Ekonometrija 1

Osme vaje:
Vpliv linearnih transformacij spremenljivk na ocene parametrov regresijske funkcije. Napovedovanje povprečne in posamične vrednosti odvisne spremenljivke.

Na osmih vajah bomo nadaljevali s proučevanjem vplivov linearnih transformacij na ocene parametrov regresijske funkcije. Pogledali si bomo tudi standardizacijo kot poseben primer linearne transformacije ter spoznali beta regresijske koeficiente. Nato pa se bomo posvetili napovedovanju (povprečne in posamične) vrednosti odvisne spremenljivke na podlagi ocenjenega regresijskega modela. Prikazali bomo postopek napovedovanja s pomočjo matričnih obrazcev in z uporabo nepravih spremenljivk.

♣-♣-♣

Primer 1: Na podlagi podatkov za Slovenijo (razdobje 1965–1989), ocenite naslednjo funkcijo porabe: $OP_t = \beta_1 + \beta_2 OD_t + \beta_3 SOC_t + u_t$, kjer je *OP* osebna poraba, *OD* so osebni dohodki, *SOC* pa socialni prejemki (datoteka `potrosnja1.dta`). Vse spremenljivke so izražene v mio DIN po cenah iz leta 1972.

Transformirajte podatke v indekse s stalno osnovo v letu 1965 in izračunajte ocene parametrov transformirane funkcije. Kaj bi se zgodilo, če bi osnovne podatke spremenili v stopnje rasti?

Izpis rezultatov obdelav v programskem paketu Stata:

```
. list year op od soc

     +-----------------------------------+
     | year        op        od       soc |
     |-----------------------------------|
  1. | 1965   1.24064   1.22162   .217641 |
  2. | 1966   1.20039   1.30085   .234338 |
  3. | 1967   1.35596    1.3316   .216654 |
  4. | 1968   1.51461   1.38599   .230007 |
  5. | 1969   1.42015   1.47827   .241759 |
     |-----------------------------------|
  6. | 1970    1.6326   1.63529   .274318 |
  7. | 1971   1.77324   1.76129   .305323 |
  8. | 1972    1.9179    1.8373     .3542 |
  9. | 1973   1.93033   1.88083   .355417 |
 10. | 1974   1.96894   1.95176   .355336 |
     |-----------------------------------|
 11. | 1975   2.03957   2.02637   .385763 |
 12. | 1976   2.12242   2.15717   .413712 |
 13. | 1977    2.2178   2.33236   .442065 |
 14. | 1978   2.30118   2.51628   .471227 |
 15. | 1979   2.36374   2.57485   .492795 |
     |-----------------------------------|
 16. | 1980   2.46614   2.55371   .499732 |
 17. | 1981   2.16908   2.34876   .456095 |
 18. | 1982   2.17416   2.27551   .450201 |
```

```
 19. | 1983   2.07052   2.08096   .413724 |
 20. | 1984   1.98772   2.04743   .393934 |
     |-----------------------------------|
 21. | 1985   2.09171   2.21289   .455336 |
 22. | 1986   2.55057   2.49932   .607234 |
 23. | 1987   2.66605   2.43745   .640811 |
 24. | 1988   2.30521   2.20272   .578541 |
 25. | 1989   2.15802   2.19513   .634889 |
     +-----------------------------------+

. 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
------------------------------------------------------------------------------

. scalar op65=op in 1
. scalar od65=od in 1
. scalar soc65=soc in 1

. gen iop=100*op/op65
. gen iod=100*od/od65
. gen isoc=100*soc/soc65

. regress iop iod isoc

      Source |       SS       df       MS              Number of obs =      25
-------------+------------------------------           F(  2,    22) =  253.82
       Model |  23899.3329     2  11949.6665           Prob > F      =  0.0000
    Residual |  1035.76295    22  47.0801339           R-squared     =  0.9585
-------------+------------------------------           Adj R-squared =  0.9547
       Total |  24935.0959    24  1038.96233           Root MSE      =  6.8615

------------------------------------------------------------------------------
         iop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         iod |   .6546597   .0859428     7.62   0.000     .4764253    .8328942
        isoc |    .166372   .0497276     3.35   0.003     .0632433    .2695007
       _cons |   21.38895    7.55347     2.83   0.010     5.724008    37.05388
------------------------------------------------------------------------------

. gen sop=100*op/op[_n-1]-100
(1 missing value generated)

. gen sod=100*od/od[_n-1]-100
(1 missing value generated)

. gen ssoc=100*soc/soc[_n-1]-100
(1 missing value generated)

. regress sop sod ssoc

      Source |       SS       df       MS              Number of obs =      24
-------------+------------------------------           F(  2,    21) =   12.35
       Model |  854.289289     2  427.144645           Prob > F      =  0.0003
    Residual |  726.605459    21  34.6002599           R-squared     =  0.5404
-------------+------------------------------           Adj R-squared =  0.4966
       Total |  1580.89475    23  68.7345543           Root MSE      =  5.8822
```

```
------------------------------------------------------------------------------
         sop |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sod |   .6821941   .3399349     2.01   0.058    -.0247392    1.389128
        ssoc |   .2406695   .2112369     1.14   0.267    -.1986217    .6799607
       _cons |  -.3450666   1.358774    -0.25   0.802    -3.170792    2.480658
------------------------------------------------------------------------------
```

■

<u>Primer 2:</u> Na podlagi podatkov v datoteki `investicije.dta` smo za razdobje 1970–1989 ocenili naslednjo investicijsko funkcijo slovenskega gospodarstva:

$$\widehat{INV} = -0,32131 - 0,15188 OS + 0,67872 DP \qquad R^2 = 0,6578$$
$$t: \quad (-1,19) \quad (-5,10) \qquad (5,65) \qquad s_e = 0,16297$$

Pri tem so *INV* investicije, *OS* so osnovna sredstva, *DP* pa je domači proizvod (vse spremenljivke so v mio DIN po cenah iz leta 1972). Izpeljite in razložite standardizirane regresijske koeficiente ter izračunajte ocene vseh parametrov transformiranega modela.

Beta koeficienti so standardizirani regresijski koeficienti, ki jih dobimo na podlagi standardiziranih vrednosti odvisne in pojasnjevalnih spremenljivk. Ker gre pri standardizaciji spremenljivk za linearne transformacije, si lahko pri izračunu beta koeficientov pomagamo z obrazci, ki smo jih uporabljali v predhodnih nalogah:

$$y_i^{Tr} = \frac{y_i - \overline{y}}{s_y} = \frac{1}{s_y} y_i - \frac{\overline{y}}{s_y} \qquad \Rightarrow \qquad a = \frac{1}{s_y} \qquad c = -\frac{\overline{y}}{s_y}$$

$$x_{ji}^{Tr} = \frac{x_{ji} - \overline{x}_j}{s_{x_j}} = \frac{1}{s_{x_j}} x_{ij} - \frac{\overline{x}_j}{s_{x_j}} \qquad \Rightarrow \qquad d_j = \frac{1}{s_{x_j}} \qquad f_j = -\frac{\overline{x}_j}{s_{x_j}}$$

$$b_j^{Tr} = \frac{a}{d_j} b_j = \frac{1/s_y}{1/s_{x_j}} b_j = \frac{s_{x_j}}{s_y} b_j; \quad j = 2,3,\ldots,k.$$

Beta koeficient nam pove, za koliko standardnih odklonov se v povprečju spremeni vrednost odvisne spremenljivke, če se, ceteris paribus, vrednost pojasnjevalne spremenljivke poveča za en standardni odklon. Uporabljamo jih predvsem pri presojanju pomembnosti posamezne pojasnjevalne spremenljivke. K pojasnitvi variance odvisne spremenljivke največ prispeva tista spremenljivka (in je s tem seveda najpomembnejša), ki ima po absolutni vrednosti največji beta koeficient.

Izpis rezultatov obdelav v programskem paketu Stata:

```
. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        year |        20      1979.5     5.91608       1970       1989
          dp |        20    5.121725    1.015669     3.1468     6.1611
```

```
         inv |        20    1.140806    .2635112     .73843    1.62434
          os |        20    13.26153    4.101662     6.6331    19.0711

. regress inv os dp

      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  2,    17) =   16.34
       Model |  .867795456     2  .433897728           Prob > F      =  0.0001
    Residual |  .451529117    17  .026560536           R-squared     =  0.6578
-------------+------------------------------           Adj R-squared =  0.6175
       Total |  1.31932457    19  .069438135           Root MSE      =  .16297

------------------------------------------------------------------------------
         inv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          os |  -.1518757   .0297665    -5.10   0.000    -.2146775   -.0890738
          dp |   .6787201   .1202086     5.65   0.000     .4251022     .932338
       _cons |  -.3213069   .2710886    -1.19   0.252    -.8932539    .2506402
------------------------------------------------------------------------------

. egen invs=std(inv)
. egen oss=std(os)
. egen dps=std(dp)

. regress invs oss dps

      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  2,    17) =   16.34
       Model |  12.4973899     2  6.24869496           Prob > F      =  0.0001
    Residual |  6.50261026    17  .382506486           R-squared     =  0.6578
-------------+------------------------------           Adj R-squared =  0.6175
       Total |  19.0000002    19  1.00000001           Root MSE      =  .61847

------------------------------------------------------------------------------
        invs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         oss |  -2.364008   .4633282    -5.10   0.000    -3.341545   -1.386471
         dps |   2.616038   .4633282     5.65   0.000     1.638501    3.593575
       _cons |   7.65e-09   .1382943     0.00   1.000    -.2917755    .2917756
------------------------------------------------------------------------------

. regress invs oss dps, noconst

      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  2,    18) =   17.30
       Model |  12.4973899     2  6.24869496           Prob > F      =  0.0001
    Residual |  6.50261026    18  .361256125           R-squared     =  0.6578
-------------+------------------------------           Adj R-squared =  0.6197
       Total |  19.0000002    20  .950000009           Root MSE      =  .60105

------------------------------------------------------------------------------
        invs |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         oss |  -2.364008   .4502741    -5.25   0.000    -3.309999   -1.418018
         dps |   2.616038   .4502741     5.81   0.000     1.670047    3.562029
------------------------------------------------------------------------------

. regress inv os dp, beta

      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(  2,    17) =   16.34
       Model |  .867795456     2  .433897728           Prob > F      =  0.0001
    Residual |  .451529117    17  .026560536           R-squared     =  0.6578
-------------+------------------------------           Adj R-squared =  0.6175
       Total |  1.31932457    19  .069438135           Root MSE      =  .16297
```

```
------------------------------------------------------------------------------
         inv |      Coef.   Std. Err.      t    P>|t|                     Beta
-------------+----------------------------------------------------------------
          os |  -.1518757   .0297665    -5.10   0.000                -2.364008
          dp |   .6787201   .1202086     5.65   0.000                 2.616038
       _cons |  -.3213069   .2710886    -1.19   0.252                        .
------------------------------------------------------------------------------
```

■

Primer 3: Sodelavci avtorevije so proučevali razmere na trgu rabljenih avtomobilov določenega tipa. Želeli so ugotoviti, kakšen vpliv imata starost (*STAR*; v mesecih) in število prevoženih kilometrov (*KM*; v 1.000 km) na ceno rabljenih avtomobilov (*CENA*; v evrih). Podatki so v datoteki `avtomobili.dta`. Ocenite regresijsko funkcijo:

$$CENA_i = \beta_1 + \beta_2 STAR_i + \beta_3 KM_i + u_i$$

ter izračunajte točkovno in intervalno napoved za ceno avtomobila, ki je star 5 let in ima 75.000 prevoženih kilometrov.

a) Napoved posamezne vrednosti v matrični obliki

1. Izračunamo točkovno napoved posamezne vrednosti odvisne spremenljivke:

$$\hat{y}_0 = \mathbf{x}_0^{\mathbf{T}} \mathbf{b} =$$

2. Izračunamo standardno napako napovedi posamezne vrednosti odvisne spremenljivke:

$$\text{var}\left(\hat{y}_0 - y_0\right) = s_e^2 \left(\mathbf{x}_0^{\mathbf{T}} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_0 + 1\right) = \mathbf{x}_0^{\mathbf{T}} \left[\text{var}-\text{cov}(\mathbf{b})\right] \mathbf{x}_0 + s_e^2 =$$

$$\text{se}\left(\hat{y}_0 - y_0\right) = \sqrt{\text{var}\left(\hat{y}_0 - y_0\right)} =$$

3. Izračunamo intervalno napoved posamezne vrednosti odvisne spremenljivke:

$$\hat{y}_0 - t_{\alpha/2}\text{se}\left(\hat{y}_0 - y_0\right) \le y_0 \le \hat{y}_0 + t_{\alpha/2}\text{se}\left(\hat{y}_0 - y_0\right); \quad \alpha =$$

Izpis rezultatov obdelav v programskem paketu Stata:

```
. regress cena star km

      Source |       SS       df       MS              Number of obs =      22
-------------+------------------------------           F(  2,    19) =   69.22
       Model |   241090396     2   120545198           Prob > F      =  0.0000
    Residual |  33087786.3    19  1741462.43           R-squared     =  0.8793
-------------+------------------------------           Adj R-squared =  0.8666
       Total |   274178182    21  13056103.9           Root MSE      =  1319.6

------------------------------------------------------------------------------
        cena |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        star |  -61.62429    20.6776    -2.98   0.008     -104.903   -18.34559
          km |  -35.24952   15.37506    -2.29   0.033    -67.42989    -3.06915
       _cons |   14284.76   509.8481    28.02   0.000     13217.64    15351.89
------------------------------------------------------------------------------

. matrix b=(e(b))'
. matrix list b

b[3,1]
               y1
 star  -61.624291
   km  -35.249522
_cons   14284.762

. matrix x0=(60\75\1)
. matrix list x0

x0[3,1]
    c1
r1  60
r2  75
r3   1

. matrix y0=x0'*b
. matrix list y0

symmetric y0[1,1]
           y1
c1  7943.5907

. matrix vce=e(V)
. matrix list vce

symmetric vce[3,3]
             star          km       _cons
 star   427.56298
   km  -285.96651   236.39249
_cons  -3100.1392  -599.45182   259945.07

. scalar se=e(rmse)
. display se
1319.6448

. matrix var0 = x0'*vce*x0 + se^2
. matrix list var0

symmetric var0[1,1]
           c1
r1  1834708.9
```

```
. matrix se0=cholesky(var0)
. matrix list se0

symmetric se0[1,1]
           c1
r1  1354.5143
```

■

b) Napoved posamezne vrednosti s pomočjo nepravih spremenljivk

1. Proučevanim spremenljivkam dodamo še nepravo spremenljivko, ki ima pri vseh n opazovanih enotah vrednost nič.

2. Opazovanim enotam dodamo še eno enoto, pri kateri ima neprava spremenljivka vrednost 1, odvisna 0, pojasnjevalne spremenljivke pa imajo enake vrednosti kot pri enoti, pri kateri izračunavamo napoved.

3. Ocenimo naslednji razširjeni regresijski model:

$$\hat{y}_i = b_1 + b_2 x_{2i} + \ldots + b_k x_{ki} + b_{k+1} D_i ,$$

pri čemer velja:

a) ocene regresijskih koeficientov b_1, b_2, ..., b_k ostanejo nespremenjene v primerjavi z osnovnim regresijskim modelom, prav tako pa tudi njihove standardne napake in *t*–statistike;
b) točkovna napoved posamezne vrednosti odvisne spremenljivke je enaka z minus ena pomnoženi oceni regresijskega koeficienta b_{k+1};
c) standardna napaka napovedi posamezne vrednosti je enaka standardni napaki regresijskega koeficienta b_{k+1}.

$$\widehat{CENA}_i = b_1 + b_2 STAR_i + b_3 KM_i + b_4 D_i$$

Izpis rezultatov obdelav v programskem paketu Stata:

```
. gen d=0

. set obs `=_N+1'
obs was 22, now 23

. replace d=1 if _n==23
(1 real change made)

. replace cena=0 if _n==23
(1 real change made)

. replace star=60 if _n==23
(1 real change made)
```

```
. replace km=75 if _n==23
(1 real change made)

. regress cena star km d

      Source |       SS       df       MS              Number of obs =      23
-------------+------------------------------           F(  3,    19) =   62.01
       Model |   323981779     3   107993926           Prob > F      =  0.0000
    Residual |  33087786.3    19  1741462.43           R-squared     =  0.9073
-------------+------------------------------           Adj R-squared =  0.8927
       Total |   357069565    22  16230434.8           Root MSE      =  1319.6

------------------------------------------------------------------------------
        cena |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        star |  -61.62429    20.6776    -2.98   0.008     -104.903   -18.34559
          km |  -35.24952   15.37506    -2.29   0.033    -67.42989    -3.06915
           d |  -7943.591   1354.514    -5.86   0.000    -10778.62    -5108.56
       _cons |   14284.76   509.8481    28.02   0.000     13217.64    15351.89
------------------------------------------------------------------------------
```

■

Sedaj izračunajte še točkovno in intervalno napoved za povprečno ceno avtomobila, ki je star 5 let in ima 75.000 prevoženih kilometrov.

c) Napoved povprečne vrednosti v matrični obliki

1. Izračunamo točkovno napoved povprečne vrednosti odvisne spremenljivke:

$$\hat{y}_0 = \mathbf{x_0^T b} =$$

2. Izračunamo standardno napako napovedi povprečne vrednosti odvisne spremenljivke:

$$\mathrm{var}\left(\hat{y}_0\right) = s_e^2\left(\mathbf{x_0^T}\left(\mathbf{X'X}\right)^{-1}\mathbf{x}_0\right) = \mathbf{x_0^T}\left[\mathrm{var-cov}(\mathbf{b})\right]\mathbf{x}_0 =$$

$$\mathrm{se}\left(\hat{y}_0\right) = \sqrt{\mathrm{var}\left(\hat{y}_0\right)} =$$

3. Izračunamo intervalno napoved povprečne vrednosti odvisne spremenljivke:

$$\hat{y}_0 - t_{\alpha/2}\mathrm{se}\left(\hat{y}_0\right) \leq E\left(y_0 \middle| \mathbf{x_0}\right) \leq \hat{y}_0 + t_{\alpha/2}\mathrm{se}\left(\hat{y}_0\right); \quad \alpha =$$

Izpis rezultatov obdelav v programskem paketu Stata:

```
. qui regress cena star km // prej ponovno odpreti datoteko //

. matrix b=(e(b))'
. matrix list b

b[3,1]
               y1
 star  -61.624291
   km  -35.249522
_cons   14284.762

. matrix x0=(60\75\1)
. matrix list x0

x0[3,1]
    c1
r1  60
r2  75
r3   1

. matrix y0=x0'*b
. matrix list y0

symmetric y0[1,1]
           y1
c1  7943.5907

. matrix vce=e(V)
. matrix list vce

symmetric vce[3,3]
             star          km       _cons
 star   427.56298
   km  -285.96651   236.39249
_cons  -3100.1392  -599.45182   259945.07

. matrix var0 = x0'*vce*x0
. matrix list var0

symmetric var0[1,1]
           c1
c1  93246.456

. matrix se0=cholesky(var0)
. matrix list se0

symmetric se0[1,1]
           c1
c1  305.36283
```

■