Propensity Score matching

Motivation: Want to know the effect of somting
① No random assignment
② Have data on pre-program characteristics that determined whether or not the individual received the treatment.

i.e. Built heath clinics based on needs, what's the effect of the program on infant mortality?

Data:

	T	inrate
1	1	10
2	1	15
3	1	22
4	1	19
5	0	25
6	0	19
7	0	4
8	0	8
9	0	6

AVG = 16.5 (T=1)

AVE: 12.4 (T=0)

+4.1

? Look like clinics are increasing infant mortality! (wrong)

	pov_rate	PC_Docs	PS	match
1	0.5	0.01	0.416	6
2	0.6	0.02	0.735	5
3	0.7	0.01	0.928	5
4	0.6	0.02	0.735	5
5	0.6	0.01	0.752	
6	0.5	0.02	0.395	
7	0.1	0.04	0.0016	
8	0.3	0.05	0.0026	
9	0.2	0.04	0.0071	

$$ATE = \frac{10+15+22+19}{4} - \frac{19+25+25+6}{4}$$

$$= -7$$

The Basic idea

1. new control group

   Based on background charateristics select a control observation (Simmilar)

2. Compute treatment effect.

Example: step1 : Logistic T pov_rate PC_Docs
step2 : use coff. estimates to calculate predicted prob. of treatment. (propensity score)

How to check the matching?

① Look at the distribution of covariate for the treatment and new control group. (They should be similar)

② Compare distributions of the propensity scores in the treatment and new control group. (Should be similar) i.e. mean variance or histogram

③ Compare distributions of the propensity score in the treatment and original control group. (•no much overlap, matching won't work •overlap perfectly, no need for matching)

Matching vs. Regression.

[solve same problem]

Matching:
① Not as sensitive to functional form of the covariates.
② easier to assess
③ large numbers of irrelevant controls.
④ straigh-forward to explain.
⑤ easier to think about key determants of program vs. determinants of actual outcome.

Regression:
① allow to estimate continuous treatment
② shows effects of all variables.
③ allow to estimate interations of treatment with covariates
④ many way to use propensity score, may not have consistant Results.