

Problem Set 1

1 Purpose

The purpose of this problem set is to assess your understanding of one key method of quantitative public opinion research: experimental design and analysis.

2 Your Task

1. In your own words, explain the “potential outcomes framework” of causal inference and explain how experiments provide a way to identify causal effects.
2. A researcher wants to understand how the provision of cash incentives to poor families with children affects the educational attainment of the children and considers two alternative designs.
 - (a) The first design involves examining the educational attainment of children whose parents are or are not eligible for public cash assistance by generating a random sample of the population, selecting individuals with incomes just above and just below the cutoff for assistance, and tracking the educational progress of their children.
 - (b) The second design involves recruiting a non-representative sample of families that are not currently eligible to receive benefits (but are close to eligibility). One half of families are randomly assigned to receive cash assistance and the other half is randomly assigned to receive nothing. Educational attainment is tracked for both groups.

Discuss the trade-offs involved in these designs, including what would be required to obtain an estimate of the causal effect of cash assistance on educational attainment.

3. Consider an experiment on 500 individuals in which one group is randomly assigned to read a treatment message from Boris Johnson supporting a “no deal” position in the ongoing Brexit negotiations and another group is assigned to a control condition that receives no information. Measures of opinions for about support for “no deal” are recorded for both groups on a 0 to 1 scale, with higher scores indicating greater favorability toward “no deal.”
 - (a) Assuming the treatment group mean score was 0.68 and the control group mean score was .51, what is the average treatment effect? Is this substantively large or small?
 - (b) Assuming the t-statistic for the mean-difference is 1.76, should we consider this effect to be statistically large and distinguishable from zero?
4. What is a randomisation distribution? What can we learn from the randomisation distribution for the sample average treatment effect?

5. The statistical power of a two-sample t -test (which is, in essence, the power of a posttest-only, two-group experimental design) is influenced by four main things: the size of each experimental group, the difference-in-means (i.e., difference in mean values of the outcome in the two groups), the variance of the outcome measure, and α (the significance level or “Type 1” error probability).
 - (a) If α (the Type 1 error probability) is 0.05, how often should we expect to find a “statistically significant” effect size when one is not present?
 - (b) If you increase the size of your treatment groups in an experiment while the expected effect size remains unchanged, what happens to the power of your experiment? Are you more or less likely to obtain a “false zero” result? What about “false positives”?
 - (c) Imagine we are expecting to find a small effect but we can only collect a small number of observations in our experiment, so the minimum detectable effect size in our study is larger than the effect size we would expect to observe given our theory. If our experiment reveals an effect that is statistically distinguishable from zero, what are the two possible interpretations of this result?
6. Sometimes experiments are “broken” due to challenges of implementing an intervention and measuring outcomes. In what ways can experiments fail? And what implication(s) does each of those points of failure have on the analysis of the experimental data and the interpretation thereof?
7. Are experiments more or less useful than other methods for generating evidence-based policy? Why? What caveats — if any — should be placed on the use of experimental evidence for policy and decision-making?

3 Submission Instructions

Please submit your answers as a PDF document via Moodle. It should be no more than 4 pages, single-spaced, in Times New Roman font size 12, on A4 paper with standard 2.54cm margins. The code for R or Stata to reproduce results should be included as an appendix, written entirely in fixed width format font (e.g., Courier New). A solution set will be provided on the course website and the activity will be discussed in class.

4 Feedback

Group feedback will be provided during class. If you would like more specific individual feedback on your work, please ask the instructor during class or office hours.