STA 402/502 Homework 5

Due: October 5th (Friday), before class

Please read the homework guidelines before working on the homework. Homework that does not follow the guidelines will be deducted points. You are to complete this assignment on your own. Remember to include an intro comment block on all programs written. Each problem should be attempted as its own program.

1. The dataset "meat" on canvas website contain the packaging and the log of bacterial growth information as described in §2.9 (page 71) and §5.3 (page 177) in the textbook. Ignore the question and information from page 71 and 72, the following exercises have some modifications and specific requirements. Our goal is to read in the dataset and then generate a report similar to that seen on page 71. Write a SAS program to do the following with the provided dataset. Include the report in your homework.

   (a) Input all six items for each data line as a single observation instead of treating them as three observations. This might involve using six variable names per line.

   (b) Use the MEAN and STD functions to calculate summary statistics as shown on page 71 for each line.

   (c) Use PUT statements to create the report.

   (d) Your output should looks similar to the output shown in the box on page 71 and cover all four packaging conditions (plastic, vacuum, mixed and Co2). Instead of x.xx, your output should show the mean and standard deviation values calculated within the DATA step.

   Hint: The SAS statement
   put varx 4.2;
   formats the printed value of the numeric variable varx so that it uses 4 character positions with 2 digits after the decimal point (which is the same format as the summary statistic on page 71).

2. Your manager has three permanent SAS data sets named Quarter1, Quarter2, and Quarter3, which she wants to merge by the variable Customer into a new permanent SAS data set called YearToDate. She also has these requirements:

   - The observations in YearToDate must be in the exact same order as they appear in the original copy of Quarter2.
   - Aside from Customer, all variables from the three data sets occur in the same order as in the original sets.
   - Aside from Customer, the variables from Quarter1 appear first in YearToDate, the variables from Quarter2 are next, and the variables from Quarter 3 are last. (Note: MERGE and SET preserve variable-name order for data sets listed in order.)
   - The original permanent SAS data sets aren't affected.

   You may assume that the given data sets have the following properties:

   - The variable names are at least three characters.
   - Each data set has a Customer variable and no other variable name is duplicated among the three sets.
   - Each value of Customer occurs exactly once in each of the three data sets. In particular, the three sets have the same number of observations.
   - The Customer values do not necessarily occur in the same order in all three sets.

   Write a SAS program to carry out this task. It must rely only on the details given above and you do not know in advance how many variables or observations there are, what order the variables appear in, or which variables are numeric or character.

   However, you must test your program on the small data sets ("Quarter1", "Quarter2" and "Quarter3") from the canvas site. Please print the resulting permanent YearToDate data set to show that your program works on those small files.

   Note: Your program may create additional temporary SAS data sets with extra variables, but the final data set YearToDate may only include variables from the original three data sets.

3. (STA 502 students only) The local school district has a SAS data set called DISTRICT that contains the rubric used for classifying teachers annual evaluations. The rubric is based on two components: a teacher score and a curriculum grade. Using these components, an overall rating for the teacher can be determined. The data for 10 elementary school teachers who are due for evaluation are stored in the SAS data set called TEACHERS. Before working on the exercises below, examine these two SAS data sets including the variable name, labels and other attributes.

   (a) Combine the two SAS data sets so that the district rating is properly assigned to each teacher. The order of variables that appear in the dataset does not matter. Print the combined dataset.

   Hint: 1. You can specify two or more variables in a BY statement. 2. Notice the difference in the variable names in the datasets. Change variable names if necessary. To change a variable name in a dataset, the following is an example.

```
data dataset (keep=newvar_a newvar_b);
set dataset;
newvar_a=var_a;
newvar_b=var_b;
run;
```

   (b) From the dataset in part a), create a new dataset that eliminate any observation that does not match to a teacher, and also sort the observations by teacher name so that administrators can easily locate each teachers rating. Print the new dataset. (Hint: you may want to search how to represent missing values for character variables in SAS.)