ECON490   **Department of Economics**
Topics in Econometrics   **UIUC**

PROBLEM SET III
DUE: WEDNESDAY MARCH 9TH, 2016

In this problem set, we look at a second non-experimental comparison that Lalonde did in his 1986 paper, where he compared the experimental trainees to a comparison group from the PSID. The PSID comparison group consists of 2490 individuals. The data are in the file psid.csv. The variables are the outcome (re78), earnings in 1978, the treatment indicator (treat), and eight covariates, age, education, married, black, hispanic, nodegree, earnings in 1974 (re74), and earnings in 1975 (re75). All earnings variables are in thousands of dollars.

1. Create two new covariates, indicators for zero earnings in 1974 (u74) and 1975 (u75). (For example, u74 = 1 if re74 ==0.)

2. Create a table with averages of the ten covariates (including the two newly created ones), by treatment status, standard deviations by treatment status, the differences in sample means, and the normalized difference.

3. Estimate the treatment effects using regression based on the following specifications. Make sure to use robust standard errors.

$$\text{reg1}: y_i = \beta_0 + \beta_1 D_i + u_i \ .$$
$$\text{reg2}: y_i = \beta_0 + \beta_1 D_i + \beta_2 re74 + \beta_3 u74 + \beta_4 re75 + \beta_5 u75 + u_i \ .$$
$$\text{reg3}: y_i = \beta_0 + \beta_1 D_i + \beta_2 re74 + \beta_3 u74 + \beta_4 re75 + \beta_5 u75 + \beta_6 black + \beta_7 hisp$$
$$+ \beta_8 age + \beta_9 married + \beta_{10} nodegree + \beta_{11} education + u_i \ .$$

Is the estimate robust to the specification?

4. Supppose we are interested in evaluating the effect of smoking vs. not smoking on disease rate. For this, we decide to compare the mean difference in the disease rate between the smokers and non-smokers. Can we estimate the treatment effect correctly by comparing the mean difference? What are possible confounders? Briefly explain.

5. You are preparing a grant to run an experiment in behavioral economics. Your experiment requires you to pay UIUC students to play the Prisoner's Dilemma under one of two environments: the standard environment (or control), and a new environment that you expect will encourage cooperation (treatment). Based on the experience of previous research, the average cost to the researcher of a student playing the control game is $30, with a variance of 4.

When students cooperate in the prisoner's dilemma, it costs the researcher more money per student (assume the variance, equal to 4, does not change under the treatment). The treatment effect of interest τ is defined as the mean increase in student earnings (i.e., cost to the researcher) under the cooperation versus the control game, so $c_T - c_C = \tau$. Since you are uncertain about how much cooperation (and therefore cost) to expect under the treatment, you decide to first run a pilot study.

Recall that the formula for the number of sample size in power calculation for a one-sided test is:

$$N = 4 \cdot \sigma^2 \cdot \frac{(\Phi^{-1}(p) - \Phi^{-1}(\alpha))^2}{\tau^2}$$

You are concerned that treatment students may cost $2 more than what control students cost (which could be the case if $\tau = 2$). Suppose that you want a test size of 5% and power of 85%. If you assign half the students to treatment, what is the minimum number of students that you would need to recruit for your pilot study to be able to detect whether $\tau \geq 2$? (It may be useful to know that if $\Phi(\cdot)$ is the cdf of the standard normal with $\Phi(0.15)^{-1} = -1.04$, $\Phi(0.85)^{-1} = 1.04$, $\Phi(0.95)^{-1} = 1.64$, and $\Phi(0.975)^{-1} = 1.96$.)

6. Consider the subclassification method where for each subgroup (x), we consider the average treatment effect (ATE(x)).

   (i) Show that $E[Y_i|D_i = 1, x] - E[Y_i|D_i = 0, x]$ has a selection bias of $E[Y_i(0)|D_i = 1, x] - E[y_i(0)|D_i = 0, x]$.

   (ii) Under what condition, this selection bias can go away?

7. Supposed that the death rate for smokers and non-smokers across age groups are given as the following.

	Death Rate Smokers	Death Rate Non-smokers	# Smokers	# Non-smokers
Old	30	19	25	140
Mid-aged	26	21	90	170
Young	22	18	235	155

   (i) Calculate the average treatment effect of smoking on death rate for each age group.

   (ii) What is the average treatment effect of smoking on death rate?