

## STAT 440 – Homework 07

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\hw07. 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\hw07 folder elsewhere.
- b. Assign the library reference **hw07** to the folder 'C:\440\hw07'. Use this library as your permanent library for this assignment. If you could not create the folder, assign the library reference **hw07** to your ...\\440\hw07 folder.

Note: If you are using a folder other than 'C:\440\hw07', you must change any pathname references in your program file to 'C:\440\hw07' 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 **HWn\_YourNetID.sas**. For example, my file for the HW07 assignment would be HW07\_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 Rich Text Format (RTF) file called **YourNetID\_HWn.rtf**. Only include your final set of output. Do not include output for every execution of your SAS program.

Once the results have been sent to the .rtf file, you may open it in Word and include your own responses in the relevant areas (as directed in the exercises).

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. You will be working with the SAS data file **nonsales** which is available in the “all data” folder as a part of classroom data on Compass and SAS OnDemand.

In a previous lesson, we saw a brute-force method to verify whether the values of the Employee\_ID variable are unique and nonmissing. It wasn't very efficient because it still required the user to read through the PROC PRINT or PROC FREQ output to find the repeated value(s). But now you have more tools at your disposal since then.

- a. Create a data set called **duplicates** that contains every Employee ID that is not unique (i.e., it has repeated values) and the number of times it appears. Print the data portion of this data set. (Include your results in the HW Report.)

If we were to then clean the duplicates, we'd need another way to identify the observations with issues.

- b. Use the data in **duplicates** to create a table containing the duplicate Employee IDs with the First and Last name of each person the IDs correspond to.
  - The table should contain the three variables mentioned above.
  - You will likely need to create a new SAS data set (i.e., table).
  - At no point in your code should you be actually typing the Employee IDs in question.

2. In this exercise, we will revisit a data set from HW4.

You will work with a data set from the National Football League containing 1890 unique observations. The SAS data set **nflrush.dat** contains 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.

*Note: This version of **nflrush.dat** does not have the quotes in the Player field.*

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

- Write a DATA step using formatted input to read the values into SAS. The output data set is to be a temporary SAS data file called **rushing\_YourNetID**.
  - Include all variables this time.
  - You are responsible for addressing any needed variable attributes such as length, format, informat, and label. (Feel free to revisit HW4 for help and adjustments from that will need to be made.)
- Print the descriptor portion of your new SAS data file once completed. (Include results in the HW Report.)
- Print a table created in SAS that reports the player with the most rushing yards for each of the teams over the six seasons covered in this data set.
  - There are 32 teams in the NFL, so there should be 32 observations in the table. If you observe more than 32, perhaps there is a typo in one or more of the Team values that needs correction.
  - If there is a tie for most rushing yards for a team over this period, include all players who achieved that team high. In that case, there would be more than 32 observations.
  - The table should include the variables Team, Player, number of Seasons with team, and Total Yards.
  - The table should be sorted in decreasing order of Total Yards.