

Ekonometrija 1

Devete vaje:

Normalna porazdelitev slučajne spremenljivke. Multikolinearnost.

Na devetih vajah bomo analizirali prisotnost multikolinearnosti in preverjali predpostavko o normalni porazdelitvi slučajne spremenljivke (slučajnih odklonov) u .



Primer 1: Za neko državo smo zbrali podatke (za obdobje 15 let) o osebni porabi (OP), plačah ($PLACE$), socialnih prejemkih (SOC), prejemkih iz tujine (TUJ) in posebnih proračunskih prejemkih (PRO) ter ocenili naslednjo funkcijo osebne porabe:

$$\widehat{OP} = 0,46 + 0,78PLACE + 0,22SOC + 0,02TUJ - 0,16PRO$$
$$t: \quad (4,5) \quad (1,2) \quad (0,4) \quad (1,0) \quad (-0,2)$$

Ocenili smo tudi regresijski model med pojasnjevalnimi spremenljivkami in ugotovili, da je 95 odstotkov variabilnosti plač pojasnjeno z variiranjem socialnih prejemkov, prejemkov iz tujine in posebnih proračunskih prejemkov.

- Obravnavajte problematiko multikolinearnosti na primeru ocenjene funkcije porabe.
 - Kakšno funkcijo porabe bi ocenjevali, da bi vsaj nekoliko omilili težave, ki jih povzroča multikolinearnost?
-

Primer 2: Na podlagi naslednjih podatkov:

$$y : 4, 8, -12, 16, 10, 24 \quad x_2 : 2, 4, -6, 2, 6, -4 \quad x_3 : 6, 10, -10, 6, 14, -6$$

smo želeli oceniti regresijski model: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$.

Ocenite regresijski model med pojasnjevalnima spremenljivkama, izračunajte njegove ostanke in determinacijski koeficient ter pojasnite, kakšne so posledice vaših rezultatov za ocenjevanje regresijskega modela: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$.

■

Primer 3: Na kmetijskem inštitutu so za zadnjih trideset let ocenjevali pridelok vina na hektar vinograda (HPV) v odvisnosti od starosti vinograda (STA), števila sončnih dni v letu (SD), količine padavin (PAD) in povprečne temperature v tretjem kvartalu leta ($T3$):

$$HPV_t = \beta_1 + \beta_2 PAD_t + \beta_3 STA_t + \beta_4 SD_t + \beta_5 T3_t + u_t.$$

Ker so bili vsi ocenjeni parcialni regresijski koeficienti zgoraj zapisanega modela statistično neznačilni, determinacijski koeficient pa izredno visok (0,985), so se odločili še za ocenjevanje regresijskega modela med pojasnjevalnimi spremenljivkami:

$$PAD_t = \beta_1 + \beta_2 STA_t + \beta_3 SD_t + \beta_4 T3_t + u_t$$

in zanj ugotovili, da je vsota kvadratov opazovanih vrednosti količine padavin, $\sum PAD_t^2$, enaka 528, vsota kvadratov ostankov 14,5, povprečna letna količina padavin pa je enaka 2.

Pojasnite, kaj so želeli preveriti na podlagi regresijskega modela med pojasnjevalnimi spremenljivkami in do kakšnega zaključka so prišli. Kaj bi na podlagi tega zaključka morali ukreniti, da bi dobili čimbolj zanesljive ocene vplivov posameznih dejavnikov na pridelek vina na hektar?

■

Primer 4: Na podlagi podatkov za 27 hiš v nekem ameriškem okraju, ki so bile prodane leta 1977 (datoteka `hise.dta`), bi radi ugotovili, kako je prodajna cena hiš (*CENA*; v 1.000 USD) odvisna od njihove površine (*POVRSINA*; v 1.000 kvadratnih čevljev), starosti (*STAROST*; v letih), števila sob (*SOBE*) in števila spalnic (*SPALNICE*).

Analizirajte problem multikolinearnosti. Kakšen regresijski model bi ocenjevali, da bi odpravili težave, ki jih povzroča multikolinearnost?

Za predlagano funkcijo cene hiš preverite, ali je izpolnjena predpostavka o normalni porazdelitvi slučajne spremenljivke u .

Izpis rezultatov obdelav v programskem paketu Stata:

```
. regress cena površina starost sobe spalnice
```

Source	SS	df	MS	Number of obs =	27
Model	4716.28866	4	1179.07216	F(4, 22) =	42.74
Residual	606.914165	22	27.5870075	Prob > F =	0.0000
				R-squared =	0.8860
				Adj R-squared =	0.8653
Total	5323.20282	26	204.73857	Root MSE =	5.2523

cena	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
povrsina	.0223931	.0037252	6.01	0.000	.0146674 .0301188
starost	-.1488602	.0816217	-1.82	0.082	-.3181331 .0204128
sobe	1.42822	2.637175	0.54	0.594	-4.040945 6.897386
spalnice	-1.735464	3.911259	-0.44	0.662	-9.84692 6.375991
_cons	6.21797	7.723658	0.81	0.429	-9.799916 22.23586

. pwcorr povrsina starost sobe spalnice, sig

	povrsina	starost	sobe	spalnice
povrsina	1.0000			
starost	-0.1781 0.3742	1.0000		
sobe	0.8457 0.0000	0.0043 0.9829	1.0000	
spalnice	0.7934 0.0000	0.1026 0.6105	0.9244 0.0000	1.0000

. collin povrsina starost sobe spalnice, corr
(obs=27)

Collinearity Diagnostics

Variable	VIF	SQRT VIF	Tolerance	R- Squared
povrsina	4.08	2.02	0.2453	0.7547
starost	1.24	1.11	0.8064	0.1936
sobe	8.98	3.00	0.1113	0.8887
spalnice	7.56	2.75	0.1323	0.8677
Mean VIF	5.46			

	Eigenval	Cond Index
1	2.7110	1.0000
2	1.0483	1.6081
3	0.1729	3.9598
4	0.0678	6.3235

Condition Number 6.3235

Eigenvalues & Cond Index computed from deviation sscp (no intercept)

Det(correlation matrix) 0.0333

. regress cena povrsina starost

Source	SS	df	MS	Number of obs =	27
Model	4708.12215	2	2354.06107	F(2, 24) =	91.85
Residual	615.080677	24	25.6283615	Prob > F =	0.0000
				R-squared =	0.8845
				Adj R-squared =	0.8748
Total	5323.20282	26	204.73857	Root MSE =	5.0624

cena	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
povrsina	.0231243	.0018073	12.79	0.000	.0193942	.0268544
starost	-.1523485	.0717955	-2.12	0.044	-.3005271	-.0041699
_cons	9.093323	4.222131	2.15	0.042	.3792732	17.80737

```
. collin povrsina starost, corr
(obs=27)
```

Collinearity Diagnostics

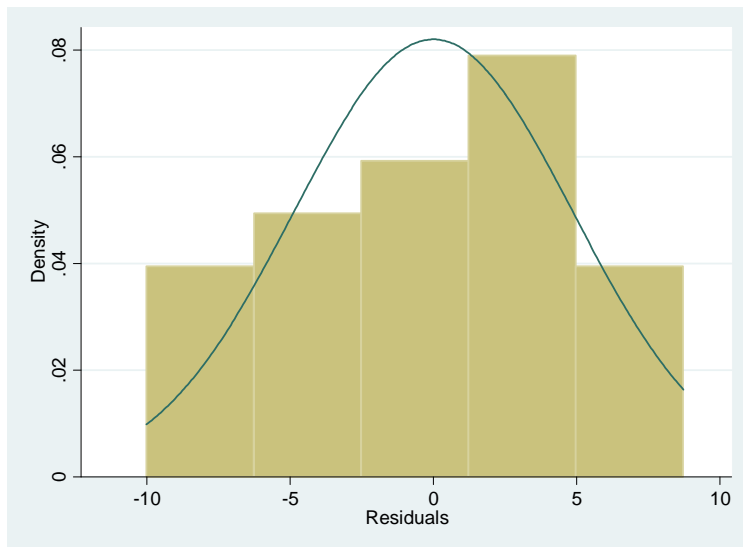
Variable	VIF	SQRT VIF	Tolerance	R-Squared
povrsina	1.03	1.02	0.9683	0.0317
starost	1.03	1.02	0.9683	0.0317
Mean VIF	1.03			

	Eigenval	Cond Index
1	1.1781	1.0000
2	0.8219	1.1972

```
Condition Number      1.1972
Eigenvalues & Cond Index computed from deviation sscp (no intercept)
Det(correlation matrix) 0.9683
```

```
. predict ecena, resid
```

```
. histogram ecena, normal
(bin=5, start=-10.017778, width=3.7491447)
```



```
. sum ecena, detail
```

Residuals				
Percentiles		Smallest		
1%	-10.01778	-10.01778		
5%	-8.911816	-8.911816		
10%	-6.921759	-6.921759	Obs	27
25%	-3.455669	-6.515256	Sum of Wgt.	27
50%	.9022909		Mean	-1.88e-08
		Largest	Std. Dev.	4.863841
75%	3.624211	5.456469		
90%	6.491363	6.491363	Variance	23.65695
95%	6.566592	6.566592	Skewness	-.3191895
99%	8.727945	8.727945	Kurtosis	2.349854

```
. return list
```

```
scalars:
```

```
      r(N) = 27
r(sum_w) = 27
r(mean) = -1.87644252071e-08
r(Var) = 23.65694875107343
r(sd) = 4.863840946317368
r(skewness) = -.3191895093658222
r(kurtosis) = 2.349854454971561
r(sum) = -5.06639480591e-07
r(min) = -10.01777839660645
r(max) = 8.727945327758789
r(p1) = -10.01777839660645
r(p5) = -8.911815643310547
r(p10) = -6.921758651733398
r(p25) = -3.455668926239014
r(p50) = .9022908806800842
r(p75) = 3.624211072921753
r(p90) = 6.491362571716309
r(p95) = 6.566591739654541
r(p99) = 8.727945327758789
```

```
. scalar opaz=r(N)
```

```
. scalar sk=r(skewness)
```

```
. scalar ku=r(kurtosis)
```

```
. scalar jb=opaz*(sk^2/6 + (ku-3)^2/24)
```

```
. display jb
```

```
.93399413
```

```
. display chi2tail(2,jb)
```

```
.62688193
```

```
. jb6 ecena
```

```
Jarque-Bera normality test: .934 Chi(2) .6269
```

```
Jarque-Bera test for Ho: normality: (ecena)
```

