SOCY7702 Statistics **Instructor: Natasha Sarkisian**

Final Paper Project

Due dates:

1. The proposal and data management draft is due on November 27 (by 11:59pm) by email.
2. The final paper is due on December 19 (by 11:59pm) by email.

For this paper, you will use GSS 2016 data to conduct your own analysis (dataset is now available on the course website). You should devise a research question (or questions) that can be answered using 4-6 variables available in the dataset. This research question (or questions) should be about group differences, relationships among variables, or effects of one variable on another.

Proposal (due November 27)

1. Introduce your research question; explain why you think it is important. You do not need to do a literature review for this, just discuss this on the basis of your knowledge, or even "common sense"). State your null and research hypotheses; justify your choice of the research hypotheses (directional vs nondirectional, and if directional, justify the choice of direction – again, no serious literature review needed though!).

2. List variables that you plan to use and explain why they can be used to measure the concepts of interest in your research question. For each variable, include codebook information from GSS codebook: http://gss.norc.org/documents/codebook/GSS_Codebook.pdf

3. If you plan to limit your sample and use a subsample rather than entire GSS sample, discuss how you plan to limit it and what variable or variables will be used in that process and how. Restrict all the following steps to your subsample (you will need to use "if" statements to restrict your sample; e.g.: tab educ if sex==2).

4. Include codebook information for each variable (using codebook command). State the level of measurement (nominal, ordinal, interval, ratio) for each variable.

5. For nominal and ordinal variables, generate frequency tables and bar charts. For interval/ratio variables, provide measures of central tendency and dispersion and get histograms and boxplots.

6. For each variable, discuss whether some data management will be required to prepare that variable for analysis. Some questions to consider:
   A. Are there missing data codes that are currently numeric codes but will need to be changed to missing?
   B. Are you planning only to use certain categories but not others and you would need to exclude them?

C. For nominal/ordinal variables, do some categories have very small frequencies and would therefore need to be combined with other categories? If so, which ones and how can you justify combining them from the substantive standpoint?
D. For interval/ratio variables, based on the boxplot, are there some extreme outliers? If there are, you may consider excluding them from analysis (for example, replace with missing). I would not recommend excluding more than 1% of your observations, however.

7. Explain what analytic technique you will use for each pair of variables involved and why. Please see the following chart for the recommended techniques based on the level of measurement of your two variables.

Variable 1 / **Variable 2**	**Nominal/ordinal**	**Interval/ratio**
Nominal/ordinal	• Chi-square	• Two samples t-test • ANOVA (if >2 groups)
Interval/ratio	• Two samples t-test • ANOVA (if >2 groups)	• Correlation • Regression

8. Write a draft do-file for these data management steps (please consult Chapters 2 and 3 of Longest book on data management commands, as well as course handouts). Open a log file and then run your do-file. You will need to submit your do-file, your log file, and all the graphs with your proposal.

Make sure to include all the eight steps in your draft. I will provide feedback on this draft; you will then work on your final paper.

Final Paper (due December 19)

The final paper should be 5-10 pages double spaced (without the appendix), it should be written in a journal format – that is, as a mini-version of a journal article. To see an example of an article using GSS to see how they describe the data and explain their variables, as well as present the results, you can look at Maselko, Joanna, and Laura D. Kubzansky. 2006. "Gender differences in religious practices, spiritual experiences and health: Results from the US General Social Survey." Social Science & Medicine, volume 62, issue 11, pages 2848-2860. https://www.sciencedirect.com/science/article/pii/S0277953605005915

Your final paper should include the following:

Introduction (1-2 pages): Description of your research question and its importance (again, no need for actual literature review); explanation and justification for your null and research hypotheses.

Methods (2-4 pages): First, include a brief description of the dataset. Make sure to read GSS methodology on the GSS website and also look up some article(s) that use GSS (e.g., the

example mentioned above) to see how they describe the dataset. Second, describe your sample. In addition, if you are selecting some subgroup for your analysis, discuss who is included and who is excluded, the reasons for these decisions, and the resulting sample size. Third, describe your variables – what questions were used in the survey, and how you recoded these variables for your analysis. Fourth, explain which types of analysis you use and why. (Include the do-files and log files with your data management steps and analysis steps and results in an appendix.)

Findings (1-3 pages):
First, organize your findings in a table or tables (see the article example for an idea for how to organize your tables). Second, describe the findings and your conclusions with regard to your hypotheses and your research question(s).

Conclusion (1-2 pages):
Describe the contributions of your study in light of the importance of your research question. Discuss whether the findings were expected or surprising. Finally, describe the limitations of your study and make suggestions for future research.

Appendix:
Include the do-file for data management, the do-file for data analysis, log files with all the output for those steps, and all the graphs that you generated.