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.

The HW Report should include the output generated from the following exercises:
1-b, 2-abcd, 3-bd

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 **hw3 template.sas** as your guide.

   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. Exam Scores

The raw data set **examscores5.dat** contains the ID and five exam scores for 144 students. The passing score for each of the five exams is 77, 79, 82, 76, and 80, respectively.

a. Read the raw data file **examscores5.dat** into two temporary SAS data sets called **passing_*NetID*** and **failing_*NetID***.
   - Use a temporary array to count the number of tests passed by each student.
   - The resulting data sets should contain seven variables: the ID, the five test scores, and the number of passed tests.
   - Those who passed at least four exams go into **passing_*NetID*** and those that passed less than four exams go into **failing_*NetID***.

b. Print the data portion of the SAS data set **failing_*NetID***. (Include results in the HW Report.)
   - Output the ID and number of passed tests.
   - Enhance the report appropriately with formats, titles, etc.

2. Baseball data

This is a revisit of a HW5 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 batting data from 1871 through the 2010 season. Each observation holds a single season of batting statistics for a single player. So each observation contains a unique combination of PlayerID and YearID.

a. Print a table that contains every player who had at least 240 hits (H) in a single season. (Include results in the HW Report.)
   - The output table should only contain each player's PlayerID, the YearID, and Hits for that season.
   - Players may appear more than once in the list.

b. Print a table that contains every player who had more than 3300 hits (H). (Include results in the HW Report.)
   - *Hint:* You'll have to total up each player's batting statistics in that category for every season he played. You don't necessarily have to create a new variable to do this when using SQL.
   - The output table should only contain each player's Player_ID, First year played in baseball, and Total Hits.

c. Print a table that contains the player with the most home runs for each of the 140 seasons. (Include results in the HW Report.)
   - The output table should only contain the player's Player_ID, First year played in baseball, and home runs for each season (row).

/ continued…

d. Print a table that finds the player who had the most hits in a single season ever. (Include results in the HW Report.)
   - The output table should only contain the player's PlayerID, the YearID, and Hits for that season.
   - This is looking for one player overall, and not one player for each season like in part c.

3. Product data

This is a revisit of a HW4 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, excluding observation numbers. (Include results in the HW Report.)

c. Using a separate DATA step, 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***, excluding observation numbers. (Include results in the HW Report.)