GSERM 2022 – Regression for Publishing

Final Examination

Instructions

1. This examination has four questions, listed at the end of this document. **You are to choose two questions and answer them.** *Please do not attempt to answer all four questions.*

2. Each of your answers should take the form of a brief empirical analysis of the data provided that answers the question(s) asked. That may include plots, tables, and any other techniques you think will be useful in answering the question. Each answer may be as short or long as you feel is appropriate, but probably need be no longer than 4-5 pages, including text, tables, and graphs. Answers will be evaluated on the basis of (a) correct application and discussion of the methods discussed in class; (b) justification of the analytic choices made; and (c) presentation and discussion of findings. *In creating your answers, you should not use data outside of that provided in the dataset.* You will *not* be evaluated on your substantive knowledge of the topic(s) at hand.

3. The exam is worth 700 possible points (350 possible points for each question answered).

4. You should submit your exam no later than 11:59 p.m. CET time on Friday, June 24, 2022. Please submit your exam to the instructor electronically, *as a PDF file*, by e-mailing it to `zorn@psu.edu`. You should also include all code necessary to replicate the analyses you conduct; this can take the form of an appendix, a separate `.R` file, or a `.Rmd` file for the entire exam answer.

Data

All four final exam questions will make use of the same data. The data are drawn from Washington University's American Panel Study (TAPS), which collected data on a panel of respondents over several monthly waves from 2012-2017. (Here, we will be looking only at single-response questions, so the data are cross-sectional.) The data are available on the course github repository, in the "Exam" folder, in a file named `GSERM-RFP-2022-Exam-Data.csv`. In addition to a respondent identifier variable (`WUSTLID`), the pool of independent variables in those data are:

- Political party identification indicators – binary variables for `Democrat` (Democratic Party, the center-left party) and `GOP` (Republican Party, the right-wing party), with independents serving as the reference category;
- `Ideology` – a seven-point Likert-type indicator variable, where higher values indicate greater political conservatism (right-wing) and lower values indicating greater progressivism (left-wing);
- `Education` – measured as a twelve-category ordinal variable with values ranging from 3 to 15, where the lowest value corresponds to a 5th-6th grade level of education and the highest reflects a doctoral degree;
- `Income` – a 15-category ordinal variable, where higher values indicate higher income levels (where each unit roughly corresponds to an increase of $10,000 in annual income);
- The respondent's `Age` in years, as of 2016;
- `Female` – a binary indicator of sex, naturally-coded; and
- Racial classifications – binary indicator variables for `White`, `Black`, and `Asian` identification (with "other" as the reference category),

The data also contain a set of survey `weights` that reflect the sampling scheme.

Questions

1. The variable `FT.Communists` reflects the respondent's placement of "communists" on a 0-100 "feeling thermometer" scale. Analyze and discuss how do respondent's feelings toward communists vary according to their demographic and political traits.

2. The variable `InterviewDuration` records the number of minutes that each respondent took to complete the (on-line) survey. Analyze and describe if and how respondents' demographic characteristics are related to their survey times, with particular attention to questions of linearity and outliers.[1]

3. The variable `HowManyKids` indicates the number of children (under the age of 18) that each respondent has in their household. Fit and discuss a model that assess the association between this variable and the respondent's demographic and political traits.

4. The TAPS survey also asked a series of yes-no / binary-response questions related to respondents' specific behaviors and preferences, including:

   - Have you ever taken the shampoo and conditioner bottles from a hotel or motel? (`StealShampoo`; 0=no, 1=yes)
   - During the past year, have you ever run out of gas while driving a car or other vehicle? (`RunOutOfGas`; 0=no, 1=yes)
   - Have you ever looked directly at the sun to see an eclipse without using a filter? (`LookedAtEclipse`; 0=no, 1=yes)
   - Have you ever stolen a street sign? (`StolenStreetSign`; 0=no, 1=yes)
   - Would you rather be attacked by a big bear or a swarm of bees? (`BeesOrBear`; 0=bees, 1=bear)

   For this question, choose <u>one</u> of these binary dependent variables, and analyze and discuss its association with respondents' demographic and political characteristics.

[1] Note that it is *not* necessary to conduct a survival analysis here; linear regression is preferred in this case.