Universidade de São Paulo / Faculdade de Filosofia, Letras e Ciências Humanas
Departamento de Ciência Política
FLP-0468 & FLS-6183
2º semestre / 2018

Problem Set# 1: Does a minimum wage increase lead to an increase in the percentage of the workforce that is formally employed?

In this problem set, we will review the lab discussing the differences between a hypothesis test of a difference in means (Diff in Diff) and how the same test can be conducted using a multiple regression model. The focus in this problem set is train your skills in writing about a statistical analysis and its interpretation. The datasets and example refers to the example discussed in Chapter 2 of *Quantitative Social Science: An Introduction*. Please prepare a brief document including output where pertinent for your answer.

Part I. Statistical Tests and their interpretation.

(a) List the variables in the two data sets (minwage.csv and minwage-reshaped.dta) and use Stata's sum command to produce descriptive stats. Please summarize the difference between both data sets. (Hint: you can use the describe command in Stata).

(b) Construct a t-test and 95 percent confidence interval for the difference in the mean rate of full-time employment in New Jersey. What are the problems with using this measure to assess the impact of the minimum wage increase on employment percentages? Why is this exercise called the naïve ttest?

(c) Now, let´s construct a t-test and 95 percent confidence interval for the difference in the average full-time employment in Pennsylvania before and after the implementation of the minimum wage increase. (Hint: This command is in the dofile, but you can easily modify the commands in the do file to run this test). What is the approximate percentage difference in formal employment in PA?

(d) Stata's testing procedure calculates the standard error of differences in averages under two alternative assumptions about variances. State these assumptions in words. (Hint: see the help file for the specific ttest command in Stata). Which calculation corresponds to the default for a ttest? Explain and check your answer.

(e) In class, we discussed that we can calculate:

$$DiffinDiff_{t-(t-1)} = \left\{(\overline{Y_t} - \overline{Y_{t-1}})\middle| T = 1\right\} - \left\{(\overline{Y_t} - \overline{Y_{t-1}})\middle| T = 0\right\}$$

Use the information in (b) and (c) to calculate the difference the proportion of full-time employment in Pennsylvania and New Jersey before and after the implementation of the minimum wage increase.

(f) Use Stata's reg command to prove the result in question e above "by computer." Please discuss in detail how to interpret each parameter in the regression that was estimated.

(g) Now, let´s examine the same test using the Stata package, diff. Please copy the table and review the results.

Part II. Research Desgins

(h) Compare the results obtained using the *difference-in-difference* approach and the multiple regression model. What are the similarities and differences?

(i) Can you think about any confounders that might affect the results?

(j) What are some of the limitations of *before-and-after designs* (comparison between pre- and posttreatment measurements)?

Part III. Random Assignment

A professor decides to run an experiment to measure the effect of time pressure on final exam scores. He gives each of his 400 students in his course the same final exam, but some students have 90 minutes to complete the exam while others have 120 minutes. Each student is randomly assigned one of the examination times based on a flip of the coin. Let Y_i denote the number of points scored on the exam by the ith student $(0 \le Y_i \le 100)$, let X_i denote the amount of time the student has to complete the exam (X_i=90 or 120), and consider the regression model $Y_i = \beta_o + \beta_1 X_i + u_i$.

(k) Explain what the term ui represents. Why will different students have different values of ui?

(l) Explain why E(ui|xi)=0 for this regression model. Are the other assumptions mentioned in Stock and Watson Chapter 4.3 satisfied? Explain.

(m) The estimated regression is $Y_i = 49 + 0.24 X_i + u_i$.

a. Compute the estimated regression's prediction for the average score of students given 90 minutes to complete the exam. Repeat for 120 minutes and 150 minutes.

b. Compute the estimated gain in score for a student who is given an additional 10 minutes on the exam.