b_2 \sum_{i=1}^{n} x_{1i} x_{2i} = \sum_{i=1}^{n} x_{1i} y_i$$

$$b_1 \sum_{i=1}^{n} x_{1i} x_{2i} + b_2 \sum_{i=1}^{n} x_{2i}^2 = \sum_{i=1}^{n} x_{2i} y_i$$

$$b_0 = \overline{Y} - b_1 \overline{X}_1 - b_2 \overline{X}_2$$

16

Ocena b_1

$$b_1 = \frac{\begin{vmatrix} \sum_{i=1}^{n} x_{1i} y_i & \sum_{i=1}^{n} x_{1i} x_{2i} \\ \sum_{i=1}^{n} x_{2i} y_i & \sum_{i=1}^{n} x_{2i}^2 \end{vmatrix}}{\begin{vmatrix} \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i} x_{2i} \\ \sum_{i=1}^{n} x_{1i} x_{2i} & \sum_{i=1}^{n} x_{2i}^2 \end{vmatrix}}$$

17

Ocena b_2

$$b_2 = \frac{\begin{vmatrix} \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i} y_i \\ \sum_{i=1}^{n} x_{1i} x_{2i} & \sum_{i=1}^{n} x_{2i} y_i \end{vmatrix}}{\begin{vmatrix} \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i} x_{2i} \\ \sum_{i=1}^{n} x_{1i} x_{2i} & \sum_{i=1}^{n} x_{2i}^2 \end{vmatrix}},$$

18

Primer

- Dati su godišnji podaci o potrošnji piva, relativnoj ceni piva i realnom dohotku na slučaj izabranih domaćinstava.
- Prema uzorku od 20 godina potrebno je oceniti parametre cenovne i dohodne elastičnosti tražnje za pivom.
- Promenljive

  Y_i - *ln (potrošnja piva)*

  X_{1i} - *ln (relativna cena piva)*

  X_{2i} - *ln (realni nivo dohotka)*
- *Model:*

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, \mathrm{i} = 1,2,...,20$$

- Uzoračka funkcija:

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i}$$

Primer: podaci

God.	Y	X1	X2
1	4.40305	0.47225	10.02578
2	4.04130	1.22026	10.58768
3	4.16044	0.97932	10.33316
4	4.18052	1.05315	10.49711
5	4.16044	0.75710	10.15131
6	4.06217	0.81697	10.12964
7	4.12228	0.83808	10.16214
8	4.17899	0.73465	10.12368
9	4.05699	1.06000	10.39294
10	4.15104	0.73827	10.04659
11	4.18814	0.79851	10.16956
12	3.87743	1.11619	10.38318
13	4.01818	1.19292	10.43950
14	3.86912	1.07717	10.25049
15	4.04305	0.72916	9.98371
16	3.94352	0.97762	10.22916
17	3.99268	0.96266	10.25149
18	3.94546	0.89116	10.11648
19	4.02356	0.77824	10.03082
20	3.95316	1.08988	10.35204

Primer: potrebne sume centriranih vrednosti

$$\sum_{i=1}^{20} x_{1i}^2 = 0.685833, \sum_{i=1}^{20} x_{2i}^2 = 0.543830,$$

$$\sum_{i=1}^{20} x_{1i}x_{2i} = 0.543319, \sum_{i=1}^{20} x_{1i}y_i = -0.305751,$$

$$\sum_{i=1}^{20} x_{2i}y_i = -0.115200, \sum_{i=1}^{20} y_i^2 = 0.3097930$$

$$\overline{Y} = 4.068577, \overline{X}_1 = 0.914177, \overline{X}_2 = \mathit{10.23282}$$

Primer: ocene parametara

$$0.685833b_1 + 0.543319b_2 = -0.305751$$
$$0.543319b_1 + 0.543830b_2 = -0.115200$$

$$b_1 = -\frac{0.103686}{0.077781} = -1.333,$$
$$b_2 = \frac{0.087112}{0.077781} \approx 1.120,$$
$$b_0 = 4.068577 + 1.333 \cdot 0.914177 - 1.120 \cdot 10.23282$$
$$b_0 = -6.174$$
$$\hat{Y}_i = -6.174 - 1.333X_{1i} + 1.120X_{2i}$$

Primer: interpretacija ocena parametara

$$\hat{Y}_i = -6.174 - 1.333X_{1i} + 1.120X_{2i}$$

- Ocene parcijalnih koeficijenata nagiba:
- -1.333, cenovna elastičnost tražnje

  Ako se **relativna cena poveća za 1%,** tada tražnja za pivom **opada za 1.33%**, pod pretpostavkom da nivo realnog dohotka ostaje nepromenjen
- 1.120, dohodna elastičnost tražnje

  Ako se **realni dohodak poveća za 1%,** onda tražnja za pivom **raste za 1.12%,** pod prepostavkom da se ne menja cena piva.

- Koeficijent determinacije

Koeficijent determinacije R^2

- Kao i u jednostavnom modelu, pokazuje udeo objašnjenog varijabiliteta u ukupnom varijabilitetu zavisne promenljive

$$R^2 = \frac{OSK}{USK} = 1 - \frac{RSK}{USK}$$

$$R^2 = \frac{\sum_{i=1}^n \hat{y}_i^2}{\sum_{i=1}^n y_i^2} = 1 - \frac{\sum_i e_i^2}{\sum_i (Y_i - \overline{Y})^2}$$

- Prilikom određivanja varijabiliteta koji je objašnjen modelom, odnosno izabranim promenljivima, vodi se računa o svakoj od ocena parcijalnih koeficijenata nagiba.
- Na primer, u modelu sa dve objašnjavajuće promenljive:

$$R^2 = \frac{b_1 \sum_{i=1}^n x_{1i} y_i + b_2 \sum_{i=1}^n x_{2i} y_i}{\sum_{i=1}^n y_i^2}$$

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n y_i^2 - b_1 \sum_{i=1}^n x_{1i} y_i - b_2 \sum_{i=1}^n x_{2i} y_i$$

Ograničenja u primeni R^2 kao pokazatelja kvaliteta regresije

1. R^2 se uvek povećava sa dodavanjem novih objašnjavajućih promenljivih:

Regresija 1: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

Regresija 2: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$

R^2 će uvek biti veći u regresiji 2, bez obzira na to kakva je eksplanatorna snaga nove objašnjavajuće promenljive.

2. R^2 je krajnje nepouzdan pokazatelj u regresionoj analizi vremenskih serija kada vrednost 0.999, ne mora nužno pokazivati ništa.

Korigovani koeficijent determinacije R^2

- Koriguje se koeficijent determinacije sa ciljem dobijanja pokazatelja koji se neće neopravdano povećavati sa rastom broja objašnjavajućih promenljivih.
- Novi pokazatelj: $\overline{R}^2$

  korigovani koeficijent determinacije
- Korekcija: svaka od suma u formuli za R^2 deli se sa korespondirajućim brojem stepeni slobode.

$$R^2 = 1 - \frac{\sum_i e_i^2}{\sum_i \left(Y_i - \overline{Y}\right)^2}$$

$$\overline{R}^2 = 1 - \frac{\sum_i e_i^2 / (n-k)}{\sum_i \left(Y_i - \overline{Y}\right)^2 / (n-1)} = 1 - \frac{n-1}{n-k}\left(1 - R^2\right)$$

n - *obim uzorka*, k – *ukupan broj parametara modela*

Primer: izračunavanje rezidualne sume kvadrata i koeficijenta determinacije u modelu sa dve objašnjavajuće promenljive

- Prethodno je ocenjena sledeća zavisnost:

$$\hat{Y}_i = -6.174 - 1.333\, X_{1i} + 1.120\, X_{2i}$$

$$\sum_i e_i^2 = \sum_i y_i^2 - b_1 \sum_i x_{1i} y_i - b_2 \sum_i x_{2i} y_i$$

$$\sum_i e_i^2 = 0.309793 - 1.333 \cdot 0.305751 + 1.120 \cdot 0.115200$$

$$\sum_i e_i^2 = 0.03125$$

$$R^2 = 1 - \frac{0.03125}{0.309793} = 0.899$$

28

Primer: korigovani koeficijent determinacije u modelu sa dve objašnjavajuće promenljive

$$\hat{Y}_i = -6.174 - 1.333X_{1i} + 1.120X_{2i},$$

$$R^2 = 0.899$$

$$\overline{R}^2 = 1 - \frac{\overbrace{n-1}^{20-1}}{\underbrace{n-k}_{20-3}}\left(1 - R^2\right)$$

$$\overline{R}^2 = 1 - \frac{19}{17}(1 - 0.899) = 0.887.$$

29

- Pretpostavke višestrukog KLRM

Pretpostavke višestrukog KLRM (I)

Redni broj pretpostavke	1.	2.	3.
Formulacija	Očekivana vrednost slučajne greške je nula	Slučajne greške su homoskedastične, odnosno poseduju istu varijansu	Slučajne greške su međusobno nekorelisane
Zapis	$E(\varepsilon_i)=0$ za svako i	$v(\varepsilon_i)=E(\varepsilon_i^2)=\sigma^2$ za svako i	$cov(\varepsilon_i,\varepsilon_j)=E(\varepsilon_i\varepsilon_j)=0.$ za svako i , j koji su različiti.

31

Pretpostavke višestrukog KLRM (II)

Redni broj pretpostavke	4.	5.	6.
Formulacija	Slučajna greška ima normalnu raspodelu	Nijedna od objašnjavajućih promenljivih nije slučajna promenljiva	Objašnjavajuće promenljive nisu međusobno linearno zavisne: nijedna se ne može izraziti kao tačna linearna kombinacija ostalih objašnjavajućih promenljivih
Zapis/naziv	$\varepsilon_i : N(0,\sigma^2)$ *za svako* i	$\text{cov}(X_{ji},\varepsilon_i)=0$ *za svako* i, $j=1,\ldots,k-1.$	Odsustvo ekstremne multikolinearnosti

32

- O ocenama parametara u dvostrukom linearnom regresionom modelu

Implikacije pretpostavki višestrukog KLRM

Ako su zadovoljene sve pretpostavke višestrukog KLRM, tada se primenom metoda ONK dobijaju:

- najbolje,
- linearne,
- nepristrasne i
- konzistentne ocene

34

Model sa dve objašnjavajuće promenljive: ocena varijanse slučajne greške modela

- Nepristrasna ocena σ^2 je:

$$s^2 = \frac{1}{n-3}\sum_{i=1}^{n} e_i^2$$

gde je rezidualna suma kvadrata $\sum_{i=1}^{n} e_i^2$ i n je obim uzorka.

Kvadratni koren, s, **je standardna greška regresije, odnosno standardna devijacija reziduala.**

- Oznake za ocene varijansi ocena nagiba: $s^2 b_1$ i $s^2 b_2$

35

Model sa dve objašnjavajuće promenljive: ocene varijansi ocena parcijalnih koeficijenata nagiba

$$v(b_1) = \frac{\sigma^2}{\left(1-r^2\right)\sum_{i=1}^{n} x_{1i}^2}, \qquad v(b_2) = \frac{\sigma^2}{\left(1-r^2\right)\sum_{i=1}^{n} x_{2i}^2}$$

$$\hat{v}(b_1) = s_{b_1}^2 = \frac{s^2}{\left(1-r^2\right)\sum_{i=1}^{n} x_{1i}^2}, \qquad \hat{v}(b_2) = s_{b_2}^2 = \frac{s^2}{\left(1-r^2\right)\sum_{i=1}^{n} x_{2i}^2}$$

$$r = \frac{\sum_{i=1}^{n} x_{1i} x_{2i}}{\sqrt{\sum_{i=1}^{n} x_{1i}^2 \sum_{i=1}^{n} x_{2i}^2}}$$

36

Standardne greške ocena parametara zavise od sledećih faktora:

Slično kao kod jednostavnog modela:

1. **Varijabilitet modela (s^2 ili s).**
2. **Suma kvadrata odstupanja podataka objašnjavajuće promenljive od aritmetičke sredine.**
3. **Obim uzorka n.**

Novo u odnosu na jednostavni model:

4. **Stepen korelisanost između objašnjavajućih promenljivih.** Što je korelacija izraženija, to su varijanse ocena veće, odnosno ocene su nepreciznije.

37

Primer: izračunavanje odgovarajućih standardnih grešaka ocena u modelu sa dve objašnjavajuće promenljive

- Ocena varijanse slučajne greške modela:

$$s^2 = \frac{\sum e_i^2}{n-3} = \frac{0.03125}{17} = 0.001838$$

- Ocena koef. korelacije

$$r = \frac{\sum_{i=1}^{n} x_{1i} x_{2i}}{\sqrt{\sum_{i=1}^{n} x_{1i}^2 \sum_{i=1}^{n} x_{2i}^2}} = \frac{0.543319}{\sqrt{0.685833 \cdot 0.543830}} \approx 0.89 \Rightarrow r^2 = 0.79$$

- Ocene varijanse ocena parcijalnih koeficijenata nagiba:

$$s_{b_1}^2 = \frac{s^2}{\left(1-r^2\right)\sum_{i=1}^{n} x_{1i}^2} = \frac{0.001838}{(1-0.79)\cdot 0.685833} = 0.01276 \Rightarrow s_{b_1} = 0.113$$

$$s_{b_2}^2 = \frac{s^2}{\left(1-r^2\right)\sum_{i=1}^{n} x_{2i}^2} = \frac{0.001838}{(1-0.79)\cdot 0.543830} = 0.01609 \Rightarrow s_{b_2} = 0.127$$

38

Primer: finalni zapis

$$\hat{Y}_i = -6.174 - 1.333 X_{1i} + 1.120 X_{2i},$$
$$\qquad\qquad\qquad (0.113) \quad (0.127)$$

$$R^2 = 0.899, \ \bar{R}^2 = 0.887.$$

39

- Elementi statističkog zaključivanja

Statističko zaključivanje u višestrukom KLRM

- Testiranje hipoteza o vrednostima parametara višestrukog KLRM
  - Pojedinačna vrednost parametra
  - Vrednosti parametara zbirno posmatrano
- Formiranje intervalnih ocena parametara višestrukog KLRM

41

Testiranje hipoteze o vrednosti pojedinačnog parametara višestrukog KLRM

- Kao i u slučaju jednostavnog modela, i u višestrukoj regresiji sa *k* parametara se koristi test-statistika oblika:

$$t_{n-k} = \frac{b_i - \beta_i^*}{s_{b_i}}, \ \mathrm{i} = 0,1,2,\ldots,\mathrm{k}\text{-}1.$$

- Hipoteze: $H_0 : \beta_i = \beta_i^*, H_1 : \beta_i \neq \beta_i^*, i = 0,1,2,\ldots,k-1.$
- Pretpostavimo da je hipoteza od interesa: $H_0 : \beta_i = 0$, protiv $H_1 : \beta_i \neq 0$.
- U uslovima validnosti nulte hipoteze test-statistika je:

$$t_{n-k} = \frac{b_i}{s_{b_i}}$$

- ***t*-odnos:** provera značajnosti pojedinačnog uticaja svake od objašnjav. promenljivih na zavisnu promen.

Pojedinačni uticaj cene i dohotka na tražnju: primena *t-odnosa*

- U prethodnom primeru smo dobili:

Ocena	-1.333	1.120
Standardna greška ocene	0.113	0.127
t-odnosi	**-11.80**	**8.82**

- Kritična vrednost *t-raspodele sa 20-3= 17* stepeni slobode i nivo značajnosti 5%: **2.11.**

Da li prihvatamo

H_0: $\beta_1 = 0$? (Ne)

H_0: $\beta_2 = 0$? (Ne)

- **Zaključak: promenljive ostvaruju uticaj koji je pojedinačno statistički značajan na nivou značajnosti 5%.**

Testiranje hipoteze o vrednosti svih parametara višestrukog KLRM

Ispitivanje kvaliteta regresije na osnovu koeficijenta determinacije – opšti slučaj

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_{k-1} X_{k-1i} + \varepsilon_i$$

- Hipoteze od interesa:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_{k-1} = 0 \Leftrightarrow H_0 : R^2 = 0$$

$$H_1 : hipoteza\, H_0 \text{ nije tacna} \Leftrightarrow H_1 : R^2 \neq 0$$

- Nulta hipoteza: regresija nije statistički značajna (zajednički uticaj objašnjavajućih promenljivih nije statistički značajan)
- Alternativna hipoteza: objašnjavajuće promenljive ostvaruju statistički značajan uticaj na kretanje zavisne promenljive (bar jedan od parametara je značajno različit od nule)

Ispitivanje kvaliteta regresije na osnovu koeficijenta determinacije – opšti slučaj

- Relevantna statistika:

$$F_{n-k}^{k-1} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

- Pravilo odlučivanja:
  - Ako je izračunata vrednost date statistike veća od kritične vrednosti *F*-raspodele sa *k-1* i *n-k* stepeni slobode, tada se nulta hipoteza odbacuje uz izabrani nivo značajnosti.
    - Objašnjavajuće promenljive ostvaruju zbirno statistički značajan uticaj na kretanje zavisne promenljive.

Osnovna ideja definisanja F test statistike

- F statistika sa **k-1** i **n-k** stepeni slobode dobija se transformacijom sledećeg količnika

$$\frac{\text{Objašnjeni varijabilitet Y}}{\text{Neobjašnjeni varijabilitet Y}}$$

Varijabilitet	Neobjašnjeni varijabilitet Y (rezidualna suma kvadrata)	Objašnjeni varijabilitet Y = Ukupni – neobjašnjeni
Broj stepeni slobode	n-k	(n-1)-(n-k)=k-1

Ispitivanje kvaliteta regresije na osnovu koeficijenta determinacije – dvostruki model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

- Relevantne hipoteze

$$H_0 : \beta_1 = \beta_2 = 0 \Leftrightarrow H_0 : R^2 = 0$$

$$H_1 : hipoteza\, H_0 \text{ nije tacna} \Leftrightarrow H_1 : R^2 \neq 0$$

- Nulta hipoteza: regresija nije statistički značajna
- Alternativna hipoteza: objašnjavajuće promenljive ostvaruju statistički značajan uticaj na kretanje zavisne promenljive
- Test- statistika:

$$F^2_{n-3} = \frac{R^2 / 2}{(1-R^2)/(n-3)}$$

1. $\varepsilon_i \sim N(0,\sigma^2) \Rightarrow \frac{\varepsilon_i}{\sigma} \sim N(0,1) \Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}\varepsilon_i^2 \sim \chi^2_n$

$$\Rightarrow \frac{1}{\sigma^2}\overbrace{\sum_{i=1}^{n} e_i^2}^{RSK} \sim \chi^2_{n-k}.$$

2. $\varepsilon_i \sim N(0,\sigma^2) \Rightarrow Y_i \sim N(E(Y),\sigma^2) \Rightarrow \frac{Y_i - E(Y)}{\sigma} \sim N(0,1)$

$$\Rightarrow \frac{1}{\sigma 2}\overbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}^{USK} \sim \chi^2_{n-1}$$

3. $USK = RSK + OSK \Rightarrow OSK = USK - OSK \Rightarrow \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 - \sum_{i=1}^{n} e_i^2$

4. $$\left.\begin{array}{l} \frac{1}{\sigma^2}\overbrace{\sum_{i=1}^{n} e_i^2}^{RSK} \sim \chi^2_{n-k} \\ \frac{1}{\sigma^2}\overbrace{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}^{USK} \sim \chi^2_{n-1} \end{array}\right\} \frac{1}{\sigma^2}\overbrace{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}^{OSK} \sim \chi^2_{\underbrace{[n-1]-[n-k]}_{k-1}}$$

$$5.\ F^{k\text{-}1}_{n\text{-}k} \sim \frac{\chi^2_{k-1}/(k-1)}{\chi^2_{n-k}/(n-k)} = \frac{\frac{1}{\sigma^2}\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}{\frac{1}{\sigma^2}\sum_{i=1}^{n}e_i^2}\frac{(n-k)}{(k-1)} =$$

$$\frac{\left[\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2 \Big/ \sum_{i=1}^{n}(Y_i-\bar{Y})^2\right]}{\left[\sum_{i=1}^{n}e_i^2 \Big/ \sum_{i=1}^{n}(Y_i-\bar{Y})^2\right]}\frac{(n-k)}{(k-1)} = \frac{R^2 \big/ (k-1)}{(1-R^2)\big/(n-k)}$$

Zajednički uticaj cene i dohotka na tražnju: primena *F-statistike*

- U prethodnom primeru smo dobili: $R^2 = 0.899$
- Realizovana vrednost *F-statistike:*

$$F^2_{17} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0.899/2}{(1-0.899)/17}$$

$$F^2_{17} = 75.7$$

- Kritična vrednost *F-raspodele sa redom 2 i 17* stepeni slobode i nivoom značajnosti 5%: **3.59**

  Da li prihvatamo

  H_0: $\beta_1 = \beta_2 = 0$? (Ne)

- **Zaključak: promenljive ostvaruju uticaj koji je zbirno statistički značajan na nivou značajnosti 5%.**

F-statistika kvaliteta regresije u jednostavnom modelu

$$\hat{Y}_i = b_0 + bX_i$$
$$(s_b)$$

$$F_{n-k}^{k-1} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

$k = 2,$ slobodan clan i jedna objasnjavajuca promenljiva

$$F_{n-2}^{1} = \frac{R^2}{(1-R^2)/(n-2)} = \frac{R^2(n-2)}{(1-R^2)}$$

$$F_{n-2}^{1} = t_{n-2}^{2} \Rightarrow \frac{R^2(n-2)}{(1-R^2)} = \left[\frac{b}{s_b}\right]^2$$

Ova dva testa se mogu koristiti alternativno u jednostavnom modelu.

F-statistika kvaliteta regresije u jednostavnom modelu: primer

- Ocenjena je zavisnost inflacije od stope rasta deviznog kursa na bazi mesečnih podataka srpske privrede u periodu: januar 2003 – januar 2007. godine (49 podatka):

$$\hat{Y}_i = 0.0077 + 0.216X_i \quad \mathrm{R}^2 = 0.140$$
$$(0.078)$$

- Testirati kvalitet regresije primenom t i F testa.
- Potvrditi vezu izmedju t i F raspodele u jednostavnom modelu.

Vežba

- Ocenjena je funkcija tražnje za kafom:

$$\hat{Y}_t = 42.22 - 3.07 X_{1t} + 2.63 X_{2t} + 0.83 X_{3t} + 0.0006t \quad R^2 = 0.83$$
$$(0.73) \qquad (0.91) \qquad (0.23) \qquad (0.002)$$

Y_t - tražnja za kafom, X_{1t} - cena kafe, X_{2t} - cena čaja,

X_{3t} - raspoloživi dohodak, t=1,2,..., 20.

U 2017. godini: $Y_t = 50$, $X_{1t} = 1.63$, $X_{2t} = 0.95$, $X_{3t} = 5$.

Vežba II

- Testirati statističku značajnost regresije.
- Kakva će biti reakcija tražnje, ako se cena kafe poveća za 15%?
- Za koliko procenata treba smanjiti cenu kafe, ako se cena čaja smanji za 20%, da bi nivo tražnje za kafom ostao nepromenjen?