PSTAT 130 Final (2hr) – Summer 2017

Name _____   Perm ID_____

I confirm that I will not use SAS, notes, books, cheat sheet, cell-phones and I will not cheat in PSTAT130 midterm Exam. (sign here) _____.

Multiple Choice Questions:

1. If you submit the following program, what information will appear in the **cardiac** data set?

  data work.cardiac(drop=age group);
   set clinic.fitness(keep=age weight group);
   if group=2 and age>=40;
  run;

a. Includes information about group=2

b. Includes information about group=2 and age greater than or equal to 40

c. Includes weights of group=2 and age greater than or equal to 40

d. Includes weights and more of group=2 and age greater than or equal to 40

2. For the observation shown below, what is the result of the IF-THEN statements?

Status	Type	Count	Action	Control
YES	1	12	A	Go

  if status ='YES' and type=1
   then Count+1;
  if status='NO' or action='A'
   then Control='NoGo';

a. Count = 13, Control=Go

b. Count = 12, Control = NoGo

c. Count = 13, Control = NoGo

d. Count = 12, Control = Go

3. Which set of statements is the most efficient equivalent to the code shown below?

  if code='l' then Standing='Freshman';
  if code='2' then Standing='Sophomore';
  if code='3' then Standing='Junior';
  if code='4' then Standing='Senior';

a. if code='l' then Standing='Freshman';
   else if code='2' then Standing='Sophomore';
   else if code='3' then Standing='Junior';
   else Standing='Senior';

b. if code='l' then Standing='Freshman';
   else if code='2' then Standing='Sophomore';
   else if code='3' then Standing='Junior';
   else if Standing='Senior';

c.   if code='l' then Standing='Freshman';
        else if code='2' then Standing='Sophomore';
        else if code='3' then Standing='Junior';
        else if code='4' then Standing='Senior';

d.   if code='l' then Standing='Freshman';
        else if code='2' then Standing='Sophomore';
        else if code='3' then Standing='Junior';
        else if then Standing='Senior';

4. Which program will combine **Actors.Props1** and **Actors.Props2** to produce **Actors.Props3**?

Actors.Props1		+	**Actors.Props2**		=	**Actors.Props3**	
Actor	**Prop**		**Actor**	**Prop**		**Actor**	**Prop**
Curly	Anvil		Tim	Ladder		Curly	Anvil
Larry	Hammer		Cho	Plier		Larry	Hammer
Moe	Poker					Moe	Poker
						Tim	Ladder
						Cho	Plier

a. **Data actors.props3;**
      merge actors.props1 with actors.props2;
          by Actor;
    run;

b. **Data actors.props3;**
      set actors.props1 with actors.prop2;
   by actor;
   run;

c. **Data actors.props3;**
      set actors.props1 actors.props2;
   run;

d. **Data actors.props3;**
      merge actors.props1 actors.props2;
      by Actor;
   run;

5. If you merge data sets **Sales.Reps**, **Sales.Close**, and **Sales.Bonus** by **ID**, what is the total number of observations in the new dataset?

Sales.Reps		Sales.Close		Sales.Bonus	
ID	**Name**	**ID**	**Sale**	**ID**	**Bonus**
1	Nay Rong	1	$28,000	1	$2,000
2	Kelly Windsor	2	$30,000	2	$4,000
3	Julio Meraz	2	$40,000	3	$3,000
4	Richard Krabill	3	$15,000	4	$2,500
		3	$20,000		
		3	$25,000		
		4	$35,000		

a. 6

b. 7

c. 8

d. 9

6. What happens if you merge (Data1st and Data2nd) the following data sets by variable SSN?

1st		2nd		
SSN	**Age**	**SSN**	**Age**	**Date**
029-46-9261	39	029-46-9261	37	02/15/95
074-53-9892	34	074-53-9892	32	05/22/97
228-88-9649	32	228-88-9649	30	03/04/96
442-21-8075	12	442-21-8075	10	11/22/95
446-93-2122	36	446-93-2122	34	07/08/96
776-84-5391	28	776-84-5391	26	12/15/96
929-75-0218	27	929-75-0218	25	04/30/97

a. The merged data will have 4 variables and 7 rows.

b. The values of Age in the 1st data set will be listed next to Age of the 2nd data set.

c. The DATA step fails because the two data sets contain same-named variables that have different values.

d. The values of Age in the 2nd data set overwrite the values of Age in the 1st data set.

7. When you merge two data sets in SAS, you must

a. Sort them first on at least one common variable and use a SET statement

b. Do not sort them but list them both in a SET statement in the Data step

c. Sort them first on at least one common variable then include MERGE statement in the Data step

d. Set them one next to another

8. The default statistics produced by the PROC FREQUENCY procedure of one variable include

a. Frequency, Counts, Cumulative Frequency, Cumulative Counts

b. Mean, Median, Minimum, Maximum

c. Frequency, Percent, Cumulative Frequency, Cumulative Percent

d. Mean, Standard deviation, Minimum, Maximum

9. Which PROC FREQ step most likely produced this two-way table?

Table of Eyes by Hair						
Eyes(Eye Color)	**Hair(Hair Color)**					
Frequency **Percent** **Row Pct** **Col Pct**	**black**	**dark**	**fair**	**medium**	**red**	**Total**
blue	1 3.70 11.11 33.33	2 7.41 22.22 33.33	2 7.41 22.22 33.33	2 7.41 22.22 33.33	2 7.41 22.22 33.33	9 33.33
brown	2 7.41 20.00 66.67	2 7.41 20.00 33.33	2 7.41 20.00 33.33	2 7.41 20.00 33.33	2 7.41 20.00 33.33	10 37.04
green	0 0.00 0.00 0.00	2 7.41 25.00 33.33	2 7.41 25.00 33.33	2 7.41 25.00 33.33	2 7.41 25.00 33.33	8 29.63
Total	3 11.11	6 22.22	6 22.22	6 22.22	6 22.22	27 100.00

a. proc freq data=color;
        var Eyes Hair Eyes/Hair;
    run;

b.   **proc freq data=color;**
        tables Eyes Hair Eyes/Hair;
    run;

c.   **proc freq data=color;**
        tables Eyes Hair Eyes*Hair;

```
            run;
    d.    proc freq data=color;
                var Eyes Hair Eyes*Hair;
            run;
```

10. The data set **Survey.Students** includes several variables. Which is a poor candidate for PROC FREQ analysis?

a. GPA

b. major

c. Gender

d. Year of study

11. The default statistics produced by the PROC MEANS procedure of one variable includes

a. n, Mean, Median, mode, Counts

b. n, Mean, Median, Minimum, Maximum

c. n, Frequency, Percent, Mean, Median

d. n, Mean, Standard deviation, Minimum, Maximum

12. What results the following SAS program produce?

proc tabulate data=energy;

    class region type;

    var expenditures;

    table region, type*expenditures;

run;

a. sums of expenditures for each region and each type.

b. averages of expenditures for each region and each type.

c. counts of expenditures for each region and each type.

d. totals and means of expenditures for each region and each type.

13. In PROC REPORT, you use the _____ statement to select variables to be summarized and you use the _____ statement to format and label the variables.

a. var, format and label

b. var, options

c. columns, options

d. columns, define

14. Can PROC REPORT produce group summary statistics?

a. Yes it produces them for all character variables

b. Yes it produces them for all variables

c. Yes it produces them for all numerical variables

d. Yes it produces them for all date variables

15. The INDEX function

a. Indexes the character string or character variable

b. extracts a portion of a character string

c. returns the position of specific character   in a character string

d. removes any trailing blank spaces from a character string

16. Suppose phonenumber variable contains values such as (806) 111-2222, (805) 222-3333, etc..
The following function substr(scan(Phonenumber,1),2,3) in SAS outputs

a. 111, 222, etc.

b. 806, 805, etc.

c. 2222, 3333, etc.

d. (806) , (805), etc.

17. The variable health **idcode** contain values of fixed length such as *123 F A-, 321 M B+, 456 F O, etc.*. The 3rd word in **idcode** identifies the blood-type of the individual (A- ,B+, etc.). How do you assign these values to a new variable named **bloodtype**?

a. bloodtype=scan(idcode,3);

b. bloodtype=substr(idcode,7,2);

c. bloodtype=index(idcode,'+') or index(idcode,'-');

d. a and b;

18. What does function round(123.45,1) do in SAS?

a. Rounds the first argument to 123.4

b. Rounds the first argument to 12345 times .01.

c. Rounds the first argument to 123.

d. None of the above

19. The function which converts numerical variable to character variable and the function which converts character variable to numerical variable are

a. PUT, 2PUT

b. CPUT, NPUT

c. PUT, INPUT

d. INPUT, PUT

20. Which of the following statements is <u>true</u> about output of the program shown below?

```
    data work.invest;
        do year=2000 to 2015;
                Capital+10000;
                capital+(capital*.08);
                output;
        end;
    run;
```

a. The last value for Year is 2016.

b. The OUTPUT statement drops the last value of the capital in DO loop.

c. The Capital at the end of 2015 is 150000+.08(150000) .

d. The DO loop performs 16 iterations.

CCACB
DCCCA
DADAC
BDCCD

Programming Questions:
1. **Controlling record OUTPUT**

    Write a single data step, based on the **employees** data set in the temporary library, that performs the following operations:
    a. Create a dataset called **FltCrew** in the temporary library that includes all records that have 'FLTAT' or 'PILOT' in the variable **JobCode**. Drop the **EmpID** and **Salary** variables from this dataset.
    b. Create a dataset called **GrndCrew** in the temporary library that includes all records that have 'MECH' or 'GRCRE' in the variable **JobCode**. Drop the **EmpID** and **Salary** variables from this dataset.

    Write a Proc Print step that prints the first 10 records of the **FltCrew** datasets, and another than prints Records 5 through 15 of the **GrndCrew** dataset.

```
data FltCrew(drop=EmpID Salary) GrndCrew(drop=EmpID Salary);
set work.employees;
if jobcode contains 'FLTAT'
then output FltCrew;
if jobcode contains 'PILOT'
then output FltCrew;
if jobcode contains 'MECH'
then output GrndCrew;
if jobcode contains 'GRCRE'
then output GrndCrew;
run;
proc print data=work.FltCrew (obs=10);
run;
proc print data=work.GrndCrew (firstobs=5 obs=10);
run;
```

Alternative solution:
```
data FltCrew(drop=EmpID Salary) GrndCrew(drop=EmpID Salary);
set work.employees;
if jobcode in ('FLTAT1','FLTAT2')
then output FltCrew;
if jobcode in ('MECH'，'GRCRE')
then output GrndCrew;
run;
```

2. Accumulating Totals for a Group of Data

The data set **flymiles** contain information about frequent flyer. The each observation represents each trip a frequent flyer made with an airline. It has two variables, which are the frequent flyer number (**ID**) and the number of miles earned for that trip (**Miles**). The below is partial list output:

Obs	ID	Miles
1	23	120
2	12	135
3	15	202
4	12	800
5	23	268
6	41	108
7	23	300
8	12	547
9	41	852
10	12	234

Create a data set named **work.freqmiles** where each observation represents each frequent flyer. It has two variables, which are frequent flyer number (**ID**) as well as a new variable named **TotMiles**, which shows the total number of frequent flyer miles the person earned.

```
proc sort data=flymiles;
by id;
run;
data work.freqmiles(keep=ID Totmiles);
set work.flymiles;
by id;
if first.ID then TotMiles=Miles;
else TotMiles=TotMiles+Miles;
if last.ID;
run;
```

3. Reading Non-Standard Data

In the directory''X:\PStat 130\data'', the **states.dat** raw data file contains information on geographic size, population and date of statehood for each of the 50 states. The order, contents and layout of the fields are as follows:

Order	Field	Notes
1	State Name	Longest value is 16 characters
2	State Population	Written in COMMA9.
3	State Size	Written in COMMA8.
4	Date of Statehood	Written in DATE9.

Sample Records

```
Alabama! 4,447,100! 50,750! 14DEC1819
Alaska! 626,932! 570,374! 03JAN1959
```

Arizona! 5,130,632! 113,642! 14FEB1912

Arkansas! 2,673,400! 52,075! 15JUN1836

California! 33,871,648! 155,973! 09SEP1850

Colorado! 4,301,261! 103,729! 01AUG1876

Connecticut! 3,405,565! 48,45! 09JAN1788

Create a Data Step to read in the States.dat data, and create a temporary dataset **States** including variables (**state, population, size, Date**). Then produce the listing report below (don't forget to change label to **size** and **Date**):

State	Population	Size in Square Miles	Date of Statehood
Alabama	4,447,100	50,750	12/14/19
Alaska	626,932	570,374	01/03/59
Arizona	5,130,632	113,642	02/14/12
Arkansas	2,673,400	52,075	06/15/36
California	33,871,648	155,973	09/09/50
Colorado	4,301,261	103,729	08/01/76
Connecticut	3,405,565	4,845	01/09/88

HINTS: The data are formatted in the data file, they contain an exclamation point as a delimiter, and you will need to use a modifier in the INPUT statement to read in variable values of different lengths.

```
data work.states;
infile "X:\PStat 130\data2\states.dat" dlm='!';
input State : $16. Population :comma9. Size : comma7. Date : date9.;
run;
proc print data=work.states label;
label size='Size in Square Miles' Date='Date of Statehood';
format Population comma9. Size comma7.   Date mmddyy8.;
run;
```

4. In the temporary library, shops data set contains the variables customer, gender, autoshop, miles, and servicedate.

Obs	customer	gender	autoshop	miles	servicedate
1	Cramer	m	TopShop	36934.6	28SEP2006
2	Abbott	f	Peerless	8240.7	18MAR2011
3	Monroe	f	Prestigious	47340.0	15JUL2009

Use your shops data set to create a file called letters.doc. This document should contain one letter for each observation, each on its own page. Each letter tells the customer the date of their next service, which is 90 days after their initial service date. See below for the desired structure of the letters:

```
Dear Mr. Cramer ,

Thank you for bringing your car in on 09/28/2006.
At the time of service, your odometer read 36,935 miles.
We have scheduled your next service appointment for 12/27/2006.
Thank you again for being our valued customer!


Regards,

TopShop Auto
```

Note: Sign each letter with the name of the autoshop, concatenated with the word 'Auto'
(i.e. Peerless Auto, Prestigious Auto, TopShop Auto).

```
options nodate nonumber;
data _NULL_;
   set shops;
   file 'X:\...\letters.doc' PRINT;
   title;
   nextservice=servicedate+90;
   if gender='m' then Salutation='Mr.';
   else Salutation='Ms.';
   put 'Dear ' Salutation customer','
       //'Thank you for bringing your car in on '
             servicedate mmddyy10.'.'
       /'At the time of service, your odometer read '
             miles comma6. ' miles.'
       /'We have scheduled your next service appointment for '
             nextservice mmddyy10. '.'
       /'Thank you again for being our valued customer!'
       ///'Regards,'
       //autoshop 'Auto';
   put _page_;
run;
```