Joel Cabrera

Experimental Methods (01:790:676)

Professor McCabe

February 26, 2020

**Problem Set #1 – Questions and Answers**

Concept Questions:

1a. Ch. 2, Exercise 1d

1d. “Yi (0) | Di = 1” means the potential outcome for observation i that is untreated, given that it receives a hypothetical treatment. In contrast, “Yi (0) | Di = 0” means the potential outcome for observation i that is untreated, given that it does not receive a hypothetical treatment.

1b. Ch. 2, Exercise 7

7a. An unbiased estimator is an estimator that can, on average, produce true parameter estimates – that is, E() = B.

7b. Randomly assigning Villages 3 and 7 to the treatment group seems to have produced upward estimates, as it can be shown by the values of their treatment effects in Table 2.1. However, such estimates are not biased due to random assignment, as said values are no longer dependent on the treatment – that is, a village being assigned a female head.

7c. If the experimenter arbitrarily assigns Villages 3 and 7 to the treatment group, then this would not only introduce research bias, but also produce biased estimates. This is because the potential outcomes of this experiment are now dependent on the treatment such that “ E[] = E[Yi (1) | Di = 1] - E[Yi (0) | Di = 0]”, which can lead to the latter.

1c. Ch. 2, Exercise 10

10a. The excludability assumption states that, if Yi (z, d) - where z ∈ [0,1] (randomly assigned to control or treatment group) and d ∈ [0,1] (receive treatment or not), then Yi(1, d) = Yi(0, d). In other words, potential outcomes only respond to di; the value of zi is irrelevant. If the researcher informs the students in the treatment group about the purpose of the experiment, then they could adjust their behavior when receiving the treatment. This, in turn, makes z relevant, violating this assumption.  
 10b. The non-interference assumption states that Yi(d1, d2, . . ., dn) = Yi(d), where di = for all d. In other words, potential outcomes are affected solely by the treatment administered to the subject, not by treatments administered to other subjects. In this case, if students bring their newspapers to the cafeteria, then there is a possibility that students in the control group could read them, rendering them no longer part of the control group. This not only violates this assumption, but also makes estimations more complicated.

1d. Ch. 2, Exercise 12

12a. One would be hesitant to assume that “E[Yi (0) | Di = 0] = E[Yi (0) | Di = 1]” and “E[Yi (1) | Di = 0] = E[Yi (1) | Di = 1]” for a couple of reasons. The first reason is that one cannot observe both Yi (0) and Yi (1) at the same time. One could never know if both the former and latter are really true. The second and potentially last reason is that one cannot truly know the potential outcome given no receipt of the treatment and an actual receipt of the treatment at the same time. For example, for the former, one could observe the potential outcome of violent encounters when reading less than 3 hours per day, given that the prisoner is reading less than 3 hours per day. But, one cannot also observe the potential outcome of violent encounters when reading less than 3 hours per day, given that the prisoner is reading more than 3 hours per day.

12b. If 10 prisoners are randomly assigned to read for 3 hours per day for 1 week (treatment group), and the remaining prisoners are also randomly assigned to not read at all during the same week (control group), it can be said that the excludability assumption is satisfied here. Prisoners in their respective groups are only responding to di; random assignment accounts for this. Afterwards, the researcher is able to obtain unbiased estimates from this experiment.

12c. The non-interference assumption, as stated in 10b, essentially states that potential outcomes are affected solely by the treatment administered to the subject, not by treatments administered to other subjects. Since it is assumed that all prisoners in the treatment group do, indeed, read for 3 hours per day for 1 week, and that prisoners in the control group do not read at all during the week, it can be said that this experiment has this assumption satisfied.

12d. The non-interference assumption plays a role when all prisoners are required to read for 3 hours a day in that we must assume that not only are all prisoners randomly assigned to the treatment group, but also are not to deviate from the amount of time spent on reading. Should we be able to do both of these, the non-interference assumption is satisfied.

1e. Ch. 3, Exercise 1a, 1b, 1c

1a. A standard error is the standard deviation of the sampling distribution of - that is, it is a measure of spread that captures the standard deviation from the mean of a population based on a sample. It is estimated by diving the standard deviation of a sample by the square root of the sample size.

1b. Randomization inference is used to test the sharp null hypothesis (Yi (1) = Yi (0) for all i) for no effect for any subject by allowing us to assume that said hypothesis is true, then simulate the sampling distribution under the null distribution (obtains all possible random assignments), and finally assess the distribution relative to the ATE that one observes under the random assignment of the sample.

1c. A 95% confidence interval is a range of values that has a 95% chance to contain the population parameter of interest. In 95% of samples, then, the parameter estimate will fall in this interval.

1f. Ch. 9, Exercise 1b

1b. An interaction effect is the change in the treatment effect that occurs from 1 subgroup to the next. This, the, allows us to find ATEs for subgroups (A.K.A. conditional ATEs).

Empirical Application:

2a. Interpreting “μq **=** E[Yqi (V)] - E[Yqi (D)]”: the population mean for all direct questions (q) asked is the difference between the mean of the potential outcome for all q and subjects (i) that were in the Veiled Report (V) condition and the mean of the potential outcome for all q and i that were in the Direct Report (D) condition. (Note: V is the treatment, D is the control). This equation is used to obtain the estimated ATE based on the difference between the average values of the treatment and control groups.

Given these facts, it can be said that E[Yqi (V)] = E[Yqi (D)] means truth reporting because the equation means that the potential outcome for all q and i that were in the Veiled Report (V) condition is equal to the mean of the potential outcome for all q and i that were in the Direct Report (D) condition, indicating no difference between the two groups of subjects. The total numbers of statements in the list for each group, then, are equivalent to each other.

2b. The dimensions of the data are 2516x12, indicating that there are 2516 subjects and 12 variables in total. There are 1246 and 1270 subjects in the Veiled and Direct conditions, respectively.

2c. Approximately 16% of subjects said “No” to the direct question: Would you be happy to have an openly lesbian, gay, or bisexual manager at work?

2d. (See R script for new variable that combines the observed outcomes from the Direct and Veiled condition). (Note: Variable should be Yi = Y5iV + Y5iD = C5iV + d5i + C5iD)

2e. Given this dataset, we test the null hypothesis of no average effect for the sensitive question “Would you be happy to have an openly lesbian, gay, or bisexual manager at work?”. The null and alternative hypotheses are formalized below.

H0: There is truthful reporting. (E[Y5i (V)] = E[Y5i (D)])

HA: There is some sensitivity. (E[Y5i (V)] ≠ E[Y5i (D)])

In R, I find that the average observed outcomes for the Veiled and Direct conditions are approximately 3.05 and 3.17, respectively. The difference between these two means is -0.12, meaning that people are 12% less likely to report that they would be happy to have an openly LGBT manager. Conducting a t-test between these two kinds of conditions, we are given a p-value of 0.001768. Since the p-value is much less than 0.05 – the default level of statistical significance, we can reject H0 – that there is truthful reporting. Because of this, it is suggested that there is, indeed, some sensitivity in how people answer questions regarding having LGBT managers at the workplace. Furthermore, given that we have the values of the difference in means and the t-statistic, they can be used to calculate the standard error of the difference in means, which is around 0.04.

Finally, performing a right-tailed alternative t-test on the two kinds of outcomes for Veiled and Direct conditions, we are given a p-value of 0.991. the H0 and HA can be found below. Since the p-value is greater than 0.05 – the default level of statistical significance, we fail to reject H0 – that there is truthful reporting. Because of this, it is suggested that people do, indeed, truthfully report on how they feel regarding LGBT managers at the workplace. (Q: Doesn’t this conclusion contradict the one made based on the 2-tailed t-test?)

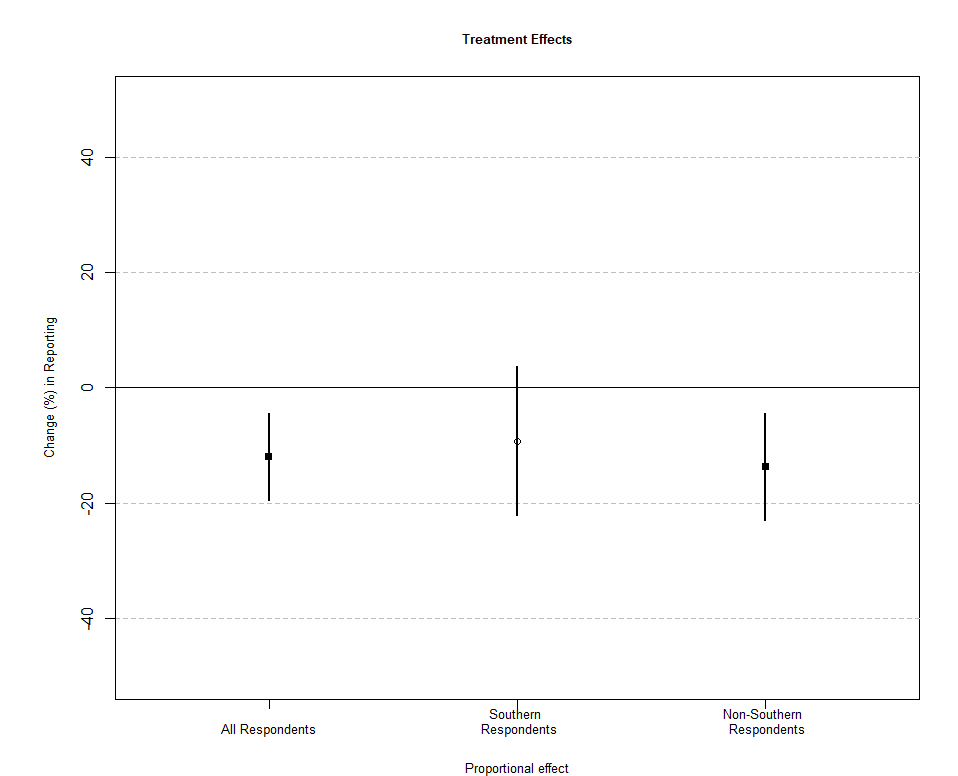
H0: There is truthful reporting. (E[Y5i (V)] ≯  E[Y5i (D)])

HA: There is some sensitivity. (E[Y5i (V)] > E[Y5i (D)])

2f. The first and second estimates of the proportions who directly said that they would not be happy with an LGBT manager at work and if asked in a veiled way are around 16.22 and 12.03, respectively. Adding these values together gives us an estimate approximately 28.25, which represents the true proportion of people who indicate they are not happy with an LGBT manager at work.

2g. We repeat the 2-sided hypothesis test for just southern respondents and then just among non-southern respondents. The t-test for the first subgroup results in a p-value of 0.16. Because it is greater than 0.05, we fail to reject the null hypothesis. Because of this statistical significance, it is suggested that southerners do, indeed, truthfully report on how they feel regarding LGBT managers at the workplace. The t-test for the second subgroup results in a p-value of around 0.004. Because it is less than 0.05, we reject the null hypothesis. Because of this statistical significance, we can say there is, indeed, some sensitivity in how non-southerners answer questions regarding having LGBT managers at the workplace.

We also provide a plot that graphically displays the estimates for the change in reporting, among just southern respondents, and among just non-southern respondents – each with its own 95% confidence interval. The plot can be found below:



From left to right, we see the estimates with their corresponding 95% intervals. Each estimate has a negative change in reporting; our previous proportions can confirm this. Notice, however, that the 95% confidence interval for southern respondents exceeds the 0% line. Because of this, we can say that such an estimate is not statistically significant at the 5% level, which can be confirmed by our previous answer for statistical significance for southern respondents.

2h. We now test whether a single CATE is different from zero and whether two CATEs are different from each other. Performing a multivariate regression of the observed outcomes of subjects from the Veiled and Direct treatment conditions on southern respondents, whether a subject was administered the Veiled or Direct condition, and the interaction between southern respondents and the administration of the Veiled or Direct condition, we are given the estimates for the coefficients on said variables and the p-value for the F-statistic. The latter has a p-value of around 0.005. Because it is less than 0.05, we can reject the null hypothesis. Because of this statistical significance, we can say that there is, indeed, some sensitivity in how people answer questions regarding having LGBT managers at the workplace (there is a difference).

However, suppose that we find a difference from this regression. It should be iterated, however, that living in the south does not cause our treatment effect to be different. This is because there are many other factors that could have caused the treatment effect to be different. In other words, we have to account for omitted variable bias first before we can make a claim about causality in this case. Furthermore, we have only performed only one kind of statistical test for this difference. Multiple robustness checks should also be performed in order to see if the results from said test appear to be the same.

2i. It is important to avoid items in list experiments such that none of them or all of them tend to apply to the respondents. This is because having either the former or latter scenario can complicate analyses for the researcher. For the latter, subjects might be disinclined to not respond to any items that they find sensitive, leaving the researcher with little to no information regarding the proportion of subjects who hold a sensitive trait. For the former, if subjects find none of the items sensitive, then there would be no information for the researcher to analyze.