STAT 440 – Homework 6

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\hw6. 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\hw6 folder elsewhere.
b. Assign the library reference **hw6** to the folder 'C:\440\hw6'. Use this library as your permanent library for this assignment. If you could not create the folder, assign the library reference **hw6** to your …\440\hw6 folder.
   Note: If you are using a folder other than 'C:\440\hw6', you must change any pathname references in your program file to 'C:\440\hw6' 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 HW6 assignment would be HW6_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*_HW*n*.rtf**. Only include your final set of output. Do not include output for every execution of your SAS program. Use the template file **hw5 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. You will be working with the SAS data files **inventory** (which contains the model ID and price of various products) and **purchase** (which contains the model ID, quantity purchased, and customer who purchased the product).

   a. Merge the **inventory** and **purchase** data sets to create a new SAS data set called **purchase_price_*NetID*** based on the Model number.
      - There are some models in the **inventory** data set that were not purchased (and, therefore, are not in the **purchase** data set). Do not include these product models in the new data set.
      - Add the Price value found in the **inventory** data set to each observation in the **purchase** data set.
      - Compute a new variable called TotalCost that calculates the total invoice cost for each Model purchased.

   b. Print the data portion of **purchase_price_*NetID*** including all variables, excluding observation numbers. (Include results in the HW Report.)

   c. Using a separate DATA step, create a SAS data set called **not_purchased_*NetID*** containing a list of all Models (and their Price) that were not purchased.
      - Note that only models that were purchased will appear in the **purchase** data set.

   d. Print the data portion of **not_purchased_*NetID***, excluding observation numbers. (Include results in the HW Report.)

   e. Repeat parts (a) and (c), but this time create both new data sets in a single DATA step. Print the data portion of both new data sets including all variables, excluding observation numbers. (Include results in the HW Report. Use titles so that we can identify this output as distinct from parts (b) and (d).)

2. The SAS data sets **bat_AL.sas7bdat**, **bat_NL.sas7bdat**, and **bat_earlyyears.sas7bdat** contain a complete history of Major League Baseball's (MLB) batting data from 1871 through the 2010 season. Each observation holds a single season (i.e., year) of batting statistics for a single player. So each observation contains a unique combination of PlayerID and YearID.

   a. Run procedure to find the minimum and maximum value for YearID in each of the three datasets. This tells us the years that each dataset spans. (Include results in the HW Report.)

   b. Concatenate the three data sets into a data set called **batting_*NetID***.
      - Include only PlayerID, YearID, TeamID, G, AB, R, H, and RBI in the DATA step and in the final output data set.

   c. Print the full descriptor portion of the new data set. Are the results as you expected given that you ran a concatenation? (Include your results and response in the HW Report.)

   d. Merge the data sets as necessary to create a new data set called **bothleagues_*NetID*** which lists all players who played at least part of one season in both the American League (AL) and the National League (NL) from 1970 to 1989.

   e. Print the full descriptor portion of the new data set. (Include your results in the HW Report.)

   f. Using any of the existing SAS data sets, create a data set called **MostHits_*NetID*** that contains all players who had over 200 hits (H) in a single season for a single team.
      - Include only variables for PlayerID, YearID, TeamID, and H.

   g. Print the data portion of **MostHits_*NetID***. (Include your results in the HW Report.)
      - If a player accomplished the feat in more than one season, there should be a separate observation for each occurrence. That is, each unique combination of Player ID and YearID.
      - If there are no individuals at all who satisfy the criteria, print the descriptor portion of the data set instead, and it will note that there are 0 observations.