Portfolio 4
PA 606: Seminar in Quantitative Techniques
(Spring 2020)

25 points

Due: Saturday, May 9, 2020

Topics:

Multivariate (Multivariable) Relationships – Analyzing the Effect of Predictors on the Outcome (Analyzing Relationships while Controlling for Spuriousness):
Ordinary Least Squares (OLS) Regression

Data Sources:

`GSS2014` (2014 General Social Survey data set located in the `vannstats` package)

- variable names and descriptions, as well as the codebook (linked), can also be found in the `vannstats` package information

Overview:

In this assignment, you'll be examining the strength of the relationship between three or more variables. Specifically, variation in one variable (our key independent variable) is related to variation in our outcome variable (our dependent variable), while controlling for other factors (additional independent variables). Another way of phrasing this is "analyzing the effect of predictors – while controlling for others – on an outcome".

Answer each of the questions below, **in full sentences/paragraphs**, and show your tables/plots (if necessary) inline (interspersed within the text). For full credit, you must append (copy and paste) your R Script at the end of this portfolio as the final page(s).

Problems

Data Set Information

- Describe the data set, including the name and who administered the data/survey (e.g. which survey research firm), year administered, the unit of analysis, and the number of observations.

OLS Regression

In this test, you'll be trying to explain or predict someone's beliefs about the **functioning of U.S. democracy in the future** (`dem10fut`), based on two other attitudinal measures as predictors: 1) their **confidence**

U.S. in financial institutions (`confinan`), and their perceptions of how **widespread corruption is in the U.S.** (`corruptn`).

- Describe the variables. Using the codebook and the list of variable descriptions for the `GSS2014` data set, for each variable, describe its text (e.g. the question asked for each variable), the level of measurement, and the values/categories within the variable.
- Describe the meaning of the coding of each variable, in terms of what higher (or lower) scores mean.
- Define a research question for the variables (e.g. "*Is variation in X related to/associated with variation in Y; Is there a mean difference in Y by categories of X*").
- Define the null hypothesis (H_0) and the alternative hypothesis (H_1) for this test.
- Describe the various assumptions of OLS Regression and how you would assess them. Next, show all necessary tables/plots that demonstrate your assessment of whether or not you've met the assumptions of the test. (*Note: if showing tables, create them, do not simply copy and paste from your output.*)
- Run the analysis. Fully and correctly report the test. If you find significance, describe the significant variables, whether the overall model is significant, and the model fit (R^2).

Extra Credit: Building out the Regression Model

- Select one additional variable to include in the model, and re-run the OLS model with the additional variable included.
  - *Note: Given that OLS (and most other tests) deal with missingness using listwise deletion, some variables may not yield results. In this case, you will get an error, and you should select another variable.*

  Describe how the models changed – Discuss changes (in sign and magnitude) for the previously included variables (`confinan` and `corruptn`) due to the inclusion of your third predictor (X_3), as well changes in the omnibus/overall F, and in the R^2.