Niclas Knecht (niclas.knecht@u-bordeaux.fr)
*Econometrics*
Spring 2024

**Problem Set 1**

**Problem 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:

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

a) What kind of factors are contained in $u$? Are these likely to be correlated with the level of education?

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

**Problem 2**

In the simple linear regression model:

$$y = \beta_0 + \beta_1 \, x + u,$$

Suppose that $\mathbb{E}[u] \neq 0$. Let $\alpha_0 = \mathbb{E}[u]$. Show that the model can always be rewritten with the same slope, but a new intercept and error, where the new error has an expected value of zero.

**Problem 3**

In the simple linear regression model:

$$y = \beta_0 + \beta_1 \, x + u,$$

where $Var[u] = \sigma^2$ and $\bar{x}$ is the mean of $x$, show that

$$Var[\hat{\beta}_1] = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})},$$

**Problem 4 (Stata)**

Use **td1_grades.dta**. This data 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. Estimate the relationship between GPA and ACT using OLS.

a) 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 5 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 the ACT? Explain.

**Problem 5**

Consider the savings function:

$$sav = \beta_0 + \beta_1 inc + u,$$

$$u = \sqrt{inc} \cdot e,$$

where $e$ is a random variable with $\mathbb{E}[e] = 0$ and $Var[e] = \sigma_e^2$. Assume that $e$ is independent of *inc*.

a) Show that $\mathbb{E}[u|inc] = 0$, i.e., that the key zero conditional mean assumption is satisfied. (*Hint*: If $e$ is independent of *inc*, then $\mathbb{E}[e|inc] = \mathbb{E}[e]$.)

b) Show that $Var[u|x] = \sigma_e^2 \cdot inc$, i.e., that the homoskedasticity assumption is violated. In particular, the variance of *sav* increases with *inc*. (*Hint*: $Var[e|inc] = Var[e]$, if $e$ and *inc* are independent.)

c) Provide a discussion that supports the assumption that the variance of savings increases with family income.

**Problem 6 (Stata)**

Use the data from **td1_education.dta**. Estimate the relationship between IQ ($IQ$) and education (*educ*) by supposing a linear relationship.

a) What is the average increase of IQ in case of a 1- and 2-year increase in education?

b) Compute the fitted values of $IQ$ and the residuals. Produce a scatter plot with the residuals on the y-axis and fitted values on the x-axis.

c) Compute the Residual Sum of Squares (SSR), the Explained Sum of Squares (SSE), and the $R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$.

d) Compute $IQ$ for an individual with 10 years of education.

e) Compute the standard error for $\hat{\beta}_1$.

f) Produce a scatter plot between the relationship of *educ* and $IQ$ and fit a line with the estimated model.

Now estimate the linear relationship between *wage* and $IQ$.

g) What is the estimated increase in *wage* due to 15 additional IQ points?

h) Suppose you are interested in the effect of an additional point of IQ on the percentage increase on wage. What model do you suggest?

i) What is the percentage increase of wage associated with an increase of 15 IQ points?

**Problem 7**

Find $\mathbb{E}[\mathbb{E}[\mathbb{E}[y|x_1, x_2]|x_1]]$.

**Problem 8**

If $\mathbb{E}[y|x] = a + bx$, find $\mathbb{E}[yx]$ as a function of moments of $x$. (Reminder: The $k$-th moment of a random variable $A$ is defined as $\mathbb{E}[A^k]$.)

**Problem 9 (Stata)**

The data set in **td1_ceos.dta** contains information on chief executice officers for U.S. corporations. The variable *salary* is annual compensation, in thousands of dollars, and *ceoten* is prior number of years as company CEO.

a) Find the average salary and the average tenure in the sample.

b) How many CEOs are in their first year as CEO (that is, $ceoten = 0$)? What is the longest tenure?

c) Estimate the simple regression model

$$log(salary) = \beta_0 + \beta_1 \, ceoten + u$$

and report your results. Note that there is a log-transformation of *salary* in the data (*lsalary*). What is the (approximate) predicted percentage increase in salary given on more year as CEO?