
Problem Set 5

Social Experiment

Read in the data set `experiment.raw` into TSP. It contains data on a social experiment which took place in the U.S. between 1975 and 1977. The experiment allowed some unemployed to participate in a training of which we will now evaluate the effect. The data set contains the following variables in the given order:

<i>treated</i>	Dummy variable for treatment (1=participation in treatment)
<i>age</i>	Age
<i>educ</i>	Years of education
<i>black</i>	Dummy for skin color (1=black)
<i>mar</i>	Dummy for marriage (1=married)
<i>nodegree</i>	Dummy for no schooling degree (1=no schooling degree)
<i>re74</i>	Real income 1974
<i>re75</i>	Real income 1975
<i>re78</i>	Real income 1978
<i>hisp</i>	Dummy for hispanic origin (1=hispanic)

In the following we will check if participants and non-participants are appropriately randomized and whether the non-participants constitute a valid control group. Therefore we first check whether both groups display comparable characteristics:

1. Compare participants and non-participants with respect to their characteristics *before* the experiment started. Also do a t-test for the similarity of characteristics in both groups. (It suffices if you do this for one of the variables.)
2. Use a probit model to check whether the treatment was actually assigned randomly. What is the dependent and what are the independent variables?
3. Did participation in the training measure have an effect on the real income in 1978? Hence, estimate the average treatment effect (ATE). Assume that the randomization was valid. What does this imply and, thus, what do you need to estimate?
4. What are the caveats of calculating the ATE and what measure would you propose instead?