Stat 604
Assignment 10 - SAS

OBJECTIVES: In this assignment you will practice using conditional statements and creating data from SAS data sets.

You should have all of the information you need to complete this assignment by viewing the first 6 SAS lectures.

This assignment will use the **COVID** permanent data set that was created in the previous assignment. If you had difficulty creating this data set, the professor's version will be made available after the deadline for that assignment.

To complete this assignment there will be three PDF files posted to Canvas –one PDF for the program, one PDF for the log, and one containing the SAS output. You will be required to run portions of the program incrementally. If you are using SAS Studio, this will require you to copy the log from each section to a separate document that you will convert to PDF for submission.

1. Add a header comment section to the beginning of a new program in your SAS session. Be sure to include a comment line above each section of the program that identifies the associated assignment step and a brief description of what the section is doing. Include housekeeping statements to clear titles and footnotes and suppress the printing of procedure titles.
2. Assign a libref to the **mylib** folder containing your permanent data sets. If you are going to use the professor's data set on SAS Studio, assign a separate library to the Fall2021 folder and add access=readonly to the end of the libname statement. Create a fileref to the pdf file for output.
3. Write a single SAS step that will use the Covid permanent data set as input and create three data sets as described in more detail below. Everything in this step must be done as efficiently as possible based on the information you have available.
    a. Use a conditional statement that will write out a blue note and the contents of the PDV before the set statement on only the first two iterations of the data step. The message in the note should read "PDV Before Set Statement".
    b. The three data sets will only contain rows from the state of Texas.
    c. Since all rows will be from Texas, the state and continent variables are not needed. The data source name is not to be included in the output data sets. Exclude any column whose name begins with country.
    d. The first data set will be a temporary dataset of pre-covid data based on a POSITIVE_CASES_COUNT value of 0.
    e. The second data set will be a permanent data set of covid data where POSITIVE_CASES_COUNT is not 0.
    f. The third data set will be a permanent data set of all Texas covid data.
    g. Create a variable of the percent of cases that are fatal by dividing the value of DEATH_COUNT by the value of POSITIVE_CASES_COUNT. NOTE: Since the pre-covid data set will not have any values to compute, when the positive cases count is 0, do not process the assignment of this variable or the variable created in the next step.

    h. Create a character variable containing a fatality group value based on the percent of fatal cases. About half of the observations have a fatality rate of two percent (.02) or less. Give this group a value of Low. The majority of remaining observations have a value less than 5 percent (.05). Give this group a value of Medium. The rest of the observations (with a fatality percent of 5 percent or more) will be in the High group.
    i. Use a conditional statement that will write out a blue note and the contents of the PDV immediately before the run statement on only the **first** iteration of the data step. The message in the note should read "PDV Before Run Statement".
4. Open a PDF destination to receive your output.
5. Write a PROC step that will report a list of data sets in the mylib library without reporting the descriptor portion of the data sets. Supply an appropriate title.
6. Write another PROC step that will report the descriptor portion of the temporary data set created above. Supply an appropriate title.
7. Local media outlets often refer to the area between Baylor University and TAMU as the Brazos Valley. This area encompasses McLennan, Falls, Robertson, and Brazos counties. Write a PROC step that will report the data portion of the permanent data set from step 3e for the Brazos Valley counties on a specific day. Supply a title like Brazos Valley Covid Data as of 01Sep2020 but use a macro variable instead of hard coding the date. Construct the subsetting statement so it can use the same macro variable that is used in the title. Ahead of the Title statement and PROC step, write two assignment statements for the macro variable. The first assignment will supply a value for September 1, 2020, and the second a value of September 1, 2021. Execute the first macro assignment statement then execute the Title statement and PROC step. Execute the second assignment statement along with the Title statement and PROC step again. Each execution should produce a page in the output with data from 4 observations. Be sure you capture the log from each execution.
8. Close the PDF destination.
9. Use the log and report information contained in your PDF output document to find the answers to the questions below and include the answers in a comment section at the bottom of your program file:
    a. Describe and explain the differences between the three PDVs written to the log.
    b. How many observations were read from the Covid data set that was created in the previous assignment?
    c. How many observations were written out to the temporary data set?
    d. Which county had the highest death_count on the 2020 date?
    e. Which county had the highest Highest % Fatal cases on the 2020 date?
    f. Describe and explain the changes in Percent Fatal Cases from the 2020 date to the 2021 date?
10. Save the final version of the program and convert it to a PDF file with a name like FKincheloe_HW10_prog.pdf. Convert the log to PDF.
11. Upload and submit the three documents to the assignment on Canvas.