

POLI 502 FA20: Homework 6

Due at 6 PM on Wednesday, December 2, 2020

Directions. Answer each question completely. Submit complete replication files so I can change your `setwd()` code and then reproduce all your output. Also include the completed assignment as a \LaTeX -typeset .pdf file.

Exercise 1

Load the latest Quality of Governance standard time series data.

A. Create a new data frame with the year as the unit of observation. It should include variables for:

1. the number of democracies in the system (using the Polity combined score and defining democracies as states scoring ≥ 7)
2. the average GDP per capita (use WDI's real GDP per capita variable)
3. the number of inter-state conflicts: sum the number of inter-state armed conflicts per country across all countries (using the UCDP PRIO indicator)

B. Using `lm()`, regress the number of state conflicts (the DV) on the number of democracies and the average real GDP per capita (the EVs).

Note: in your own work, you generally will not want to use `lm()` when your DV is a count. But we are overlooking that for this exercise.

C. Interpret the output from part B. What does each coefficient indicate regarding the association between each EV and the DV? What does the intercept indicate (and is it meaningful)? What do significance tests for each coefficient indicate? What do the R^2 and F test indicate?

D. Create a prediction plot to illustrate the relationship between the number of democracies and the number of interstate conflicts. Vary the count of democracies from its first quartile to its third quartile on the x-axis. Include a 95% confidence interval for the prediction line.

Exercise 2

Load the ANES 2012 time series data.

A. Create a variable equal to 1 if the respondent voted for Barack Obama in 2012, and equal to 0 otherwise (using the `prepost` summary variable).

B. Using the `ineq_incgap_x` variable, code a factor variable with three values: 1) the respondent thinks the income gap has become larger (somewhat or much); 2) the respondent thinks the income gap has become smaller (somewhat or much); 3) the respondent thinks the gap has remained about the same. Code "don't know" and "refused" as missing.

C. Using `lm()`, regress the vote variable from part A on the factor variable from part B, using "about the same" as the reference category, and on the factor variable for race where "white" is the reference category

(code "refused" etc. as missing for race).

Note: be sure the data are representative of the US population. The simplest way to do so is to omit the oversampled observations. However, you can earn bonus points by correctly implementing survey weights (which `lm()` cannot do).

Hint: you can use the `relevel()` function to set the reference category if necessary.

Hint 2: if you want to grapple with survey weights, check out the `survey` package.

D. Interpret the regression from part C (do everything I asked in exercise 1-C). Be sure to interpret dichotomous variables with respect to reference categories.

E. Examine the variance of residuals over the range of the predicted values. Does the model meet the assumption of homoskedasticity?

F. Redo the regression from part C. using `glm()` with a logit link function. Interpret the coefficients.

Again, be sure the data are representative of the US population. You can earn bonus points by correctly implementing survey weights (which `glm()` also cannot do).

From the Text

Diez et al.: Chapter 8 Exercise 8.44; Chapter 9 Exercise 9.16.