

Econometrics

Problem Set 1

Ref.: Stock, J. H., and M. W. Watson. Introduction to econometrics. Pearson, 2020.

Question 1

Suppose that a researcher, using data on class size (CS) and average test scores from 50 third-grade classes, estimates the OLS regression:

$$\widehat{TestScore} = 640.3 - 4.93CS, R^2 = 0.11, SER = 8.7$$

- A classroom has 25 students. What is the regression's prediction for that classroom's average test score?
- Last year a classroom had 21 students, and this year it has 24 students. What is the regression's prediction for the change in the classroom average test score?
- The sample average class size across the 50 classrooms is 22.8. What is the sample average of the test scores across the 50 classrooms?

Question 2

A random sample of 100 20-year-old men is selected from a population and these men's height and weight are recorded. A regression of weight on height yields

$$\widehat{Weight} = -79.24 + 4.16 * Height, R^2 = 0.72, SER = 12.6,$$

where Weight is measured in pounds and Height is measured in inches.

- a. What is the regression's weight prediction for someone who is 64 inches tall? 68 inches tall? 72 inches tall?
- b. A man has a late growth spurt and grows 2 inches over the course of a year. What is the regression's prediction for the increase in this man's weight?
- c. Suppose that instead of measuring weight and height in pounds and inches, these variables are measured in centimeters and kilograms. What are the coefficient estimates from this new centimeter–kilogram regression?

Question 3

a. Show that the sample regression line passes through the point (\bar{X}, \bar{Y}) .

Question 4

Use Growth.xlsx, which contains data on average growth rates from 1960 through 1995 for 65 countries, along with variables that are potentially related to growth.¹ A detailed description is given in Growth_Description, also available on the website. In this exercise, you will investigate the relationship between growth and trade.

- a. Construct a scatterplot of average annual growth rate (Growth) on the average trade share (TradeShare). Does there appear to be a relationship between the variables?
- b. One country, Malta, has a trade share much larger than the other countries. Find Malta on the scatterplot. Does Malta look like an outlier?

Question 4 (cont.)

- c. Using all observations, run a regression of Growth on TradeShare. What is the estimated slope? What is the estimated intercept? Use the regression to predict the growth rate for a country with a trade share of 0.5 and for another with a trade share equal to 1.0.
- d. Estimate the same regression, excluding the data from Malta. Answer the same questions in (c).
- e. Plot the estimated regression functions from (c) and (d). Using the scatterplot in (a), explain why the regression function that includes Malta is steeper than the regression function that excludes Malta. Where is Malta? Why is the Malta trade share so large? Should Malta be included or excluded from the analysis?

Question 5

A researcher runs an experiment to measure the impact of a short nap on memory. There are 200 participants, and they can take a short nap of either 60 minutes or 75 minutes. After waking up, each participant takes a short test for short-term recall. Each participant is randomly assigned one of the examination times, based on the flip of a coin. Let $\textcolor{blue}{Y_i}$ denote the number of points scored on the test by the i th participant ($0 \leq Y_i \leq 100$), let $\textcolor{blue}{X_i}$ denote the amount of time for which the participant slept prior to taking the test ($X_i = 60 \text{ or } 75$), and consider the regression model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Question 5 (cont.)

- a. Explain what the term $\textcolor{blue}{\textbf{\textit{u}}_i}$ represents. Why will different participants have different values of $\textcolor{blue}{\textbf{\textit{u}}_i}$?
- b. What is $\textcolor{blue}{\textbf{\textit{E}}(\textcolor{blue}{\textbf{\textit{u}}_i} | X_i)}$? Are the estimated coefficients unbiased?
- c. The estimated regression is $\hat{Y}_i = 55 + 0.17X_i$.
- (i) Compute the estimated regression's prediction for the average score of participants who slept for 60 minutes before taking the test. Repeat for 75 minutes and 90 minutes.
- (ii) Compute the estimated gain in score for a participant who is given an additional 5 minutes to nap.