ECONOMICS 705
Econometrics II (First Half)
Prof. Jeffrey Smith
Department of Economics
University of Wisconsin-Madison
Fall 2022

PROBLEM SET 1
Version of September 26, 2022
Due at the start of class on Monday, October 10

Writing up your answers to the empirical problems

Your write-up should consist of two portions. The first portion is just the answers to the questions, with whatever text is required to explain them. **This portion must be typed!**

The second portion, on separate pages, **consists of a Stata log file** that shows how you got the answers to the empirical questions. The log file must be clear and must include comments that will allow the reader to quickly see the command or commands leading to each answer. It should not include everything you tried – just the final set of commands employed to get the answers.

Submit completed problem sets on canvas as a single PDF file with a filename of the form "LastnameFirstnamePS#.pdf" where "#" is 1, 2, or 3, depending on the problem set.

Problem sets not turned in on Canvas using the format just described will receive no credit.

Data for empirical problems

The data come from the experimental evaluation of the National Job Training Partnership Act (JTPA) called the National JTPA Study (NJS). The JTPA program was the largest federal active labor market program from 1982 to around 2000. The experiment was conducted around 1989.

The file name is "Economics 705 Fall 2022 NJS Data.dta"

The dataset contains the following variables:

bifid: identification number
ra_stat: 1 = treatment group; 2 = control group
enroll: 1 = participated in JTPA; 0 = did not participate in JTPA
esum18i: = self-reported earnings in the 18 months after random assignment
sex: 1 = male; 0 = female
race: 1 = white; 2 = black; 3 = Hispanic; 4 = other

age: age in years at baseline
totch18: total number of children age 18 or younger
child_miss: 1 = missing *totch18*; 0 = not missing *totch18*
bfeduca: years of schooling at baseline
ed_miss: 1 = missing *bfeduca*; 0 = not missing *bfeduca*
bfyrearn: self-reported earnings in the year before random assignment
earn_miss: 1 = missing *bfyrearn*; 0 = not missing *bfyrearn*
site_num: categorical variable coded 1-16 for the 16 experimental sites

I will refer to *sex*, *race*, *age*, *totch18*, *child_miss*, *bfeduca*, *ed_miss*, *bfyrearn*, *earn_miss* and *site_num* as baseline covariates.

You can learn more about the NJS (if you are deeply curious) in

Bloom, Howard, Larry Orr, Stephen Bell, George Cave, Fred Doolittle, Winston Lin and Johannes Bos. 1997. "The Benefits and Costs of JTPA Title II-A Programs: Finidngs from the National Job Training Partnership Act Study." *Journal of Human Resources* 32: 549-576.

This article is available on canvas.

Problems

1. (5 points) Keep only adult women, where adult for this purpose means age greater than or equal to 22.

2. (5 points) Create a variable called *treat* that equals one for individuals in the experimental treatment group and zero for individuals in the experimental control group.

3. (5 points) Drop observations with missing values of *esum18i*. In real life, one would worry more about this, and so do more about this, than we are doing here. There are many missing values of *esum18i* because many individuals did not respond to the follow-up survey that provides the information used to construct it.

4. (5 points) In expectation, experiments balance the distributions of all variables not affected by the treatment, both observed and unobserved, between the experimental treatment group and the experimental control group. Summarize the age, education, and earnings last year variables separately for the experimental treatment group and the experimental control group. Do you see any large differences? Explain.

5. (5 points) Using Stata's `ttest` command, formally test the null hypothesis of equal population means in the treatment and control groups for the age, education and earnings last year variables. Describe your findings. Explain why (or why not) this is an interesting null to test.

6. (5 points) Calculate the "standardized difference" for the schooling variable. Report and discuss; in particular, is it larger or smaller than the arbitrary cutoff of 20 defined by Rosenbaum and Rubin?

7. (5 points) Using Stata's `ttest` command, estimate the mean difference in *esum18i* between the treatment and control groups. Describe and interpret your findings; be sure to note whether (or not) the estimated impact is statistically and/or substantively different from zero.

8. (5 points) Using Stata's `regress` command, calculate the experimental impact of assignment to the treatment group on *esum18i*. Do not include any other covariates in the model. What is the estimated impact? Is it statistically different from zero? How does it compare to the mean difference estimated in the preceding problem?

9. (5 points) Repeat the regression in the preceding problem but including all of the baseline covariates as conditioning variables. Discuss and interpret your findings, being sure to relate them to the estimates obtained in the preceding problem. What happens to the estimated standard errors? Why? Is this impact estimate unbiased? Why? [Be sure to think carefully about how you code the baseline variables.]

10. (5 points) What fraction of the treatment group actually participates in JTPA (i.e. has *enroll* = 1)? What fraction of the control group manages to defeat the experimental protocol and participate in JTPA?

11. (5 points) Calculate the instrumental variables estimate of the impact of actually participating in JTPA using two-stage least squares with treat as an instrument for *enroll*. What is the interpretation of the instrumental variables estimate in a common effect world? What is the interpretation of the instrumental variables estimate in a heterogeneous treatment effect world? Be sure to note any assumptions required for the interpretations you describe.

12. (5 points) Provide and justify an estimate of the complier mean of the *bfeduca* variable.

13. (5 points) Create a binary indicator for "age less than or equal to 31" called *agele31*.

14. (5 points) Using the indicator constructed in the preceding problem, construct separate experimental impacts ("subgroup impacts") for younger and older individuals (i.e. observations with *agele31* = 1 and observations with *agele31* = 0) using a linear regression model and the usual baseline covariates.

Do this two ways: first, by estimating completely separate experimental impact regressions for the younger and older individuals; second, by estimating a single experimental impact regression but including an interaction between the treatment indicator and the *agele31* variable (as well as a main effect in *agele31*).

Describe and interpret your findings. In particular, discuss the evidence for differing impacts for older and younger individuals and discuss the evidence regarding whether the two ways of estimating the subgroup effects produce different estimates. Indicate which estimator you prefer and why.

15. (5 points) Compare the standard deviations of earnings in the 18 months after random assignment in the experimental treatment and control groups. Formally test the null of equal variances using the `sdtest` command. Describe and interpret your findings as they relate to the question of homogeneous or heterogeneous treatment effects.

16. (5 points) Estimate the quantile treatment effect at the 75^{th} percentile of the distribution of earnings in the 18 months after random assignment. Compare it to the average treatment effect.

17. (5 points) Discuss two alternative interpretations of the estimate obtained in the preceding problem and the assumptions required for each interpretation.

18. (5 points) Construct an outcome variable *employ* that equals one for observations with *esum18i* greater than zero and that equals zero otherwise. Write down the marginal distributions of *employ* in the treatment and control groups.

19. (5 points) Construct the Frechét-Höffding bounds for the (0,0) cell in the joint distribution from the preceding problem. Describe your findings in words. In particular, do you think the bounds are wide or narrow? Justify your answer.

20. (5 points) Using your answers from the preceding problem, describe the distributions of treatment effects associated with the Frechét-Höffding upper and lower bounding distributions.