Quantitative Data Set Final Project

Highlighted sections are new and did not appear on the draft assignment

1. Overview

- Summary: In this assignment you'll build on the quantitative data set you created earlier—begin with what you have there—to produce an expanded data set and some results that could provide insights for your project.
- Due date: The assignment is due via email by Fri., November 21 at 5:00pm. Late assignments will incur a score penalty for each day, including weekend days, they are late.
- Turning in: Send a single email message with the 4 files due to Prof. Manna (pmanna@wm.edu). PLEASE have these files be actual objects attached to an email, NOT links in an email message to OneDrive / SharePoint files. It is MUCH EASIER for me to manage all these files when they are attached objects.
- Weight: The assignment is worth 10% of your semester grade.

2. Directions

2.1 Artificial intelligence guidance. AI tools are only permitted on this assignment for 1) improving prose that you have already written; and 2) extracting data from your sources as you get them into Stata format. The codebook file (see section 2.4) asks that you make an AI disclosure statement.

2.2 Final Stata data set. Building on the data set you created earlier in the semester, compile a revised final data set in Stata format (`.dta`) that could contribute to your research project. The data set should contain these elements.

- File name is YourLastNameFinal.dta
- At least 10 variables.
  --Have 1 variable (or 2, if needed) that is a case ID variable making clear what the unit of analysis is for your data set (i.e., survey respondent ID; year; country with year as an example of needing two variables; etc.).
  --At least 2 of your variables should be substantively meaningful variables for analysis that you created using the `generate` and/or `recode` command in Stata.
- All variables in the data set should include
  --variable labels
  --value labels for numeric variables that are nominal or ordinal (not including case IDs).
- There should be a consistent unit of analysis across all variables, meaning the rows in the data set should all be the same type of unit. For example, each row should be an individual person if you're working with survey data; or a state if you have state-level data; and so on. Don't combine different units of analysis into a single data set.
- All variables should be properly encoded and formatted so one can obtain meaningful results when running appropriate Stata analysis commands that we're learning in class.

2.3 Final Stata `.do` file. Include a proper `.do` file, accompanying your `.dta` file, with these elements.

- File name is YourLastNameFinal.do
- An abstract at the top that follows the format of the example we discussed in class on 9/23. (Feel free to include additional items in the abstract if you'd like to provide more detail.)
- Code showing how you created 2 variables using the `generate` and/or `recode` command.
- Code to attach variable labels to all variables in the data set.*
- Code to attach value labels to values for nominal or ordinal variables.*
- Code to run appropriate commands to produce two types of descriptive results (see Module 2) that could provide insights for your project. You will include one of these sets of results in an analysis writeup file (see Section 2.5).
  --A set of numeric results (summary statistics, cross-tabulation, regression, etc.).
  --A data visualization (box plot, scatter plot, histogram, etc.).

*Note: If your file already contains these because your original source attached them then simply include a comment note in the `.do` file saying that. Importantly, though, if you are using survey data, you likely will need to recode some of the values to address the presence of missing value codes. I'll post additional guidance on that issue.

2.4 Final codebook file. This file provides written documentation so readers can understand the contents of the data set before beginning to analyze the data. Set up the codebook with these sections and elements.

- Save the file as a Word document.
- File name is YourLastNameCodebookFinal.docx.
- Include page numbers.
- Overview section: Include these elements:
  --Your name
  --Date
  --Your GOVT 301 section
  --AI disclosure statement: Describe how you used AI, following the guidance in section 2.1, or attest that you did not use AI to develop the data set or elements of the codebook.
  --A brief, clearly written statement (approximately 2-4 sentences) that describes the data set. There should be enough detail that a person would generally understand what to expect when they open the data set to use it.
  --Optional: Any additional notes you think would be useful to clarify at the start.
- Data sources section: Provide documentation of the original sources and links to the places where you found the raw data. Also, if you went to a website and had to do certain steps to extract the data from the site (i.e., run some sort of query or a set of queries), describe those steps here. Basically, you want to provide details so a person could retrace all your steps and find the original data you found.
- Variables section: List all the variables one by one, including the ones you created with `generate` and/or `recode`, and for each one include:
  --variable name
  --variable label
  --basic descriptive statistics (for interval and ratio variables) or tabular summaries (for nominal and ordinal variables) for each variable.

2.5 Analysis writeup file. Create a file with these elements that briefly analyzes results from your data.

- Save the file as a Word document.
- File name is YourLastNameAnalysis.docx.
- Include your name and GOVT 301 section atop the file.
- Paste in your project research question, phrased as a question, after your name and class section.
- Paste in either the set of numeric results or the data visualization you created from your `.do` file. Make sure they appear clearly, which might require you to adjust font type, size, and formatting. Provide a clear and appropriate substantive title (i.e., Table 1. Table title or Figure 1 Figure title).
- Compose a first paragraph where you report on some aspect of your results using the method for writing about results that we learned in class on Tuesday, October 7.
- Compose a second paragraph where you explain how the results you just described could assist in answering your research question or could provide related information to help readers understand the larger context.

3. Grading

1. Nuts and bolts. You attach all four files in one email to Prof. Manna with the files named as the directions describe.

2. Stata data set. The Stata data set opens, can be run without error, and contains all elements from the directions.

3. Stata `.do` file. The file opens, runs without error, and contains all elements from the directions.

4. Codebook file. The codebook contains the three overall sections and all the specific elements from the directions. The writing is clear in the narrative sections and follows the standards for excellent writing that have been part of our previous writing assignments this semester. The formatting for the information accompanying each variable is clear and polished.

5. Analysis writeup file. The file contains the elements from the directions. The writing is clear and follows the expectations for excellent writing that have been part of our previous writing assignments this semester.