Durably reducing transphobia

Background

A study by David Broockman and Joshua Kalla measured how individual prejudice towards trans people can be impacted by an "approximately 10-minute conversation encouraging actively taking the perspectives of others". The researchers sent 56 canvassers to Miami, Florida residences and randomly assigned residents to a treatment where they would receive one of these perspective-taking conversations. More information about the study can be found in their paper, linked below:

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

We examine findings from the first wave of their experiment, where individual responses on their support for anti-discrimination laws for trans people, and feeling thermometers towards trans people, were measured before treatment assignment, and three days after.

The CSV data file, `transphobia_wave1.csv`, contains the following variables.

Name	Description
`age`	Voter file age
`female`	`1` if the voter is female, `0` if the voter is male.
`race`	Voter file race
`democrat`	`1` if the voter is a Democrat, `0` if the voter is not a Democrat.
`treatment`	"treatment" if the respondent was assigned to treatment, and "control" if assigned to placebo.
`law_pre`	Baseline support or opposition to anti-discrimination law, measured from `-3` (opposed) to `3` (in support), measured before any respondents are treated.
`law_post`	Support or opposition to anti-discrimination law, after respondents have been treated.
`law_diff`	Difference between post- and pre- support or opposition to anti-discrimination law.
`therm_pre`	Baseline feeling thermometer towards trans people (0-100), measured before any respondents are treated.
`therm_post`	Feeling thermometer towards trans people, after respondents have been treated.
`therm_diff`	Difference between post- and pre- feeling thermometer measures.

Question 1

Import the first-wave survey data from the experiment, `transphobia.csv`, and assign it to a data frame object named `wave`. Second, estimate the average treatment effect in the survey, save it to an object named `wave_ate`, and print out the results. The resulting object should be a 1x3 tibble, containing three values: treatment and control group average changes in feeling thermometers towards trans people, and the ATE that differentiates the treatment and control group averages. Use a combination of `group_by()`, `summarize()`, `pivot_wider()`, and `mutate()` to produce the desired tibble object.

Interpret the estimate substantively with respect to the experiment.

Question 2

Before creating an entire bootstrap distribution, first create one bootstrap sample from the experiment. Use `rep_slice_sample()` to produce one bootstrap sample, and assign it to an object named `boot_1`. Then use similar coding from the previous question to estimate the ATE from this bootstrap sample, and save the results to an object named `boot_1_ate`.

Print out the results. Is there any difference between the survey ATE and this bootstrap ATE?

Question 3

When a sampling distribution cannot be constructed using population data, a bootstrap distribution using a random sample from the population can serve as an estimate for that sampling distribution.

First: create a bootstrap distribution for the ATE on feeling thermometers towards trans people. This time, use `rep_slice_sample()` to produce 1000 bootstrap samples, instead of just one bootstrap sample. In the same chain of command pipes, calculate the ATE for each bootstrap sample, making sure to group by both the bootstrap sample indicator `replicate` *and* the treatment/control indicator `treatment`. Save the resulting bootstrap distribution of 1000 bootstrap ATE estimates to an object named `boot_distro`.

Second: produce a density histogram of the bootstrap distribution.

Question 4

Use the bootstrap distribution to estimate a 95% confidence interval. Specifically: use `summarize()` and `quantile()` to find the lower (the 0.025 quantile) and the upper (the 0.975 quantile) bounds of the confidence interval, naming the columns containing these values `lower_ci` and `upper_ci`, respectively. Save this confidence interval to `boot_distro_ci`.

Print out your confidence interval and interpret it substantively with respect to the context of the experiment.

> *Hint: pipe your bootstrap distribution into `ungroup(replicate)` before piping into the necessary `summarize()` function and arguments.*

Question 5

Now that we have a confidence interval for our bootstrap distribution, let's add their values to the density histogram of the bootstrap distribution. Add two `geom_vline()` commands to the plotting code used to produce the density histogram in question 3, where the mapping aesthetic 'xintercept' argument is set to the lower and upper bounds of the confidence interval, respectively. Refer to these values directly using vector notation, i.e. `boot_distro_ci$lower_ci` and `boot_distro_ci$upper_ci`.

Question 6 (optional)

Reviewing what we've learned from the last four problems, let's focus on another outcome, the change in support or opposition to anti-discrimination laws for trans people, `law_diff`.

First, estimate the ATE for `law_diff` using only the entire survey sample, and save the results to an object named `wave_ate2`.

Second, create a bootstrap distribution of 1000 bootstrap ATEs for `law_diff`, saving the resulting distribution to an object named `boot_distro2`.

Third, calculate the 95% confidence interval for your bootstrap distribution, and save the results to `boot_distro2_ci`.

Finally, produce a density histogram of this bootstrap distribution, with two vertical lines representing the lower and upper bounds of the distribution's confidence interval.

Interpret the bootstrap distribution's density histogram and confidence intervals with respect to the experiment.