STA 402/502 Homework 10

Due: November 12 (Monday), 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.

Do the following exercise and make sure you do not edit the given permanent dataset. Use ODS RTF to save the created tables (as well as corresponding titles) and include them in your homework, do not screenshot the output.

1. The United States Department of Transportation publishes statistics on many modes of transportation. The SAS data set called AIRTRAFFIC contains data on the number of flights and passengers leaving 12 major airports in the United States. The data set contains data from each airline, for each quarter, for 20 years. In addition to the variables for year, quarter, and airline, there are two variables representing number of flights and passengers for each of the 12 airports. The variable names for the flights and passengers all start with the three-letter airport code: ATL (Atlanta), BOS (Boston), DEN (Denver), DFW (Dallas Fort Worth), EWR (Newark), HNL (Honolulu), LAX (Los Angeles), MIA (Miami), ORD (Chicago), SAN (San Diego), SEA (Seattle), and SFO (San Francisco). The data are sorted by year, airline, and quarter.

   (a) Examine this SAS data set including the variable labels and attributes. For BOS (Boston), create a data set that contains variables for the sum of the flights and for the sum of the passengers for (**402**: each year; **502**: each year and each quarter), for each airline in year 1991. Use a macro variable to specify the value of the selected year.
   Print the 18th-25th observations of the dataset in your homework.

(b) Use the data set from part (a), create a variable that represents the number of passengers per flight for (**402**: year 1991; **502**: each quarters of 1991) and airline. Round this value to a whole number. Then create another dataset that contains one observation for the airline with the highest number of passengers per flight for (**402**: year 1991; **502**: each quarters of year 1991).

Print the final dataset in your homework.

(c) Convert your code for parts (a) and (b) into a macro so that it can be run for any input data having similar set up, for any year and any airport. **Make sure the table generated by the macro has appropriate label and format.** Call the macro once for each year from 2000 to 2010 for the Chicago airport. (Hint: You may want to specify the output dataset as a variable in your macro. You will get either 1 (STA 402) or 4 (STA 502) observations in the output dataset based on the macro.)

(d) Combine the 11 data sets generated from part (c) into a bigger dataset. Print the dataset with appropriate title and variable names.

2. Suppose that you have mortgage application data from a national bank with five branches in California contained in a SAS data set called LOANAPP. The Board of Directors would like some information about the approved loans. Variables in this data set include the identification information for the applicant, credit score, loan information, home price, and the down payment as a percent of the sales price.

   Write a macro that can create a table which computes the following variables for different property types in a specific branch, make sure the property types are displayed as "Primary Residence", $\cdots$, "Commercial Property" instead of 1,2,3,4:

   - The number of approved loans;
   - The mean of an arbitrary variable, for which people can define the label and format;
   - The median of an arbitrary variable, for which people can define the label and format;

   The macro should work by changing the branch number, or change the property type to mortgage type. The generated table should have a title "Summary for Branch XX", where "XX" is the branch number,

which displays automatically when you call the macro. Make sure the variables have appropriate labels.

Use this macro, generate the following two tables and include them in your homework (No additional modification or procedures are allowed other than call the macro):

(a) A table that summarizes the total number of approved loans, the mean loan amount (values include a dollar sign, separated by comma and have 2 decimal points), the median down payment percentage (values include a percentage sign and 2 decimal points) for different property types, for branch 1.

(b) A table that summarizes the total number of approved loans, the mean credit score (values are integers), the purchase price (values include a dollar sign, separated by comma and have 3 decimal points) for different mortgage types, for branch 3.

Hint: The following is a sample code that calculates the number of loan approvals with less than 5% down payment by branch, and displays the branch names as "LIV925", $\cdots$, "COR760" instead of "1,2,3,4,5".

```
proc format;
value bname  1='LIV925'
             2='SV408'
             3='SLO805'
             4='GLN626'
             5='COR760';
run;


proc sql;
select branch "Branch name" format=bname.,
sum(LoanApproved) as number "Number of loan approvals"
from loanapp where PercentDown<0.05 group by branch;
quit;
```

3. Multiple Choice Questions (2pt for each question, you may just provide the answer.)

   (a) Given the SAS data set SASUSER.HIGHWAY:

```
Steering Seatbelt Speed Status Count
-------- -------- ----- ------- -----
absent No 0-29 serious 31
absent No 0-29 not 1419
absent No 30-49 serious 191
absent no 30-49 not 2004
absent no 50+ serious 216
```

The following SAS program is submitted:

```
%macro HIGHWAY(Belt=no);
proc print data=SASUSER.HIGHWAY;
where Seatbelt="&Belt" ;
run;
%mend;
%HIGHWAY(Belt=No)
```

How many observations appear in the generated report?

A. 0

B. 2

C. 3

D. 5

(b) Given the data set SASHELP.CLASS:

```
Name Age
------- ---
Mary 15
Philip 16
Robert 12
Ronald 15
```

The following SAS program is submitted:

```
%macro MP_ONE(pname=means);
proc &pname data=SASHELP.CLASS;
run;
%mend;

%MP_ONE(print)
%MP_ONE()
```

Which PROC steps execute successfully?

A. PROC MEANS only
B. PROC PRINT only
C. PROC MEANS and PROC PRINT
D. No PROC steps execute successfully

(c) The following SAS program is submitted:

```
%macro COLS1;
Name Age;
%mend;

%macro COLS2;
Height Weight;
%mend;

proc print data=SASHELP.CLASS;
[_insert_VAR_statement_here_]
run;
```

Which VAR statement successfully completes the program to produce a report containing four variables?

A. `var %COLS1 %COLS2;`
B. `var %COLS1-%COLS2;`
C. `var %COLS1 Weight Height;`
D. `var Weight Height %COLS1;`

(d) The following SAS program is submitted:

```
%let Value=9;
%let Add=5;
%let Newval=%eval(&Value/&Add);
%put &Newval;
```

What is the value of the macro variable Newval when the %PUT statement executes?

A. 0.555
B. 1
C. 1.8
D. 2

(e) The following SAS program is submitted:

```
proc contents data=TESTDATA.ONE;
run;
```

Which SQL procedure step produces similar information about the column attributes of TESTDATA.ONE?

A. proc sql; contents from TESTDATA.ONE; quit;
B. proc sql; describe from TESTDATA.ONE; quit;
C. proc sql; contents table TESTDATA.ONE; quit;
D. proc sql; describe table TESTDATA.ONE; quit;