GLOBAL POVERTY & IMPACT EVALUATION (Fall 2009)

Problem Set 1: Randomized Evaluations

The goal of this problem set is to analyze data from a randomized impact evaluation. We will be using data from a paper by Thornton (AER 2008) titled, "The Demand and Impact of Learning HIV Status". The questions on this problem set will lead you through a series of exercises that are standard practice when analyzing data from a randomized evaluation.

Please submit a do file, log file, and 1 page write up for the problem set. Please work on your own (no group work). Written responses should be *brief*. Questions with a "*" indicate that they are optional.

Download the data from the AER website:
http://www.aeaweb.org/articles.php?doi=10.1257/aer.98.5.1829
UC Berkeley students should have ungated access to the data set. If you have problems accessing it, you might need to set your proxy server settings.

Preliminaries:
Thornton (2008) examines whether if varying the cost of HIV testing can increase the number of people who get their test results. There are two interventions: 1) cash payments to individuals who receive their HIV test results and 2) the distance a person needs to travel to obtain their HIV test results. Both interventions are randomly assigned on an individual level.

Key Variables:
any = 1 if randomly assigned cash incentive to obtain HIV test result
under = 1 if randomly assigned distance to get test result is under 1.5 km
Ti = the randomly assigned cash incentives (amounts) to obtain HIV test results

Going forward, those who were assigned a cash incentive or have to travel less then 1.5 km to get their test results are known as the "treatment" group and those who are not receiving a cash incentive or have to travel further then 1.5 km are known as the "control" group.

<u>Questions</u>
Part I: Summary Statistics
The first step is to look at summary statistics of your sample. This will tell you the sample population that you are analyzing. We will also see if there are differences between the treatment and control group.

Q1) Present summary statistics for the study sample. What is the average age? What percentage of males are in the study? What percentage of people are infected with HIV?
Q2) Present summary statistics separately for those in the control and treatment group. Are there major differences in any of the variables (i.e. age, education, HIV rates)?
Q3*) Do a test to see whether differences in age, HIV rates, and marriage are statistically different between the treatment and control group. Do you see any differences (using a p-value of .05 as the statistically significant threshold)? Are you concerned by any of the differences? How could they affect the analysis?

* Indicates the question is optional and not is not required to get credit for the problem set.

Part II: Analysis using graphs

We can create simple graphs that can help us see the effects of the treatment.

Q4) Generate a bar graph, where the X-axis represents the control and treatment group, and the y-axis is the percentage in each group that learn their HIV status. Let the treatment group in this question be anyone who receives a cash incentive to obtain their HIV test results.

Q5) Now, generate the same bar graph, but this time varying the amount of cash that people receive in treatment (use the "Ti" variable for the x-axis).

Part III: Analysis using linear regression

Using OLS regression analysis is one of the most common tools used to estimate the effect of a treatment or randomly assigned intervention.

Q6) Run the following OLS regression, where getting your HIV test result is the dependent variable, and receiving a cash incentive as your co-variate.

$$Got\ Test\ Result = \alpha + \beta(Any\ Cash\ Incentive) + \epsilon$$

What is your estimate of β ? Is it statistically significant (at the 5% level)? What happens when you include additional control variables (age, male, education, marriage)? Does your estimate of β change?

Q7) Now run a similar regression, but this time replace "Any Cash Incentive" with "Cash Amount". What is your estimate of β ? Is it statistically significant? What happens when you include additional control variables (age, male, education, marriage)? Does your estimate of β change?

$$Got\ Test\ Result = \alpha + \beta(Cash\ Amount) + \epsilon$$

Q8) Now interpret your findings. Based on your estimates from Q4, what can you say about the effect of offering cash incentives on people learning their HIV status? Would you say that this is a big or small effect? Now look at your estimates from Q5. Does a doubling of the cash incentive from $1 to $2 have a big effect on people's willingness to get their HIV test results? Does this surprise you?

Part IV: Heterogeneous Treatment Effects

We might be interested in whether the treatment has different effects for sub-populations. For example does giving cash incentives have a different effect for men and women?

Q9*) Create an interaction term which interacts the treatment (any) with the gender variable (male). Run the following regression

$$Got\ Test\ Result = \alpha + \beta(Any\ Cash\ Incentive) + \gamma(Male) + \delta(Any\ Cash\ Incentive\ X\ Male) + \epsilon$$

What is your estimate of δ ? Is it statistically significant? How do you interpret this result?

Q10*) Create an interaction term that interacts treatment (any) with education (educ2004). Run the following regression

$$Got\ Test\ Result = \alpha + \beta(Any\ Cash\ Incentive) + \gamma(Education) + \delta(Any\ Cash\ Incentive\ X\ Education) + \epsilon$$

What is your estimate of δ ? Is it statistically significant? How do you interpret this result?

Part V: Policy Implications

Q11) Based on your findings from Part III, if the goal of the government is to increase the number of people who know their HIV status, what type of policy would you recommend?

Q12*) Based on your findings from Part IV, are there certain groups that cash incentives should target?

Part VI: Research Design

Q12*) Read the Thornton (2008) article, especially pages 1829-1839. If you could run a similar evaluation, what would you change and why? If you implemented this evaluation in an urban area, what challenges might you face?

* Indicates the question is optional and not is not required to get credit for the problem set.