STAT 480 Assignment #1

(Due July 9th, Thursday, 11:59 PM)

Read the following instructions carefully before you start your homework.

Instructions: For this homework, you only need to turn in an electronic .sas file on ANGEL. There is a dropbox in folder `Lessons > Homework > Homework#1` on AGNEL, where you can submit your .sas file. Your SAS codes should be well organized and commented so that your codes are readable. Use the following naming convention for the file you upload – `LastNamehw1.sas`. For instance, my .sas file name should be `liuhw1.sas`. If there is some errors in your SAS program (it always happens), check the log window. Two common mistakes for SAS beginners are: 1) forget to add $ when the variable is a character variable; 2) forget to add a semicolon at the end of each statement.

This homework is designed to help you become more familiar with reading data into SAS. I recommend you to complete your homework with school computers. You are required to print the results for each part if it is possible. (Use PROC print).

1. **Read in-stream data**

(a) Create two new folders `MyLibrary1` and `MyLibrary2` in your computer with the following pathnames:

`X:/STAT480/SAS library/MyLibrary1`

`X:/STAT480/SAS library/MyLibrary2`

Specify two reference names for these two folders with LIBNAME statement. (In the statement `LIBNAME lib1 location-of-folder`, `lib1` is the reference name. Reference names can be any names you like.)

(b) Read the following in-stream data and store it in `MyLibrary1` (To store this dataset in `MyLibrary1`, you need to specify the library name in DATA step). You should use CARDS statement in this problem. Remember Statement DATALINES and statement CARDS are the same, you can simply use CARDS to replace DATALINES.

(c) In SAS, the default length for a character variables is 8. If the length is longer than 8, only the first 8 characters will be read. For example, we can only see the first 8 letters of variable `Name` in output. We can use length statement to specify a longer length for character variables.

Name	Gender	GPA	SatMath	SatVerbal
BeckAnderson	Male	2.67	700	700
DupontJules	Female	2.98	700	500
GavettiNicole	Female	2.67	470	470
EastonMatthew	Male	3.6	710	560
MattiaceJohn	Male	3.76		
ZhangChen	Female	3.86	610	720

`length variable-name $ new-length;` For example,

`length variable-name $ 15;`

Specify a new length for variable `Name` and add the legnth statement **BEFORE** input statement in your codes in part(b), so that you can read these names completely. Every time you refer to a dataset in a permanent library (not in the temporary Work library), you should use a two-level name, that is, library-name.datasetname).

(d) Add a title for the dataset. Take a look at the html output, the title for this dataset is 'The SAS System'. We can use title statement to give a new title for the dataset.

`Title title-of-dataset;`

You can add the title statement before the PROC step. Now, give the title 'Class-Survey' to the dataset.

(e) Use Set statement to assign the dataset above to a new dataset.

2. **Read .txt and .csv files**

(a) Download the .txt file `ClassSurvery.txt` from ANGEL (in folder 'Data') and place it in the following folder in your computer

`X:/STAT480/Data`

If you cannot download the .txt file, you can simply copy and paste the whole text into a .txt file and name it `ClassSurvey.txt`

(b) Use FILENAME statement to assign a reference name to point to this .txt file.

(c) Read the file into SAS and store it in `MyLibrary2`. In this part, you should pay attention to:

- What is the delimiter that is used to separate values?
- Is there any missing observation?
- Which row does the first observation lie in?
- The length of character variables.

(d) Add a title for this dataset.

(e) Observe the html output carefully, can you find anything interesting/abnormal? (It is possible that there is nothing interesting/abnormal.)

(f) Repeat parts (a) - (e) for file `ClassSurvery.csv`. You can download this file from ANGEL.