Introduction to Causal Inference

Solutions to Quiz 1

Teppei Yamamoto

Friday, July 15 2016

Instructions:

- Write your name in the space provided below before you begin.
- You have **15 minutes** to complete the quiz.
- The exam is closed book and calculators are not allowed.
- Please turn off your phone before you begin.
- Note that I have allocated more space than we anticipate you will need for each problem. Just answer the questions as best you can, don't try to fill the available space.
- Good luck!

Your Name: ______________________________

Problem 1

Are the following statements true or false? Justify your answers briefly.

a) While the population average treatment effect (PATE) must be estimated with statistical uncertainty, the sample average treatment effect (SATE) can be calculated without sampling variability because it is defined for the given sample we observe in the data.

False. A SATE is still an estimate of an unobserved estimand. In particular, sampling variability enters into the SATE because the randomization scheme for an experiment is not fixed. Imagine if we ran the same experiment 100 times on the same sample; the randomization vector would probably look different each time, and so would the estimated SATE.

b) Under the complete randomization of treatment assignment, the average treatment effect is identified because the randomization guarantees statistical independence between the treatment indicator D_i and the observed outcome Y_i.

False. We must assume $\{Y_{0i}, Y_{1i}\} \perp\!\!\!\perp D_i$ (i.e. that D_i is statistically independent of *potential* outcomes) in order to identify the average treatment effect. The identification assumption in the question assumes that D_i is independent of Y_i, the observed outcome, which we hope is not true in order for our treatment to have an effect!

c) All of the following scenarios are examples of SUTVA violations: i) treatments may be transmitted from treated to untreated units, ii) units' outcomes are affected by the treatment assignment of nearby units, and iii) resources used to treat one set of subjects diminish resources that would be otherwise available to other subjects.

True. All of these scenarios are examples of SUTVA violations. i) and ii) violate the non-interference assumption because here the potential outcomes for unit i would rely on the treatment/control status of other units. For non-interference to hold no matter which subjects are allocated to treatment or control, a given subject's potential outcomes must remain the same. ii) can be thought of as violating both non-interference or the consistency assumption.

Problem 2

Consider a randomized control trial with four observations, of which two units were randomly assigned to treatment via complete randomization. We use $D_i \in \{0,1\}$ and Y_i to denote the treatment (1 for treatment and 0 for control) and the observed outcome for unit i, respectively.

a) The table below shows the data observed from this experiment, augmented with columns for potential outcomes and the treatment effect for each unit. Fill in all the empty cells in the table based on the observed information, denoting unknown information with "?".

i	D_i	Y_i	Y_{1i}	Y_{0i}	τ_i
1	1	8	8	?	?
2	0	2	?	2	?
3	0	4	?	4	?
4	1	6	6	?	?

b) Define the population average treatment effect for the treated (ATT) using the above notation and propose an unbiased estimator for this estimand. Then, using the data in the table, estimate this quantity.

The estimand:

$$ATT = \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1]$$

Given randomization of treatment, an unbiased estimator is the difference-in-means for the treated and control. Recall we are given randomization of D_i, so the $ATE = ATT$. Thus, for m, the number of units treated:

$$\widehat{ATT} = \frac{1}{m}\sum_{i}^{N} Y_{1i}D_i - \frac{1}{N-m}\sum_{i}^{N} Y_{0i}(1-D_i)$$

Finally, plugging in, we get our estimate:

$$\frac{8+6}{2} - \frac{2+4}{2} = 4$$