Problem Set 3: Correlation and Regression

1 Purpose

The purpose of this problem set is to assess your understanding of one key method of quantitative public opinion research: cross-sectional and panel regression analysis.

2 Your Task

3 Overview

You should review the methodological material from Week 3, then apply the technique of regression analysis to empirical data set (the European Social Survey).

1. The structure of this problem set is slightly different from Problem Sets 1 and 2 in that it is a "lab" activity to give you a sense of how to conduct basic quantitative analysis in Stata or R, how to understand the mechanics of multivariate regression analysis, and how to interpret the results.

2. Download the European Social Survey (ESS) Round 7 (2014) dataset.[1] You will need to register an email address to do so. Alternatively, a copy of the dataset in Stata format has been provided on Moodle. Also download the documentation (again, a copy is provided on Moodle). Load the data into Stata or R. Your choice.

3. From the documentation and/or the data itself, identify an attitudinal variable that you would like to consider as an outcome or dependent variable to be explained. Then, identify one particularly important variable that you think might explain that outcome. Discuss why you think these two variables might be related. Calculate the correlation coefficient for the relationship and report it in your paper. Is the correlation large or small?

4. Then, without reference to the data or codebook, identify 3-4 (or more) other factors that you think might explain this outcome and/or your key explanatory variable. Draw a causal graph ("directed acyclic graph") showing the possible causal relationships between these variables. Include this in your paper.

5. With reference to your graph, identify variables in the dataset that might operationalize the potential explanatory variables you have identified in your graph. In natural language, describe each of the variables from your graph, why you think they might be causally relevant, and describe what - if any - variables in the ESS dataset can be used to operationalize the constructs. Discuss any weaknesses or limitations in the way these variables are measured.

6. If there are any variables in your graph that are not available in the dataset (and there should be some), describe what relationship you think they might have on your outcome and/or other explanatory variables and discuss what problems this may introduce into your analysis.

7. Estimate a regression equation based upon your graph, with the outcome variable regressed on the explanatory variables. Include the output of the regression model in your paper (you may want to use a `fixed width` font to do so; or properly format the output as a table). Interpret the results. Are they substantively significant? Are they statistically significant? Is it reasonable to interpret the effect of any or all of the explanatory variables as causal? Why or why not?

---

[1] From `http://www.europeansocialsurvey.org/data/download.html?r=7`

4 Submission Instructions

Please submit your answers as a PDF document via Moodle. It should be no more than 4 pages, single-spaced, in Times New Roman font size 12, on A4 paper with standard 2.54cm margins. This problem set is self-assessed. A solution set will be provided on the course website and the activity will be discussed in class.

5 Feedback

Group feedback will be provided during class. If you would like more specific individual feedback on your work, please ask the instructor during office hours.