

PLSC 502: “Statistical Methods for Political Research”

Exercise A

The purpose of this exercise is to “get your hands dirty” measuring, collecting, coding, and assembling data. In particular, you’ll be building four semi-original datasets, each with somewhat different characteristics. This exercise has four parts; specific instructions are below.

I. Cross-Sectional Data

1. Build a database where the unit of analysis is the faculty office – specifically, offices held by tenured or tenure-track political science faculty at Penn State (that is, in Pond Lab).¹
2. Operationalize, measure, and code four variables:
 - Whether the office in question is on the first, second, or third floor of Pond Lab,
 - Whether the faculty member in that office is teaching or tenure-line faculty,
 - The rank (assistant, associate, or full professor) of each office’s inhabitant, and
 - The gender self-identification (male or female) of the office’s occupant.

Obtain this information from any available sources. Be sure to make your decisions about operationalization, etc. as transparent as possible, and document your sources.

II. Time Series Data

1. Count the number of letters in your first name (k), and pick the k th state from the [alphabetical list of U.S. states](#) (so, if your name was “Christopher,” you’d select the 11th state, which happens to be Georgia).
2. Go [here](#), grab the Excel file at the bottom of the page, and use the information in that file to create a time-series dataset of your state’s values on the following variables:
 - Population, 2000-2010,
 - Annual SNAP participant counts, 1997-2012, and
 - Annual SNAP participant benefits, 1969-2012.

Note that the documentation for the spreadsheet is [here](#).

III. Time-Series Cross-Sectional (“Panel”) Data

1. Return to the state you used in the time-series exercise, above, and build a panel database of “your” state’s current members of Congress. Record one observation for each such member for each Congressional session (2-year term) they have been in office, as of 2022.
2. For each such member in each session, collect and code data on:
 - that member’s political party identification in that term,
 - the number of terms s/he has served, and
 - the party of the state’s sitting governor in that term/session.

Once again, obtain this information from any sources you find available, and note your sources in the write-up of the exercise.

¹Note that some of the faculty currently have temporary offices in other buildings; please endeavor to assign faculty’s “permanent” offices to them as best you can.

IV. Relational Data

1. Go [here](#), roll the die once, and add 2000; that's your year.
2. Navigate to the website of the amazing [Correlates of State Policy](#) project. Use the information / data there to build a relational dataset of the 50 U.S. states. In this dataset, each row of the data should be one dyad / pair of states. Thus, your dataset should have a total of $\frac{50 \times 49}{2} = 1225$ rows.
3. In your dataset, code / include the following data; data should be from the year indicated by step (1):
 - Whether (=1) or not (=0) that pair of states both have the same value on the *Income Taxes* variable,
 - Whether (=1) or not (=0) that pair of states both have the same value on the *Electronic Waste Recycling Program* variable, and
 - The absolute value of the difference between the two states' values on the *Top 1% Share of Income* indicator.

Note that step (3) *can* be done “by hand,” but you’re probably better off writing code to do it. HINT: In the latter case, the [combn](#) command may be of some value, although there are other ways to tackle the problem too.

General Matters

1. For each part listed above, the data you create should take the form of comma-separated value files (`.csv` files), *not* Excel spreadsheets.
2. Be sure to include unit identifiers in each dataset.
3. In addition to the datasets, include in your email a PDF file briefly outlining what you did in each of the four sections above. In the latter, include (a) the coding scheme for each variable you coded, (b) its level of measurement, and (c) the number of non-missing observations you collected/assembled data on.
4. Submit all four datasets in electronic format – via e-mail attachment – to Nathan (nam@psu.edu) *and* to me (zorn@psu.edu) *before the end of the Fall 2023 term*.
5. This is an ungraded (but required) exercise.