

Primer 1: V vzorec smo izbrali 35 evropskih držav in zanje pridobili naslednje podatke (datoteka `tobak.dta`):

- ♦ odstotek prebivalcev, starejših od 65 let (*STAR*);
- ♦ poraba cigaret na prebivalca (*TOBAK*; izražena v številu zavitkov cigaret);
- ♦ število umrlih na 100.000 prebivalcev (*SMRTNOST*).

- a) Ocenite linearni regresijski model, v katerem je *SMRTNOST* odvisna spremenljivka, preostali dve pa sta pojasnjevalni, in preverite ustreznost specifikacije.
- b) Ocenite dvojno-logaritemski model in presodite, ali je primernejši od linearnega.
- c) Ocenite linearni regresijski model, če je spremenljivka *SMRTNOST* izražena kot število umrlih na 1.000 prebivalcev, odstotek prebivalcev, starejših od 65 let, pa kot delež.

Izpis rezultatov obdelav v programskem paketu Stata:

```
. * Ocenjevanje linearnega modela
```

```
. regress smrtnost star tobak
```

Source	SS	df	MS	Number of obs =	35
Model	316665.524	2	158332.762	F(2, 32) =	33.26
Residual	152312.816	32	4759.7755	Prob > F =	0.0000
				R-squared =	0.6752
				Adj R-squared =	0.6549
Total	468978.34	34	13793.4806	Root MSE =	68.991

smrtnost	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
star	52.79328	6.971745	7.57	0.000	38.5923	66.99426
tobak	1.441107	.501654	2.87	0.007	.4192715	2.462943
_cons	54.76546	104.3763	0.52	0.603	-157.8422	267.3731

```
. ereturn list
```

```
scalars:
```

```

e(N) = 35
e(df_m) = 2
e(df_r) = 32
e(F) = 33.26475414229081
e(r2) = .6752241987489192
e(rmse) = 68.99112626929494
e(mss) = 316665.5238198494
e(rss) = 152312.8161249855
e(r2_a) = .6549257111707266
e(ll) = -196.2838620887884
e(ll_0) = -215.9647151864189
e(rank) = 3

```

```
matrices:
```

```

e(b) : 1 x 3
e(V) : 3 x 3

```

```
functions:
```

```
e(sample)
```

```

. scalar opaz=e(N)
. scalar par=e(rank)
. scalar nvklin=e(rss)

. display opaz, par, nvklin
35 3 152312.82

```

```

. * Ramsey RESET test linearnega modela

```

```

. predict smrtnoste, resid
. predict smrtnosthat, xb

. scatter smrtnoste smrtnosthat

. gen smrtnost2=smrtnosthat^2
. gen smrtnost3=smrtnosthat^3
. gen smrtnost4=smrtnosthat^4

. regress smrtnost star tobak smrtnost2

```

Source	SS	df	MS			
Model	345515.417	3	115171.806	Number of obs =	35	
Residual	123462.923	31	3982.67493	F(3, 31) =	28.92	
Total	468978.34	34	13793.4806	Prob > F	= 0.0000	
				R-squared	= 0.7367	
				Adj R-squared	= 0.7113	
				Root MSE	= 63.108	

smrtnost	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
star	224.7894	64.2223	3.50	0.001	93.80719	355.7717
tobak	5.739992	1.661854	3.45	0.002	2.35062	9.129365
smrtnost2	-.0018243	.0006778	-2.69	0.011	-.0032067	-.0004419
_cons	-1155.779	459.7978	-2.51	0.017	-2093.543	-218.0152

```

. regress smrtnost star tobak smrtnost2 smrtnost3

```

Source	SS	df	MS			
Model	350470.52	4	87617.63	Number of obs =	35	
Residual	118507.82	30	3950.26066	F(4, 30) =	22.18	
Total	468978.34	34	13793.4806	Prob > F	= 0.0000	
				R-squared	= 0.7473	
				Adj R-squared	= 0.7136	
				Root MSE	= 62.851	

smrtnost	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
star	943.1825	644.6101	1.46	0.154	-373.287	2259.652
tobak	25.29777	17.54074	1.44	0.160	-10.52521	61.12074
smrtnost2	-.0169109	.0134873	-1.25	0.220	-.0444556	.0106337
smrtnost3	5.48e-06	4.90e-06	1.12	0.272	-4.52e-06	.0000155
_cons	-4428.451	2957.723	-1.50	0.145	-10468.93	1612.025

```

. display invFtail(2,30,0.05)
3.3158295

```

```

. display Ftail(2,30,4.28)
.02316133

```

```
. regress smrtnost star tobak smrtnost2 smrtnost3 smrtnost4
```

Source	SS	df	MS	Number of obs =	35
Model	353040.915	5	70608.183	F(5, 29) =	17.66
Residual	115937.425	29	3997.84223	Prob > F =	0.0000
				R-squared =	0.7528
				Adj R-squared =	0.7102
Total	468978.34	34	13793.4806	Root MSE =	63.228

smrtnost	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
star	7569.722	8289.582	0.91	0.369	-9384.376 24523.82
tobak	206.1113	226.188	0.91	0.370	-256.495 668.7176
smrtnost2	-.2300152	.2661156	-0.86	0.394	-.7742827 .3142523
smrtnost3	.0001644	.0001982	0.83	0.414	-.0002411 .0005699
smrtnost4	-4.39e-08	5.47e-08	-0.80	0.429	-1.56e-07 6.81e-08
_cons	-24950.19	25765.73	-0.97	0.341	-77647.02 27746.64

```
. display invFtail(1,30,0.05)
4.1708768
```

```
. display Ftail(3,29,3.03)
.04522609
```

```
. qui regress smrtnost star tobak
. estat ovtest // Stata vkljuci potence do cetrtega reda //
```

```
Ramsey RESET test using powers of the fitted values of smrtnost
Ho: model has no omitted variables
      F(3, 29) =      3.03
      Prob > F =      0.0451
```

```
. * Ocenjevanje potencnega modela
```

```
. gen lsmrtnost=log(smrtnost)
. gen lstar=log(star)
. gen ltobak=log(tobak)
```

```
. regress lsmrtnost lstar ltobak
```

Source	SS	df	MS	Number of obs =	35
Model	.526056623	2	.263028311	F(2, 32) =	47.67
Residual	.176566817	32	.005517713	Prob > F =	0.0000
				R-squared =	0.7487
				Adj R-squared =	0.7330
Total	.70262344	34	.020665395	Root MSE =	.07428

lsmrtnost	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lstar	.7856827	.0935545	8.40	0.000	.5951183 .9762471
ltobak	.2834557	.0692842	4.09	0.000	.1423285 .424583
_cons	3.449101	.3855704	8.95	0.000	2.663719 4.234482

```
. scalar nvklog=e(rss)
. display nvklog
.17656682
```

```
. * Izracun primerljivega multiplega determinacijskega koeficienta
```

```
. qui regress smrtnost star tobak  
. predict ylinhat, xb
```

```
. gen lylinhat=log(ylinhat)
```

```
. correlate lsmrtnost lylinhat  
(obs=35)
```

```
-----+-----  
          | lsmrtn~t lylinhat  
-----+-----  
lsmrtnost | 1.0000  
lylinhat  | 0.8489  1.0000
```

```
. return list
```

```
scalars:
```

```
      r(N) = 35  
      r(rho) = .8489370622461852
```

```
matrices:
```

```
      r(C) : 2 x 2
```

```
. scalar r1p=r(rho)^2
```

```
. scalar r1barp=1-(1-r1p)*((opaz-1)/(opaz-par))
```

```
. display r1p, r1barp
```

```
.72069414 .70323752
```

```
. qui regress lsmrtnost lstar ltobak  
. predict yloghat, xb
```

```
. gen antilyloghat=exp(yloghat)
```

```
. correlate smrtnost antilyloghat  
(obs=35)
```

```
-----+-----  
          | smrtnost antily~t  
-----+-----  
smrtnost  | 1.0000  
antilyloghat | 0.8341  1.0000
```

```
. return list
```

```
scalars:
```

```
      r(N) = 35  
      r(rho) = .8340791174085507
```

```
matrices:
```

```
      r(C) : 2 x 2
```

```
. scalar r2p=r(rho)^2
```

```
. display r2p
```

```
.69568797
```

```
. scalar r2barp=1-(1-r2p)*((opaz-1)/(opaz-par))
```

```
. display r2p, r2barp
```

```
.69568797 .67666847
```

```
. * Box-Coxov test (originalni postopek)
```

```
. sum lsmrtnost
```

Variable	Obs	Mean	Std. Dev.	Min	Max
lsmrtnost	35	6.763665	.1437546	6.317526	7.02153

```
. return list
```

```
scalars:
```

```
      r(N) = 35
      r(sum_w) = 35
      r(mean) = 6.763664872305734
      r(Var) = .0206653952894608
      r(sd) = .1437546357146817
      r(min) = 6.317525863647461
      r(max) = 7.021530151367188
      r(sum) = 236.7282705307007
```

```
matrices:
```

```
      r(C) : 2 x 2
```

```
. scalar lsmrtnostmean=r(mean)
```

```
. display lsmrtnostmean
```

```
6.7636649
```

```
. scalar gmean=exp(lsmrtnostmean)
```

```
. display gmean
```

```
865.80947
```

```
. means smrtnost
```

Variable	Type	Obs	Mean	[95% Conf. Interval]	
smrtnost	Arithmetic	35	874.12	833.776	914.464
	Geometric	35	865.8095	824.0931	909.6377
	Harmonic	35	856.6999	812.9898	905.3771

```
. return list
```

```
scalars:
```

```
      r(ub_h) = 905.3770659078801
      r(lb_h) = 812.989836787356
      r(Var_h) = 3.33772696849e-08
      r(mean_h) = 856.6998722903223
      r(ub_g) = 909.6376520150456
      r(lb_g) = 824.0930530006964
      r(Var_g) = .020665392547539
      r(mean_g) = 865.8094881516742
      r(N_pos) = 35
      r(ub) = 914.4639925009122
      r(lb) = 833.7760081966325
      r(Var) = 13793.48058661279
      r(mean) = 874.1200003487724
      r(N) = 35
```

```
. scalar gmean=r(mean_g)
```

```
. display gmean
```

```
865.80949
```

```

. scalar lstat=(opaz/2)*abs(log(nvklin/(gmean^2)/nvklog))
. display lstat
2.4573123

. display invchi2tail(1,0.05)
3.8414588

. display chi2tail(1,lstat)
.11697879

. display nvklin/(gmean^2), nvklog
.20318507 .17656682

```

```

. * Box-Coxov test (modificirani postopek)

```

```

. gen smrtnostt=smrtnost/gmean

```

```

. regress smrtnostt star tobak

```

Source	SS	df	MS	Number of obs =	35
Model	.422431286	2	.211215643	F(2, 32) =	33.26
Residual	.203185079	32	.006349534	Prob > F =	0.0000
Total	.625616365	34	.018400481	R-squared =	0.6752
				Adj R-squared =	0.6549
				Root MSE =	.07968

smrtnostt	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
star	.0609756	.0080523	7.57	0.000	.0445737	.0773776
tobak	.0016645	.0005794	2.87	0.007	.0004843	.0028447
_cons	.0632535	.1205535	0.52	0.603	-.1823059	.3088129

```

. scalar nvklint=e(rss)

```

```

. scalar lstatmod=(opaz/2)*abs(log(nvklint/nvklog))
. display lstatmod
2.4573135

```

```

. display invchi2tail(1,0.05)
3.8414588

```

```

. display chi2tail(1,lstatmod)
.1169787

```

```

. display nvklint, nvklog
.20318508 .17656682

```

```
. * Linearne transformacije
```

```
. regress smrtnost star tobak
```

Source	SS	df	MS	Number of obs =	35
Model	316665.524	2	158332.762	F(2, 32) =	33.26
Residual	152312.816	32	4759.7755	Prob > F =	0.0000
				R-squared =	0.6752
				Adj R-squared =	0.6549
				Root MSE =	68.991
Total	468978.34	34	13793.4806		

smrtnost	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
star	52.79328	6.971745	7.57	0.000	38.5923 66.99426
tobak	1.441107	.501654	2.87	0.007	.4192715 2.462943
_cons	54.76546	104.3763	0.52	0.603	-157.8422 267.3731

```
. gen smrtnosttr=smrtnost/100
```

```
. gen startr=star/100
```

```
. gen tobaktr=tobak
```

```
. regress smrtnosttr startr tobaktr
```

Source	SS	df	MS	Number of obs =	35
Model	31.6665583	2	15.8332792	F(2, 32) =	33.26
Residual	15.2312794	32	.475977481	Prob > F =	0.0000
				R-squared =	0.6752
				Adj R-squared =	0.6549
				Root MSE =	.68991
Total	46.8978377	34	1.37934817		

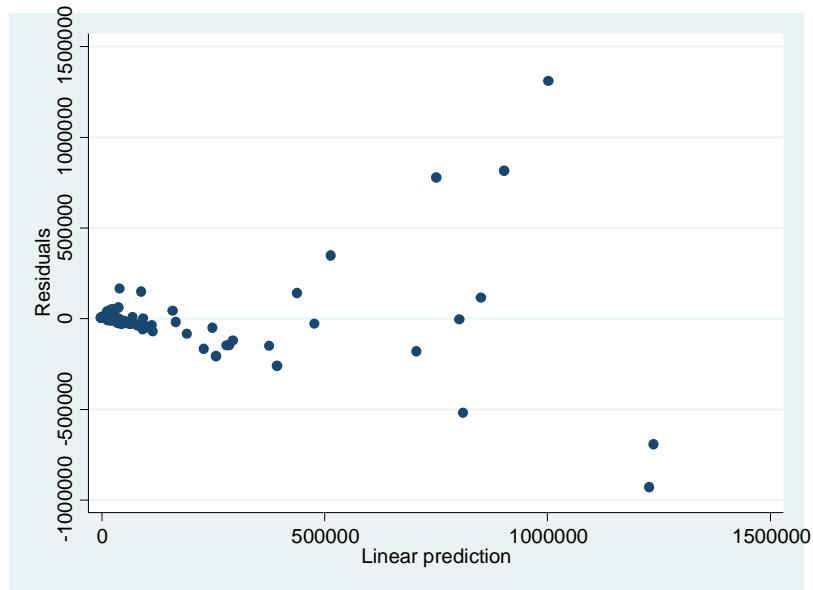
smrtnosttr	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
startr	52.79329	6.971745	7.57	0.000	38.59231 66.99427
tobaktr	.0144111	.0050165	2.87	0.007	.0041927 .0246294
_cons	.5476531	1.043763	0.52	0.603	-1.578423 2.67373



Primer 2: Analiziramo produkcijske funkcije za podjetja industrijske dejavnosti v RS v podskupini DL 30.02 – *Proizvodnja računalnikov in druge opreme za obdelavo podatkov* na podlagi podatkov iz zaključnih računov za leto 1998 (datoteka proizvod.dta).

- S pomočjo Ramseyevega RESET testa proučite pravilnost specifikacije za linearno produkcijsko funkcijo.
- S pomočjo Box-Coxovega testa ugotovite, ali je na osnovi našega vzorca podatkov ustreznejši linearni ali potenčni (dvojno-logaritemsko linearni) model.

a) Ramseyev RESET test:



Razširimo linearno produkcijsko funkcijo z drugo in tretjo potenco ocen odvisne spremenljivke, jo ocenimo in testiramo domnevo, da sta vrednosti obeh regresijskih koeficientov dodanih spremenljivk enaki 0:

$$\hat{Q}_i = 94370,4 - 22732,6 L_i - 0,6854 K_i + 7,40 \cdot 10^{-6} \hat{Q}_i^2 - 4,98 \cdot 10^{-12} \hat{Q}_i^3$$

(2,53)	(-3,84)	(-0,55)	(6,29)	(-7,06)
(0,014)	(0,000)	(0,582)	(0,000)	(0,000)

$n = 81$	$R_N^2 = 0,7578$	$\bar{R}_N^2 = 0,7451$
$s_e = 195.927,0$	$NVK_N = 0,29 \cdot 10^{13}$	$F(4, 76) = 59,4636$ (0,0000)

$$H_0: \beta_4 = \beta_5 = 0$$

$$H_1: \exists \beta_j \neq 0; j = 4, 5$$

$$F = \frac{(R_N^2 - R_0^2) / m}{(1 - R_N^2) / (n - k_N)} = \frac{(0,7578 - 0,5756) / 2}{(1 - 0,7578) / (81 - 5)} = 28,586$$

$$F_{k(m_1 = 2, m_2 = n - k = 76, \alpha = 0,05)} = 3,12$$

$F > F_k$, zavrnemo ničelno domnevo

Ker znaša vrednost F -statistike 28,5863, lahko na podlagi vzorčnih podatkov z zanemarljivo stopnjo značilnosti zavrnemo domnevo, da sta oba regresijska koeficienta dodanih spremenljivk enaka 0. Na podlagi Ramseyevega RESET testa smo torej ugotovili, da je osnovni linearni model smiselno razširiti, vendar bi bilo treba natančneje ugotoviti na kakšen način in s katerimi pojasnjevalnimi spremenljivkami.

b) Box-Coxov test:

H_0 : modela sta enakovredna

H_1 : modela nista enakovredna

$$\bar{y}_G = \left(\prod_{i=1}^n y_i \right)^{\frac{1}{n}} = e^{\frac{1}{n} \sum_{i=1}^n \ln y_i} = 47.024,1$$

$$l = \frac{n}{2} \left| \ln \frac{NVK_L / \bar{y}_G^2}{NVK_{LL}} \right| = 168,079$$

$$\chi_k^2 (m=1, \alpha=0,05) = 3,84146$$

$l > \chi^2$, zavrnamo ničelno domnevo

$NVK_L / \bar{y}_G^2 > NVK_{LL}$, boljši je dvojnologaritemsko-linearni model

Izpis rezultatov obdelav v programskem paketu Stata:

. * Ramseyev RESET test:

. regress q l k

Source	SS	df	MS	Number of obs =	81
Model	6.9350e+12	2	3.4675e+12	F(2, 78) =	52.90
Residual	5.1130e+12	78	6.5551e+10	Prob > F =	0.0000
Total	1.2048e+13	80	1.5060e+11	R-squared =	0.5756
				Adj R-squared =	0.5647
				Root MSE =	2.6e+05

q	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
l	9687.383	3640.852	2.66	0.009	2439.003 16935.76
k	2.27941	.7553228	3.02	0.003	.775678 3.783142
_cons	-11875.29	34865.13	-0.34	0.734	-81286.43 57535.85

. predict qe, resid

. predict qhat, xb

. scatter qe qhat

. gen q2=qhat^2

. gen q3=qhat^3

. regress q l k q2 q3

Source	SS	df	MS	Number of obs =	81
Model	9.1306e+12	4	2.2826e+12	F(4, 76) =	59.46
Residual	2.9174e+12	76	3.8387e+10	Prob > F =	0.0000
Total	1.2048e+13	80	1.5060e+11	R-squared =	0.7578
				Adj R-squared =	0.7451
				Root MSE =	2.0e+05

q	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
l	-22732.56	5925.006	-3.84	0.000	-34533.23	-10931.88
k	-.6854294	1.238961	-0.55	0.582	-3.153034	1.782175
q2	7.40e-06	1.18e-06	6.29	0.000	5.06e-06	9.74e-06
q3	-4.98e-12	7.06e-13	-7.06	0.000	-6.39e-12	-3.58e-12
_cons	94370.32	37316.77	2.53	0.014	20047.54	168693.1

```
. display invFtail(2,76,0.05)
3.1169818
```

```
. display Ftail(2,76,28.586292)
5.535e-10
```

```
. qui regress q l k
. estat ovtest // Stata vkljuci potence do cetrtega reda //
```

Ramsey RESET test using powers of the fitted values of q

Ho: model has no omitted variables

F(3, 75) = **26.33**
 Prob > F = **0.0000**

```
. * Box-Coxov test (originalni postopek):
```

```
. regress q l k
```

Source	SS	df	MS	Number of obs =	81
Model	6.9350e+12	2	3.4675e+12	F(2, 78) =	52.90
Residual	5.1130e+12	78	6.5551e+10	Prob > F =	0.0000
Total	1.2048e+13	80	1.5060e+11	R-squared =	0.5756
				Adj R-squared =	0.5647
				Root MSE =	2.6e+05

q	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
l	9687.383	3640.852	2.66	0.009	2439.003	16935.76
k	2.27941	.7553228	3.02	0.003	.775678	3.783142
_cons	-11875.29	34865.13	-0.34	0.734	-81286.43	57535.85

```
. scalar opaz=e(N)
. scalar par=e(rank)
. scalar nvklin=e(rss)
```

```
. regress lq ll lk
```

Source	SS	df	MS	Number of obs =	81
Model	178.261263	2	89.1306313	F(2, 78) =	190.75
Residual	36.44752	78	.467275898	Prob > F =	0.0000
Total	214.708783	80	2.68385978	R-squared =	0.8302
				Adj R-squared =	0.8259
				Root MSE =	.68358

lq	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ll	.9645479	.1199229	8.04	0.000	.7257997	1.203296
lk	.1885438	.0673358	2.80	0.006	.0544886	.322599
_cons	7.546026	.4617465	16.34	0.000	6.62676	8.465293

```

. scalar nvklog=e(rss)

. display opaz, par, nvklin, nvklog
81 3 5.113e+12 36.44752

. means q

Variable | Type      Obs      Mean      [95% Conf. Interval]
-----+-----
      q | Arithmetic  81    180519.3    94709.47    266329.2
      | Geometric   81    47024.12   32733.99    67552.64
      | Harmonic    81    15706.63   11426.1     25115.61
-----+-----

. scalar gmean=r(mean_g)
. display gmean
47024.117

. scalar lstat=(opaz/2)*abs(log(nvklin/(gmean^2)/nvklog))
. display lstat
168.0791

. display invchi2tail(1,0.05)
3.8414588

. display chi2tail(1,lstat)
1.944e-38

. display nvklin/(gmean^2), nvklog
2312.2462 36.44752

```

