Problem Set 2

1. State the assumptions to use the matching research design.
2. Use Pset2data.dta to solve this question. The data includes an outcome variable Y, treatment dummy D, and covariate X. Do matching using the covariate X.

(a) Plot the histogram of X by the treatment dummy D. Comment on whether any assumption for matching is violated.

1. Report what fraction of observations can be exactly matched on X.

(c) Do the exact matching on X and report the ATE with its standard error. Use the observations which can be exactly matched only.

1. Do the nearest neighbor matching on X using 1 nearest neighbor. Use all observations in this calculation. Do not adjust for the bias. Report the ATE with its standard error.
2. Repeat the (d) but with bias adjustment option. Report the ATE and its standard error.
3. Do the Coarsened Exact Matching on X. Use the automatic binning done by “cem” command. Report what fraction of observations can be matched after binning.
4. Estimate the ATT using the CEM weight. Report the standard error.

(NOTE : the STATA “cem” command gives a weight variable to compute the ATT. If you want to compute the ATE, use “cem\_strata” variable with “teffects nnmatch” command.)

1. Compute the propensity score using a logit model. Plot the histogram of the propensity score by treatment D.
2. Use the Imbens-Hirano estimator to compute the ATE with the propensity score you computed in (h). Report the ATE with its standard error.