UNIVERSITI TEKNOLOGI MARA
FINAL EXAMINATION

COURSE	:	**STATISTICAL COMPUTING**
COURSE CODE	:	**STA705**
EXAMINATION	:	**JUNE 2016**
TIME	:	**3 HOURS**

INSTRUCTIONS TO CANDIDATES

1. This question paper consists of two (2) parts : PART A (2 Questions)
   PART B (5 Questions)
2. Answer ALL questions in both parts using separate Answers Booklets. Start each answer on a new page.
3. Do not bring any material into the examination room unless permission is given by the invigilator.
4. Please check to make sure that this examination pack consists of :
   i) the Question Paper
   ii) two Answer Booklets – provided by the Faculty
5. Answer ALL questions in English.

DO NOT TURN THIS PAGE UNTIL YOU ARE TOLD TO DO SO

This examination paper consists of 11 printed pages

PART A

QUESTION 1

a) Create the following vectors using the **rep** function:

```
8 8 8 10 10 10 11 11 11 11 11 3 3 3 3 3 3 3
```

(3 marks)

b) Use the built-in function (**sort, order** and **rank**) in the following sequence:

```
86 100 25 56 75 35 42
```

Obtain the following results.

i) Sort _____

ii) Order _____

iii) Rank _____

(5 marks)

c) Using ***generate factor levels***, produce the R command of the following output:

```
   Grade Colour Location
1      A   Blue    Inner
2      A   Blue    Inner
3      A   Blue    Inner
4      A   Blue    Inner
5      A    Red    Inner
6      B    Red    Inner
7      B    Red    Inner
8      B    Red    Inner
9      B  Green    Inner
10     B  Green    Inner
11     C  Green    Outer
12     C  Green    Outer
13     C  White    Outer
14     C  White    Outer
15     C  White    Outer
16     D  White    Outer
17     D  Black    Outer
18     D  Black    Outer
19     D  Black    Outer
20     D  Black    Outer
```

(6 marks)

d) Produce the following data set which is obtained from the R random generator.

```
> data.1
 [1] 0 0 1 0 0 1 0 1 1 1 0 1 0 1 0 0 1 0 0 0 1 1 0 0 1 1 1 1 0 0
```

(3 marks)

e) Give the R command to compute the probability of success for `data.1`.

(2 marks)

f) Produce the R function to generate 50 samples of probability of success for the data obtained in part (d).

```
Find.pr=function(dtx,S)

{
      n=
      prob.b=numeric(S)
      set.seed(100)
      for (b in 1:S)
      {
       }

}
```

The output is obtained as follows:

```
> Find.pr(data.1,50)
$data.1
 [1] 0 0 1 0 0 1 0 1 1 1 0 1 0 1 0 0 1 0 0 0 1 1 0 0 1 1 1 1 0 0

$sample.prob
 [1] 0.633 0.467 0.533 0.533 0.367 0.433 0.533 0.533 0.567 0.533 0.233 0.467
[13] 0.533 0.533 0.433 0.500 0.367 0.500 0.633 0.333 0.533 0.500 0.467 0.500
[25] 0.567 0.533 0.433 0.433 0.367 0.467 0.333 0.567 0.567 0.400 0.600 0.367
[37] 0.400 0.367 0.500 0.433 0.333 0.500 0.400 0.367 0.600 0.533 0.533 0.567
[49] 0.633 0.367
```

(6 marks)

QUESTION 2

Ozone pollution data collected in Rennes, France during the summer of 2001. The variables available are `maxO3` (maximum daily ozone), `T12` (temperature at midday), `wind` (direction), `rain` and `Wx12` (projection of the wind speed vector on the east-west axis at midday). The information on the "***ozone***" dataset is given as follows:

```
> str(ozone)
'data.frame':   112 obs. of  13 variables:
 $ maxO3 : int  87 82 92 114 94 80 79 79 101 106 ...
 $ T9    : num  15.6 17 15.3 16.2 17.4 17.7 16.8 14.9 16.1 18.3 ...
 $ T12   : num  18.5 18.4 17.6 19.7 20.5 19.8 15.6 17.5 19.6 21.9 ...
 $ T15   : num  18.4 17.7 19.5 22.5 20.4 18.3 14.9 18.9 21.4 22.9 ...
 $ Ne9   : int  4 5 2 1 8 6 7 5 2 5 ...
 $ Ne12  : int  4 5 5 1 8 6 8 5 4 6 ...
 $ Ne15  : int  8 7 4 0 7 7 8 4 4 8 ...
 $ Wx9   : num  0.695 -4.33 2.954 0.985 -0.5 ...
 $ Wx12  : num  -1.71 -4 1.879 0.347 -2.954 ...
 $ Wx15  : num  -0.695 -3 0.521 -0.174 -4.33 ...
 $ maxO3y: int  84 87 82 92 114 94 80 99 79 101 ...
 $ wind  : Factor w/ 4 levels "East","North",..: 2 2 1 2 4 4 4 2 2 4 ...
 $ rain  : Factor w/ 2 levels "Dry","Rainy": 1 1 1 1 1 2 1 1 1 1 ...
```

```
> ozone
         maxO3   T9  T12  T15 Ne9 Ne12 Ne15     Wx9    Wx12    Wx15 maxO3y  wind  rain
20010601    87 15.6 18.5 18.4   4    4    8  0.6946 -1.7101 -0.6946     84 North   Dry
20010602    82 17.0 18.4 17.7   5    5    7 -4.3301 -4.0000 -3.0000     87 North   Dry
20010603    92 15.3 17.6 19.5   2    5    4  2.9544  1.8794  0.5209     82  East   Dry
20010604   114 16.2 19.7 22.5   1    1    0  0.9848  0.3473 -0.1736     92 North   Dry
20010605    94 17.4 20.5 20.4   8    8    7 -0.5000 -2.9544 -4.3301    114  West   Dry
20010606    80 17.7 19.8 18.3   6    6    7 -5.6382 -5.0000 -6.0000     94  West Rainy
20010607    79 16.8 15.6 14.9   7    8    8 -4.3301 -1.8794 -3.7588     80  West   Dry
20010610    79 14.9 17.5 18.9   5    5    4  0.0000 -1.0419 -1.3892     99 North   Dry
20010611   101 16.1 19.6 21.4   2    4    4 -0.7660 -1.0261 -2.2981     79 North   Dry
20010612   106 18.3 21.9 22.9   5    6    8  1.2856 -2.2981 -3.9392    101  West   Dry
20010613   101 17.3 19.3 20.2   7    7    3 -1.5000 -1.5000 -0.8682    106 North   Dry
20010614    90 17.6 20.3 17.4   7    6    8  0.6946 -1.0419 -0.6946    101 South   Dry
20010615    72 18.3 19.6 19.4   7    5    6 -0.8682 -2.7362 -6.8944     90 South   Dry
20010616    70 17.1 18.2 18.0   7    7    7 -4.3301 -7.8785 -5.1962     72  West Rainy
20010617    83 15.4 17.4 16.6   8    7    7 -4.3301 -2.0521 -3.0000     70 North   Dry
20010618    88 15.9 19.1 21.5   6    5    4  0.5209 -2.9544 -1.0261     83  West   Dry
….
20010916    71 15.4 17.7 16.6   4    5    5 -3.8302  0.0000  1.3892     69 North   Dry
20010917    60 13.7 14.0 15.8   4    5    4  0.0000  3.2139  0.0000     71 North Rainy
20010918    42 12.7 14.3 14.9   8    7    7 -2.5000 -3.2139 -2.5000     60 North Rainy
20010919    65 14.8 16.3 15.9   7    7    7 -4.3301 -6.0622 -5.1962     42  West Rainy
20010920    71 15.5 18.0 17.4   7    7    6 -3.9392 -3.0642  0.0000     65  West   Dry
20010921    96 11.3 19.4 20.2   3    3    3 -0.1736  3.7588  3.8302     71  East Rainy
20010922    98 15.2 19.7 20.3   2    2    2  4.0000  5.0000  4.3301     96  East   Dry
20010923    92 14.7 17.6 18.2   1    4    6  5.1962  5.1423  3.5000     98 North   Dry
20010924    76 13.3 17.7 17.7   7    7    6 -0.9397 -0.7660 -0.5000     92  West Rainy
20010925    84 13.3 17.7 17.8   3    5    6  0.0000 -1.0000 -1.2856     76 South   Dry
20010927    77 16.2 20.8 22.1   6    5    5 -0.6946 -2.0000 -1.3681     71 South Rainy
20010928    99 16.9 23.0 22.6   6    4    7  1.5000  0.8682  0.8682     77 South   Dry
20010929    83 16.9 19.8 22.1   6    5    3 -4.0000 -3.7588 -4.0000     99  West Rainy
20010930    70 15.7 18.6 20.7   7    7    7  0.0000 -1.0419 -4.0000     83 South   Dry
```

a) Write the R command to produce the given output.

Output:

```
      T12              maxO3             wind       rain         Wx12
 Min.   :14.00   Min.   : 42.00   East :10   Dry  :69   Min.   :-7.878
 1st Qu.:18.60   1st Qu.: 70.75   North:31   Rainy:43   1st Qu.:-3.565
 Median :20.55   Median : 81.50   South:21              Median :-1.879
 Mean   :21.53   Mean   : 90.30   West :50              Mean   :-1.611
 3rd Qu.:23.55   3rd Qu.:106.00                         3rd Qu.: 0.000
 Max.   :33.50   Max.   :166.00                         Max.   : 6.578
```

(2 marks)

b) Write the R command to produce the following plots.

Figure 1

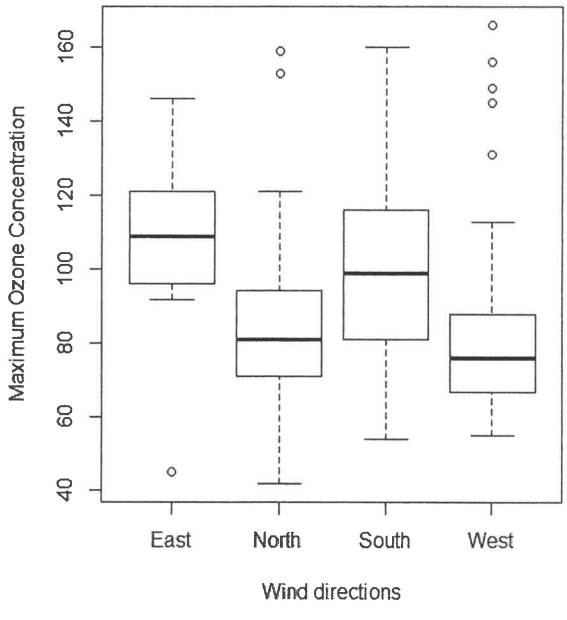
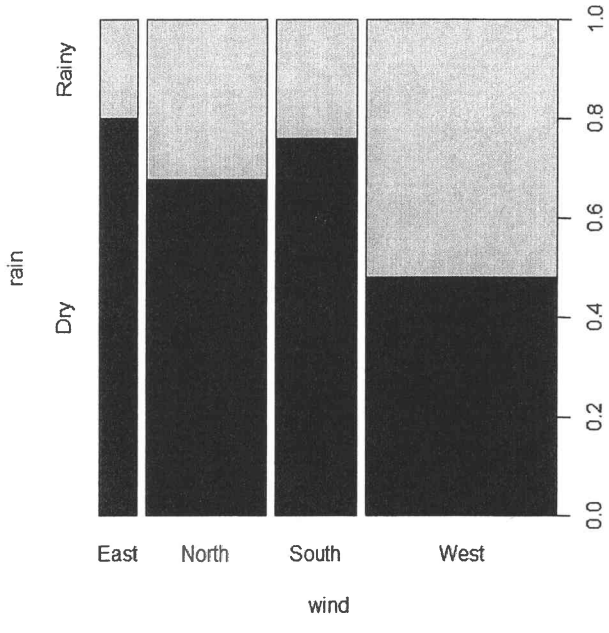


Figure 2



(5 marks)

c) Comment on the plot obtained in Figure 2.

(2 marks)

d) Produce the R function to obtain the following descriptive statistics.

```
$mean
              East  North   South   West
mean.T12    23.670 20.065  23.729 21.080
mean.maxO3 105.600 86.129 102.524 84.700
mean.Wx12    3.747 -0.641  -0.900 -3.583

$skewness
              East North South   West
skew.T12    -0.039 0.721 0.469  0.771
skew.maxO3 -0.777 1.288 0.420  1.508
skew.Wx12    0.322 0.578 0.102 -0.350
```

(5 marks)

e) Write the R command to select the "***ozone***" data sets contain both variables **T12** and **maxO3** which are greater than or equal to 25 and 120, respectively. The data with the variable which less than these values should be excluded. The required data should be obtained as follows:

```
          T12 maxO3  wind rain    Wx12
20010623 28.6   146 South  Dry -1.9284
20010624 25.2   121 North  Dry -0.5209
20010625 32.7   146  East  Dry  6.5778
20010703 30.1   139 South  Dry  2.0000
20010725 26.9   149  West  Dry -0.9397
20010726 27.7   153 North  Dry  1.5000
20010727 28.3   159 North  Dry  1.2856
20010728 27.6   149  West  Dry -1.5321
20010729 29.6   160 South  Dry -0.6840
20010730 30.5   156  West  Dry -1.8794
20010801 29.5   126  East  Dry  3.7588
20010823 28.4   131  West  Dry -1.9696
20010824 27.2   166  West  Dry -0.8660
20010825 33.5   159 South  Dry  0.6946
```

(3 marks)

f) Produce the R command to obtain the following result based on the output in part (e).

	T12	maxO3	wind	rain	Wx12
20010625	32.7	146	East	Dry	6.5778
20010801	29.5	126	East	Dry	3.7588
20010703	30.1	139	South	Dry	2.0000
20010726	27.7	153	North	Dry	1.5000
20010727	28.3	159	North	Dry	1.2856
20010825	33.5	159	South	Dry	0.6946
20010624	25.2	121	North	Dry	-0.5209
20010729	29.6	160	South	Dry	-0.6840
20010824	27.2	166	West	Dry	-0.8660
20010725	26.9	149	West	Dry	-0.9397
20010728	27.6	149	West	Dry	-1.5321
20010730	30.5	156	West	Dry	-1.8794
20010623	28.6	146	South	Dry	-1.9284
20010823	28.4	131	West	Dry	-1.9696

(2 marks)

g) Produce the R function to divide the projection of the wind speed vector on the east-west axis at midday into two categories (Positive and Negative). The required data should be obtained as follows:

	T12	maxO3	wind	rain	Wx12	Category
20010625	32.7	146	East	Dry	6.5778	Positive
20010801	29.5	126	East	Dry	3.7588	Positive
20010703	30.1	139	South	Dry	2.0000	Positive
20010726	27.7	153	North	Dry	1.5000	Positive
20010727	28.3	159	North	Dry	1.2856	Positive
20010825	33.5	159	South	Dry	0.6946	Positive
20010624	25.2	121	North	Dry	-0.5209	Negative
20010729	29.6	160	South	Dry	-0.6840	Negative
20010824	27.2	166	West	Dry	-0.8660	Negative
20010725	26.9	149	West	Dry	-0.9397	Negative
20010728	27.6	149	West	Dry	-1.5321	Negative
20010730	30.5	156	West	Dry	-1.8794	Negative
20010623	28.6	146	South	Dry	-1.9284	Negative
20010823	28.4	131	West	Dry	-1.9696	Negative

(6 marks)

PART B

QUESTION 1

Output A shows a temporary data set called STAFF that contains the 10 rows and 5 columns. Based on Output B, it is shown that label and format are assigned on two variables. Identify these variables and write a program to include the permanent label and format and print the result as shown in Output B.

Output A

STAFF Dataset

Name	IdNumber	Salary	Site	HireDate
Capalleti, Jimmy	2355	21163	BR1	30JAN09
Chen, Len	5889	20976	BR1	18JUN06
Davis, Brad	3878	19571	BR2	20MAR84
Leung, Brenda	4409	34321	BR2	18SEP94
Martinez, Maria	3985	49056	US2	10JAN93
Orfali, Philip	0740	50092	US2	16FEB03
Patel, Mary	2398	35182	BR3	02FEB90
Smith, Robert	5162	40100	BR5	15APR06
Sorrell, Joseph	4421	38760	US1	19JUN11
Zook, Carla	7385	22988	BR3	18DEC10

Output B

STAFF with a Format for the Variables Salary and Site

Name	IdNumber	Salary in U.S. Dollars	Site	HireDate
Capalleti, Jimmy	2355	$21,163	Birmingham UK	30JAN09
Chen, Len	5889	$20,976	Birmingham UK	18JUN06
Davis, Brad	3878	$19,571	Plymouth UK	20MAR84
Leung, Brenda	4409	$34,321	Plymouth UK	18SEP94
Martinez, Maria	3985	$49,056	Miami USA	10JAN93
Orfali, Philip	0740	$50,092	Miami USA	16FEB03
Patel, Mary	2398	$35,182	York UK	02FEB90
Smith, Robert	5162	$40,100	INCORRECT CODE	15APR06
Sorrell, Joseph	4421	$38,760	Denver USA	19JUN11
Zook, Carla	7385	$22,988	York UK	18DEC10

(10 marks)

QUESTION 2

a) Explain the results generated by the following program. Your explanation must describe the process involved in each line of sas statements.

```
data normal (keep= x);
call streaminit(4321);
do i = 1 to 100;
   x = rand("Normal");
   output;
end;
run;
```

(6 marks)

b) Using the program in (a), add relevant sas statements to show the use of macro variables to simulate 10 data samples.

(5 marks)

QUESTION 3

Examine the following program. The data step creates a sas data set called CHANGED, by reading instream data. The proc step displays the output that is the data portion of CHANGED. Fill in the blank spaces in the data step with sas statements (using ARRAY and DO LOOP), that will replace each missing value with zero (0), with the number of columns (variables) maintains at three. The result is shown in the Output C.

```
options nodate pageno=1 linesize=80 pagesize=60;

data changed _____;
input Reference Usage Introduction;
_____;
_____;
_____;
_____;

datalines;
45 63 113
. 75 150
62 . 98
;
proc print data=changed;
title 'Number of Books Sold';
run;
```

Output C

Number of Books Sold

Obs	Reference	Usage	Introduction
1	45	63	113
2	0	75	150
3	62	0	98

(10 marks)

QUESTION 4

Examine the given program (PROGRAM A). This program uses PROC SQL to create a sas data set called CUSTORDER and prints the following output. Your task is to write a new the program that uses data step and proc step to yield the same results as generated by this program in Output D.

PROGRAM A:

```
libname proclib 'C:\SAS Lesson';
proc sql;
create table proclib.custorder
    (IdNum char(4),Gender char(1),ProductType char(3),
     Price num,Order_Qty num,OrderDate num informat=date9.
format=ddmmyy10.);
insert into proclib.custorder
    values('1639','F','PT1',263.62,25,'26JUN2015'd)
    values('1065','M','ME3',524.25,30,'26JAN2015'd)
    values('1400','M','ME1',297.69,50,'05NOV2015'd)
    values('1561','M',null,365.14,10,'30NOV2015'd)
    values('1221','F','FA3',.,.,'22SEP2015'd);
title 'PROCLIB.CUSTORDER Table';
select * from proclib.custorder;
```

OUTPUT D

PROCLIB.CUSTORDER Table

IdNum	Gender	ProductType	Price	Order_Qty	OrderDate
1639	F	PT1	263.62	25	26/06/2015
1065	M	ME3	524.25	30	26/01/2015
1400	M	ME1	297.69	50	05/11/2015
1561	M		365.14	10	30/11/2015
1221	F	FA3	.	.	22/09/2015

(10 marks)

QUESTION 5

Given the following program, answer each question:

a) State the libref.

b) What is the function of the 'SELECT' clause.

c) State the name of the two tables.

d) State the status (either permanent or temporary) of the tables in (c).

e) What is the purpose of this statement: idnumber=idnum.

f) State the number of rows and columns created in the output table.

```
libname proclib 'SAS-library';
proc sql number;
title 'Information for Certain Employees Only';
select Lname, Fname, City, State, IdNumber, Salary, Jobcode
      from proclib.staff, proclib.payroll
      where idnumber=idnum and idnum in ('1919', '1400', '1350', '1333');
```

(9 marks)

END OF QUESTION PAPER