STAT 440 – Homework 08

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\hw08. 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\hw08 folder elsewhere.
b. Assign the library reference **hw08** to the folder 'C:\440\hw08'. Use this library as your permanent library for this assignment. If you could not create the folder, assign the library reference **hw08** to your …\440\hw08 folder.
   Note: If you are using a folder other than 'C:\440\hw08', you must change any pathname references in your program file to 'C:\440\hw08' 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 HW08 assignment would be HW08_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 **HW*n_YourNetID*.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.

   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 set **orders** which contains a group of orders from U.S. customers.

   a. Suppose you want to reward customers who spent $400 or more on retail or internet purchases and $70 or more on catalog purchases.
      - Create three temporary SAS data sets called **discount_ret**, **discount_cat**, and **discount_int**.
      - The value of Order_Type indicates whether the sale was retail (=1), catalog (=2), or internet (=3).
      - The variable Total_Retail_Price contains the amount that the customer spent on each individual order.
      - Create a variable named TotSales to hold the total sales to each customer by order type.
      - A given customer can output to more than one data set if s/he spent enough money in multiple areas.
      - Keep the variables Customer_ID, Customer_Name, and TotSales for all data sets. Additionally keep gender for the catalog data set, and hold on to the birth date variable for the internet data set.

   b. Print the data portion of all three output SAS data sets. (Include results in the HW Report.)
      - Enhance the report appropriately with formats, titles, etc.

   c. Repeat part (a) by creating the three data sets listed in those directions. But for this exercise you will also create a data set called **top_buyers_*NetID***. This data set must contain all customers with purchases totaling greater than $500 across all three platforms (retail, catalog, internet).

   d. Print the data portion of **top_buyers_*NetID*** and **discount_int**. (Include results in the HW Report.)
      - Enhance the report appropriately with formats, titles, etc.

2. You will be working with the raw data set **importexport87-15.dat** which contains three variables pertaining to the United States trade in goods with the rest of the world from 1987 to 2014. The data were originally downloaded from the Census Bureau website (http://www.census.gov/foreign-trade/balance/c0015.html).

   The variables, in field order from left to right, are…
   - The date marking the beginning of a month's worth of data.
   - The total value of exports from the U.S. in millions of dollars.
   - The total value of imports to the U.S. in millions of dollars.

   a. Write a program to read in the data from **importexport87-15.dat** into a SAS data set called **trade_*NetID***.
      - The SAS data set should contain a new variable called Balance representing the difference in value to the U.S. between exports and imports, i.e. Exports – Imports.
      - Remember to use formats when appropriate.

   b. Print the first 24 observations in the data portion of **trade_*NetID***. Suppress observation numbers. (Include results in the HW Report.)

   c. Write a new program that reads in the data from **importexport87-15.dat** and ultimately creates a SAS data set called **yearlyimports_*NetID*** which has only one observation per year.
      - The SAS data set should contain a new variable called YearTotal representing the total value of imports for each year.
      - The SAS data set should contain a new variable called YearAvg representing the average value of imports per month for each year.
      - This data set should contain the variables Year (current year), YearTotal, and YearAvg.

   d. Print the descriptor portion of **yearlyimports_*NetID***. (Include results in the HW Report.)

   e. Print the data portion of **yearlyimports_*NetID***. (Include results in the HW Report.)

   f. By any means you see fit, write a program that lists the average value of imports per year for each decade (the 80's, 90's, etc.). Note that not every decade is fully represented with 10 years of data in this data set. Identify in the report which decade has the smallest average and which has the largest average. (Include results in the HW Report.)