ST555 Homework 5

When you have completed this HW, submit via Moodle the following:

I. Your SAS program (.sas) that contains the answers to the fill in the blanks as comments. Be sure to include the question number and letter for each comment.
II. Your SAS log (.log)
III. Your Results, generate your results using ods as a .pdf or .rtf file

1) In this problem, you will use the file 'Supplier.csv' which contains information about Orion Star suppliers.

   a. Write a DATA step to read these data into SAS. There is an issue here that may not have been discussed in class. The file 'Supplier.csv' contains a header row that specifies the names of the variables. You need to tell SAS to skip the header row, and start reading in data from the second row. You can do this by using the option firstobs=2 in the infile statement.

   b. Make sure the character variables are not truncated.
   How many characters are needed to avoid truncating

      i. Supplier Name? _____

      ii. Address? _____

      iii. Country? _____

   c. Create a new <u>permanent</u> SAS dataset based on Supplier.csv with the following specifications:

      i. Create three new variables, 'Discount', 'DiscountType', and 'Region' where:
      -- If Country is 'CA' or 'US',
      -- Discount is equal to 0.10
      -- DiscountType is equal to 'Required'
      -- Region is equal to 'North America'.
      -- If Country is any other value,
      -- Discount is equal to 0.05
      -- DiscountType is equal to 'Optional'
      -- Region is equal to 'Not North America'.

      ii. Ensure that the values of 'DiscountType' and 'Region' are not truncated.

      iii. Only include the supplier name, country, and the 3 variables created in part b.

   d. Write any procedure that will print the number of observations within each 'Region' value.

      How many observations are in North America? _____

      How many observations are not in North America? _____

2) Displayed below is the first 5 rows of the file 'Prices.dat':

210200100009*09JUN2007*31DEC9999*$15.50*$34.70
210200100017*24JAN2007*31DEC9999*$17.80
210200200023**31DEC9999*$8.25*$19.80
210200600067*27OCT2007*31DEC9999*$28.90
210200600085*28AUG2007*31DEC9999*$17.85*$39.40

The variables are all numeric and are named 'ProductID', 'StartDate', 'EndDate', 'UnitCostPrice', and 'UnitSalesPrice'. Note the missing values for some observations.

a. Write a DATA step to read in these data. Note that you will need to use SAS informats for nonstandard values.

b. Add labels so that 'ProductID' has the label 'Product ID', 'StartDate' has the label 'Start of Date', 'EndDate' has the label 'End of Date', 'UnitCostPrice' has the label 'Cost PPU' (FYI: PPU is price per unit), and 'UnitSalesPrice' has the label 'Sales PPU'.

c. Format 'StartDate' and 'EndDate' so that it displays as ex 07/01/2014.

d. Format the 'UnitCostPrice' and 'UnitSalesPrice' variables so that they display a $ sign and two digits after the decimal point.

e. Print the dataset with labels and ensure that the data were read in correctly.

f. Does the 25th observation contain any missing values?_____

g.

3) Create a dataset called “salesmgmt” using the raw data file named “managers.dat”. The variables provided are ID, first name, last name, gender, salary, job title, country, birthdate, and hire date.

a. Provide results that show you’ve read the dataset into SAS correctly.

b. Generate the report below.

Orion Star Managers

Obs	ID	Last	Title	HireDate	Salary
1	120102	Zhou	Sales Manager	01JUN1989	.
2	120103	Dawes	Sales Manager	01JAN1974	87975
3	120261	Highpoint		01AUG1987	243190
4	121143	Favaron	Senior Sales Manager	01JUL1997	95090
5	121144	Capachietti	Sales Manager	.	83505
6	121145	Lansberry	Sales Manager	01APR1976	84260

c. Recreate the report in b. but display Salary with a dollar sign ($) and commas in the appropriate places

4) Use the SAS Help facility or online documentation to investigate the IMPORT and EXPORT procedures. Describe the chief differences from using the SAS/ACCESS LIBNAME statement. In this description, include the following: which method guarantees accessing the most recent data in the workbook or database?; which method makes a SAS copy of the data?; which method allows for easier programming (e.g. are name literals needed)?

5) Creating Group Introduction Dataset:
   a. take the information in the forum for your Group Introductions and create one SAS dataset call it "Orion.GroupIntro_Take2" so it will be a permanent SAS dataset in your Orion library folder.
   b. The number of observations in your dataset should be equal to the number of group members in your dataset (3 or 4) and you should have as many variables as it takes to include all introductory material. *[Note this is reworded from last time as many of you did not include enough variables]*
   c. Each observation should have the same data value for the "GroupTrait" variable.
   d. Use the Print procedure to view how SAS displays your GroupIntro dataset.

      Copy and paste the Output here

      _____

6) How long did it take you to complete this homework? _____