## Day 6: Propensity Score Matching

1. **Basic Idea**

Selection bias in some analysis.

---The policy treatment is not randomly assigned; some units are favored over others.

----------------------------------------------------

Step 1: select covariates for estimating propensity scores.

---An easier option is to use your control variables, depending on your research context, other variables are preferred.

Step 2: estimate propensity scores with a logit or probit model.

---logit model is more preferred. the dependent variable is a dummy variable indicating treatment status.

Step 3: match with the given propensity scores and calculate treatment effect.

---kth nearest matching with a given caliper.

-------------------------------------------------------

1. **STATA**

ssc install psmatch2

----*the propensity score matching package in STATA*

Please download the data matching\_earnings from github

----*this is a cross-section sample data describing with a treatment and incomes*.

global treatment TREAT

global ylist REDIFF

global xlist AGE EDUC MARR

global breps 5

----*recommended before every regression analysis, global your independent variable, dependent variable and control variables.*

bysort $treatment: summarize $ylist $xlist

----*summarize our variables by treatment, check if there is a difference between different groups.*

psmatch2 TREAT $xlist, outcome($ylist) logit neighbor(2) ties common ate caliper(0.05)

----*TREAT is the treatment indicator we had earlier*

----*followed by outcome is our dependent variable*

----*we use logit model*

----*ties indicate if multiple units from control group share the same propensity score, we took their average.*

----*ate reports both average treatment effect, average treatment effect on the treated and average treatment effect on the untreated*.

----*caliper is set to 0.05. for someone from treatment group, if no one from the control group has the propensity score within the calipered range, it will be dropped from analysis.*

pstest, both graph saving(balancing\_assumption, replace)

psgraph, saving(common\_support, replace)

----*psmatch2 generates a list of new variables:*

*pscore is the propensity score calculated*

*treated is treatment status, same as our TREAT*

*support indicates whether it was matched*

*weight: for control group, it is calculated based on how many times it was matched. weight = k/m, where k is how many times the certain unit is matched, and m is how many control group units each treatment group matches.*

*id: id assigned to each unit*

*n1/n2: whom this unit is matched with*

*nn: how many units from control group a unit in the treatment group is matched with*

*pdif: the absolute difference between propensity scores of matched pairs.*

----------

reghdfe $ylist TREAT $xlist if \_weight != . , noabsorb

est store m1

reghdfe $ylist TREAT $xlist if \_support == 1, noabsorb

est store m2

gen weight = \_weight \* 2

replace weight = 1 if treated == 1 & \_weight != .

reghdfe $ylist TREAT $xlist [fweight = weight], noabsorb

est store m3

esttab m1 m2 m3