Problem Set 1
MGMT 737

Spring 2021

1. Part I: IRB. Complete the Yale IRB certification for "Social-Behavioral-Educational Researchers - Basic/Refresher" and send me your CITI certification form. To complete this training, visit https://bmsweb-h.yale.edu/ords/tms/tms_enrollments.offerings?p_crs_id=1931&p_std_id= (you must be connected to Yale via VPN). Congratulations, you are now able to get experiments approved and run your own!

2. Part II: Randomization. This analysis will use the Dehijia and Wahba sample from the Lalonde dataset of the NSW experiment. The dataset is `lalonde_nsw.csv`. The outcome variable is `re78` (real earnings in 1978). The treatment indicator is `treat`. The remaining variables are potential covariates.

   (a) Calculate the average treatment effect of the policy $E(\tau_i)$ using a simple difference in means.

   (b) Test the null of $\tau_i = 0$ for all i using a randomization test.
   N.B. Hold fixed the number of treated and control (e.g. assume the treatment count would be held fixed) and permute the labels randomly 1000 times – you do not need to fully do every permutation (there would be too many). Report the quantile that your estimate from the previous question falls.

   (c) Run a regression using robust standard errors (you may use canned software) of the outcome on the treatment dummy, and compare the p-values from this test to the previous answer.

3. Part III: Propensity Scores. This analysis will use the dataset from Part II as well as the PSID dataset from Dehijia and Wahba, `lalonde_psid.csv`. These datasets have identical variables. This dataset is a sample of observations from the Panel Survey of Income Dynamics that can be used as controls.

   (a) Using `age`, `education`, `RE74` and `RE75`, construct a propensity score using the *treated* group in `lalonde_nsw.csv` and the control sample of `lalonde_psid.csv`. Use a logit regression model to do so (you may use a canned routine to run the regression). Report the average p-score for the treated and control samples, and plot the propensity score densities for the treatment and control groups.

   (b) Using your p-score estimates, calculate the IPW estimate for the average treatment effect.

   (c) Compare this to a linear regression using the PSID and NSW treatment sample, using `age`, `education`, `RE74` and `RE75` as controls.

   (d) Extra credit: In their paper, Dehijia and Wahba use a larger model: Age, Age squared, Age cubed, School, School squared, Married, No degree, Black, Hisp, RE74, RE75, U74, U75, School × RE74. Feel free to calculate replicate this for the above for extra credit. U is an indicator for RE equal to zero.