UNIVERSITI TEKNOLOGI MARA
FINAL EXAMINATION

COURSE	:	**STATISTICAL COMPUTING**
COURSE CODE	:	**STA705**
EXAMINATION	:	**JUNE 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 from PART A and PART B in the question paper.
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) a blank A4 paper – 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 17 printed pages

PART A

STUDENT ID NO.	:	
PROGRAMME CODE	:	
GROUP	:	
LECTURER'S NAME	:	

PART A

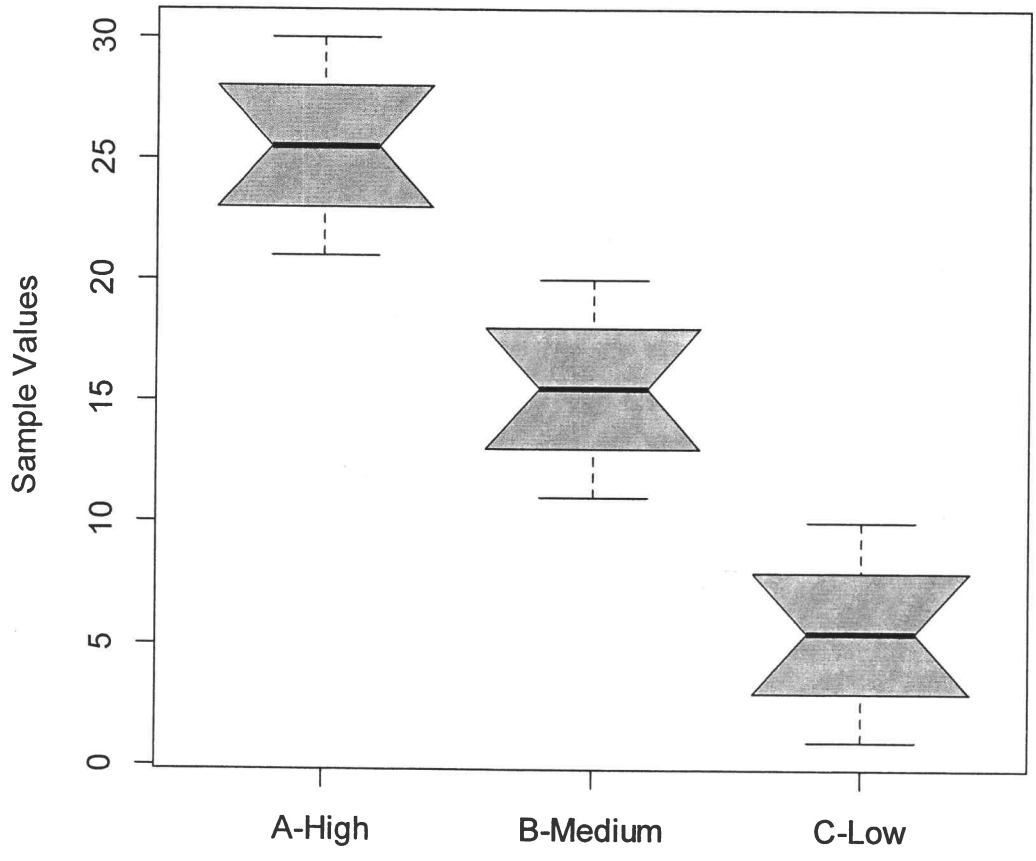
QUESTION 1

a) Write the R command to obtain the following dataset which is produced from the R random generators.

```
> dt1
   groupA groupB groupC
1       4     17     26
2       3     18     27
3       5     13     25
4       1     19     30
5       9     15     23
6       6     14     21
7      10     11     24
8       2     12     29
9       8     16     22
10      7     20     28
```

(4 marks)

b) Produce the following output based on the data obtained in part (a).



(5 marks)

c) The information on the ***"CO2"*** dataset obtained from the R system is given as follows:

```
> str(CO2)
Classes 'nfnGroupedData', 'nfGroupedData', 'groupedData' and
'data.frame':
84 obs. of  5 variables:
 $ Plant    : Ord.factor w/ 12 levels "Qn1"<"Qn2"<"Qn3"<..: 1 1 1 1 1
 $ Type     : Factor w/ 2 levels "Quebec","Mississippi": 1 1 1 1 1 1
 $ Treatment: Factor w/ 2 levels "nonchilled","chilled": 1 1 1 1 1 1
 $ conc     : num  95 175 250 350 500 675 1000 95 175 250 ...
 $ uptake   : num  16 30.4 34.8 37.2 35.3 39.2 39.7 13.6 27.3 37.1 ...
```

Write the R command to produce the following output.

```
> yy
     Plant            Type         Treatment       conc          uptake
 Qn1    : 7   Quebec     :42   nonchilled:42   Min.   :  95   Min.   : 7.70
 Qn2    : 7   Mississippi:42   chilled   :42   1st Qu.: 175   1st Qu.:17.90
 Qn3    : 7                                    Median : 350   Median :28.30
 Qc1    : 7                                    Mean   : 435   Mean   :27.21
 Qc3    : 7                                    3rd Qu.: 675   3rd Qu.:37.12
 Qc2    : 7                                    Max.   :1000   Max.   :45.50
 (Other):42
```

```
> xx
              conc      uptake
skewness  0.7201458 -0.1040551
kurtosis -0.6826587 -1.3482674
```

(4 marks)

d) Complete the following R function to produce 30 bootstrap samples based on the correlation between the two variables, ***conc*** and ***uptake***.

```
Find.samples<-function(dtx,S=1000)

{     dt<-as.data.frame(dtx)

      [                                                    ]

      alpha.hat<-cor(dt1,dt2)
      n<-nrow(dtx)
      alpha.b<-numeric(S)
      set.seed(100)
      for (b in 1:S)
      {
      [                                                    ]
      [                                                    ]
      [                                                    ]
      [                                                    ]
      [                                                    ]
        }

      [                                                    ]
      [                                                    ]
      [                                                    ]
      [                                                    ]
      [                                                    ]

}
```

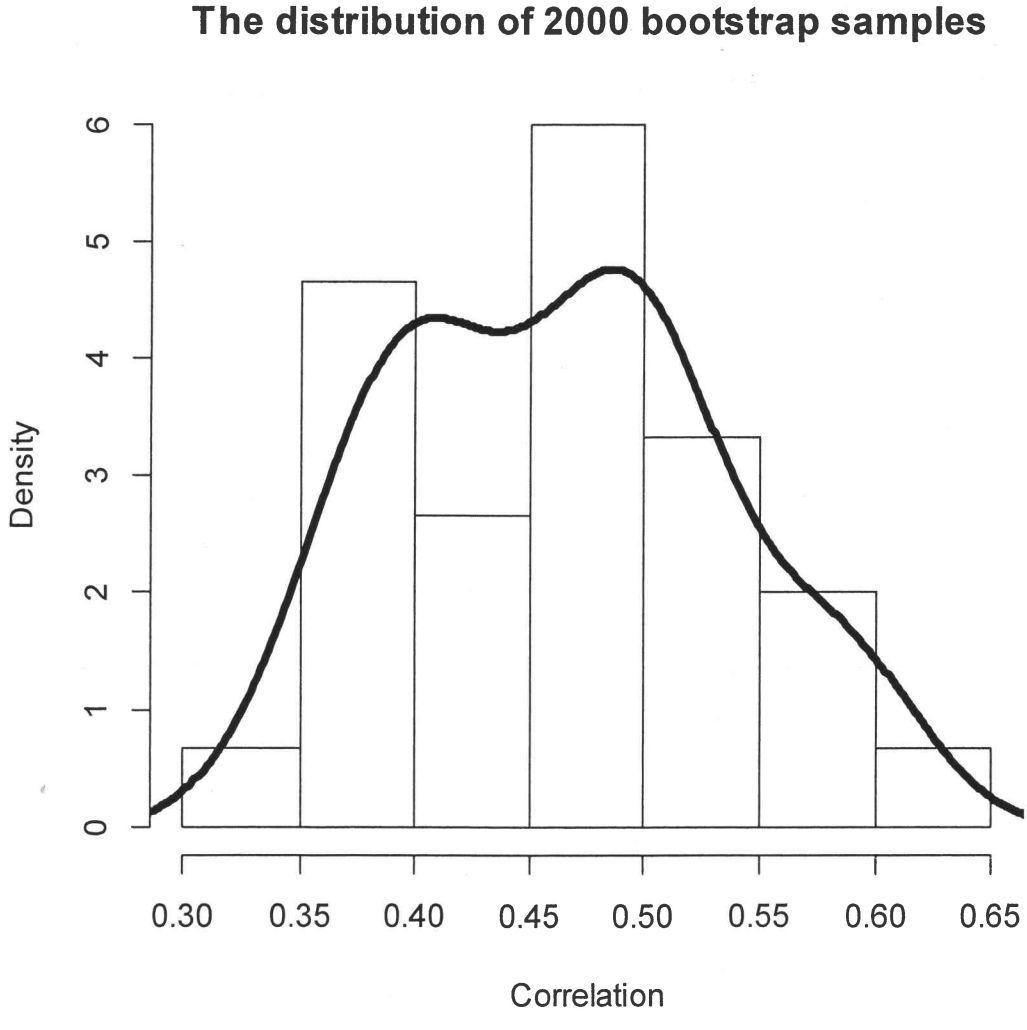
(12 marks)

The output is obtained as follows:

```
$sample.boot

 [1] 0.483 0.359 0.382 0.484 0.604 0.399 0.371 0.347 0.504 0.497 0.485 0.395
[13] 0.386 0.520 0.454 0.464 0.409 0.492 0.428 0.514 0.419 0.521 0.586 0.522
[25] 0.415 0.466 0.571 0.452 0.573 0.397

$plot
```



QUESTION 2

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

```
> str(Mroz)
'data.frame':   753 obs. of  8 variables:
 $ lfp : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
 $ k5  : int  1 0 1 0 1 0 0 0 0 0 ...
 $ k618: int  0 2 3 3 2 0 2 0 2 2 ...
 $ age : int  32 30 35 34 31 54 37 54 48 39 ...
 $ wc  : Factor w/ 2 levels "no","yes": 1 1 1 1 2 1 2 1 1 1 ...
 $ hc  : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ lwg : num  1.2102 0.3285 1.5141 0.0921 1.5243 ...
 $ inc : num  10.9 19.5 12 6.8 20.1 ...
```

a) Write the R command to produce the given output. (Hint: use ***sample***)

(3 marks)

Output:

```
> data1
    lfp k5 k618 age  wc  hc       lwg    inc
34  yes  0    1  51  no  no 1.2729658 17.100
109 yes  0    1  44  no  no 1.2241757 17.669
123 yes  0    2  43  no yes 1.4789581 10.900
320 yes  0    1  47  no  no 1.1987233 14.700
362 yes  0    0  54 yes  no 2.1892564 18.220
483  no  2    1  33 yes yes 1.0503179 40.250
495  no  1    1  32  no  no 1.1927967 14.800
541  no  0    0  54  no  no 0.8151277 13.000
592  no  1    3  39  no yes 1.1914352 24.760
749  no  0    2  40 yes yes 1.0828638 28.200
```

b) Obtain the R command to produce the following output based on ***data1*** in part (a):

```
     age  wc  hc       lwg   inc
34    51  no  no 1.2729658 17.10
123   43  no yes 1.4789581 10.90
320   47  no  no 1.1987233 14.70
483   33 yes yes 1.0503179 40.25
495   32  no  no 1.1927967 14.80
541   54  no  no 0.8151277 13.00
749   40 yes yes 1.0828638 28.20
```

(3 marks)

c) The information on the variable "***inc***" which is obtained from the "***Mroz***" dataset can be divided into two status such as ***low_inc*** (where ***inc*** < median) and ***high_inc*** (where ***inc*** ≥ median). Complete the R program to produce the given output.

```
attach(Mroz)

status.1<-function(score)
{ if (score<median(inc)) {print("low_inc")}
    [                                    ]
  else print("INVALID SCORE")}

sta.1<-function(dt)
{  kp<-numeric(length(dt))
   for (i in 1:length(dt))
    [                                    ]
  as.data.frame(result)
}

Result<-sta.1(inc)
    [                                    ]
}
```

(8 marks)

Output:

```
$Result

      inc    level
1   10.91  low_inc
2    19.5 high_inc
3   12.04  low_inc
4     6.8  low_inc
5    20.1 high_inc
6   9.859  low_inc
7   9.152  low_inc
8    10.9  low_inc
9  17.305  low_inc
10 12.925  low_inc
11   24.3 high_inc
12   19.7 high_inc
13     15  low_inc
14   14.6  low_inc
….

743  21.62 high_inc
744 23.426 high_inc
745     26 high_inc
746   7.84  low_inc
747    6.8  low_inc
748   5.33  low_inc
749   28.2 high_inc
750     10  low_inc
751  9.952  low_inc
752 24.984 high_inc
753 28.363 high_inc

$Frequency_table

high_inc  low_inc
     377      376
```

d) Explain the use of the following R statements.

i) `rank` _____

ii) `sort` _____

iii) `order` _____

(5 marks)

e) Complete the following table by using the commands in part (d).

	Cost	ranked	sorted	ordered
1	137			
2	256			
3	118			
4	355			
5	212			
6	99			
7	111			
8	502			
9	212			
10	146			

(6 marks)

PART B

STUDENT ID NO.	:	
PROGRAMME CODE	:	
GROUP	:	
LECTURER'S NAME	:	

PART B

QUESTION 1

A study based on 196 persons selected in a probabilistic sample within two sectors in a city is conducted where the information collected are given below:

Variable Number	Variable Name	Description
1	Identification Number (ID)	1-196
2	Age (AGE)	Age of person (in years)
3	Socioeconomic status (ECO)	1=upper, 2=middle, 3=lower
4	Sector (SECTOR)	1=sector 1, 2=sector 2
5	Disease status (DISEASE)	1=with disease, 0=without disease
6	Saving account status (SAVING)	1=has saving account, 0=does not have saving account

a) Write a SAS program to read data in **disease_outbreak.txt** file and create a permanent SAS dataset.

(5 marks)

b) Add SAS statements to produce output that list all the respondents having the disease and has saving account.

(6 marks)

c) Include SAS statements to create a new variable **GROUP** as follows:

AGE	GROUP
< 40	JUNIOUR
≥ 40	SENIOUR

(7 marks)

d) Create another data set that contains **ID**, **ECO**, **DISEASE**, **SAVING** and **GROUP** for the first 20 respondents in your SAS program.

(7 marks)

QUESTION 2

An insurance company wants to make evaluation on their staff performance for year 2014. Besides that, this company also wants to determine who deserve to get bonus for year 2014. In order to determine employee's job performance, several criteria of employees has been taking into account. The Job_Performance file that contain those criteria are described in the table below:

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

a) Write a SAS program to read in data using **CARDS** statement and print out the list as given in the table.

(5 marks)

b) Create a new variable called **Score** which is based on:

$$50 \times \frac{\text{Attendance}}{300} + 25 \times \frac{\text{No\_Customers}}{300} + 25 \times \frac{\text{Experience}}{10}.$$

Hence, include a SAS statement to produce the summary statistics for **Score.**

(6 marks)

c) Write additional SAS statements in the program to produce a list of employees to be given bonus for their excellent performance of the year with score 70 and above and give a title **Bonus for the year 2014** to the output.

(7 marks)

d) Write SAS statements to prepare two separate dataset for those getting and not getting bonus for the year 2014. Print out the list for those getting the bonus showing variables **Employee_ID**, **Gender** and **No_Customers** only.

(7 marks)

END OF QUESTION PAPER