UNIVERSITI TEKNOLOGI MARA
FINAL EXAMINATION

COURSE	:	**STATISTICAL COMPUTING**
COURSE CODE	:	**STA705**
EXAMINATION	:	**DECEMBER 2015**
TIME	:	**3 HOURS**

INSTRUCTIONS TO CANDIDATES

1. This question paper consists of two (2) parts : PART A (2 Questions)
   PART B (2 Questions)
2. Answer ALL questions in both parts using separate Answers Booklets.
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) Write the R command using **seq** to obtain the following result:

```
10 12 14 16 18 20
```

(2 marks)

b) Produce the R command to obtain the following output by using the build-in function **rep** and **seq**.

```
10 12 14 16 18 20 10 12 14 16 18 20 10 12 14 16 18 20 25 50
25 50 25 50 25 50 25 50 25 50
```

(3 marks)

c) Give the R command to reconstruct the result obtained in part (b) into the following form.

```
> x1
     [,1] [,2] [,3] [,4] [,5]
[1,]   10   12   14   16   18
[2,]   20   10   12   14   16
[3,]   18   20   10   12   14
[4,]   16   18   20   25   50
[5,]   25   50   25   50   25
[6,]   50   25   50   25   50
```

(3 marks)

d) Use loop procedure to generate 20 sets of random samples (with set.seed=100) based on the data provided in part (a). Samples of data are shown as below:

```
[1] 12 12 16 10 14 14
[1] 18 14 16 12 16 20
[1] 12 14 18 18 12 14
....
[1] 10 16 20 18 10 12
[1] 20 10 20 18 12 20
```

(4 marks)

e) Complete the following R function to produce 20 sample means based on the results obtained in part (d).

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

{
      n=  [                    ]
      mean.b=numeric(S)
      set.seed(100)
      for (b in 1:S)
      {
        [                                    ]
      }
        [                                    ]
}
```

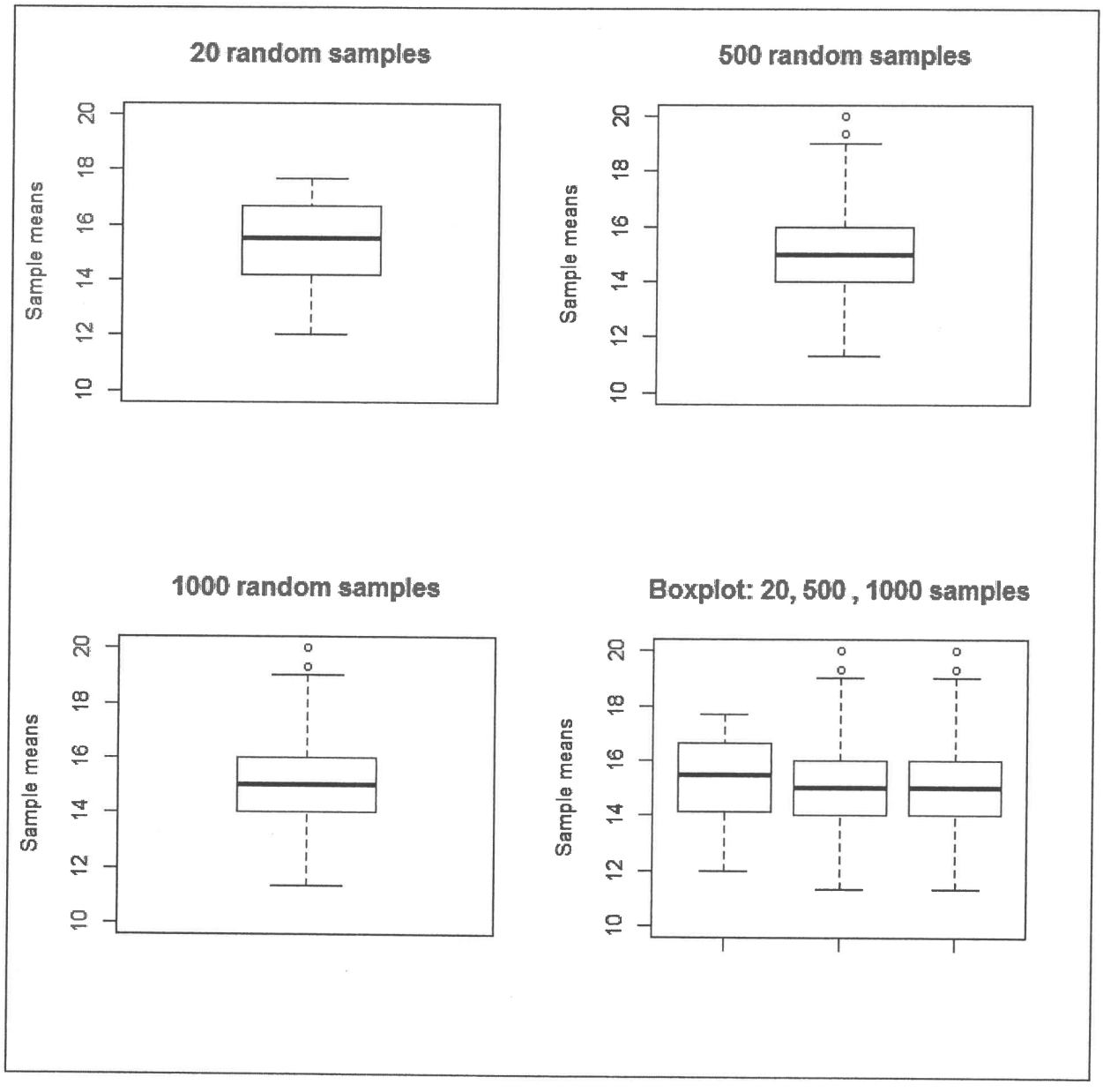
(6 marks)

The output is obtained as follows:

```
> Find.samples(a1,20)
$data
[1] 10 12 14 16 18 20

$sample.mean
 [1] 13.000 16.000 14.667 16.667 15.333 17.667 15.000 17.333 12.000 13.000
[11] 16.000 14.000 17.667 16.667 14.333 15.667 13.333 16.000 14.333 16.667
```

f) Use the function obtained in part (e) to get the sample means generated based on 500 and 1000 sets of random samples. Write the R function to produce the following plots:



(7 marks)

QUESTION 2

The information on the "***road***" dataset obtained from the R system is given as follows:

```
> str(road)
'data.frame':   26 obs. of  6 variables:
 $ deaths : int  968 43 588 640 4743 566 325 118 115 1545 ...
 $ drivers: int  158 11 91 92 952 109 167 30 35 298 ...
 $ popden : num  64 0.4 12 34 100 ...
 $ rural  : num  66 5.9 33 73 118 73 5.1 3.4 0 57 ...
 $ temp   : int  62 30 64 51 65 42 37 41 44 67 ...
 $ fuel   : num  119 6.2 65 74 105 78 95 20 23 216 ...
```

```
> road
        deaths drivers  popden rural temp  fuel
Alabama    968     158    64.0  66.0   62 119.0
Alaska      43      11     0.4   5.9   30   6.2
Arizona    588      91    12.0  33.0   64  65.0
Arkanas    640      92    34.0  73.0   51  74.0
Calif     4743     952   100.0 118.0   65 105.0
Colo       566     109    17.0  73.0   42  78.0
Conn       325     167   518.0   5.1   37  95.0
Dela       118      30   226.0   3.4   41  20.0
DC         115      35 12524.0   0.0   44  23.0
Florida   1545     298    91.0  57.0   67 216.0
Georgia   1302     203    68.0  83.0   54 162.0
Idaho      262      41     8.1  40.0   36  29.0
Ill       2207     544   180.0 102.0   33 350.0
Ind       1410     254   129.0  89.0   37 196.0
Iowa       833     150    49.0 100.0   30 109.0
Kansas     669     136    27.0 124.0   42  94.0
Kent       911     147    76.0  65.0   44 104.0
Louis     1037     146    72.0  40.0   65 109.0
Maine     1196      46    31.0  19.0   30  37.0
Maryl      616     157   314.0  29.0   44 113.0
Mass       766     255   655.0  17.0   37 166.0
Mich      2120     403   137.0  95.0   33 306.0
Minn       841     189    43.0 110.0   22 132.0
Miss       648      85    46.0  59.0   57  77.0
Mo        1289     234    63.0 100.0   40 180.0
Mont       259      38     4.6  72.0   29  31.0
```

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

Output:

```
$result1
        deaths rural fuel
Alabama    968    66  119
Calif     4743   118  105
Florida   1545    57  216
Georgia   1302    83  162
Ill       2207   102  350
Ind       1410    89  196
Iowa       833   100  109
Kent       911    65  104
Louis     1037    40  109
Maine     1196    19   37
Mich      2120    95  306
Minn       841   110  132
Mo        1289   100  180

$result2
        deaths rural fuel
Ill       2207   102  350
Mich      2120    95  306
Florida   1545    57  216
Ind       1410    89  196
Mo        1289   100  180
Georgia   1302    83  162
Minn       841   110  132
Alabama    968    66  119
Louis     1037    40  109
Iowa       833   100  109
Calif     4743   118  105
Kent       911    65  104
Maine     1196    19   37

$result3
           deaths     popden   temp
mean     1000.654   595.735 43.692
median    799.500    66.000 41.500
maximum  4743.000 12524.000 67.000
sd        946.842  2437.947 13.013
coef.var    0.946     4.092  0.298
skewness    2.362     4.496  0.443
kurtosis    6.713    19.043 -1.062
```

(10 marks)

b) Given that the Kendall's correlation between variables ***deaths*** and ***popden*** is 0.2923, determine the correct input of **ABC** and the value of **XXX**.

```
ORDINARY NONPARAMETRIC BOOTSTRAP
Call:
boot(data = ABC, statistic = cor.kd, R = 2000)
Bootstrap Statistics :
original    bias    std. error
t1* XXX   0.0041662   0.1932577
```

(2 marks)

Based on the above information, write the R function to get the statistic, **cor.kd**.

(3 marks)

c) Using the R program given below, spot the errors and correct them.

```
Prog.1=function(dta)
{
     if (dta>=0&&dta<40) status="Low_Temp"
     else if (dta>=40&&dta>60) status="Moderate_Temp"
     else if (dta>=60&&data<100) status="High_Temp"
     else if "Invalid DATA"
}

Prog.2=function(dt)
{  kp=numeric(length(dt))
   for (i in 1:dt)
   {  kp=prog.1(dta(i))
      Result=cbind(data=dt,status=kp)
      result=as.data.frame(result)
      tb=table(result)  }

par(mfrow=c(1,2))
gf1=hist(tb,main="Figure 1")
gf2=piechart(tb,col=gray(seq(0.5,1.0,length=5)),
main="Figure 2")
list(freq=tb,plots=c(gf1,gf2))
}
```

(10 marks)

Output

```
$freq
kp
       High_Temp        Low_Temp     Moderate_Temp
               5              11                10

$plots
```

Figure 1

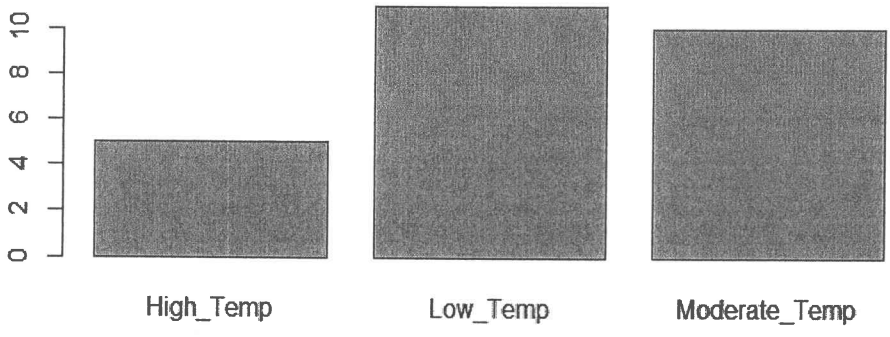
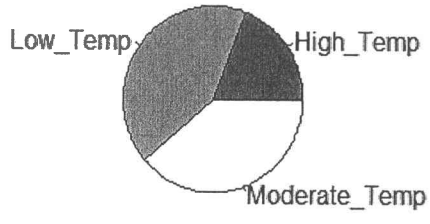


Figure 2



PART B

QUESTION 1

An insurance company wants to evaluate their staff performance to determine who deserve to be given bonus for a particular year. In order to obtain the employee's job performance score, several criteria of employees has been considered. Those criteria are listed in the table below:

Employee_ID	Attendance	No_Customers	Experience	Gender
8867	250	176	5	M
8771	212	56	7	F
7012	118	101	6	F
8588	274	124	2	M
9101	197	95	4	M
6781	210	201	8	M
7864	201	198	9	F
8897	141	64	5	F
5691	167	174	5	M
5821	174	105	6	F

a) Write a SAS program to read the data and store it in a temporary SAS dataset.

(4 marks)

b) Prepare a SAS format for the variable gender where M stands for MALE and F stands for Female to be used in all printout.

(6 marks)

c) Include a SAS statement to calculate the performance score based on the sum of the following weightage:

50% of overall attendance (out of 300 days),
25% of number customers (out of 300 customers)
25% of years experience (out of 10 years)

(5 marks)

d) Include SAS statements to group the variable Score as follows:

Score	Status
< 70	Without Bonus
≥ 70	With Bonus

(5 marks)

e) Include SAS statements to store the listing of employees who are eligible for the bonus in a separate dataset that contains only information on Employee ID, Gender and Score.

(5 marks)

QUESTION 2

a) Give example to illustrate how to use the following SAS commands:
   i) Drop
   ii) Firstobs
   iii) Infile
   iv) Double Trailing @@
   v) Merge

(10 marks)

b) i) Based on the following SAS program spot 5 errors and correct them.

```
Data qmt100;
Input name gender $ test quiz final;
carrymark : (test*0.3) + (quiz*0.1);
Totmark = carrymark + (final*0.6),
lines;
Nuriz   Female 70 10  95
Dinur   Male   85  9  96
Salleh  Male   60  7  88
Sapiah  Female 43  5  50
Eliza   Female 71  8 78
;
run;
proq print;
run;
```

(5 marks)

ii) Based on the **Output 1** for the corrected program in part (i) given below, add commands in your program to produce **Output 2**.

(10 marks)

Output 1

Obs	name	gender	test	quiz	final	carrymark	Totmark
1	Nuriz	Female	70	10	95	22.0	79.0
2	Dinur	Male	85	9	96	26.4	84.0
3	Salleh	Male	60	7	88	18.7	71.5
4	Sapiah	Female	43	5	50	13.4	43.4
5	Eliza	Female	71	8	78	22.1	68.9

Output 2

Obs	name	gender	test	quiz	final	carrymark	Totmark
1	Sapiah	Female	43	5	50	13.4	43.4
2	Eliza	Female	71	8	78	22.1	68.9
3	Salleh	Male	60	7	88	18.7	71.5
4	Nuriz	Female	70	10	95	22.0	79.0
5	Dinur	Male	85	9	96	26.4	84.0

Analysis Variable : Totmark				
N	Mean	Std Dev	Minimum	Maximum
5	69.3600000	15.7000955	43.4000000	84.0000000

Obs	name	Totmark
1	Sapiah	43.4
2	Eliza	68.9
3	Salleh	71.5
4	Nuriz	79.0
5	Dinur	84.0

END OF QUESTION PAPER