Stata Project #1:
Descriptive Statistics

Due Date: 11:55 p.m. on February 9, 2017 (via Sakai)

Introduction and Directions

The purpose of these Stata projects is to give you the opportunity to apply your developing knowledge of Stata to a small-scale research project of your own design. There will be three such projects over the course of the semester, each one building on the last.

The goal of this first project is for you to gain experience navigating statistical codebooks, finding variables of interest to you, and then generating appropriate measures of central tendency and dispersion in the Stata environment. Please following the directions in each of the sections below and answer the associated prompts. Once you are finished, upload your answers to the three prompts through Sakai in a Word document labeled "LASTNAME_project1.docx." In addition to the Word document, you must also upload the .do file you used to generate your descriptive statistics. You will want to make frequent reference to your "Stata Handout #1" document for details on how to set up and navigate your Stata session.

The dataset is located in the "Project #1" sub-folder in Sakai, which is nested within the "Project Materials" folder.

Developing a Research Question

We will use a subset of the 2012 General Social Survey dataset. Take a look at Appendix G in your textbook (pp. 417 to 424 in the 4th edition) to familiarize yourself with the variables.

Once you have explored the data a bit, identify one or two "dependent variables" that interest you and another eight or nine "independent variables" that interest you. We haven't talked about the distinction between dependent and independent variables yet, but for now, just think of the dependent variables as "outcomes" and the independent variables as "causes." So, if your independent variable was "hours of TV watched per day" and your dependent variable was "mean number of hours of sleep per night," then you would be suggesting that you think a person's TV consumption might, in some way, affect the number of hours they sleep (but not vice versa). As such, you will want to make sure you pick independent variables that you realistically think might have some sort of effect on the dependent variable(s) you choose. You can also treat your dependent and/or independent variables as "indicators" of other concepts: for example, perhaps you want to use measures of general happiness, marital happiness, and job satisfaction as indicators of life satisfaction.

Finally, your dependent variables **must** be interval-ratio variables (we'll find out why later in the semester). The independent variables can be a mix of interval-ratio, ordinal, and nominal levels of measurement. By the time you are finished exploring the codebook, you should have a total of ten variables (one or two dependent and eight or nine independent). In total, you must have five interval-ratio variables and another five consisting of a mix of nominal and ordinal variables (with at least one of each).

Develop one or two research questions based on the hypotheses you have about how the independent variables might be related to the dependent variables. You may want to refer back to Wallace's "Wheel of Science" from our first class lecture for more details on how to do this.

Prompt 1: Provide one or two sentences describing your research question(s).

Generating Descriptive Statistics

Use your Stata Handout #1 to explore some measures of central tendency and dispersion for your variables.

Prompt 2: Fill in the table below with your descriptive statistics. (Note that the Range, IQR, and Standard Deviation columns will not be appropriate for some of your variables given their levels of measurement. If they are not appropriate, put "NA" in the box.)

	Variable Name	Level of Measurement	Central Tendency	Range	IQR	Standard Deviation
1		Nominal				
2		Ordinal				
3						
4						
5						
6		Interval				
7		Interval				
8		Interval				
9		Interval				
10		Interval				

Interpretation

Use your results to try to get a grasp on what the "typical case" might look like according to these data.

Prompt 3: For each of your ten variables, write a sentence or two describing the "typical" American adult—as indicated by your measure of central tendency—and also how much variation there is around this measure (as indicated by your measure of dispersion).