

# *SOC 5050: Problem Set 01*

*Christopher Prener, Ph.D.*

*August 29<sup>th</sup>, 2016*

## *Directions*

Please complete all steps below. For PARTS 1 and 2, enter your final answers in the accompanying Markdown file. Included a scanned copy of your work by hand for questions 4 through 6. You should also include your final do-file, log-file, and plots from PARTS 2 and 3. All requested documents should be uploaded to your GitHub assignment repository by 4:20pm on Monday, September 5<sup>th</sup>, 2016.

### *Part 1: Standard Deviation by Hand*

Answer the questions for PART 1 using the table below.

TABLE 1. 2010 Missouri Congressional Election Results

District	Population	Winner	Party	Incumbent	Turnout
1	587,000	Clay	1	1	184,779
2	706,600	Akin	0	1	265,632
3	625,300	Carnahan	1	1	203,085
4	680,000	Hartzler	0	0	225,056
5	634,000	Cleaver	1	1	191,423
6	700,000	Graves	0	1	221,912
7	722,000	Long	0	0	222,431
8	657,000	Emerson	0	1	195,999
9	683,000	Luetkemeyer	0	1	210,358

*Notes:* Party value labels are 0 = Republican and 1 = Democrat;

Incumbent value labels are 0 = No and 1 = Yes

1. What type of data (numeric/string) is the Winner variable? Why?
2. What type of data and level of measurement (binary/nominal/ordinal/continuous) is the Turnout variable? Why?
3. What type of data and level of measurement is the Party variable? Why?
4. What is the median for the Population variable? What does that value mean?

5. What are the mode and mean for the Party variable? What do each of those values mean?
6. What is the standard deviation of the Turnout variable? What does that value mean?

### *Part 2: Descriptive Statistics in Stata*

Use the do-file structure provided with this assignment as a base for executing the code required to answer the questions in PARTS 2 and 3. All code should be included and executed in a single do-file. Both PARTS 2 and 3 use the 2013 National Health Interview Survey dataset (36147-0003-Data.dta). You'll need to copy and paste these data onto your desktop. The do-file structure will change the name of the dataset to ps-01-data.dta.

7. What type of data (numeric/string) and level of measurement (binary/nominal/ordinal/continuous) is the variable REGION?  
Include a justification for your answer.
8. What type of data and level of measurement is the HHX variable?  
Include a justification for your answer.
9. What type of data and level of measurement is the variable R\_MARITL?  
Include a justification for your answer.
10. What type of data and level of measurement is the variable LCTIME11?  
Include a justification for your answer.
11. What type of data and level of measurement is the variable PSAL?  
Are there any apparent issues with this variable? Include a justification for your answer.
12. Provide all of the relevant descriptive statistic(s) for the variable PSAL based on its level of measurement. Include a justification for both the statistics you provided and those that you did not.
13. Provide all of the relevant descriptive statistic(s) for the variable LCTIME11 based on its level of measurement. Include a justification for both the statistics you provided and those that you did not.

*Part 3: Descriptive Plots in Stata*

14. Create the appropriate plot(s) for the variable PSAL based on its level of measurement. Include a justification for both the plot(s) you provided and those that you did not.
15. Create the appropriate plot(s) for the variable LCTIME11 based on its level of measurement. Include a justification for both the plot(s) you provided and those that you did not.

*Grading Rubric*

*Part 1* Each question is worth one point towards the thirty point total for the assignment. Half of the credit comes from stating the correct answer and half comes from your accompanying justification. Partial credit on problems requiring math by hand will be awarded based on the work you submit.

*Part 2* Each question is worth one point towards the thirty point total for the assignment. Half of the credit comes from stating the correct answer and half comes from your accompanying justification.

*Part 3* Each question is worth one point towards the thirty point total for the assignment. Credit will be divided equally among the number of plots submitted for each question. This grade reflects the selection of the appropriate plots as well as your accompanying justification.

*Stata Do-File* The overall quality of the Stata do-file is worth twelve points. This grade will be based on the clarity, organization, and layout of your do-file.

*Design* An additional three points are based on the layout and design of each of your figures. This grade will be based on the use of schemes as well as customization of the plots (titles, subtitles, and notes).

### *Document Details*

Document produced by Christopher Prener, Ph.D. for the Saint Louis University course SOC 5050 - QUANTITATIVE ANALYSIS: APPLIED INFERENCE STATISTICS. See the [course wiki](#) and the repository [README.md](#) file for additional details.



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).