STA 402/502 Homework 6

Due: October 15th (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.

1. Refer to the motivating example we discussed on Friday during class, each patient is scheduled to have three follow up visits. The visits should be every 30 days starting from the baseline visit date. Create three new variables corresponding to the scheduled follow up visit dates for the patients. The baseline and the three follow up visit dates should display according to the following four format (use the date October 5, 2018 as an example) listed below respectively. In your report, include the first 5 observations from the dataset.

   - 05OCT2018
   - 05/10/18
   - Friday, October 05, 2018
   - 2018-10-05

2. This question is to distinguish between the WHERE and IF statement when merging datasets. WHERE statement will subset the data before merging and IF statement will subset the data after the datasets have been merged.

   In the program after this introduction, replace the gaps marked by triangles $\triangle 1$ and $\triangle 2$ with lines of data so that:

   - The data sets DS1 and DS2 each have exactly four observations and no missing values;
   - The data set DS_Where has exactly one observation for each value $w = 1, 2, 3, 4, 5$, which appear in that order.

- Design your data sets so that the programs output shows the following properties:
- The $w = 1$ observations are identical in DS_Where and DS_If, with no missing values;
- The $w = 2$ observations are identical in DS_Where and DS_If, with a missing value for y;
- The $w = 3$ observations are identical in DS_Where and DS_If, with a missing value for z;
- DS_Where has a $w = 4$ observation with at least one missing value, but the $w = 4$ observation in DS_If has no missing values;
- DS_If has no observation with $w = 5$.

A copy of the program (where_if.sas) can be downloaded with this assignment on Canvas. You may not change any aspect of the program except for inserting data as specified above .

(a) Include all the printed output as well as the updated code in your homework.

(b) Use the same datasets ds1 and ds2, create a merged dataset by w, which consists of observations only appeared in data ds2. Include the printed dataset.

```
data ds1;
input w x y;
datalines;

△1 <<<======Replace this with data

run;

data ds2;
input w x z;
datalines;

△2 <<<====== Replace this with data

run;

data  ds_where;
```

```
merge ds1 ds2;
by w;
where x<7;
run;

data ds_if;
merge ds1 ds2;
by w;
if x<7;
run;

proc print data=ds1 noobs;
title "First dataset";
run;
proc print data=ds2 noobs;
title "Second dataset";
run;
proc print data=ds_where noobs;
title "Merged with WHERE";
run;
proc print data=ds_if noobs;
title "Merged with IF";
run;
```

3. A gourmet pizza restaurant is considering adding new toppings to its menu. Each month they survey ten customers to rate their preferences for three different toppings. The SAS data set called PIZZARATINGS contains data for respondent ID, topping type, and rating.

   (a) Create a format to be used on the Rating variable so that the data values are presented as n/a (missing values), never (1), might (2), at least once (3), occasionally (4), and often (5). Apply this format to the dataset.

   (b) For each topping type, count the number of observations for each type of response (including the group "missing values"). (Hint: You may use PROC FREQ procedure with some option, so that the percentage is not included while the missing value is included in the table.)

   (c) Calculate the average rating for each topping and include the result in the homework.

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

   (a) The following SAS program is submitted:

```
data WORK.TEST;
set WORK.DRIVER;
if Jobcode='Driver2' then Description='Senior Driver';
else Description='Unknown';
run;
```

   The value for the variable "Jobcode" is: DRIVER2. What is the value of the variable "Description"?

   A. DRIVER2
   B. ' ' (missing character value)
   C. Senior Driver
   D. Unknown

   (b) Given the contents of the raw data file 'EMPLOYEE.TXT':

```
----+----10---+----20---+----30--
Xing           2 19 2004 ACCT
Bob            5 22 2004 MKTG
Jorge          3 14 2004 EDUC
```

   The following SAS program is submitted:

```
data WORK.EMPLOYEE;
infile 'EMPLOYEE.TXT';
input
@1  FirstName $
@15 StartDate
@25 Department $;
run;
```

   Which SAS informat correctly completes the program?

   A. date9.
   B. mmddyy10.
   C. ddmmyy10.
   D. mondayyr10.

   (c) The following is a SAS program.

```
proc format;
value score 1-50 =‘Fail’
51-100 = ‘Pass’;
run;

data course;
input exam;
format exam score.;
datalines;
50.5
;
run;
```

What is the value for exam?

A. Fail
B. Pass
C. 50.5
D. No output

(d) Given the SAS data set WORK.ONE:

```
N   BeginDate
-   ---------
1   09JAN2010
2   12JAN2010
```

The following SAS program is submitted:

```
data WORK.TWO;
set WORK.ONE;
Day=<insert code here, choose from the options below>;
format BeginDate date9.;
run;
```

The data set WORK.TWO is created, where Day would be 1 for Sunday, 2 for Monday, 3 for Tuesday, ... :

```
WORK.TWO

N   BeginDate   Day
-   ---------   ---
1   09JAN2010     1
2   12JAN2010     4
```

Which expression successfully completed the program and creates the variable Day?

A. day(BeginDate)

B. weekday(BeginDate)

C. dayofweek(BeginDate)

D. getday(BeginDate,today())

(e) Given the SAS data set WORK.EMP_NAME:

```
Name  EmpID
----  -----
Jill   1864
Jack   2121
Joan   4698
John   5463
```

Given the SAS data set WORK.EMP_DEPT:

```
EmpID  Department
-----  ----------
2121  Accounting
3567  Finance
4698  Marketing
5463  Accounting
```

The following program is submitted:

```
data WORK.ALL;
merge WORK.EMP_NAME(in=Emp_N)
WORK.EMP_DEPT(in=Emp_D);
by Empid;
if (Emp_N and not Emp_D) or (Emp_D and not Emp_N);
run;
```

How many observations are in data set WORK.ALL after submitting the program?
(Note that the statement

```
if Emp_N and not Emp_D
```

is equal to

```
if Emp_N=1 and Emp_D=0 )
```

A. 1

B. 2
C. 3
D. 5