

# Econ 322: Econometrics

## 2nd Assignment: Simple Linear Regression

Due Date: 23:00, 10/19/2023, EST

Read the following instructions carefully.

- You can submit your reports unlimited times before the deadline. Do wait until the last minute and tell me that you have an internet problem.
- Report the required results in a word or PDF document and submit it **together with your code**. In the report, make your answers concise and to the point. My TA has a lot of submissions to grade. So make sure that your answers are easy to spot, in case my TA may miss them.
- Name your submitted files in the following format: initial\_first\_full last name\_A2 (so that we know this is your answer to the second assignment).
- You can refer to the sample codes I gave in my lecture notes and the uploaded Python codes.

In the data file **Earnings\_and\_Height.csv**, you will find data on earnings, height, and other characteristics of a random sample of U.S. workers. We are interested in two variables: **height**, which gives height without shoes (in inches), and **earnings**, which is annual labor earnings, expressed in 2012 dollar. In the “sex” column, 1=Male, 0 = Female.

Note: In the survey, labor earnings are reported in 23 brackets (for example, 26,000–30,00). For each of these brackets researchers estimated a value of average earnings based on information in the Current Population, and these average values were assigned to all workers with incomes in the corresponding bracket. The earnings values for 1994 were converted to 2012 dollar using the consumer price index.

You can continue to work on the .py file.

**Q1 (50 points).** What are the mean values of **height** and **earnings**, respectively? Round up the numbers up to 2 decimal points. (This is a question that suppose to give some “free” points. Make sure you can get the points.)

**Q2 (10 points).** Draw a scatterplot of annual earnings (Earnings) on height (Height).

**Q3 (20 points).** Run a regression of Earnings on Height.

- (a) What is the sample size?

- (b) What is the R-squared?
- (c) What is the estimated slope?
- (d) Use the estimated regression to predict earnings for a worker who is 70 inches tall

**Q4 (20 points).** Run a regression of Earnings on Height, using data for female workers only. (figure out how to select a subset of a dataframe based on my intro to Python lecture.)

- (a) What is the sample size?
- (b) What is the R-squared?
- (c) What is the estimated slope?
- (d) 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?