AMERICAN UNIVERSITY - CENTRAL ASIA 06.10.2022; 10.10.2022
Econometrics I (ECO 608)
Nurgul Tilenbaeva

Seminar 3

Using OLS in Stata

1 Running a Regression Using OLS

1. Import the Stata data file `"reg01"` from the e-course platform.
2. Run the linear regression given by

   $income = \beta_0 + \beta_1 educ + \beta_2 jobexp + \beta_3 race + u$
3. Interpret the OLS coefficients (intercept and slope coefficients). When doing it, make sure to discuss the sign of the coefficient, its magnitude and ceteris-paribus interpretations, as well as statistical significance of the coefficients. For reference, `income` is measured in `thousands of US dollars`, `educ` and `jobexp` in `years`, and `race` takes the value of `1` for `black` and `0` for `white`.
4. What is the estimated effect on `income` if an individual engages in part-time one-year master degree studies, while working at the same time?
5. How do you interpret the R-squared obtained from running this regression?

2 Inference and Hypothesis-Testing

1. Test the null hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_1 : \beta_1 > 0$ at `5% significance level`. Make the calculations by hand drawing a graph and using the corresponding table. What do the results of hypothesis-testing suggest?
2. Test the null hypothesis $H_0 : \beta_3 = 0$ against the alternative $H_1 : \beta_3 < 0$ at `10% significance level`. Make the calculations by hand drawing a graph and using the corresponding table. What do the results of hypothesis-testing suggest?

3. Test the null hypothesis $H_0 : \beta_2 = 0$ against the alternative $H_1 : \beta_2 \neq 0$ at `1% significance level`. Make the calculations by hand drawing a graph and using the corresponding table. What do the results of hypothesis-testing suggest?

4. Test the null hypothesis $H_0 : \beta_1 = 2$ against the alternative $H_1 : \beta_1 \neq 2$ at `5% significance level`. Make the calculations by hand drawing a graph and using the corresponding table. What do the results of hypothesis-testing suggest?

5. Test the null hypothesis $H_0 : \beta_1 = \beta_2$ against the alternative $H_1 : \beta_1 \neq \beta_2$ at `10% significance level`. As it is difficult to make the calculations by hand, use Stata for this hypothesis-testing. What do the results suggest?

6. Test the null hypothesis $H_0 : \beta_1 = 0, \beta_2 = 0$ against the alternative $H_1 : H_0$ *is not true* at `1% significance level`. Use Stata for this hypothesis-testing. What do the results suggest?

3 Review of Formulas

1. The following is the regression output linking academic performance of the school `api00` to a number of variables, including whether it is a year-round school `yr_rnd`, the average class size in kindergarten through 3rd grade `acs_k3`, the percentage of students receiving free meals `meals`, the percentage of teachers who have full teaching credentials `full`, etc. Please fill in the gaps in the regression output.

```
      Source |       SS       df       MS              Number of obs =     395
-------------+------------------------------           F(    ,      ) =  232.41
       Model |  6740702.01     9   748966.89           Prob > F      =  0.0000
    Residual |  1240707.78   385  3222.61761           R-squared     =  0.8446
-------------+------------------------------           Adj R-squared =
       Total |  7981409.79   394  20257.3852           Root MSE      =  56.768

------------------------------------------------------------------------------
       api00 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ell |  -.8600707   .2106317            0.000    -1.274203   -.4459382
       meals |  -2.948216   .1703452   -17.31   0.000
      yr_rnd |  -19.88875               -2.15   0.032    -38.09218    -1.68531
    mobility |             .4362053    -2.98   0.003    -2.158995   -.4437089
      acs_k3 |    1.3187   2.252683            0.559
      acs_46 |   2.032456   .7983213     2.55   0.011
        full |    .609715               1.28   0.201    -.3258169    1.545247
        emer |             .6054086    -1.17   0.244     -1.89694    .4837018
      enroll |             .0167921    -0.72   0.469    -.0451798    .0208517
       _cons |   778.8305   61.68663    12.63   0.000     657.5457    900.1154
------------------------------------------------------------------------------
```

4 Gauss-Markov Assumptions

1. Is Assumption MLR1. Linear in Parameters satisfied for the following models? Explain.

   $y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + u$

   $y = \beta_0 + \beta_1 x_1 + \beta_2 log(x_2) + u$

   $y = \beta_0 + \beta_1^2 x_1 + \beta_2 x_2^2 + u$

   $y = \beta_0 + \beta_1 x_1 + log(\beta_2) x_2 + u$

2. Is Assumption MLR3. No Perfect Collinearity satisfied for the following models? Explain.

   $cons = \beta_0 + \beta_1 income + \beta_2 income^2 + u$

   $income = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 kyrgyz + \beta_4 russian + \beta_5 uzbek + u.$ The assumption is that the population of interest is represented by Kyrgyz, Russian and Uzbek ethnicity.

   $gpa = \beta_0 + \beta_1 attend + \beta_2 expend_{educ} + \beta_3 expend_{health} + \beta_4 expend_{education\ and\ health} +$

u

3. Under which Gauss-Markov Assumptions OLS estimators are unbiased estimators of the population parameters?

4. If the errors exhibit heteroskedasticity (i.e. violate MLR5. Homoskedasticity), does it lead to biased OLS estimates?

5. Under which Gauss-Markov Assumptions OLS estimators are the best linear unbiased estimators (BLUEs) of the population parameters?