ST555 Homework 4

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

1) 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.
2) Your SAS log (.log)
3) Your Results, generate your results using ods as a .pdf or .rtf file

1. **Defining Ranges in User-Defined Formats:** Utilize the Orion nonsales.sas7bdat dataset and follow these instructions:

   a. Create a character format named $GENDER that displays gender codes as follows:

F	Female
M	Male
Any other value	Invalid code

   b. Create a table that provides the frequency and percentage of employees in each of the 3 gender codes.

   c. Create a numeric format named SALRANGE that displays salary ranges as follows:

At least 20,000 but less than 50,000	Below $50,000
At least 50,000 but less than 100,000	Between $50,000 to $100,000
At least 100,000 and up to 500,000	$100,000 or more
missing	Missing salary
Any other value	Invalid salary

   d. Create a table that provides the frequency and percentage of employees in each of the 4 categories for salary range.

   e. Create a new permanent dataset called "Asst" and only include employees who have the word "Assistant" in their Job_Title.

   How many employees meet this criterion? _____

   f. Create a report with a listing of the employees in the dataset "Asst" and make sure the print out includes the following:

      i. Apply the user-defined formats from part (a) and (b) to the Gender and Salary variables, respectively.

      ii. Display Employee_ID as the identifier in place of the Obs column.

iii. Create a label for Employee_ID that eliminates the underscore and puts the word Employee on the first row of the column heading and ID on the second row of the column heading.

iv. Create a label for Job_Title that eliminate the underscore and puts the word Job on the first row of the column heading and Title on the second row of the column heading.

2. Exploring Format Storage Options

User-defined formats are stored in the **formats** catalog in the **work** library, **work.formats**. Use the SAS Help Facility or product documentation to explore permanent format catalogs in PROC FORMAT.

What option enables you to store the formats in a permanent library? _____

What option causes SAS to look for formats in permanent libraries? _____

3. Creating a SAS Data Set: Follow the steps below to create a dataset "delays" that subsets observations based on three conditions.

a. Write a DATA step to create work.delays using orion.orders as input.

b. Create a new variable, Order_Month, and set it to the month of Order_Date.

c. Create a new variable, Days4Delivery, that calculates the number of days from the order date to the delivery date.

d. Use a WHERE statement and a subsetting IF statement to select only the observations that meet all of the following conditions:

- Delivery_Date values that are more than four days beyond Order_Date
- Employee_ID values that are equal to 99999999
- Order_Month values occurring in the summer defined as June through August

e. The new data set should include only Employee_ID, Customer_ID, Order_Date, Delivery_Date, Days4Delivery, and Order_Month.

f. Add permanent labels for Order_Date, Delivery_Date, and Order_Month that read Date Ordered, Date Delivered, and Month Ordered, respectively.

g. Add permanent formats to display Order_Date and Delivery_Date as MM/DD/YYYY.

h. Add a PROC CONTENTS step to verify that the labels and formats were stored permanently.

i. Print out the results with Employee_ID displayed as the identifier in place of the Obs column.

4. In this problem, you will work with the Football.xls excel file. The excel file contains two sheets: footballrecruiting and Colleges. The first sheet contains football recruiting information and the second sheet contains information on all colleges that offer a bachelors degree in the U.S. [To read the dataset into SAS, Install PC Files Server OR utilize proc import]

a. Use an appropriate LIBNAME statement to read these two worksheets into SAS. Write a PRINT procedure and only print the first 10 observations of each worksheet.

b. Create two SAS datasets in the Work library based on the two worksheets and name them 'Colleges' and 'Recruit'.

c. Write an appropriate procedure to display the descriptive portion of the 'Colleges' dataset. What format does the 'Address' variable have? _____

d. Create a new temporary SAS dataset based on 'Colleges' that contains:

1) only private schools in Texas.
2) Drop the 'Address' and 'Type' variables.
3) Format the 'Students' and 'Undergrads' variables so that a comma separates the thousands place.
4) From the Log window, how many observations and variables are there?

Observations _____ Variables _____

e. Create a new SAS temporary dataset based on 'Recruit' with the following specifications:

1) Do not include the variable 'Points' in this new dataset.
2) Apply the following labels: 'Conf' should have the label 'Conference', 'Off' should have the label 'Offense', 'Def' should have the label 'Defense', and 'SpT' should have the label 'Special Teams'.
3) Only include schools ranked in the top 20 and in conferences that end in 'C'.
4) How many observations and variables are there?

Observations _____ Variables _____

f. Print the dataset in (e) with the labels. Which school is ranked number 1? _____

g. Use a LIBNAME statement to deassign the library created in (a).

5. How long did it take you to complete this homework? _____