Who is your TA (check one)?

🞏 De Andre 🞏 Corey 🞏 Jingkai 🞏 Neeraj

**SOC 3811/5811 – STATA Assignment #1**

PART 1. FOR ALL STUDENTS IN SOC **5811**

Go the GSS Data Explorer website at: <https://gssdataexplorer.norc.org/>. You may need to register for a (free) account. Once you are logged in, go to “MyGSS” and create a new project. When prompted, specify that you want “Stata (Control + System)” files (and not files for other software packages). Your cart will automatically include three variables, including “year.” Add the variable “wordsum” to your cart. Wordsum is a 10-item vocabulary test. (Before going any further, please go read about the GSS’s WORDSUM test here: <https://en.wikipedia.org/wiki/Wordsum>.) Then, create a data extract that includes the variables “year” and “wordsum.”

Once the extract is ready, the GSS Data Explorer will provide Stata code for reading in the two variables. Be sure to declare missing values to be missing!Then do the following:

1. Create a table below with frequency and percentage distributions of wordsum (for all years, not for each separate year).
2. Create a table below that reports the mean, median, and standard deviation of wordsum scores for each GSS year.
3. Answer this question in words: How has the distribution of wordsum scores changes over the years?
4. Paste below the Stata syntax you used to read in the GSS extract and perform the analyses above.

PART 2. FOR ALL STUDENTS IN SOC **3811**

I recently ran an internet survey of a random sample of 2,181 Americans. I collected information about each person’s age, education, race/ethnicity, and sex. I also asked them four questions about whether they had engaged in certain delinquent/deviant behaviors. The data that resulted from this survey are contained in the data file for this assignment. Below I include “meta-data” --- information about each variable, including (1) the name of the variable; (2) the variable’s column locations in the data file; (3) a brief description of the variable; and (4) the label associated with each numeric value. None of the variables have any missing data.

Use the example STATA syntax file (on the course web site) and modify it to accomplish the following goals. When you are done, type or paste your answers for questions #2 through #4 below into the Word file and turn it in via Canvas.

1. Read the data file into STATA
2. Produce frequency distributions for each of the variables for the entire sample.
3. Produce frequency distributions of the four deviance/delinquency variables separately for men and women
4. Answer these questions based on the results above:
   1. What percentage of people did not complete high school?
   2. What percentage has ever smoked pot?
   3. On which of the four deviance/delinquency variable are men’s and women’s answers most different? That is, for which of the four do men and women give the most divergent answers?

**DESCRIPTION OF VARIABLES IN THE DATA FILE FOR THIS ASSIGNMENT FOR SOC 3811**

**ARRESTED** (Column 1-10)

"Not counting minor traffic violations, have you ever been arrested and booked for breaking the law?"

1=Yes

2=No

**POT** (Column 11-22)

"Have you ever used marijuana, for example: grass or pot, in your lifetime?"

1=Yes

2=No

**THEFT** (Column 23-34)

"Have you ever stolen something from a store or something that did not belong to you worth less than 50 dollars?"

1=Yes

2=No

**DRUNK** (Column 35-46)

"Have you ever driven a vehicle while you were under the influence of alcohol?"

1=Yes

2=No

**AGE** (Column 47-58)

Respondent's Age in Years

18=18 Years old

...

93=93 Years old

**EDUCATION** (Column 59-70)

Respondent's Education

1=No formal education

2=1st, 2nd, 3rd, or 4th grade

3=5th or 6th grade

4=7th or 8th grade

5=9th grade

6=10th grade

7=11th grade

8=12th grade NO DIPLOMA

9=HIGH SCHOOL GRADUATE - high school DIPL

10=Some college, no degree

11=Associate degree

12=Bachelors degree

13=Masters degree

14=Professional or Doctorate degree

**RACE** (Column 71-82)

Respondent's Race/Ethnicity

1=White, Non-Hispanic

2=Black, Non-Hispanic

3=Other, Non-Hispanic

4=Hispanic

5=2+ Races, Non-Hispanic

**SEX** (Column 83-94)

Respondent's Sex

1=Male

2=Female