QTM 385 - Experimental Methods

Assignment 06

Instructions

As in previous assignments, here are 10 questions related to our recent classes. This time, they cover attrition. Please answer all question and export the file as a PDF or html file. If you use R or Jupyter Notebook, make sure to include both the code and the output in the document.

If you have any questions about the assignment, feel free to email me at danilo.freire@emory.edu.

Good luck!

Questions

1. Sometimes experimental researchers exclude subjects from their analysis because the subjects (1) appear to understand what hypothesis the experiment is testing, (2) seem not to be taking the experiment seriously, or (3) fail to follow directions. Discuss whether each of these three practices is likely to introduce bias when the researcher compares average outcomes among non-excluded subjects. (From Gerber & Green, 2012, p. 245)

2. Suppose a researcher studying a developing country plans to conduct an experiment to assess the effects of providing low-income households with cash grants if they agree to keep their children in school and take them for regular visits to health clinics. The primary outcome of interest is whether children in the treatment group are more likely to complete high school. A random sample of 1,000 households throughout the country is allocated to the treatment group (cash grants), and another sample of 1,000 households is allocated to the control group.

(a) Suppose that halfway through the project, a civil war breaks out in half of the country. Researchers are prevented from gathering outcomes for 500 treatment and 500 control subjects living in the war zone. What are the implications of this type of attrition for the analysis and interpretation of the experiment?

(b) Another identical experiment is performed in a different developing country. This time the attrition problem is as follows: households that were offered cash grants are more likely to live at the same address years later, when researchers return in order to measure outcomes. Of the 1,000 households assigned to the treatment group, 900 are found when researchers return to measure outcomes, as opposed to just 700 of the 1,000 households in the control group. What are the implications of this type of attrition for the analysis and interpretation of the experiment? (from Gerber & Green, 2012, p. 246)

3. In the RAND Health Insurance Experiment (HIE), attrition varied significantly across the different insurance plans (free vs. cost-sharing). Discuss how differential attrition rates across treatment subgroups can lead to misleading conclusions about the average treatment effect (ATE), even if overall attrition appears modest. How does this relate to the concept of $MIPO|X$? How can this be tested?

4. Explain why relying solely on Inverse Probability Weighting (IPW) to adjust for attrition can be problematic. What assumptions are required for IPW to produce unbiased estimates? What are the consequences if these assumptions are violated? Is it better to present the estimate with and without IPW? Justify your answers with further readings Gerber & Green (2012, p. 244) has some suggested readings that you can use.

5. Building on the idea of minimising attrition during an experiment, propose three specific and practical data collection strategies (beyond simply "collect more data") that might help reduce attrition in a longitudinal study. Justify why these strategies might be effective.

6. How is the validity of IPW estimates dependent on the assumption of MIPO|X? Discuss what happens if this assumption is violated.

7. For the upcoming questions, you will be using a dataset called attrition_data.csv, which is available in the course repository. This contains the results from an experiment with attrition in both the covariates and the treatment. There are two outcome variables in the dataset. Using the `attritevis` package, as shown in class, please answer do the following:

Load the dataset and show the number of cases of attrition and their proportion. Estimate the balance across attrition for `outcome2`. What do you observe?

1. Plot the attrition timeline with the total and the control and treatment groups. Use the parameters y, `treatment_q` and `outcome_q` to show the results. What do you observe?

2. Estimate Manski bounds for the outcomes. Interpret the results.

3. Finally, estimate Lee bounds and contrast the result with the previous two. Which of the two methods would you use? Justify your answer and elaborate on the possible violation of the monotonicity assumption.