790:676: Experimental Methods

Due: 2/26 11:59PM on Canvas

Problem Set: Common Experimental Concepts and Analyses

Instructions

You will submit, on Canvas, two documents (or a single compiled R Markdown or KnitR document):

- An R script (.R) file with your code. Follow the best practices by titling your script and using # comments to explain your steps. This code should be clean. I should be able to run your code to verify that the code produces the answers you write down. (If you have decided not to use R, please provide an equivalent log of the steps you took in SPSS or Stata.)
- A word document or pdf with answers to the written questions. This document should also have a title with your name on it.

You are allowed to use any and all resources to help you answer the questions with the exception of your colleagues' solutions. You may work together on the problem set, but every individual should write up his or her problem set completely independently.

Concept Questions

1. Please complete the following exercises from Gerber and Green (2012).
   (a) Chapter 2, Exercise 1(d)
   (b) Chapter 2, Exercise 7
   (c) Chapter 2, Exercise 10
   (d) Chapter 2, Exercise 12
   (e) Chapter 3, Exercises 1(a), 1(b), 1(c)
   (f) Chapter 9, 1(b)

Empirical Application

2. **List Experiment Analysis.** We will replicate a portion of the analyses conducted in "The Size of the LGBT Population and the Magnitude of Antigay Sentiment Are Substantially Underestimated" by Katherine B. Coffman, Lucas C. Coffman, and Keith M. Marzilli Ericson in *Management Science*, published in 2016. The authors conduct a set of list experiments to study the size of the LGBT population and anti-LGBT sentiment in the United States.

   The goals of this problem are

   - to get you confident in reading, understanding, and applying mathematical notation,
   - expose you to another experimental technique,

- and allow you practice conducting difference-in-mean tests and quantifying uncertainty using actual social science data.

List experiments (or the Item Count Technique) are one of several methods that researchers use in situations where they are interested in asking people about sensitive subjects. (List experiments are discussed briefly in Chapter 2 of the assigned Mutz (2011) reading for Week 5.) We will read about other sensitive question techniques at the end of the course. For now, you may read the first part of the article to familiarize yourself with the authors' research question, hypotheses, and design.

In list experiments, the researcher is able to ascertain the proportion of subjects who hold a sensitive trait, but is unable to identify if any particular subject holds the sensitive trait. This design feature is supposed to make it more likely that subjects will reveal sensitive beliefs and behaviors on surveys than if they were asked a question directly.

In the study here, the list experiment technique is modified. An additional design feature also allows the researcher to be able to estimate "how sensitive" the sensitive survey item is. We estimate how reporting the sensitive item (our outcome) changes as a result of manipulating whether the item is asked directly (control) vs. asked in a veiled setting (treatment). Our null hypothesis is that there is truthful reporting, meaning there is no difference. The alternative hypothesis is that there is some sensitivity (there is a difference).

Review Table 1 in the article for the experimental design. Each experiment consists of two conditions.

- In a "direct report" condition, respondents are provided with a list (get it- list experiment!) of four items. Notably, this list does not contain a sensitive item. Respondents then answer "Please fill in the bubble that corresponds to the total number of statements above that apply to you." This outcome c_{qi}^D is recorded as a number from 0 to 4 in the direct report condition.
- In the "veiled report" condition, subjects receive the same list except one additional sensitive item is added to the list. Respondents indicate the number of statements but, again– **importantly**– not which statements apply to them. This is an outcome (c_{qi}^V) from 0 to 5.
- In the Direct Report condition, subjects are also asked about the sensitive item in a direct question (outside of the list). For example, subjects in the Direct Report condition were asked "Do you consider yourself to be heterosexual?" The answer to each of these questions is "No" or "Yes" (d_{qi}), where one of the answers is always thought to be sensitive.

The researchers define the

- The observed outcome for those in the Direct Report condition as $Y_{qi}^D = d_{qi} + c_{qi}^D$ where d_{qi} is equal to one or zero based on subject i's answers to a direct question (q), and c_{qi}^D is the number of the four statements reported as true.
- In the Veiled Report condition, the observed outcome is just $Y_{qi}^V = c_{qi}^V$ because subjects were not asked the direct question.

Under truthful reporting, which is the null hypothesis, the authors state that $E[Y_{qi}(D)] = E[Y_{qi}(V)]$ and that the quantity of interest is: $\mu_q = E[Y_{qi}(V)] - E[Y_{qi}(D)]$.

The data for this study are in `lgbt2.dta` on Canvas. The variables are as follows where "5" corresponds to the sensitive item "Would you be happy to have an openly lesbian, gay, or bisexual manager at work?"

- `in5` is c_{5i}^D. It is missing (NA) for anyone in the veiled condition.
- `s5_direct` is d_{5i} where 1 is "Yes" and 0 is "No." It is missing (NA) for anyone in the veiled condition.
- `s_private_sum_5` is c_{5i}^V. It is missing (NA) for anyone in the direct condition.
- `privatetreatment` is the treatment indicator where 0 means the subject was in the Direct Report (D) condition, and 1 means the subject was in the Veiled Report (V) condition.
- `south` is an indicator variable (1=lives in south, 0=does not live in south)

Questions

(a) Explain what $\mu_q = E[Y_{qi}(V)] - E[Y_{qi}(D)]$ means. Explain why $E[Y_{qi}(D)] = E[Y_{qi}(V)]$ means truthful reporting.

(b) Load the data. How many subjects were in the Veiled vs. Direct Report conditions?

(c) What proportion of subjects reveal the sensitive item for the **direct question**: Would you be happy to have an openly lesbian, gay, or bisexual manager at work? Here, the sensitive answer is considered to be "No." Note: if you have missing data (`NA` in a column), you have to add `na.rm = TRUE` to functions like `mean(x, na.rm = TRUE)` to tell R to ignore missing data, which will be the approach we take here.

(d) Create a single variable for the outcome Y_i that combines the observed outcomes from the Direct and Veiled treatment conditions into one variable. Hint: $Y_i = c_{5i}^V$ **if** respondents are in the veiled condition, and $Y_i = d_{5i} + c_{5i}^D$ **otherwise**.

- Slides 34-36 in Week 2 provide examples of creating new variables in R. Chapter 2 in Imai QSS also provides some examples `http://assets.press.princeton.edu/chapters/s2-11025.pdf`.

(e) Carry out a two-sided hypothesis test for a null hypothesis of no average effect (i.e., no change in reporting or "truthful reporting") for the sensitive question "Would you be happy to have an openly lesbian, gay, or bisexual manager at work?"

- Report the average observed outcome for each condition,
- The estimate for change in reporting,
- The standard error of this estimate,
- and the p-value for the hypothesis test. Explain what the p-value means, and provide a conclusion about the null hypothesis. Assume we are using conventional thresholds for statistical significance.
- In addition, what is the p-value for the same test, except this time where the alternative hypothesis is that $E[Y_{qi}(V)] > E[Y_{qi}(D)]$? Note: We can modify our t-test function in R to specify the type of test. By default, it assumes a two-sided test. We can actually, instead, add an argument to specify the alternative: `t.test(y1, y0, alternative = "two.sided")`, `t.test(y1, y0, alternative = "less")`, or `t.test(y1, y0, alternative = "greater")`.

(Hint: Don't overthink this. You should be able to use the same tools for hypothesis testing that we have used in class.)

Note: Because the sensitive item is "No" in this case, this affects our interpretation of the difference in means. In this case, if the mean in the veiled condition is less than the direct condition, we can interpret it to mean that there is a change in reporting, and that the item is sensitive: The idea is that this difference would mean more reporting of the senstive item "No" in the veiled condition (because the veiled condition has compelled fewer people to say that the statement that they would be happy to have an openly LGB manager applies to them), and thus, the overall average count of items is smaller.

Also note: Your answers may be slightly different from those in the published paper, as the authors estimated change in reporting using a regression analysis with several covariates. We are looking at the raw experimental effect instead.

(f) We now have an estimate of the proportion of people who say, directly, they would not be happy with an openly lesbian, gay, or bisexual manager at work. We also have an estimate of the additional proportion of people who indicate they are not happy with an openly lesbian, gay, or bisexual manager if asked in a veiled way. Combining these two estimates, what is our estimate of the true proportion of people who would not be happy with an LGB manager?

(g) Repeat the two-tailed hypothesis test just among southern respondents and then just among non-southern respondents. Create a plot that displays the estimate for change in reporting for the full sample, the estimate for just southern respondents, and the estimate for just non-southern respondents. Display the 95% confidence intervals around each estimate. Make sure you use clear labels and sensible dimensions for your figure. Provide a brief interpretation.

(h) We now address the distinction between testing whether a single CATE is different from zero and testing whether two CATEs are different from each other. Conduct a statistical test to determine if there are differences between Southerners and Non-Southerners in change in reporting on this question. Now regardless of what you find, let's assume you do find a difference: is this causal? In other words, if we do find a difference, can we conclude that living in the south **causes** our treatment effect to be different? Explain your conclusion.

(i) A rule of thumb in constructing the lists in list experiments is that you should avoid a list where none of the statements or all of the statements tend to apply to your respondents. Why is this important?