

Econ 272 / GSB 607
Spring 2025

Department of Economics
Stanford University

PROBLEM SET I

DUE: SUNDAY APRIL 6th, 2025, 11PM.

Be concise but also clear what numbers you are reporting, and answer in full sentences. You should also hand in supporting code, but all the answers should be in a pdf or word file.

Use the experimental data from the LaLonde/Dehejia/Wahba data posted on the course website, `lalonge_experimental.txt`. There are two variables, the binary treatment indicator (1 for the individuals who were assigned to the training program, and 0 for those who were not), and an outcome variable, earnings after the program in 1978 in 1,000s of dollars

1. Computational Exercises

- (a) Test the null hypothesis that there is no effect of the training program whatsoever on earnings the year after the program using a t-statistic. Also report the p-value based on a normal distribution.
- (b) Approximate the exact p-value based on the test statistic equal to the difference in averages by treatment status,

$$T = \bar{Y}_t - \bar{Y}_c.$$

Use 1,000,000 draws from the randomization distribution to approximate the p-value. Include a histogram of the randomization distribution of the test statistic under the null hypothesis.

- (c) Use the statistic based on the difference in average ranks by treatment status.
- (d) Estimate the average treatment effect and construct a 95% confidence interval using the Neyman variance estimator, and compare that to the homoskedastic variance.

2. Theoretical Calculations

- (a) Suppose I plan to test the null hypothesis of no effect whatsoever, using the difference in means as the test statistic. I will test at the 5% level, and I want my power to be at least 80%, for an effect size of 1.79 (thousand dollar), assuming homoskedasticity, and using the sample standard deviation from earnings in 1978 in part (1), 6.63, as the population standard deviation. What is the sample size needed?
- (b) Suppose I plan to test the null hypothesis of no effect whatsoever, using the difference in fractions with positive earnings as the test statistic. I will test at the 5% level, and I want my power to be at least 80%, for an effect size of 0.11, assuming homoskedasticity, and using the sample standard deviation from the indicator for positive earnings in 1978 in part (1), 0.46, as the population standard deviation. What is the sample size needed?
- (c) How do you interpret the difference in the sample size needed for the two test statistics?