Stata Project #2:
Inferential Statistics and Measures of Association

Due Date: 11:55 p.m. on April 4, 2017 (via Sakai)

Introduction and Directions

The goal of this second project is for you to gain experience testing your own research questions—specifically those that you developed in the first Stata Project.

Please follow the directions in each of the four sections below and answer the associated prompts. Once you are finished, upload your answers to the four prompts through Sakai in a Word document labeled "LASTNAME_project2.docx." In addition to the Word document, you must also upload the .do file you used to run your hypothesis tests and measures of association. You will want to make frequent reference to your "Stata Handout #1" and "Stata Handout #2" documents for details on how to set up and navigate your Stata session and perform hypothesis tests. The Stata code for the measures of association are scattered throughout the PowerPoint slides, as well as homework assignments #5 and #6.

You will use the same dataset from the previous project (i.e., a selection from the 2012 General Social Survey). The dataset is located in both the "Project #1" and "Project #2" sub-folders in Sakai, which are nested within the "Project Materials" folder. The codebook is also available in each of these folders if you need to refresh your memory on what these variables are measuring.

Restating Your Research Question(s)

To help you get your head back into the material, please restate your research question(s) from the first Stata project.

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

Testing Hypotheses

Use Stata Handout #2, the Stata questions on your homework assignments, the slides, and your textbook to run some hypothesis tests. The dependent variables in your research questions should be your dependent variables in your tests.

Up to this point, we have only talked about hypothesis tests where the independent variable is either dichotomous, multinomial (i.e., a nominal variable with more than two categories), or ordinal (if we ignore the rank-ordering of the variable and treat it as purely categorical). As such, your hypothesis tests will need to be specific to your nominal and ordinal independent variables (we'll come back to your interval-ratio variables in hypothesis testing later in the semester).

I leave it to your discretion to decide on how many hypothesis tests you need to run. However, I expect you to run the number of hypothesis tests necessary to fully assess your research questions (again, with the exception of your interval-ratio independent variables).

***Note*: From now on, if you are running a one- or two-sample tests with an interval-ratio variable in Stata, I want you to only use a *t*-test—not a *z*-test. In other words, if you find yourself wanting to use the `-ztest-` command, just go ahead and use the `-ttest-` command instead. Recent releases of Stata (at least for Mac) have started requiring the dependent variables issued to the `-ztest-` command to be dichotomous; so, depending on type of machine and the Stata version you are using, the `-ztest-` command may not give you the correct result. Remember, though, that the *t*-distribution is asymptotically normal, so the `-ttest-` command will give you an unbiased result regardless of your sample size.**

> **Prompt 2: Formulate, execute, and interpret a series of hypothesis tests based on your research questions. Be sure to follow the "five-step" model for each of your tests.**

Measures of Association

Use some measures of association to assess the substantive significance of your *statistically significant* test results. (Friendly bit of advice: since your dependent variables are measured at the interval-ratio level, only one of the measures of association we have talked about will be useful.)

If you feel the need to "break out" your multinomial independent variables so that you can make contrasts like you would with a dichotomous variable, you can do the following:

```
gen new_var_name = (old_var_name==#)
```

Where `new_var_name` is the name you give to the new dichotomous variable you are generating, `old_var_name` is the name of the original multinomial variable, and # is the number referring to the category of interest in `old_var_name`. This will create a dichotomous variable that is "1" if an observation is coded # and "0" if not. For example, say I have a religious affiliation variable called "relig" that is coded as follows: 1 = Protestant, 2 = Catholic, 3 = Jewish, 4 = Muslim, 5 = Hindu, 6 = Buddhist, and 7 = Other. Now say I wanted to make a dichotomous variable called "protestant," where an observation gets a "1" if they are coded as Protestant and a "0" if otherwise. To make this dichotomous variable (also called a "dummy variable"), I just have to do the following:

```
gen protestant = (relig==1)
```

Similarly, if I wanted a dichotomous variable that was "1" if Catholic and "0" if otherwise, I would do this:

```
gen catholic = (relig==2)
```

And so on.

> **Prompt 3: For each of your statistically significant tests, use a measure of association to assess the substantive significance of the relationship. Write a sentence interpreting the strength and/or direction of the relationship.**

Communication

Finally, let's exercise some of our communication skills.

> **Prompt 4: Now assume that you are asked to communicate these findings to a layperson. Take a step back from the statistical lingo and tell this person what they can expect to observe in the population.**