**Homework #4**

**Submission: please submit a MS word document that includes the homework questions, your response to each question, post related code that you ran for each question (copy and paste from SAS reserving the color and font0, and any screenshots when output is provided from SAS.**

Question #1 If your project data was not approved in Homework 3, or if you are changing the data, please re-answer all question #1 in homework 3. You have to have data set approved before starting the project. Otherwise, points will be deducted from the project report.

Question #2 (50 pts):

The files Student0405\_Females.csv, Student0405\_Males.csv, and Student0405\_Additional.csv contain information about statistics classes at a large state university in the northeastern United States. The files Student0405\_Females.csv and Student0405\_Males.csv contain the following variables:

Location Variable Description

C1 ID Unique ID

C2         Sex               Coded Female, Male  
C3         GPA            Grade point average as reported by student  
C4         ReligImp      How important student feels religon is in their life.  
C5        MissClass    Number of classes student reports missing in typical week  
C6          Seat             Usual classroom seat for student (Front, Middle, Back)

The Student0405\_Females.csv contains information only on females while the Student0405\_Males.csv contains information only on males. The file Student0405\_Additional.csv contains the additional variables:

Location Variable Description

C1 ID Unique ID

C6        PartyDays    Days per month student goes to parties  
C7          Study            Hours per week of studying, as reported by student

1. Import the files into three different data sets named Males, Females, and Additional that correspond with the names of the CSV file (For example the data set Females corresponds with the CSV file Student0405\_Females.csv). (10 pts).
2. Concatenate (append) the Males and Females data into a new data set named Students (10 pts).
3. Merge the Additional information onto the Students data set using the variable ID as the key. Only keep rows from the additional data set that match the rows in the Students data set. (10 pts).
4. Create a new variable called AdjustedGPA that is the GPA reported in the file minus the average GPA for everyone (10 pts).
5. Create another variable called LetterGrade that reports an A if the GPA is 3.5 or higher, a B is the GPA is greater than or equal to 2.5 but less than 3.5, a C if the GPA is greater than or equal to 1.5 but less than 2.5, a D if it is greater than or equal 0.5 but less than 1.5, and an F if the GP is below 0.5 (10 pts).

Question #3: Using the final data set from Question 2, complete the following exercises (50 pts).

1. Report the average and standard deviation of the variable AdjustedGPA. Round the results to two decimal places (10 pts).
2. Calculate the percentage of students that responded to each level of the question ReligImp. What percentage of students feel religion is very important, fairly important, and not important (10 pts)?
3. Calculate the percentage of female students who sit in the front row and compare it with the percentage of male students who sit in the front row (10 pts)
4. Calculate the percentage of students that miss 0 classes, 1 to 3 classes, or more than 3 classes. (10 pts)
5. Examine the classes missed as defined in part d for each of the seat locations. (10 pts).
6. Calculate the Average GPA for the three seat locations (5 pts).