

Ekonometrija 1

Tretje vaje:

Izračun in razlaga parametrov ocenjenega regresijskega modela.

Na tretjih vajah bomo na podlagi podatkov iz prejšnjih vaj ocenili še preostale parametre, ki se običajno izračunavajo pri ocenjevanju regresijskih modelov. V nadaljevanju bomo komentirali rezultate primera ocenjevanja regresijskega modela s paketom Stata.



Primer 1: V primeru iz prejšnjih vaj smo na podlagi spodaj navedenih količin za deset enot ocenjevali regresijske koeficiente v modelu $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$:

$$\begin{aligned} \bar{y} = 2,4; \quad \bar{x}_2 = 0,4; \quad \bar{x}_3 = 0,8; \quad \sum y_i^2 = 184; \quad \sum x_{2i} x_{3i} = -2; \\ \sum x_{2i}^2 = 10; \quad \sum x_{3i}^2 = 12; \quad \sum y_i x_{2i} = 40; \quad \sum y_i x_{3i} = -2. \end{aligned}$$

Na primeru regresijskih koeficientov na kratko ponovite postopek pri statističnem preizkušanju domnev. Ali so ocenjeni regresijski koeficienti statistično značilno različni od nič? Izračunajte mere primernosti vzorčnega regresijskega modela in ugotovite, ali model kot celota zadovoljivo pojasnjuje varianco odvisne spremenljivke.

Zapišite ocenjeni regresijski model z vsemi izračunanimi parametri. Izračunajte ocene elastičnosti funkcije in pojasnite v čem so temeljne razlike med njimi in ocenami parcialnih regresijskih koeficientov. ■

Primer 2: V reviji Naše stanovanje so se odločili za raziskavo trga nepremičnin na območju ljubljanske mestne občine. V ta namen so oblikovali vzorec 30 stanovanj, ki so naprodaj in zbrali podatke (shranjeni so v datoteki `stanovanje.dta`) o njihovi ceni (*CENA*; v mio SIT), površini (*M2*; v kvadratnih metrih), starosti (*STAR*; v letih) in oddaljenosti od centra mesta (*ODD*; v kilometrih).

Ocenite linearno multiplo regresijsko funkcijo, v kateri boste proučevali vpliv navedenih pojasnjevalnih spremenljivk na ceno stanovanj in razložite dobljene rezultate. Izračunajte tudi ocenjene vrednosti odvisne spremenljivke in ostanke, ter jih izpišite skupaj z opazovanimi vrednostmi odvisne spremenljivke.

Izpis rezultatov obdelav v programskem paketu Stata:

```
. describe
```

Contains data from stanovanje.dta

```
obs:      30
vars:      4                      DD Mmm YYYY HH:MM
size:      330
```

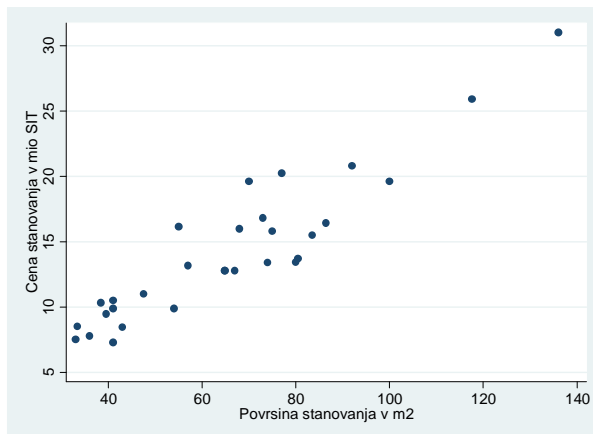
```
-----
                storage  display  value
variable name  type      format  label  variable label
-----
cena           float    %8.0g          Cena stanovanja v mio SIT
m2            float    %8.0g          Povrsina stanovanja v m2
odd           byte      %8.0g          Oddaljenost stanovanja od centra mesta v km
star          byte      %8.0g          Starost stanovanja v letih
-----
```

Sorted by:

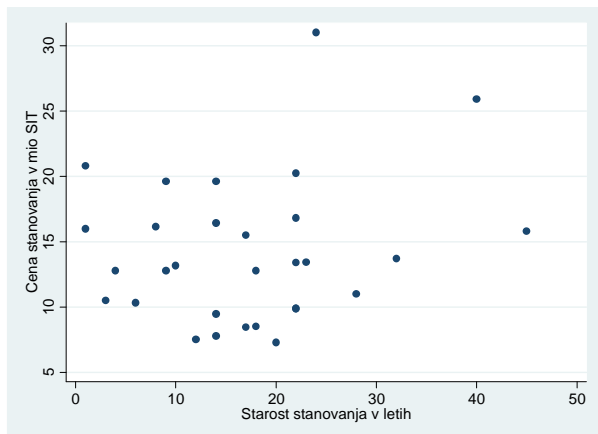
```
. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
n	30	15.5	8.803408	1	30
cena	30	14.21717 ^①	5.528791 ^②	7.3	31
m2	30	65.65 ^③	25.23567	33	136
odd	30	9.866667	7.233701	1	28
star	30	17.03333	10.49953	1	45

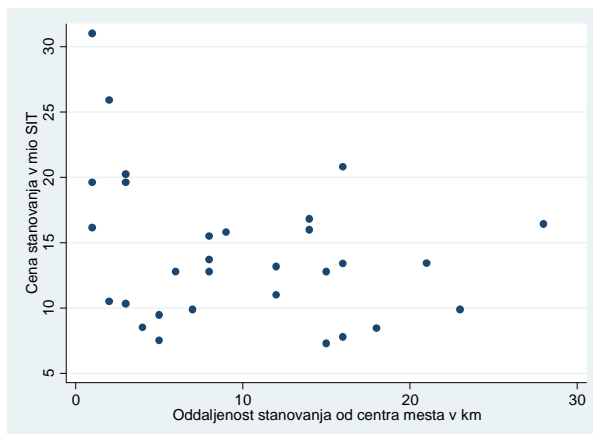
```
. scatter cena m2
```



```
. scatter cena star
```



```
. scatter cena odd
```



```
. regress cena m2 star odd
```

Source	SS	df	MS	Number of obs =	30
Model	805.156707	3	268.385569	F(3, 26) =	85.83 ⁶
Residual	81.3017384 ¹⁵	26	3.12698994 ⁵	Prob > F =	0.0000 ⁷
				R-squared =	0.9083 ⁸
				Adj R-squared =	0.8977 ⁹
Total	886.458445 ¹⁷	29	30.5675326	Root MSE =	1.7683 ⁴

cena	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
m2	.2051172 ¹⁰	.0137199	14.95 ¹¹	0.000 ¹²	.1769156 .2333188
star	-.0486017 ¹³	.0328782	-1.48	0.151	-.1161838 .0189804
odd	-.1635526 ¹⁴	.0455379	-3.59	0.001	-.2571572 -.0699481
_cons	3.192791 ¹⁵	1.083781	2.95	0.007	.9650468 5.420536

```
. estat vce
```

Covariance matrix of coefficients of regress model

e(V)	m2	star	odd	_cons
m2	.00018824			
star	-.00013912	.00108098		
odd	.00004938	-.00002756	.0020737	
_cons	-.01047515	-.00900722	-.02323306	1.1745823

```
. estat ic
```

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	30	-93.35871	-57.52271 ¹⁸	4	123.0454 ¹⁹	128.6502 ²⁰

Note: N=Obs used in calculating BIC; see [R] BIC note

```
. predict e, resid
. predict cenahat, xb
. list cena cenahat e
```

	cena	cenahat	e
1.	13.16	12.43582	.7241774
2.	16	14.80242	1.197579
3.	15.5	18.18543	-2.685427
4.	20.8	19.39813	1.401871
5.	13.7	16.83079	-3.130794
6.	8.46	8.242655	.2173457
7.	16.815	14.80737	2.007628
8.	7.5	8.560676	-1.060675
9.	12.795	15.43281	-2.637815
10.	19.614	22.86053	-3.246534
...			
29.	10.5	11.12969	-.6296857
30.	12.784	13.60394	-.8199354
Mean	14.2172	14.21717	9.20e-09

Rešitev primera 2

Zapis regresijskega modela:

$$CENA_i = \beta_1 + \beta_2 M2_i + \beta_3 STAR_i + \beta_4 ODD_i + u_i$$

$$\widehat{CENA}_i = 3,19279 + 0,20512M2_i - 0,04860STAR_i - 0,16355ODD_i$$

$$t: \quad (2,95) \quad (14,95) \quad (-1,48) \quad (-3,59)$$

$$p: \quad (0,007) \quad (0,000) \quad (0,151) \quad (0,001)$$

$$n = 30 \quad R^2 = 0,9083 \quad \bar{R}^2 = 0,8977 \quad s_e = 1,76833 \quad KV = 0,124 \quad F = 85,83 \\ (0,000)$$

Razlaga nekaterih parametrov modela:

- $s_e = 1,76833$ (⊕): standardna napaka ocene regresije ali ocena standardnega odklona slučajne spremenljivke:

Ocena standardnega odklona slučajne spremenljivke, s katerim merimo velikost odstopanj opazovanih vrednosti cene stanovanj od ocenjenih vrednosti, dobljenih na podlagi multiple regresijske funkcije, je enaka 1,768 mio SIT.

- $KV = \frac{s_e}{\bar{y}} = \frac{1,7683}{14,2172} = 0,124$: ocena koeficienta variacije:

Koeficient variacije je relativna mera primernosti vzorčnega regresijskega modela, njegova ocena pa v konkretnem primeru znaša 0,124, kar pomeni, da standardna napaka ocene regresije predstavlja 12,4 odstotka ocenjene povprečne vrednosti cene stanovanj.

- Vrednost F -statistike in stopnja značilnosti pri preizkušanju ničelne domneve:

$$H_0: \beta_j = 0; \quad \forall j = 2, 3, 4$$

$$H_1: \beta_j \neq 0; \quad \exists j = 2, 3, 4$$

Vrednost F -statistike pri preizkušanju ničelne domneve, da so vsi parcialni regresijski koeficienti hkrati enaki 0, je enaka 85,83 (⊕), točna stopnja značilnosti pa je zanemarljiva ($p=0,000$, ⊕). Zato lahko zavrnilo ničelno domnevo pri zanemarljivi stopnji značilnosti in sprejmemo sklep, da je vsaj eden izmed parcialnih regresijskih koeficientov statistično značilno različen od 0. Na podlagi vzorčnih podatkov lahko torej trdimo, da vzorčni regresijski model kot celota zadovoljivo pojasnjuje variabilnost cene stanovanj oziroma izraža linearno povezanost med spremenljivkami.

- $R^2 = 0,9083$ (⊕): ocena multiplega determinacijskega koeficienta:

Ocena multiplega determinacijskega koeficienta je enaka 0,9083, kar pomeni, da je skoraj 91 odstotkov variabilnosti cene stanovanj pojasnjeno z linearno odvisnostjo cene stanovanj od njihove površine, starosti in oddaljenosti od centra mesta.

- $b_2 = 0,205117$ (10): ocena (prvega) parcialnega regresijskega koeficienta:

Če se površina stanovanj poveča za en kvadratni meter, se cena v povprečju poveča za 205.117 SIT, pri čemer ostaneta preostali pojasnjevalni spremenljivki, torej starost stanovanj in oddaljenost od centra mesta, nespremenjeni.

- Vrednost t -statistike in stopnja značilnosti pri preizkušanju ničelne domneve: $H_0 : \beta_2 = 0$

Vrednost t -statistike pri preizkušanju ničelne domneve, da je drugi parcialni regresijski koeficient enak 0, je enaka 14,95 (11), točna stopnja značilnosti pa je zanemarljiva ($p=0,000$, 12). Zato lahko zavrnamo ničelno domnevo pri zanemarljivi stopnji značilnosti in sprejmemo sklep, da je parcialni regresijski koeficient b_2 statistično značilno različen od 0. Na podlagi zbranih podatkov lahko torej trdimo, da velikost stanovanj vpliva na njihovo ceno.

- Ocena (prve) parcialne elastičnosti pri povprečnih vrednostih:

$$\hat{\varepsilon}_{\bar{y}, \bar{x}_2} = b_2 \frac{\bar{x}_2}{\bar{y}} = 0,20512 \cdot \frac{65,65}{14,2172} = 0,947; \quad \bar{y} = 14,2172 \quad \bar{x}_2 = 65,65$$

Če se površina stanovanj pri povprečni vrednosti poveča za en odstotek, se cena pri povprečni vrednosti poveča za 0,947 odstotka, pri čemer ostaneta starost in oddaljenost od centra mesta nespremenjeni.

- Vrednost logaritma funkcije verjetja (13):

$$\ln L = -\frac{n}{2} \left[\ln(2\pi) + \ln \left(\frac{NVK}{n} \right) + 1 \right] = -57,52$$

- Akaikejev informacijski kriterij (14):

$$AIC = -2 \ln L + 2k = 123,05$$

- Schwarzov kriterij ali Bayesianski informacijski kriterij (15):

$$SC = BIC = -2 \ln L + k \ln(n) = 128,65$$

■