**STA 402/502 Final Project**

Fall 2018

**Instructions:**

The goal of this project is to practice using the tools discussed during class, from textbook, from software supporting document or other resources. The project will be programming intensive. The outcome of the project is a well formatted report that explains the analysis you have done with supporting code and other materials. Each student will work on an individual project.

The project can be one of the following types:

1. Data analysis project:
2. You may choose from the datasets provided by the instructor, read in the data to SAS, do data cleaning/manipulating/merging, create customized tables, graphs and conduct statistical analysis.
3. You may also obtain a dataset based on your own interest. The datasets should be comparable to the given datasets in terms of scale (a large number of observations or columns), or is a messy dataset that needs a lot of work on data cleaning. Datasets should not come directly from a textbook exercise, nor a dataset that requires minimal programming to manage and analyze.
4. Other projects:

The projects may also start without a given dataset. I could be the techniques we have not discussed during class, such as web scrapping or data mining. Or simulation study of either a system or certain statistical methods.

Students are encouraged to talk with the instructor about the projects they choose. Especially for those who would like to use their own data or work on projects without an existing dataset.

**Project Phases and Grading:**

The final project will be 100 points in total and contains the following three phases:

**Project Plan (10 point):** By Friday, November 16, 9:10am, submit a 1-2 pages project plan, which includes a description of the problem, a brief summary of the dataset if you would like to work on a data analysis project, or a summary of your specific research material, your research questions and analysis plan. If you would like to use your own data of interest, please also include the data source (e.g. website and the data file).

**Progress Update (20 point):** By Friday, November 30, 9:10am, submit a 2-4 pages report describing the progress you have made on the project. Please include what you have completed and what you plan to do next, also include additional graphs, tables and reference as supplements. The code should be uploaded separately as a SAS file. Please also address any questions you meet during the analysis process. A rubric for grading will be posted later.

**Final Report (70 point):** By Wednesday, December 12, 12:00pm noon, submit a 4-8 pages final report. Additional graphs, tables and references are not limited to the page restriction. The SAS code and dataset (if you choose to use your own data) should also be submitted separately.

The following are some basic requirement of the final report and code. More detailed requirement and rubric for the final project will be available later.

* Your final report should be well formatted with clear section title, table/graphic labels, etc.
* The report should not only contain the programming procedures you have done, but should also address the research questions and your answers to those research questions based on the analysis.
* After I specify the original dataset and working directory, your program should work and all the results in your reports will be generated if I click on the “Running” button.
* The code should not be repeated simple SAS procedures. At least one macro function should be used.
* The code is well indented and with necessary comments.