PLSC 4782: Final exam

Seamus Wagner

April 27, 2022

Final is open Wednesday April 27, 2022 at 8:00 am EST through Tuesday May 3, 2022 at 3:00 pm EST. As a reminder, senior grades are due at 5:00 pm EST on Tuesday May 3, 2022, so do not push the deadline. Submit your final through the assignments section on Carmen for plagiarism review in addition to the quiz portion for the timer. **You must complete this for your final to be graded!**

Section 1: Conceptual questions

For questions 1-5, answer each prompt in a few sentences. This should be accomplished in two to five sentences and should be thorough enough to address the topic. Explain/answer the question in such a way that someone who did not take this course can understand. hint (This typically means that pulling language from slides is not sufficient since they contain jargon and are not designed for someone who has been taking the course.)

Q1

Explain what censoring is and why it is an issue with duration data?

Q2

Give one example (that wasn't brought up in class) of a research design where you would want to use either fixed effects **or** random effects.

Q3

What is a counterfactual, why is it useful for causal inference?

Q4

What is cross-validation? Why would a researcher use it?

Q5

What makes a differences in differences design a good option? When would a researcher use one?

Section 2: Analytic portion

In this section, answer the question in a brief statement with any relevant tables or figures included. Be sure to include all of your code somewhere in your submission for this assignment. This section is not intended to evaluate how quickly you can code, I will provide data and you should have code to fully answer each question. For instance, if I give you a categorical outcome in the response, your code if probably either in your answer to the corresponding problem set or provided by me in the lab for that week. I will also identify all packages you will need to answer the problem. I will say tidyverse for all of them solely for ggplot2, if you prefer base R plots that is totally fine. Hint (the only updates to code should be the names of variables or possibly data sets or very brief mutations (like make a given variable binary.))

Q6

Download the data “finalsdata.csv" and load it. You will need the “nnet" package and a package for outputting a table (texreg,stargazer,etc). These data are originally from the General Social Survey in 2012. I have made modifications to the data to make them accessible and to change some of the codings so make sure to use the file provided. This also means these are not *real* so any inferences are for class purposes and you should not take these substantively seriously.
Evaluate the effect of the predictor list on votes for Obama versus McCain using a multinomial model. Interpret coefficients and standard errors of the model. Vote fro Obama and McCain is your DV, the rest are IV. Feel free to make any transformations you think re necessary (log, as.factor, etc).

- Income = numeric - income in USD of respondents
- wordsum = integer -Number of words correct in vocabulary test
- wrkstat = character - Labor force status
- age = integer - Age in years

- authoritarianism = integer - scale of authoritarian attitudes, higher = more authoritarian attitudes
- con_gvt = integer - confidence in government overall
- conarmy = character -confidence in military
- degree = character - respondent's highest degree
- pres08 = Vote for Obama or McCain

Q7

Run another model but exclude income and authoritarianism. What does this exclusion do to the model fit, how do they compare along AIC/BIC/Deviance/Log-Likelihood fits. (Hint: texreg/htmlreg/screenreg will display all of the needed values by default)