Durably Reducing Transphobia Revisited

A recent paper by Broockman and Kalla, explored in an earlier exercise of *QSS*, asked if individual minds could be changed with respect to contentious political topics, particularly transgender rights. Broockman and Kalla developed a field experiment where individuals were randomly assigned to receive different treatments and their attitudes toward transgender rights were measured before and after the treatment. The paper that describes their findings will be the basis of this excercise:

Broockman, David, and Joshua Kalla (2016). "Durably reducing transphobia: A field experiment on door-to-door canvassing." *Science* 352(6282): 220-224.

Broockman and Kalla were focused on whether perspective-taking – "imagining the world from another's vantage point" – might be particularly effective in reducing intergroup prejudice. Therefore, participants in the treatment group were visited by a canvasser—some of whom were transgender and some of whom were not—and asked to think about a time when they were judged unfairly and were guided to translate that experience to a transgender individual's experience. A placebo group had conversation with canvassers about recycling.

The data you will use, `broockman_kalla_2016.csv`, contains all the variables you will need. The names and descriptions of variables are shown below. You may not need to use all of these variables for this activity. We've kept these unnecessary variables in the dataset because it is common to receive a dataset with much more information than you need. Some of the items appeared on multiple surveys; the `#` sign below will be replaced with the survey number in your analysis:

- the baseline (pre-treatment) survey is survey 0;
- the 3-day survey is survey 1;
- the 3-week survey is survey 2;
- the 6-week survey is survey 3, and;
- the 3-month survey is survey 4.

Name	Description
`vf_age`	Voter file age
`vf_racename`	Voter file race
`vf_female`	`1` if the voter is female, `0` if the voter is male.
`vf_democrat`	`1` if the voter is a Democrat, `0` if the voter is not a Democrat.
`respondent_t#`	`1` if voter responded to the survey at wave 0, 1, 2, 3, or 4 of the study, `0` if voter does not.
`miami_trans_law_t#`	Support or opposition to antidiscrimination law (measured from `-3` (opposed) to `3` (in support))
`therm_trans_t#`	Feeling thermometer towards trans people (0-100).
`treat_ind`	`1` if individual was assigned to treatment, `0` if assigned to placebo.
`exp_actual_convo`	`1` if treatment was actually delivered, `0` if it was not.
`canvasser_trans`	`1` if canvasser identified as transgender or gender non-conforming, `0` if canvasser did not.
`canvasser_id`	Canvasser identifier.
`hh_id`	Household identifier.

Question 1

First, load `broockman_kalla_2016.csv` into R and explore the data. How many observations are there? This file includes all of the people that the researchers *attempted* to contact. As we've learned, experiments also experience this noncompliance. However, many of these people did not even respond to the baseline survey before the experiment was conducted. As such, we are not be able to estimate how the treatment changed these individuals' attitudes, since we don't know where they were to begin with. To deal with these people who were never reached, subset your data so that you are only looking at people who have a valid (non-NA) answer for `therm_trans_t0`, which was collected in the pre-treatment baseline study.

How many observations are you left with? How many are assigned to treatment? Placebo? What do these numbers tell us about the randomization process? You should use this subset for the rest of this questions in this part of the exam.

Question 2

To examine the effect of these conversations, Broockman and Kalla compared how people in the treatment and placebo groups evaluated transgender people on a specific type of survey question called a "feeling thermometer," where respondents indicate where their feelings fall on a scale of 0 (very cold) to 100 (very warm). Further, to measure whether the changes persisted, Broockman and Kalla also conducted follow-up surveys up to 3 months after the original treatment. Thus, Brookman and Kalla's study required reaching participants in multiple ways: both for a face-to-face interaction where the treatment would be delivered and through multiple follow-up surveys where the long-term effects of the treatment could be measured.

We would like to further check that noncompliance did not meaningfully affect our randomization, since the basis of calculating our treatment effect depends on this. To do this, we can conduct a *placebo test* to ensure that the average level of support (our outcome) is not meaningfully different between the treatment and control groups *before* treatment was administered (during the baseline survey). Conduct a t-test on support before treatment was administered ($t = 0$) between the treatment and control groups among the subset of people you calculated in question 1. Construct a 95% confidence interval around your estimate. If the null hypothesis is no difference between the two groups, can you reject the null hypothesis?

Question 3

To simplify our analysis, we will focus on the average treatment effects among the treated (ATT). Kalla and Broockman experienced noncompliance in their study, where individuals assigned to the treatment group refused to engage in conversations with canvassers. Because accounting for this noncompliance statistically is beyond the scope of this class, we will look only at the cases where treatment was actually delivered. In contrast, Broockman and Kalla estimate the average treatment effect of compliers (adjusting for compliance rates). They also include many covariates; we will only include a few. For these reasons and others, our estimates may differ slightly.

Next, let's see if the attitudes in the treatment and placebo changed after the treatment was administered and for how long these differences lasted. Perform four difference-in-means calculations—one for each wave ($t = 1, 2, 3, 4$). When performing each test, remove any NAs in that wave. (These NAs might arise from participants failing to complete different waves of the surveys—a process that social scientists call "attrition"—or for other reasons.) For example, for wave 1, use all cases that don't have missing outcome data in wave 1, and in wave 2, use all cases that don't have missing outcome data in wave 2 (in the `t.test()` function, you can accomplish this with the `na.action = na.omit` argument). Construct 95% confidence intervals for your estimates. At each stage, do we reject or retain the null hypothesis that no difference between the two groups exists? You are welcome to use `t.test()` to calculate these differences and the confidence intervals, but pick one comparison and calculate the estimates by hand (the techniques in Section 7.1 of *QSS* may be helpful).

Question 4

Let's focus on the difference-in-means test at time $t = 1$. If the null hypothesis is that there is no difference between the groups, what would it mean to make a type I error? What would it mean to make a type II error? If we did something to increase the statistical power of the study, would we increase or decrease the probability of a type II error?

Question 5

Finally, let's approximate the estimation strategy used by Broockman and Kalla in their analysis. To estimate the average treatment effect, they use a regression framework. Regress the feeling thermometer dependent variable (measure at time $t = 3$) on treatment assignment. To further alleviate concerns of imbalances between treatment and control groups, adjust for an individual's feeling thermometer scores at time $t = 0$, her age, gender, race, and political party. Further, to account for possible differences between the more than 50 canvassers, please include in your model a fixed effect (i.e., dummy variable) for each canvasser. What is the estimated treatment effect (be sure to mention units)? Is this estimate statistically significant at the 0.05 level? Conduct the same analysis for surveys three months after treatment ($t = 4$) and compare the effects and statistical significance. Provide a substantive interpretation of the results.