Introduction to Causal Inference

Solutions to Quiz 2

Teppei Yamamoto

Saturday, July 16 2016

Instructions:

- Write your name in the space provided below before you begin.
- You have **20 minutes** to complete the quiz.
- You may answer questions either in English or Japanese.
- 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: _____

Are the following statements true or false? Why? For each question, state whether you think the statement is true or false, and explain why you think so in a few sentences.

a) If data on a pre-treatment covariate is available for the subjects in a planned experiment, block randomization on that covariate is preferable to complete randomization followed by regression adjustment of the covariate.

True. Block randomization guarantees a perfect balance on the pre-treatment covariate between the treated and control groups. In contrast, complete randomization might leave the covariate imbalanced by chance particularly when the sample size is not large, and a post-hoc adjustment of the variable might cause bias that could have been entirely avoided by the block randomization.

b) A key identification assumption in a randomized experiment is that there is no interference between units. However, if interference can be assumed to occur only within clusters of units, we can correct for the violation of the no-interference assumption by adjusting standard errors using methods such as cluster-robust standard errors.

False. Cluster robust SEs cannot correct violation of SUTVA. It can only address the possible correlation between potential outcomes of the units belonging to the same clusters, which is different from interference (i.e. treatment on unit A directly affecting the outcome of unit B belonging to the same cluster).

c) External validity of an experiment is compromised when the sample average treatment effect (ATE) diverges from the population ATE. Therefore, results from an experiment are guaranteed to be externally valid if one has a true random sample from the population of interest.

False. External validity of an experiment also hinges on whether the treatment used in the experiment corresponds to the theoretical causal variable. If the treatment in your experiment is a poor representation of the real-world counterpart you are ultimately interested in, you may not be able to draw valid inference about the real world based on your experimental result.

d) A researcher is analyzing the effect of a treatment in a randomized experiment and uses a two-sample t-test with unequal variances to reject the null hypothesis of zero average treatment effect. The researcher could have tested the same null hypothesis with a randomization test like Fisher's exact test to avoid any large sample approximations.

False. A randomization test will test the sharp null of no effect instead of the zero average treatment effect.

e) When you have experimental data, it is better to run a simple regression of the outcome variable on the binary treatment variable than to use the difference-in-means estimator. This is because the ordinary least squares (OLS) regression yields the most efficient estimator and thus the smallest standard error for the average treatment effect.

False. The OLS slope coefficient is numerically identical to the difference-in-means estimator when the treatment is binary and no other predictors are included in the model. Therefore, their sampling variances are also identical.