**STA502 Homework 6**

**Lina Lee**

**Question 1**

**Code:**

/\*header: Exercise Number 1

Purpose: 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

Input: [visits.sas7bdat](https://miamioh.instructure.com/courses/83265/files/10108277/download?verifier=9cnHEyLSIlqGCuou9opOwrcqHM2qJqEWB9bg5euH&wrap=1)

\*/

ods rtf file="C:\Users\linal\Desktop\2018\STA502\HW6\problem1" bodytitle style=journal;

/\*Create three new variables corresponding to the scheduled follow up visit dates for the patients,

and display according to the four format\*/

data visits;

set day16.visits;

followup1=VisitDt+30;

followup2=VisitDt+60;

followup3=VisitDt+90;

format VisitDt DATE9.;

format followup1 DDMMYY.;

format followup2 WEEKDATE.;

format followup3 YYMMDDD10.;

run;

/\*print the first 5 observations from the dataset\*/

proc print data=visits(obs=5);

run;

ods rtf close;

**Results:**

***Follow-up date***

| *Obs* | *ID* | *VisitDt* | *Gender* | *Visit* | *B\_Cholesterol* | *followup1* | *followup2* | *followup3* |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *1* | 178772 | 31DEC2013 | Male | 0 | 228 | 30/01/14 | Saturday, March 1, 2014 | 2014-03-31 |
| *2* | 300732 | 31DEC2013 | Female | 0 | 236 | 30/01/14 | Saturday, March 1, 2014 | 2014-03-31 |
| *3* | 372409 | 31DEC2013 | Male | 0 | 239 | 30/01/14 | Saturday, March 1, 2014 | 2014-03-31 |
| *4* | 418589 | 31DEC2013 | Female | 0 | 232 | 30/01/14 | Saturday, March 1, 2014 | 2014-03-31 |
| *5* | 472192 | 31DEC2013 | Female | 0 | 223 | 30/01/14 | Saturday, March 1, 2014 | 2014-03-31 |

**Question 2**

**Code:**

/\*header: Exercise Number 1

Purpose: create the dataset which satisfy requirements by completing datalines.

Input: [where\_if.sas](https://miamioh.instructure.com/courses/83265/files/10108191/download?verifier=7AxcO4VhybHvhbLzdZr6PHa1w02NSoZG0kOgLrSq&wrap=1)\*/

/\* This program shows the difference between WHERE and IF in subsetting the context of a data-set MERGE. \*/

data ds1;

input w x y;

datalines;

1 1 0

3 1 0

4 7 0

5 0 0

;

run;

data ds2;

input w x z;

datalines;

1 1 0

2 1 0

4 3 3

5 7 3

;

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;

**Code:**

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;

/\*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 in\_1\_alt;

merge ds1(in=indata1)

ds2(in=indata2);

by w;

if indata2=1;

run;

proc print data=in\_1\_alt;

title=”merged by w”

run;

**Results: (a) all the printed output are below and the updated code is in above code**

***First dataset Second dataset***

| *w* | *x* | *y* |
| --- | --- | --- |
| 1 | 1 | 0 |
| 3 | 1 | 0 |
| 4 | 7 | 0 |
| 5 | 0 | 0 |

| *w* | *x* | *z* |
| --- | --- | --- |
| 1 | 1 | 0 |
| 2 | 1 | 0 |
| 4 | 3 | 3 |
| 5 | 7 | 3 |

***Merged with WHERE Merged with IF***

| *w* | *x* | *y* | *z* |
| --- | --- | --- | --- |
| 1 | 1 | 0 | 0 |
| 2 | 1 | . | 0 |
| 3 | 1 | 0 | . |
| 4 | 3 | . | 3 |
| 5 | 0 | 0 | . |

|  | *x* | *y* | *z* |
| --- | --- | --- | --- |
| 1 | 1 | 0 | 0 |
| 2 | 1 | . | 0 |
| 3 | 1 | 0 | . |
| 4 | 3 | 0 | 3 |

| *bs* | *w* | *x* | *y* | *z* |
| --- | --- | --- | --- | --- |

| *Obs* | *w* | *x* | *y* | *z* |
| --- | --- | --- | --- | --- |
| *1* | 1 | 1 | 0 | 0 |
| *2* | 2 | 1 | . | 0 |
| *3* | 4 | 3 | 0 | 3 |
| *4* | 5 | 7 | 0 | 3 |

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

***Merged by W***

**Question 3:**

**Code:**

ods rtf file="C:\Users\linal\Desktop\2018\STA502\HW6\problem3" bodytitle style=journal;

%let data\_folder=C:\Users\linal\Desktop\2018\STA502WHW6;

/\*number3\*/

data pizza;

set tmp1.pizzaratings;

run;

/\*(a) Create a format to be used on the Rating variable\*/

proc format;

value Rating .="n/a"

1="never"

2="might"

3="at least once"

4="occasionally"

5="often";

run;

data pizza2;

set pizza;

format rating Rating.;

run;

/\*For each topping type, count the number of observations for each type of response\*/

proc freq data=pizza2;

table topping\*rating / missprint;

run;

/\*Calculate the average rating for each topping\*/

proc means data=pizza2 maxdec=2;

class topping;

var rating;

run;

ods rtf close;

ods rtf file="C:\Users\linal\Desktop\2018\STA502\HW6\problem3" bodytitle style=journal;

%let data\_folder=C:\Users\linal\Desktop\2018\STA502WHW6;

/\*number3\*/

data pizza;

set tmp1.pizzaratings;

run;

/\*(a) Create a format to be used on the Rating variable\*/

proc format;

value Rating .="n/a"

1="never"

2="might"

3="at least once"

4="occasionally"

5="often";

run;

/\*For each topping type, count the number of observations for each type of response\*/

proc freq data=pizza;

table topping\*rating / missprint;

run;

/\*Calculate the average rating for each topping\*/

proc means data=pizza maxdec=2;

class topping;

var rating;

run;

ods rtf close;

|  |
| --- |
| ***The FREQ Procedure*** |

| *Table of Topping by Rating* | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Topping(Type of topping)* | *Rating(Customer rating (1 = would never order to 5 = would order often))* | | | | | | |
| *Frequency Percent Row Pct Col Pct* | *n/a* | *never* | *might* | *at least once* | *occasionally* | *often* | *Total* |
| *Arugula* | 80 . . . | 9 2.50 22.50 13.43 | 6 1.67 15.00 8.70 | 7 1.94 17.50 9.46 | 9 2.50 22.50 11.69 | 9 2.50 22.50 12.33 | 40 11.11 |
| *Eggplant* | 70 . . . | 15 4.17 30.00 22.39 | 7 1.94 14.00 10.14 | 6 1.67 12.00 8.11 | 14 3.89 28.00 18.18 | 8 2.22 16.00 10.96 | 50 13.89 |
| *PineNut* | 20 . . . | 10 2.78 10.00 14.93 | 26 7.22 26.00 37.68 | 24 6.67 24.00 32.43 | 20 5.56 20.00 25.97 | 20 5.56 20.00 27.40 | 100 27.78 |
| *Shrimp* | 30 . . . | 16 4.44 17.78 23.88 | 19 5.28 21.11 27.54 | 20 5.56 22.22 27.03 | 22 6.11 24.44 28.57 | 13 3.61 14.44 17.81 | 90 25.00 |
| *Squash* | 40 . . . | 17 4.72 21.25 25.37 | 11 3.06 13.75 15.94 | 17 4.72 21.25 22.97 | 12 3.33 15.00 15.58 | 23 6.39 28.75 31.51 | 80 22.22 |
| *Total* | . . | 67 18.61 | 69 19.17 | 74 20.56 | 77 21.39 | 73 20.28 | 360 100.00 |
| *Frequency Missing = 240* | | | | | | | |

***The MEANS Procedure***

| *Analysis Variable : Rating Customer rating*  *(1 = would never order to 5 = would order often)* | | | | |
| --- | --- | --- | --- | --- |
| *Type of topping* | *N Obs* | *N* | *Mean* |
| Arugula | 120 | 40 | 3.08 |
| Eggplant | 120 | 50 | 2.86 |
| PineNut | 120 | 100 | 3.14 |
| Shrimp | 120 | 90 | 2.97 |
| Squash | 120 | 80 | 3.16 |

**Question 4**

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 \_le '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 sub-

mitting 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

6

B. 2

C. 3

D. 5