

Problem Set 3

Due: October 6

Please upload the problem sets to the G drive folder used in the practice sessions. Make sure to place the file in your own subfolder (the one with your name and id number), and to name the file with the problem set number i.e., PS_1, PS_2, etc.

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:

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

- (a) A classroom has 35 students. What is the regression's prediction for that classroom's average test score?
- (b) Last year a classroom had 24 students, and this year it has 21 students. What is the regression's prediction for the change in the classroom's average test score?
- (c) The sample average size across the 50 classrooms is 22.8. What is the sample average of the test scores across the 50 classrooms? (*Hint*: review the formulas for the OLS estimators).
- (d) What is the sample standard deviation of test scores across the 50 classrooms? (*Hint*: review the formulas for the R^2 and SER)

Question 2

In this question, you will investigate the relationship between earnings and height using the Stata data file earnings and height, which contains data on earnings, height, and other characteristics of a random sample of U.S.

workers. A detailed description is given in a Description file.¹ Both files are in the course's website, in the Data Sets section.

- (a) What is the median value of height in the sample?
 - i. Estimate average earnings for workers whose height is at most 67 inches.
 - ii. Estimate average earnings for workers whose height is greater than 67 inches.
 - iii. On average, do taller workers earn more than shorter workers? How much more? What is a 95% confidence interval for the difference in average earnings?
- (b) Construct a scatterplot of annual earnings (*Earnings*) on height (*Height*). Notice that the points on the plot fall along horizontal lines. (There are only 23 distinct values of Earnings). Why? (*Hint*: Carefully read the detailed data description.)
- (c) Run a regression of *Earnings* on *Height*.
 - i. What is the estimated slope?
 - ii. Use the estimated regression to predict earnings for a worker who is 67 inches tall, for a worker who is 70 inches tall, and for a worker who is 65 inches tall.
- (d) Suppose height were measured in centimeters instead of inches. Answer the following questions about the *Earnings* on *Height* (in cm) regression.
 - i. What is the estimated slope of the regression?
 - ii. What is the estimated intercept?
 - iii. What is the R^2 ?
 - iv. What is the standard error of the regression?

¹These data were used by Professors Anne Case (Princeton University) and Christina Paxson (Brown University) in their paper "Stature and Status: Height, Ability, and Labor Market Outcomes," *Journal of Political Economy*, 2008, 116(3): 499–532.

- (e) Run a regression of *Earnings* on *Height*, using data for female workers only.
- i. What is the estimated slope?
 - ii. A randomly selected woman is 1 inch taller than the average woman in the sample. Would you predict her earnings to be higher or lower than the average earnings for women in the sample? By how much?
- (f) Repeat (e) for male workers.
- (g) Do you think that height is uncorrelated with other factors that cause earnings? That is, do you think that the regression error term, say u , has a conditional mean of zero, given *Height* (X)?