Ekonometrija 1

Sedme vaje:
Preverjanje specifikacije regresijskega modela. Vpliv linearnih transformacij spremenljivk na ocene parametrov regresijske funkcije.

Na sedmih vajah bomo najprej izvedli Ramseyev RESET test, ki je namenjen preverjanju pravilnosti specifikacije regresijskega modela ter Box-Coxov test, ki ga uporabljamo pri odločanju med linearnim in dvojno logaritemskim modelom. Nato bomo proučili še vpliv linearnih transformacij na ocene parametrov regresijske funkcije.

♣-♣-♣

Primer 1: Na podlagi spodnjih podatkov (datoteka `stroski.dta`) o obsegu proizvodnje določene dobrine (Q; v kosih) in o celotnih stroških (C; v denarnih enotah) ocenite linearno funkcijo celotnih stroškov.

Q	1	2	3	4	5	6	7	8	9	10
C	150	220	250	260	265	280	295	350	410	520

S pomočjo ustreznega testa preverite ustreznost takšne specifikacije in pojasnite temeljno slabost uporabljenega testa. Kakšno specifikacijo funkcije celotnih stroškov predlagate?

Izpis rezultatov obdelav v programskem paketu Stata:

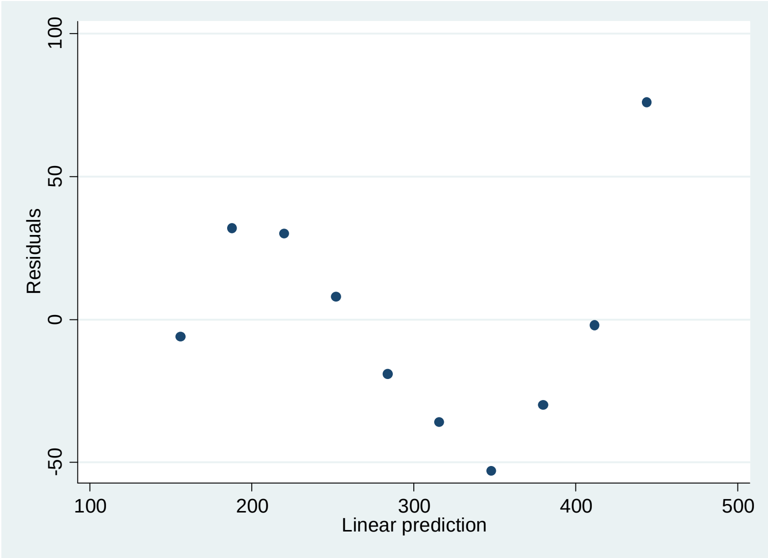
```
. regress c q

      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  1,     8) =   51.32
       Model |      84480     1       84480           Prob > F      =  0.0001
    Residual |      13170     8     1646.25           R-squared     =  0.8651
-------------+------------------------------           Adj R-squared =  0.8483
       Total |      97650     9       10850           Root MSE      =  40.574

------------------------------------------------------------------------------
           c |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           q |         32   4.467051     7.16   0.000     21.69896    42.30104
       _cons |        124   27.71732     4.47   0.002     60.08374    187.9163
------------------------------------------------------------------------------

. predict ce, resid
. predict chat, xb
```

```
. scatter ce chat
```




```
. gen c2=chat^2
. gen c3=chat^3

. regress c q c2 c3

      Source |       SS       df       MS              Number of obs =      10
-------------+------------------------------           F(  3,     6) =  915.80
       Model |  97437.2086     3  32479.0695           Prob > F      =  0.0000
    Residual |  212.791375     6  35.4652292           R-squared     =  0.9978
-------------+------------------------------           Adj R-squared =  0.9967
       Total |       97650     9       10850           Root MSE      =  5.9553

------------------------------------------------------------------------------
           c |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           q |   365.2392   27.13037    13.46   0.000     298.8536    431.6249
          c2 |  -.0402067   .0029539   -13.61   0.000    -.0474347   -.0329787
          c3 |   .0000481   3.27e-06    14.71   0.000     .0000401    .0000561
       _cons |   584.8318    30.1545    19.39   0.000     511.0464    658.6172
------------------------------------------------------------------------------

. display invFtail(2,6,0.05)
5.1432528

. display Ftail(2,6,180.95)
4.338e-06
```

■

<u>Primer 2:</u> Na podlagi podatkov za Slovenijo (razdobje 1965–1989), ocenite linearno in dvojno logaritemsko funkcijo porabe (datoteka `potrosnja1.dta`):

$$OP_t = \beta_1 + \beta_2 OD_t + \beta_3 SOC_t + u_t \,,$$
$$\ln OP_t = \beta_1 + \beta_2 \ln OD_t + \beta_3 \ln SOC_t + u_t \,,$$

pri čemer so spremenljivke naslednje (vse v cenah iz leta 1972):

- OP_t: osebna poraba v mio DIN;
- OD_t: osebni dohodki v mio DIN;
- SOC_t: socialni prejemki v mio DIN.

Razložite postopek primerjave razlagalne moči regresijskih modelov na podlagi vrednosti determinacijskih koeficientov. Ali lahko zavrnemo ničelno domnevo, da sta modela empirično enakovredna?

Izpis rezultatov obdelav v programskem paketu Stata:

```
* Primerjava multiplih determinacijskih koeficientov:

. gen lop=log(op)
. gen lod=log(od)
. gen lsoc=log(soc)

. regress op od soc

      Source |       SS       df       MS              Number of obs =      25
-------------+------------------------------           F(  2,    22) =  253.82
       Model |  3.67855587     2  1.83927794           Prob > F      =  0.0000
    Residual |    .1594234    22  .007246518           R-squared     =  0.9585
-------------+------------------------------           Adj R-squared =  0.9547
       Total |  3.83797927    24  .159915803           Root MSE      =  .08513

------------------------------------------------------------------------------
          op |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          od |   .6648524   .0872809     7.62   0.000     .4838429    .8458619
         soc |   .9483866    .283467     3.35   0.003      .360512    1.536261
       _cons |   .2653599   .0937114     2.83   0.010     .0710144    .4597054
------------------------------------------------------------------------------

. ereturn list

scalars:
                  e(N) =  25
               e(df_m) =  2
               e(df_r) =  22
                  e(F) =  253.8154037665093
                 e(r2) =  .9584616308434277
               e(rmse) =  .0851264833658064
                e(mss) =  3.678555870555993
                e(rss) =  .159423399745036
               e(r2_a) =  .9546854154655575
                 e(ll) =  27.7148810369553
               e(ll_0) =  -12.04934047424811
               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)
```

```
. regress lop lod lsoc

      Source |       SS       df       MS              Number of obs =      25
-------------+------------------------------           F(  2,    22) =  279.30
       Model |  1.11851042     2  .559255212           Prob > F      =  0.0000
    Residual |  .044051471    22   .00200234           R-squared     =  0.9621
-------------+------------------------------           Adj R-squared =  0.9587
       Total |  1.16256189    24  .048440079           Root MSE      =  .04475

------------------------------------------------------------------------------
         lop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lod |   .6994262   .1117262     6.26   0.000     .4677203    .9311321
        lsoc |   .1758547   .0753965     2.33   0.029     .0194919    .3322175
       _cons |   .3604694   .1452971     2.48   0.021     .0591417     .661797
------------------------------------------------------------------------------

. scalar nvklog=e(rss)

. display opaz, par, nvklin, nvklog
25 3 .1594234 .04405147


. qui regress lop lod lsoc
. predict yloghat, xb

. gen antilyloghat=exp(yloghat)

. correlate op antilyloghat
(obs=25)

             |       op antily~t
-------------+------------------
          op |   1.0000
antilyloghat |   0.9793   1.0000

. return list

scalars:
                  r(N) =  25
                r(rho) =  .9792916410344565

matrices:
                  r(C) :  2 x 2

. scalar r2p=r(rho)^2

. display r2p
.95901212


* Box-Coxov test (originalni postopek):

. means op

    Variable |    Type        Obs        Mean       [95% Conf. Interval]
-------------+----------------------------------------------------------
          op | Arithmetic      25    1.985546        1.820478   2.150614
             |  Geometric      25    1.942643        1.773935   2.127395
             |   Harmonic      25     1.89542        1.724275   2.104283
------------------------------------------------------------------------

. return list

scalars:
               r(ub_h) =  2.104283291203963
               r(lb_h) =  1.724275335263516
              r(Var_h) =  .0160941132749612
```

```
               r(mean_h) =  1.895420251551919
                 r(ub_g) =  2.127395190730542
                 r(lb_g) =  1.773934675261189
                r(Var_g) =  .0484400787132108
               r(mean_g) =  1.942642555083358
                r(N_pos) =  25
                   r(ub) =  2.150614442816441
                   r(lb) =  1.820477571496302
                  r(Var) =  .1599158029292095
                 r(mean) =  1.985546007156372
                    r(N) =  25

. scalar gmean=r(mean_g)

. display gmean
1.9426426

. scalar lstat=(opaz/2)*abs(log(nvklin/(gmean^2)/nvklog))
. display lstat
.52366949

. display invchi2tail(1,0.05)
3.8414588

. display chi2tail(1,lstat)
.46928056

. display nvklin/(gmean^2), nvklog
.04224412 .04405147


* Box-Coxov test (modificirani postopek):

. gen opt=op/gmean

. regress opt od soc

      Source |       SS       df       MS              Number of obs =      25
-------------+------------------------------           F(  2,    22) =  253.82
       Model |  .974746187     2  .487373094           Prob > F      =  0.0000
    Residual |  .042244111    22  .001920187           R-squared     =  0.9585
-------------+------------------------------           Adj R-squared =  0.9547
       Total |   1.0169903    24  .042374596           Root MSE      =  .04382

------------------------------------------------------------------------------
         opt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          od |   .3422413   .0449289     7.62   0.000     .2490643    .4354182
         soc |   .4881941   .1459182     3.35   0.003     .1855782      .79081
       _cons |   .1365974   .0482391     2.83   0.010     .0365556    .2366392
------------------------------------------------------------------------------

. scalar nvklint=e(rss)

. scalar lstatmod=(opaz/2)*abs(log(nvklint/nvklog))
. display lstatmod
.52367237

. display invchi2tail(1,0.05)
3.8414588

. display chi2tail(1,lstatmod)
.46927934

. display nvklint, nvklog
.04224411 .04405147
```

■

Primer 3: Za 10 evropskih držav smo proučevali odvisnost števila avtomobilov na 1.000 prebivalcev (*AVTO*) od bruto domačega proizvoda na prebivalca (v USD) (*BDP*) in števila izdanih vozniških dovoljenj v milijonih (*DOV*) ter ocenili naslednjo regresijsko funkcijo:

$$\widehat{AVTO} = 17,5701 + 0,02187BDP + 2,78452DOV \qquad R^2 = 0,9223$$
$$t: \quad (0,39) \quad (6,18) \qquad (3,23) \qquad s_e = 36,34$$

Zapišite regresijsko funkcijo z ocenami vseh pripadajočih parametrov, če bi bila odvisna spremenljivka število avtomobilov na 100 prebivalcev, število vozniških dovoljenj izraženo v tisočih, bruto domači proizvod na prebivalca pa v evrih (upoštevajte 1 EUR = 1,15 USD).

Primer 4: Na podlagi podatkov za preteklo leto smo za 100 slučajno izbranih gospodinjstev ocenili naslednjo funkcijo odvisnosti višine izdatkov za poletni dopust od realnega razpoložljivega dohodka:

$$\widehat{\ln DOP} = -11,1 + 1,45 \ln RRD \qquad R^2 = 0,935 \qquad s_e = 0,284$$
$$(-4,1) \quad (7,3)$$

Pri tem smo z *DOP* označili izdatke za dopust v 1.000 SIT, z *RRD* pa realni razpoložljivi dohodek prav tako v 1.000 SIT. Kakšne bi bile ocene parametrov te funkcije, če bi bili obe spremenljivki izraženi v SIT (namesto v 1.000 SIT, kot do sedaj)?

■