PublPol 303D **Winter 2021**

Problem Set VI

Due: 9am, Monday February 14th, 2022

Use the data set lalonde_cps_clean.dta. We are interested in estimating the average effect of the job training program. We know from the experimental data that this is approximately $2,000. The question we ask is the basic question addressed originally by Lalonde. Could we have figured this out only using the non-experimental data?

Explore regression and matching methods for estimating the average effect for the treated individuals.

You may work on your own or in a group of 2. If you choose to work with a partner, ensure it is not a partner you have worked with the past 2 weeks.

1. Report the difference in means in the outcome, re78, for the treatment and control group, and a 95% confidence interval for the population difference.
2. Is this difference a credible estimate of the average treatment effect?
3. Read in the data and test for each of the ten exogenous covariates / pretreatment variables (age educ black hispanic nodegree married re74 u74 re75 u75) whether their means differ between the treatment and the control group. In the light of these results, revisit the question in (2).
4. Use a regression analysis to estimate the average effect of the program.
5. Make the analysis more flexible by including interactions between all the variables.
6. Use a matching estimator to estimate the average effect for the treated.
7. Use a logistic regression model to estimate the probability of being in the treatment.
8. Use the predict command to get the propensity score.

9. Test whether the average propensity score is different in the treatment and control groups.

10. Create histograms for the propensity score in the treatment and control groups separately.

11. Use matching on the estimated propensity score to estimate the average effect on the treatment.

12. Summarize your overall findings, and give your prefered non-experimental estimate of the average effect of the training program on the treated individuals.