ECON 104 **Truth in Numbers**

Homework 1

Please evaluate the effect on annual earnings of the job training provided to participants in the National Supported Work (NSW) Demonstration study. In this study a sample of men were randomly assigned to either receive or not receive job training. Baseline earnings were measured in 1975 which is the year before the they were randomly assigned to either the treatment or control group. Earnings were also measured in 1978 which is several years after the treatment started. More detail on the NSW are available in Robert Lalonde, "Evaluating the Econometric Evaluations of Training Programs," American Economic Review, Vol. 76, pp. 604-620.

The dataset for this homework is called **NSW.csv** and includes the following variables: treatment indicator ***treat*** (1 if treated, 0 if not treated), ***age*** (age in years), ***education*** (years of education), ***black*** (1 if black, 0 otherwise), ***hispanic*** (1 if Hispanic, 0 otherwise), ***married*** (1 if married, 0 otherwise), ***nodegree*** (1 if no high school degree, 0 otherwise), ***re75*** (earnings in 1975), and re78 (earnings in 1978).

1. Provide evidence that the randomization worked by comparing the means of the sample characteristics in the treatment and control groups. Please create a clean table that includes columns with the means of each group, the difference between the two groups and the p-value of the difference. The table should be comprehensible on its own.

2. Is the table you produced in answer to question 1 consistent with the randomization being correctly implemented? Why or why not?

3. Estimate the difference in the mean earnings for the treatment and control group. How big an effect did the training program have on earnings. Is it statistically significant? Is it large in a practical sense?

4. Create a carefully labeled table where each column corresponds to a regression. The

first column contains the parameters of a the regression $re78 = B_0 + B_1 treat + u$. In each following column you add one more covariate to the regression.

5. What effect does adding covariates have on your estimate of the treatment effect? What does this tell you about the relationship between the covariates and the outcome?

6. Will comparing the mean earnings of the group that got the training with the mean earnings of the group that didn't get us an unbiased estimate of the causal effect of the job training on wages? Why or why not?

7. Please attach all your code.