

Prof. C. M. Dalton
ECN 209A
Spring 2015
Homework 2

Due Thursday, Feb 19th

For full credit: Questions must be solved completely on separate, clean papers and stapled. You may work in groups, but make sure to list collaborators.

Question 1 Suppose you are interested in the relationship between education and wage. More specifically, you believe the relationship is captured by the following linear regression model:

$$Wage = \beta_0 + \beta_1 Education + u$$

For each of the following, describe the symbol's meaning. Then, list the estimated counterpart you would find once you run the above regression on your data. List both the estimated symbol and meaning. β_0 , β_1 , u , $E(Wage | Education)$

Question 2 Suppose that a researcher, using data on class size (CS) and average test scores from 99 third-grade classes, estimates the OLS regression:

$$\widehat{TestScore} = 546.42 + (-6.11) CS$$

$$\text{Where } R^2 = 0.06 \text{ and } SER = 12.1$$

- Teacher Chen's classroom has 25 students. What would the regression predict for this classroom's average test score?
- Suppose another 25-student classroom has an observed test score of 400. Is this the same as your answer in a? What explains the difference? Give one reason for the difference that would not cause a problem in our OLS estimation.
- Suppose next year Teacher Chen's classroom will increase from 25 to 27 students. What is the predicted change in the test score?
- What is the value of the Total Sum of Squares (TSS) in this regression? (*Hint: Use the formulas for SER and R^2*)

Question 3 Your textbook presented you with the following regression output:

$$\widehat{TestScore} = 698.9 - 2.28 \times STR$$

$$n = 420, R^2 = 0.051, SER = 18.6$$

- How would the slope coefficient change, if you decided one day to measure test scores in 100s, i.e., a test score of 650 became 6.5? Would this have an effect on your interpretation?
- Do you think the regression R^2 will change? Why or why not?

Question 4 In this exercise, you will use data to investigate the relationship between the number of completed years of education for young adults and the distance from the student's high school to the nearest four-year college.

Download the dataset "hw2_data.xls" on the course website and load the xls sheet into Stata. This dataset contains data from a random sample of high school seniors interview in 1990 and re-interviewed in 1996. Download the "hw2_data_codebook.pdf" codebook for a definition of the variables.

- a. Write down the OLS regression model you will run, writing the model in terms of your variables. (i.e. what is your dependent variable, independent variable, etc.)
- b. For both variables, verify that the variable seems to be measuring what we think it is. Perform a few simple analyses using Stata to satisfy this. Paste this Stata output into a document where you describe what this test tells you about the variable or what this test confirms about what we think it is measuring. (or write the explanation below the Stata printout).
- c. Given your results in b, do you have any recommendations for dealing with this data? Make those changes if you can support them.
- d. Run a regression to obtain the predicted values for the model in a. List and interpret your parameters. "Interpret" means reporting: 1. Significance 2. Sign of the parameter 3. Meaning of the number.
- e. How well does your regression explain your question? Support your statement.
- f. If you made changes to your dataset, run the above regression on the original uncorrected data. Do the parameters change? If so, explain why you would expect the parameters to change in that way.

Question 5 Use the cps92_12.dta dataset to run a regression that explains average hourly earnings by the age of a worker and if that worker holds a bachelor degree.

- a. Stata gives you output reporting whether your estimated coefficients are statistically significant. Which components of the Stata output give information on statistical significance of the slope coefficient?

Pick one of these Stata output pieces and calculate the value by hand to prove you are just as smart as a computer. (You just have other things to do besides hand-calculate 42,000 dataset results every time!)

- b. How well does your regression explain the average hourly earnings of a worker? Support your claim.
- c. Interpret your coefficient from the regression. "Interpret" means reporting: 1. Significance 2. Sign of the parameter 3. Meaning of the number.