Problem Set 4a

Prof. Conlon

Due: 4/4/20

Packages to Install

The packages used this week are

- ggplot2
- data.table (data tables are computationally efficient and IMHO easier to work with)
- rdd (package for regression discontinuity designs)
- estimatr (tidy version of lm)
- knitr (make pretty tables using kable command)
- extraDistr (package with extra distributions)

Problem 1 (Coding Exercise)

Using the Lalonde dataset and the **cobalt** package finish the exercise from the slides.

That is:

Consider three possible matching techniques

1. Caliper matching on a single variable (pick the best one)
2. 4 nearest neighbor matching.
3. Propensity Score matching using a logit
4. Propensity Score matching using a kernel

For each matching approach:

a. Create a balance table. For each pretreatment covariate, include comparisons for treated and untreated units in terms of the mean and standard deviation. Report a test, for each covariate, of the hypothesis that the difference in means between treatment conditions is zero.

b. For each covariate, plot its distribution under treatment and control

c. Estimate the ATT and/or ATE of participating in the job training program

d. Can you estimate both ATE or ATT? Why or why not?

Problem 2 (Coding Exercise)

The dataset for this exercise comes from a paper by Benjamin Olken entitled "Monitoring Corruption: Evidence from a Field Experiment in Indoneisa". The paper evaluates an attempt to reduce corruption in road building in Indonesia. The treatment we focus on was "accountability meetings". These meetings were held at a village level, and project officials were probed to account for how they spent project funds. Before construction began, residents in the treated villages were encouraged to attend these meetings. The dataset is called "olken.csv".

The outcome we care about is **pct.missing**, the difference between what officials claim they spent on road construction and an independent measure of expenditures. Treatment is given by **treat.invite** such that:

$$\text{treat.invite} = \begin{cases} 1 & \text{if village received intervention} \\ 0 & \text{if village was control} \end{cases}$$

We have the following four pre-treatment covariates:

- head.edu : the education of the village head
- mosques : mosques per 1000 residents
- pct.poor : the percentage of households below the poverty line
- total.budget : the budget for each project

We now have the following questions:

a. Create a balance table. For each pretreatment covariate, include comparisons for treated and untreated units in terms of the mean and standard deviation. Report a test, for each covariate, of the hypothesis that the difference in means between treatment conditions is zero.

b. For each covariate, plot its distribution under treatment and control (either side-by-side using facet_grid or overlap).

c. Given your answers to part a and b, do the villagers seem similar in their pre-treatment covariates?

d. Regress the treatment on the pre-treatment covariates. What do you conclude?

e. Using the difference-in-means estimator, estimate the ATE and its standard error.

f. Using a simple regression of outcomes on treatment, estimate the ATE and its standard error. Compare your answer in (f) to (e).

g. Using the same regression from part (f), include pre-treatment covariates in your regression equation (additively and linearly). Report estimates of treatment effects and its standard error. Do you expect (g) to differ from (f) and (e)? Explain your answer.