PD Dr. Alexander Spermann
Ute Leuschner

Tutorial for Microeconometrics
University of Freiburg
Summer term 2011

---

# Problem Set 2
# Social Experiment

Read in the data set “nsw_sub.dta” into TSP. It contains data on a social experiment which took place in the U.S. in the mid-1970s. The purpose of the experiment was to evaluate the effect of a job training program on the earnings of disadvantaged male workers. Training status of the individuals was randomly assigned. The experimental data contains the following variables in the given order:

| **Variable name** | **Variable description** |
|---|---|
| train | Dummy variable for treatment (1 if participation in treatment) |
| age | Age in years |
| educ | Years of education |
| black | Dummy for skin color (1 if black) |
| hisp | Dummy for Hispanic origin (1 if hispanic) |
| married | Dummy for marriage (1 if married) |
| nodegree | Dummy for no schooling degree (1 if no schooling) |
| re74 | Real income 1974 |
| re75 | Real income 1975 |
| re78 | Real income 1978 |

In the following we will check if participants and non-participants are appropriately randomized and whether the non treated individuals form a valid control group.

1. Compare participants and non participants with respect to their characteristics before the experiment started (*pre-treatment variables*). Are there significant differences in the means of the covariates for the treatment and control groups? To analyze this perform statistical tests for comparing sample means and sample proportions.

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 job training program have an effect on post-training earnings? In order to show this, estimate the average treatment effect (ATE) by taking the mean of the 1978 income of the participants and subtracting the mean of 1978 income for the non participants. Under which assumption this gives us the “true” treatment effect? Is the estimated treatment effect significant? What would you suggest to do in order to answer the last question?

4. Suppose that randomization did not work and that participants and non participants differ with respect to their observed characteristics. How would you estimate the ATE now?

5. Under which assumptions social experiments solve the fundamental evaluation problem?