Homework 1

RII Workshop

Austin Cutler

Instructions

I'm expecting to do two homework assignments for this workshop. While the workshop is not graded, you will get the most out of it if you make an honest effort on the homework. That said, I do not expect (or want) this to take an unreasonable amount of time (>10 hours). I do, however, want you to make a genuine attempt at solving each problem on your own. If there are any problems that you are unable to do, then come to the next session with questions!

Question 1

1. Load the USJudgeRatings data, which is built in for R
2. Using a loop, calculate the *minimum* of each variable in the data, storing the minimums in a vector
3. Which variable has the lowest minimum? Which has the highest?

Question 2

For this problem, we're going to use data from an experiment I conducted with Edward Hohe. Please do not share this data with anybody, as this is part of still ongoing research. You can click here to access the data. The data is `pes_cleaned_data.rds`, and is from the FSU module in the inaugural Primary Election Survey, which was run in 2024 (the codebook can be found here, although the variables you'll need for this problem are named obvious things). There are 1603 observations, and 129 variables.

The two variables you'll need for this exercise are `tr` and `out_feel`. `tr` is a treatment with three possible values: *con* if they were in the control group, *policy* if they received a brief definition of polarization during the survey that emphasized policy disagreement, and *partisan* if they received a brief definition of polarization that instead emphasized social identity/partisan animosity. `out_feel` is a feeling thermometer measure of respondent's attitudes towards their out-party (this will be NA for pure independents). Using this data, please do the following:

1. Load the data into R, and remove any unnecessary variables (you can leave in more than just the two mentioned above, but there are a lot of variables that don't need to be in there). You'll also want to remove observations where `out_feel` is NA to make the next steps a little easier.
2. Using a for loop, estimate the mean of `out_feel` for each group.
3. Using 1000 bootstrap iterations, estimate 95% confidence intervals for each group mean.
4. Make a graph showing the group means with their confidence intervals. Based on these results, *substantively* does it appear that either treatment influenced how people feel towards their out-party?

Question 3

For this problem, we're going to practice simulating data and extracting regression coefficients. We have three variables, y, x, and z. X is a dummy variable, and z is a continuous variable that follows a normal distribution with mean 2, standard deviation of 1.5. The value of y is determined by the following population regression model:

$$y \sim 5 + .75x + 3z + \epsilon$$

Where ϵ is normally distributed with a mean of 0. For this problem, do the following:

1. Simulate x from the binomial distribution and z from the normal distribution using the 'r' functions we discussed in class with n = 100
2. Get the values of y based on the regression equation above
3. Combine x, z, and y into a dataframe or tibble, and estimate a linear regression model
4. Compare the estimated regression coefficients to the true population parameters
5. Using the KTW method, estimate 95% confidence intervals. Is the true population value captured by these intervals?
6. Repeat 1.-3. but increase n to be 1000.[1]
7. Create coefficient plots showing the regression coefficients with the 95% confidence interval from the KTW method, broken down by sample size. *Do not use an existing package to make the coefficient plots*, instead extract the coefficients from your model object, and make a dataframe with the variable names, coefficients, upper and lower bound. You can use `ggplot()` or `plot()` for this problem.
8. Discuss how sample size impacts your results

[1] Bonus: create a loop to save the results from the regression models changing the value of n