Final Project

Directions:

- Submit one program file, one output file and one log file for the whole Lab in 'J:\Classes\Student46'.
- Use comment statements to separate your answers. Use titles where appropriate and for tables and reports.
- Make sure that your log file and output file contain only one run.
- Clear the screen after each run and present only the last one in 'J:\Classes\Student46'.
- Name your file according to your last name, example: lastname_prog.sas
- Use the **SAS Help and Documentation** if needed (see lecture 1).

Part 1:

Considering the SAS data set Medical:

1. Compute the *frequencies* for the:
    a. days of the week,
    b. Date of the visit
2. Supply a format for the days of the week and months of the year. Save it in a format library called *FmtVisit*. Provide a display of *FmtVisit*.
3. Using the variable *VisitDate*, create a temporary SAS data set, call it *Interval* with the variables in Medical plus a new variable named Quarter, representing the number of quarters since January 1, 2006. Provide the data portion of this new data set.
4. You want to see each patient in the Medical data set on the same day of the week, 5 weeks after they visited the clinic. Provide a listing of the patient number, the visit date and the date for the return visit.
5. Now, you want to see each patient on the same day of the month, 6 months after they visited the clinic. Provide a listing of the patient number, the visit date, and the return visit.

Part 2:

In this part, you will work on the College data set. Use the file **College.csv** in **'J:\CLASSES\STAT46'**.

Note: whenever you create user_defined formats, store them in a format library and provide a printout of the library at the end of Part2.

1. Import the College.csv file into SAS.
2. Compute general statistics: mean,median, minimum and maximum as well as the number of missing and non-missing values for the variables: GPA and SchoolRank.
3. Report the mean and median of GPA and ClassRank broken down by school size (SchoolSize) once using a BY statement and another time using a CLASS statement.
4. Report the mean GPA for the following categories of ClassRank:

   0 – 50 = **Bottom Half**,

   51 – 74 = **3rd quartile**,

   75 – 100 = **Top Quarter**,

   By creating an appropriate format. Do not use a DATA step.

5. Create four summary data sets containing the number of nonmissing and missing values and the mean, median, minimum and maximum for ClassRank and GPA broken down by Gender and SchoolSize.
   a. The first data set (Grand) should contain the statistics for all subjects,
   b. the second data set (By Gender) should contain the statistics broken down by gender.
   c. The third data set (BySize) should contain the statistics broken down by SchoolSize,
   d. And the fourth data set (Cell) should contain the statistics broken down by Gender and SchoolSize.

Do this by using PROC MEANS (with a CLASS statement) and one DATA step.