Stat 604
Assignment 11 - SAS

OBJECTIVES: In this assignment you will practice using functions to alter data from SAS data sets.

All tools needed for completing this assignment have been covered in the first 8 SAS lessons. However, one of the concepts is covered in more detail in the first few minutes of Lesson 9.

This assignment will use the **"All Texas" -** permanent data set that was created in the previous assignment. If you had difficulty creating this data set, the professor's version, named **alltx.sas7bdat,** is available on Canvas and in the Fall2021 folder on SoDA. You will also use the **tabled1x.sas7bdat** data set that is posted to Canvas and SoDA. The **tabled1x** data set was created from an Excel file downloaded from the Bureau of Labor Statistics. The numbers shown represent the number of seasonally adjusted jobs in thousands in each month listed by industry sector and state. It will be very important for you to know the structure of this data set to be able to successfully complete this assignment. Read each numbered step completely before starting to write the code for the step.

1. The first Covid19 case in Texas was reported on February 12, 2020. Below the program header, include a macro assignment statement to create a macro variable that contains this date in a manner that can be used throughout the program in data step statements and in titles.
2. Include housekeeping statements to clear titles and footnotes and suppress the printing of procedure titles.
3. Assign a libref to the **mylib** folder containing your permanent data sets. Downloaded homework files must be in a separate folder from the mylib folder. Assign a libref to the homework data folder and add access=readonly to the end of the statement to prevent accidental corruption of the original data. Create a fileref to the pdf file for output.
4. Write a single SAS step that will use the **"All Texas"** permanent data set as input and create a permanent "Jobs" data set in mylib with the following modifications:
    a. Change the way the following variables are displayed without changing the underlying data: Percent Fatal Cases as a percentage with 3 decimal places, Report_Date like 10/29/21, death_count and positive_cases_count with comma separators and no decimal places.
    b. Convert the County_FIPS_Number variable to character. It must have the same name in the output data set and use no more spaces than necessary. There is to be no note in the log about numeric to character conversion.
    c. Create a new variable that contains the full weekday name of the Report_Date. This can be done with a slight modification to one of the conversion expressions demonstrated in the lecture slides
    d. Create a new "Covid Week" variable that contains the week number of the Report_Date relative to the date of the first Covid case. In other words, all dates reported in the same week as Feb. 12, 2020, will be week 0. Those in the prior week will be -1, etc. Use the macro variable in this expression so we can change the reference point if we want.
5. Write a single step that will use the **tabled1x** data set as input and create a permanent data set in mylib with the following modifications:
    a. For efficiency, do not read into the PDV any observations that have a missing state value.

    b. Some of the state names have a footnote number appended to them in the form of a number enclosed in a set of parentheses. We want the value in the variable named State to contain only the actual name of the state. But we want to preserve the original value. Use a data set option to change the name of the original state variable. When the original state value ends with the number in parentheses, assign the portion of the value prior to the parenthesis to the State variable. Otherwise, assign the original value to the State variable.
    c. Use a variable list in the mean function to create a new variable that is the average of the values in Aug_2017 and Aug_2018. Make sure the name will not cause a "circular" reference should variable lists be used on the new data set.
    d. Include a statement that will delete the row and return to the top of the data step when the new average value is missing.
    e. Use a variable list in the sum function to create a new variable with the total of jobs from all of the 2017 months.
    f. Use a variable list in the sum function to create a new variable with the total of jobs from all of the 2018 months.
6. Close all output destinations. Open a PDF destination to receive your output. Suppress the creation of bookmarks in the PDF file.
7. Write a PROC step that will report the descriptor portion of the first permanent data set created above in step 4. Use "Texas Covid History" as the first title and "Descriptor Portion" as the second title.
8. Produce a report from this permanent data set where the county_fips_number is 48029 and the covid week value is between -1 and 1. This fips number is from Bexar County where the first Covid case in Texas was reported. Change only the second title to be "Bexar County Data around 12Feb2020". Use the macro variable instead of the literal date to construct the title.
9. Print the descriptor portion of the permanent data set created in step 5. The printout must list the variables in creation order. Use "2017-2018 Jobs Data" as the first title and "Descriptor Portion" as the second title.
10. Print all of the data portion of this data set and change the second title to "Data Portion".
11. Close the PDF destination.
12. Use the information you discovered about the downloaded data, the log and the 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. On what day of the week was the first case reported in Bexar County?
    b. What was the Positive_Cases_Count on Saturday of Covid Week 1 in Bexar County?
    c. How many observations are in the tabled1x data set, how many were read in by the data step and how many were written out?
    d. How does the average number of August Government jobs in the District of Columbia compare with Texas? (I know it's hard to follow the split table. Use the Obs value to link the two sections together.)
13. Save the final version of the program and convert it to a PDF file with a name like FKincheloe_HW11_prog.pdf. Convert the log to PDF.
14. Upload and submit the three documents to the assignment on Canvas.