POLI 502 FA20: Homework 5
Due at 6 PM on November 11, 2020

Directions. Answer each question completely. Submit complete replication materials (R script and necessary raw data) 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

A. Using the latest Quality of Governance time series data (easy to find online), aggregate as appropriate (e.g., using summarize within tidyverse-dplyr) to create a dataset at the region-year level of analysis (use the ht_region and year identifiers). It should record for each region, each year, the number of states experiencing inter-state conflicts (use the UCDP-PRIO inter-state armed conflict variable). However, rather than a simple count, code the armed conflict variable in three categories: (1) 0 states experiencing inter-state armed conflict, (2) 1 state experiencing inter-state armed conflict, (3) 2 or more states experiencing inter-state armed conflict. Hint: pay attention to the ucdp variable coding.

B. At what level of measurement is the armed conflict variable? Describe exactly what it is measuring. For example, in several observations, the conflict variable records that exactly one state experiences inter-state conflict. How is this possible given that at least two states must fight in order for an interstate conflict to exist?

C. If we want to know whether the *number of conflicts* (where, for example, the 2003 Iraq War would be one) varies by region, what might be problematic with the conflict variable that you coded?

D. Create a crosstab for region and the armed conflict variable. Conduct a chi-squared (χ^2) test, acting as though we meet assumptions regarding independent observations (though we likely do not).

E. Explain the result of the χ^2 test. What is the null hypothesis, and do you reject it? What are the limitations regarding what the chi-squared test can tell us about the dependence between these two variables?

Exercise 2

Open the ANES 2012 time series data, saving it as an R object called anes (available in the assignments folder on Blackboard). You will use these data for Exercise 2.

A. Use the party ID variable (pid_x) to code a dichotomous variable equal to 1 for Democrats (including strong D, weak D, and independent leaning D) and 0 for Republicans (equivalent coding). Exclude true independents and the various missing values (refused/no answer/etc.) from the new measure.

B. Recode the education variable (dem_edu) as a numerical variable, removing all missing values and the "other" category. You should be left with an indicator that varies from 1 to 16.

C. Conduct a difference of means test to examine whether mean education differs for Democrats and Republicans. Do you reject the null hypothesis? Does the substantive difference in mean education between parties seem important?

D. What is the level of measurement for the education variable you created? Could this level of measurement potentially cause problems in the test described in part C?

E. Conduct an ANOVA to examine whether mean education varies by party identification, this time using the full 7-point scale (pid_x again, but omitting the various missing values).

F. What do the results from the ANOVA tell us? What is the null hypothesis, and do you reject it? What are the limitations regarding what ANOVA can tell us about the difference between groups?

From the Text

Note: Do all math in R and submit the (annotated) code. Show all your work!

Diez et al.: Chapter 6 Exercise 6.32; Chapter 7 Exercises 7.30 and 7.44