PUBPOL 639: QUIZ 1 SOLUTIONS
Winter 2011

1. Which of the following is necessary for a randomized trial to produce a good estimate of the causal effect for participants in a study? Circle all that apply.

a. Random sampling of participants for the study

No. Participants in the study are considered the population of interest here. Randomly sampling them from a larger population is not necessary to estimate the causal effect. Random sampling is important, however, if we want to take the causal effect for these participants and extrapolate to a broader population. Since this first proposition was a bit confusing given the question, students who circled this answer are not penalized.

b. Random assignment of participants into treatment and control groups

Yes. Random assignment guarantees the independence of the potential outcome from the treatment status, which allows to estimate the causal effect using the observed difference in outcomes between the two groups.

c. Ability to observe (and thus control for) all factors that may influence the outcome we are studying.

No. Since random assignment makes potential outcome and treatment status independent, it creates two groups that are the same on average (if they are big enough). Then, these "factors" are the same on average between the groups so we do not have to control for them.

2. Explain the term "Fundamental Problem of Causal Inference". Use the potential outcomes framework and terminology if that would be useful, but you do not have to.

The causal effect of a treatment on an individual is defined as the difference between the potential outcome resulting from the treatment and the potential outcome resulting in the absence of the treatment, which can be written mathematically as: Treatment effect = $Y_{1,i} - Y_{0,i}$. *The "Fundamental Problem of Causal Inference" refers to the fact that we cannot observe both potential outcomes* Y_{0i} *and* Y_{1i} *for a given individual (an individual is treated, or is not). So we can never observe or estimate directly the treatment effect on an individual. A randomized controlled trial will, however, let us estimate the average treatment effect for the group of individuals receiving treatment:*

$$\text{Average treatment effect} = E[Y_{1i} - Y_{0i} \mid Di = 1]$$

3. You are given data from a properly executed randomized controlled trial. You have individual-level data on the outcomes (Y) as well as information about who was assigned to the treatment and control groups (D).

a. How can you use these data to estimate the average causal effect of the treatment?

$$E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0] = \{ E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 1]\} + \{ E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]\}$$

The average difference in observed outcomes = average causal effect + selection bias

Random assignment of treatment cancels selection bias so that the average causal effect can be estimated by the average observed difference in outcomes between the treatment and control groups.

b. How can you estimate the statistical significance of the effect discussed in (3a)? If possible, give an equation.

The statistical significance of the causal effect can be estimated testing the following hypothesis:

H_0: no difference in population outcomes = $\mu_{treatment} - \mu_{control} = 0$

against the alternative hypothesis:

H_1: there is a difference in population outcomes = $\mu_{treatment} - \mu_{control} \neq 0$

This test can be done using the following t-statistic:

$$t = \frac{\overline{Y}_{treatment} - \overline{Y}_{control}}{SE\left(\overline{Y}_{treatment} - \overline{Y}_{control}\right)}$$

The effect is considered statistically significant if the associated p-value is less than or equal to an alpha level of .05.

c. If you learned that the data instead did not come from a randomized controlled trial, under what condition would the approach you took to (3a) still be valid?

If for some reason you knew that there was no selection bias, then this simple comparison would be OK. Even though an actual randomized controlled experiment was not done, you may have some institutional knowledge that "treatment" was assigned in a way that approximated a randomized trial. For instance, you may think that the exact location and timing of tornadoes was pretty random so you can estimate the effect of these by comparing a difference in means. Since this assumption is not testable, you would need to provide other evidence that it is likely to hold [e.g. Do places that get tornadoes look like places that don't?]

4. Following the notation we discussed in class, which of the following is observed in actual data? Circle all that apply.

a. $E[Y_{0i} \mid D_i = 0]$

Yes, this is the average outcome of the control group

b. $E[Y_{0i} \mid D_i = 1]$

No, this is the average of the outcome that treated individuals would have had if they had not been treated

c. $E[Y_{1i} \mid D_i = 0]$

No, this is the average of the outcome that non-treated individuals would have had if they had been treated

d. $E[Y_{1i} \mid D_i = 1]$

Yes, this is the average outcome of the treatment group