technische universität dortmund

Fakultät Wirtschaftswissenschaften

Problem Set 1.1

Labour Economics, Winter Term 2025/26

Submit by Sunday, 02 November, 22:45h ***on Moodle!***

Learning objectives

- Create and interprete descriptive statistics
- Conduct ordinary least squares (OLS) regressions
- Interpret omitted variables bias (OVB)

Tasks

Get familiar with R. You can find some books in Moodle under *Readings*. In case it is your first time using R we recommend having a look at chapter 1 of *Using R for Introductory Econometrics*, watching one of the many useful YouTube videos, or doing some of the free R exercises on `https://www.codeacademy.com/` or `https://www.datacamp.com/`. Finally, we find AI tools and coding assistants (e.g. ChatGPT, Claude) to be useful in suggesting solutions to many coding problems.

Download the data `ps1_clean_data.Rda` and open it in R Studio. The data contains various variables. $hours$ is the dependent variable indicating the number of hours worked per day. $motivation$ is the intrinsic motivation of individuals for a successful career. This is compared to the average motivated individual where higher values are associated with higher motivation. $education$ displays the number of years spent on education. $wage$ is the wage per hour. $wage\_premium$ is a dummy that indicates whether an individual received a randomly assigned wage increase of 35% or not. This could be because they were drafted into an income support program like the one in Canada discussed in lecture but will be important only in Problem Set 1.2.

a) Generate the log of wages (ln_wage). Produce a table with descriptive statistics for $education$, $motivation$, $hours$ and ln_wage. Also calculate correlations between these variables and plot the density of log wages as well as a histogram for the years of education. Briefly comment on your results.

b) Compare descriptive statistics for individuals who have a $motivation$ above or equal to zero versus below. Do the same thing for $education$, 14 years or more (some college) versus below. Again comment on what you find.

c) Plot a scatter of the hours worked against ln_wage. What do you notice?

d) Do a simple regression of $hours$ on ln_wage. Add the regression line to the plot from c). Include the 95% confidence interval and interpret your results.

e) Now add $education$ to your regression and explain to what extent and why results change.

f) Finally, do the full multivariate regression of $hours$ on ln_wage, $education$, and $motivation$. Compare your results to before, i.e., out of d)–e), what is your preferred regression specification and results?

Notes: You can work in teams of 1–3 students. Please upload your code as well as a pdf-file with discussions on what you found in the data in response to the tasks above. It should be clear which lines of code and answers in the `.pdf` refer to which question.