STA 402/502 Homework 7

Due: October 22 (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.

1. The United States Bureau of Labor Statistics publishes various indexes that measure average prices of consumer goods in urban areas. The SAS data set called GAS contains data on the average price of unleaded regular gasoline (per gallon) for recent years by month in the United States. The variables in this file are year, month, and average gasoline price.

   (a) Review the names, labels, and attributes of the variables in the SAS data set GAS. Record the label and variable type for GasPrice as a comment in your program.

   (b) Report the minimum, maximum and average gasoline price per year. Present the price statistics to two decimal places. (Hint: use PROC REPORT)

   (c) Calculate the average and standard deviation of gasoline prices per quarter per year. Report price statistics to three decimal places and add dollar sign in front of the price statistics. (Hint: first create a new variable indicating the quarters for each year: January-March: first quarter, April-June: second quarter, $\cdots$, September-December: fourth quarter. Use PROC REPORT)

2. The World Health Organization (WHO) collected data in countries across the world regarding the outbreak of swine flu cases and deaths in 2009. The data in the SAS data set called SFF includes information

on cases and deaths per country by month during the epidemic. This question is to practice using RETAIN.

(a) Review the names, labels, and attributes of the variables in the SAS data set SFF. Record the name and length of the character variables as a comment in your program.

(b) Create a variable that shows the cumulative sum of the number of cumulative deaths reported on the first day of the month for October across all countries. Create a new dataset only contain the following three variables: number of cumulative deaths reported on the first day of the month for October, continent and the created variable. Print the 75th to 80th observation of this new dataset in your homework. (Hint: use the sum(A,B) to calculate the summation of A and B in your code.)

(c) Count the number of countries within each continent. Create and print a dataset containing only the continent name and the count. (The dataset only contains 6 observations)

(d) To find potential errors in the data, create a new dataset containing countries that reported a first death date,but reported no first case date. This dataset should include only the variables continent, country, first case date, last reported number of cases, and first death date.

(e) (STA 502 only) Create a new dataset containing the following variables (this dataset only has 6 observations). Then use PROC PRINT to print the dataset so that the column names are variable labels instead of variable names. (Hint: use proc print data=datasetname label; to print the dataset with labels as header.)

- Continent name.
- A variable that counts the number of countries per continent that reported no cases for the number of cumulative cases reported on the first day of August. Label this variable as "Number of countries reported no cases".
- A variable that counts the number of countries per continent that reported at least one case for the number of cumulative cases reported on the first day of August. Label this variable as "Number of countries reported cases".
- A variable that calculates the summation of cumulative cases per continent for the number of cumulative cases reported on the first day of August.

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

   (a) Given the following SAS program:

```
proc format;
value agegrp low-12 ='Pre-Teen'
13-high = 'Teen';
run;

proc means data=SASHELP.CLASS;
var Height;
class Sex Age;
format Age agegrp.;
run;
```

   Which statement in the proc means step needs to be modified or added to generate the following results:

Analysis Variable : Height					
Sex	Age	N Obs	Minimum	Maximum	Mean
F	Pre-Teen	3	51.3	59.8	55.8
	Teen	6	56.5	66.5	63.0
M	Pre-Teen	4	57.3	64.8	59.7
	Teen	6	62.5	72.0	66.8

   A. var Height / nobs min max mean maxdec=1;
   B. proc means data=SASHELP.CLASS maxdec=1 ;
   C. proc means data=SASHELP.CLASS min max mean maxdec=1;
   D. output nobs min max mean maxdec=1;

   (b) Consider the following data step:

```
data WORK.NEW;
set WORK.OLD(keep=X);
if X < 10 then X=1;
else if X >= 10 AND X LT 20 then X=2;
else X=3;
run;
```

In filtering the values of the variable X in data set WORK.OLD, what value new value would be assigned to X if its original value was a missing value?

A. This step does not run because of syntax errors.
B. X would get a value of 3.
C. X would retain its original value of missing.
D. X would get a value of 1.

(c) Given the following code:

```
proc print data=SASHELP.CLASS(firstobs=7 obs=15);
where Sex='M';
run;
```

How many observations will be displayed?

A. 8
B. 9
C. 8 or fewer
D. 9 or fewer

(d) The following SAS program is submitted:

```
data date;
Month="01";
Yr="1960";
X=mdy(Month,01,Yr) ;
run;
```

What is the value of the variable X?

A. the numeric value 0
B. the character value 01011960
C. a missing value due to syntax errors
D. the step will not compile because of the character argument in the mdy function.

(e) The following SAS program is submitted:

```
data WORK.TEST;
set WORK.MEASLES(keep=Janpt Febpt Marpt);
array Diff{3}  Difcount1-Difcount3;
array Patients{3} Janpt Febpt Marpt;
run;
```

What new variables are created?

A. Janpt, Febpt, and Marpt
B. Diff1, Diff2 and Diff3
C. Difcount1, Difcount2 and Difcount3
D. Patients1, Patients2 and Patients3