STAT 440 – Homework 11

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\hw11. 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\hw11 folder elsewhere.
b. Assign the library reference **hw11** to the folder 'C:\440\hw11'. Use this library as your permanent library for this assignment. If you could not create the folder, assign the library reference **hw11** to your …\440\hw11 folder.
   Note: If you are using a folder other than 'C:\440\hw11', you must change any pathname references in your program file to 'C:\440\hw11' 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 HW11 assignment would be HW11_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. This is a revisit of a HW6 data set, but now you'll use PROC SQL for each part.

   The SAS data set **batting** contains 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, YearID, and TeamID.

   a. Print the descriptor portion of the data set. (Include your results and response in the HW Report. You'll have to copy and paste it from the Log.)

   b. Print a table that contains every player who had at least 245 hits (H) in a single season with one team. (Include results in the HW Report.)
      - The output table should only contain each player's PlayerID, the YearID, the TeamID, and Hits for that season.
      - Players might appear more than once in the list.

   c. Print a table that shows the first (minimum) and last (maximum) year for each baseball league (LgID) represented in the data set. (Include results in the HW Report.)

   d. Print a table that contains the player with the most home runs for each of the 140 seasons as long as their season home run total is at least 50. (Include results in the HW Report.)
      - The output table should only contain the player's PlayerID, the YearID, and home runs for each year/season (row).
      - If there was a tie for a given season, print all players who achieved the most home runs for that year.

2. This is a revisit of a HW6 exercise, but now you'll use PROC SQL for each part.

   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, temporary SAS data set called **purchase_price_*NetID*** based on the Model number.
      - Add the Price value found in the **inventory** data set to each observation in the **purchase** data set.
      - 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.
      - 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. (Include results in the HW Report.)
   c. Using a separate query, create a list of all Models (and the Price) that were not purchased in a permanent SAS data set called **not_purchased_*NetID***.
   d. Print the data portion of **not_purchased_*NetID***. (Include results in the HW Report.)