POL SCI 231B (Spring 2022): Problem Set 2

Prof. Thad Dunning / GSI Clara Bicalho
Dept. of Political Science
University of California, Berkeley

Distributed Friday, February 25; due Sunday, March 7 at 11:59 PM

Please email only one copy of your solution set per group to clara.bicalho@berkeley.edu. Make sure to *include "Problem Set 2 [GROUP NAME]" in the email subject or body field.*

Remember to work out the problems on your own, before you meet with your group to agree on solutions.

Before working problems 1-6, read Chapter 6 of Freedman, Statistical Models (2nd Edition) and peruse the paper by Gibson reprinted at the back of the book.

1. Provide answers for exercises 2-4 in Freedman, Statistical Models (2nd edition), Chapter 6, Exercise set C (p. 90), using the methods of of StatsMods section 6.1 (discussed in Lecture 5). Show your work and explain your reasoning (i.e., why the methods explained in class should work). You should use R to do calculations when appropriate; be sure to print all results at each step. Note: your results may not exactly replicate those in Figure 2 on StatsMods p. 88 or Figure 1 in Gibson's paper. We will return to this later. In this part, simply focus on using and explaining the methods discussed in Lecture 5 to estimate (using the OLS assumptions as necessary) the following:

   Exercise 2 — the path coefficients;

   Exercise 3 — the SD of δ in equation (10) on p. 89 of StatsMods; and

   Exercise 4 — the SEs of the path coefficients and their difference. (Here, please also calculate the t-ratios and comment on which if any are statistically significant).

   (Numbering follows StatsMods Ch. 6, Exercise set C).

2. Show using math and R that for the path coefficients in Exercise 2 in Question 1 above, it does not matter whether we stipulate $n = 26$ states or $n = 36$ states.

3. The notes to Figure 1 in Gibson's paper as reprinted in StatsMods (p. 324) say "Boldfaced entries are bivariate correlation coefficients, with pairwise missing data deletion." Is this right? Explain. Also, explain the difference between "pairwise" and "listwise" missing data deletion.

4. After reading the last paragraph on p. 90 of StatsMods, read https://www.stat.berkeley.edu/users/census/repgibson.pdf and download the Gibson data from Tables 1 and A1, which is saved as a text file at https://www.stat.berkeley.edu/users/census/gibson.txt (read the first link for an explanation of how the data are organized). Then:

   (a) See if you can calculate the three bivariate correlation coefficients (with pairwise missing data deletion) given in Figure 1 in the paper using R;

   (b) See if you can replicate the regression results given at the end of the first document (i.e., the standardized, weighted regression) using R.

   Notes: be sure to filter out observations with "-1" (these are the missing data in the file). Refer to the documentation for lm for the weights. The scale function in the mosaic package might be useful for standardizing.

5. Read sections 6.4 and 6.5 of Statistical Models, then write down a response schedule similar to e.g. equations (14) or (16) for the Gibson study of political repression. Interpret your response schedule carefully. What assumptions are embedded in it? Do any of the assumptions raise theoretical, conceptual, or empirical difficulties?

6. Consider a question we partially discussed in class, as well as two extensions. Explain your answers in your own words.

(a) In your view, which of these two statements is closer to the truth?
   Regression analysis can demonstrate causation;
   Regression analysis assumes causation but can be used to estimate the size of a causal effect—if the assumptions of the regression models are correct.
   Pick one of the two statements and defend your choice in detail. Or if you think neither is correct, argue for why not.

(b) Does your answer change, depending on whether one is analyzing experimental or observational data? Explain.

(c) Does your answer change if we substitute "Analysis under the Neyman potential outcomes model" for "Regression analysis" in (a)? Explain.

Before working problem 7, read Gerber & Green Chapter 10.

7. Now, provide answers for Gerber & Green Ch. 10 Exercise 1 (a)-(e).