AMERICAN UNIVERSITY - CENTRAL ASIA 26.09.2022
Econometrics I (ECO 608)
Nurgul Tilenbaeva

Seminar 2

The Simple Regression Model

1. Let `kids` denote the number of children ever born to a woman, and let `educ` denote years of education for the woman. A simple model relating fertility to years of education is

$$kids = \beta_0 + \beta_1 educ + u,$$

where `u` is the unobserved error.

   (a) What kinds of factors are contained in `u`? Are these likely to be correlated with level of education?

   (b) Will a simple regression analysis uncover the ceteris paribus effect of education on fertility? Explain.

2. The following table contains the `ACT` scores and the `GPA` (grade point average) for eight college students. Grade point average is based on a four-point scale and has been rounded to one digit after the decimal.

Student	*GPA*	*ACT*
1	2.8	21
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	29
6	3.0	25
7	2.7	25
8	3.7	30

   (a) Estimate the relationship between `GPA` and `ACT` using OLS; that is, obtain the intercept and slope estimates in the equation

$$\hat{GPA} = \hat{\beta}_0 + \hat{\beta}_1 ACT$$

Comment on the direction of the relationship. Does the intercept have a useful interpretation here? Explain. How much higher is the `GPA` predicted to be if the `ACT` score is increased by five points?

(b) Compute the fitted values and residuals for each observation, and verify that the residuals (approximately) sum to zero.

(c) What is the predicted value of `GPA` when `ACT`=20?

(d) How much of the variation in `GPA` for these eight students is explained by `ACT`? Explain.

3. Use the data in `SLEEP75` from Biddle and Hamermesh (1990) to study whether there is a tradeoff between the time spent sleeping per week and the time spent in paid work. We could use either variable as the dependent variable. For concreteness, estimate the model

$$sleep = \beta_0 + \beta_1 totwrk + u$$

where `sleep` is minutes spent sleeping at night per week and `totwrk` is total minutes worked during the week.

(a) Report your results in equation form along with the number of observations and R^2. What does the intercept in this equation mean?

(b) If `totwrk` increases by 2 hours, by how much is `sleep` estimated to fall? Do you find this to be a large effect?

4. Use the data in `CHARITY` [obtained from Franses and Paap (2001)] to answer the following questions:

(a) What is the average gift in the sample of 4,268 people (in Dutch guilders)? What percentage of people gave no gift?

(b) What is the average mailings per year? What are the minimum and maximum values?

(c) Estimate the model

$$gift = \beta_0 + \beta_1 mailsyear + u$$

by OLS and report the results in the usual way, including the sample size and R-squared.

(d) Interpret the slope coefficient. If each mailing costs one guilder, is the charity expected to make a net gain on each mailing? Does this mean the charity makes a net gain on every mailing? Explain.

(e) What is the smallest predicted charitable contribution in the sample? Using this simple regression analysis, can you ever predict zero for `gift`?