STAT 440 – Homework 04

Students are encouraged to work together on homework. However, sharing or copying any part of the homework is an infraction of the University's rules on Academic Integrity.

Final submissions must be uploaded to our Compass 2g site on the Homework page. No email, hardcopy, or late submissions will be accepted.

Getting the program file ready

a. Create a folder on the hard drive with the following pathname – C:\440\hw04. Save all data files accompanying this assignment in that folder. If you cannot create the folder because you are working on a university computer and don't have permission, create the …\440\hw04 folder elsewhere.
b. Assign the library reference **hw04** to the folder 'C:\440\hw04'. Use this library as your permanent library for this assignment. If you could not create the folder, assign the library reference **hw04** to your …\440\hw04 folder.
   Note: If you are using a folder other than 'C:\440\hw04', you must change any pathname references in your program file to 'C:\440\hw04' before submitting your homework.

Submitting your work to Compass 2g

You are to submit two (and only two) files for your homework submission.

1. Your SAS program file which should be saved as **HW*n_YourNetID*.sas**. For example, my file for the HW04 assignment would be HW04_dunger.sas. All program statements and code should be included in one program file.

2. Your Report including all relevant output to address the exercises. For this homework, use ODS to send your results to a Portable Document Format (PDF) file called **HW*n_YourNetID*.pdf**. For example, my file for HW01 would be HW01_dunger.pdf. Only include your final set of output. Do not include output for every execution of your SAS program. Use the template file **hw01 template.sas** as your guide.

You have an unlimited number of submissions, but only the last one will be viewed and graded. Homework submissions must always come as a pair of files, as described above.

1. In this exercise, you will work with a data set from the National Football League containing 1890 unique observations. The raw data sets **nflrush.dat** and **nflrush_quotes.dat** contain data from the 2010-2015 seasons of professional American football for every player who recorded rushing yards (ran with the ball instead of being thrown the ball). If the player moves the ball forward, he receives positive rushing yards. If he's knocked backward, he records negative rushing yards for that attempt.

   The first record in the raw data set contains the column headers.

Variable	Name	Description
1	Season	
2	Player	Full name
3	Team	Player's team; max of 3 letters
4	Games	Number of games in which that player appeared that season
5	Att	Number of rushing attempts
6	Yds	Number of rushing yards
7	Avg	Average rushing yards per attempt, rounded to nearest 0.01
8	YPG	Average rushing yards per game, rounded to nearest 0.1
9	Lg	Longest rushing attempt
10	TD	Number of rushing touchdowns (i.e., scores, goals)
11	FD	Rushing first-downs

   a. Write a DATA step using formatted input to read the values from **nflrush.dat** into SAS. The output data set is to be a temporary SAS data file called **rushing_*YourNetID***.
      - Do not include variables for number of games, attempts, and yards per game.
      - You are responsible for addressing any needed variable attributes such as length, format, informat, and label.

   b. Print the descriptor portion of your new SAS data file once completed. (Include results in the HW Report.)

   c. Use PROC SORT to generate a list of players who had the most rushing touchdowns in a single season from 2013-2015. Store the results in a data set of your choosing. Print the top 10 including only the player's name, number of rushing touchdowns, and season in which the total was achieved. (Include results in the HW Report.)

   d. Write a DATA step using list input to read the values from **nflrush_quotes.dat** into SAS. The output data set is to be a temporary SAS data file called **localnfl_*YourNetID***.
      - This dataset should only include players from St. Louis (Stl), Chicago (Chi), Indianapolis (Ind), and Green Bay (GB).
      - Do not include variables for number of games, attempts, and yards per game.
      - You are responsible for addressing any needed variable attributes such as length, format, informat, and label.

   e. Print the descriptor portion of your new SAS data file once completed. (Include results in the HW Report.)

f. Use PROC SORT to generate a list of players who had the most rushing yards in a single season. Store the results in a data set of your choosing. Print the top 10 including only the player's name, team, number of rushing yards, and season in which the total was achieved. (Include results in the HW Report.)

2. The raw data set **employee_roster4.dat** contains many variables regarding the employees of the Orion corporation. It's the same data values seen in previous homework, but quite a different layout. Be sure to look at the raw data to do some preliminary examination. Missing values may exist.

Line	Variable Order
1	ID, Name, Country
2	Company, Department, Section, Organization Group, Job Title, Gender
3	Salary, Birth Date, Hire Date, Termination Date

a. Read in the raw data file and create a SAS data file called **low_earners4_*NetID***.
   - Subset the observations so to only Orion employees who make under $25,000 and work in the Sales department.
   - You are responsible for addressing any needed variable attributes such as length, format, informat, and label.

b. Print the descriptor portion of **low_earners4_*NetID***. (Include results in the HW Report.)

c. Print the data portion of **low_earners4_*NetID*** including only each employee's name, gender, department, job title, and salary in that order. (Include results in the HW Report.)