UNIVERSITI TEKNOLOGI MARA
FINAL EXAMINATION

COURSE	:	**STATISTICAL COMPUTING**
COURSE CODE	:	**STA705**
EXAMINATION	:	**DECEMBER 2014**
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 all two (2) parts 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 one blank A4 paper – provided by the Faculty

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

This examination paper consists of 23 printed pages

PART A

STUDENT ID NO.	:	
PROGRAMME CODE	:	
GROUP	:	
LECTURER'S NAME	:	
COURSE CODE	:	STA705

PART A

QUESTION 1

a) Write the R command to produce the following dataset which is generated from the Uniform distribution.

```
       [,1]   [,2]   [,3]   [,4]   [,5]   [,6]   [,7]   [,8]
[1,] 10.988 20.011 47.098 17.186 22.468 37.101 20.102 46.257
[2,] 12.168 31.286 47.230 26.008 32.981 25.236 28.952 31.076
[3,] 30.768 14.728 20.160 25.264 42.756 12.846 43.717 14.355
[4,] 19.565 18.695 21.582 25.661 37.707 18.220 47.428 19.490
[5,] 38.723 14.111 15.571 33.846 19.862 30.527 43.687 44.016
[6,] 23.264 24.266 25.704 41.449 49.524 37.934 17.787 19.985
[7,] 41.727 27.730 16.065 23.982 28.836 14.328 35.095 44.396
```

(4 marks)

b) Write the R command to produce a portion of the dataset in part (a).

```
       [,1]   [,2]   [,3]   [,4]
[1,] 20.011 17.186 37.101 20.102
[2,] 14.728 25.264 12.846 43.717
[3,] 14.111 33.846 30.527 43.687
[4,] 27.730 23.982 14.328 35.095
```

(3 marks)

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

	Type	Pressure	Weight
1	Small	Low	<100kg
2	Small	Low	<100kg
3	Small	Low	<100kg
4	Small	Low	100kg-200kg
5	Medium	Low	100kg-200kg
6	Medium	Low	100kg-200kg
7	Medium	High	201-250kg
8	Medium	High	201-250kg
9	Large	High	201-250kg
10	Large	High	>250kg
11	Large	High	>250kg
12	Large	High	>250kg

(7 marks)

d) Based on the output obtained in part (c), write the R command to produce the following:

```
, , 1

     [,1]  [,2]
[1,] "Low" "Low"
[2,] "Low" "Low"
[3,] "Low" "NA"

, , 2

     [,1]   [,2]
[1,] "High" "High"
[2,] "High" "High"
[3,] "High" "High"
```

(3 marks)

e) Give the R command to compute a Monte Carlo estimate of

$$\int_3^9 e^{-2x}\, dx$$

by generating 1000 samples from a uniform distribution using set seed 123.

(6 marks)

f) Provide the R command to compute the exact value of the integral in part (e).

(2 marks)

QUESTION 2

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

```
> str(Animals)
'data.frame':   28 obs. of  2 variables:
 $ body : num  1.35 465 36.33 27.66 1.04 ...
 $ brain: num  8.1 423 119.5 115 5.5 ...
>
```

```
> Animals
                    body  brain
Mountain beaver     1.350    8.1
Cow               465.000  423.0
Grey wolf          36.330  119.5
Goat               27.660  115.0
Guinea pig          1.040    5.5
Dipliodocus     11700.000   50.0
Asian elephant   2547.000 4603.0
Donkey            187.100  419.0
Horse             521.000  655.0
Potar monkey       10.000  115.0
Cat                 3.300   25.6
Giraffe           529.000  680.0
Gorilla           207.000  406.0
Human              62.000 1320.0
African elephant 6654.000 5712.0
Triceratops      9400.000   70.0
Rhesus monkey       6.800  179.0
Kangaroo           35.000   56.0
Golden hamster      0.120    1.0
Mouse               0.023    0.4
Rabbit              2.500   12.1
Sheep              55.500  175.0
Jaguar            100.000  157.0
Chimpanzee         52.160  440.0
Rat                 0.280    1.9
Brachiosaurus   87000.000  154.5
Mole                0.122    3.0
Pig               192.000  180.0
>
```

a) The information on the variable ***ratio*** is referred to the *brain-to-body weight ratio* based on $(\frac{\text{brain}}{\text{body}})$ in the dataset which can be divided into three groups A, B and C. Complete the R program to produce the output given on page 8.

```
prg.1<-function(data)
{
     if (data >=0&& data <5) status<-"C"
     [                                  ]
}
prg.2<-function(dtx)
{
   [                                    ]
   kp<-numeric(length(dt))
   for (i in 1:length(dt))
   {
   [                                    ]
   as.data.frame(result)
}
```

(9 marks)

Output:

```
> a2=prg.2(Animals)
> a2
                    body  brain ratio   group
Mountain beaver    1.350    8.1  6.00      B
Cow              465.000  423.0  0.91      C
Grey wolf         36.330  119.5  3.29      C
Goat              27.660  115.0  4.16      C
Guinea pig         1.040    5.5  5.29      B
Dipliodocus    11700.000   50.0  0.00      C
Asian elephant  2547.000 4603.0  1.81      C
Donkey           187.100  419.0  2.24      C
Horse            521.000  655.0  1.26      C
Potar monkey      10.000  115.0 11.50      B
Cat                3.300   25.6  7.76      B
Giraffe          529.000  680.0  1.29      C
Gorilla          207.000  406.0  1.96      C
Human             62.000 1320.0 21.29      A
African elephant 6654.000 5712.0  0.86      C
Triceratops      9400.000   70.0  0.01      C
Rhesus monkey       6.800  179.0 26.32      A
Kangaroo           35.000   56.0  1.60      C
Golden hamster      0.120    1.0  8.33      B
Mouse               0.023    0.4 17.39      B
Rabbit              2.500   12.1  4.84      C
Sheep              55.500  175.0  3.15      C
Jaguar            100.000  157.0  1.57      C
Chimpanzee         52.160  440.0  8.44      B
Rat                 0.280    1.9  6.79      B
Brachiosaurus   87000.000  154.5  0.00      C
Mole                0.122    3.0 24.59      A
Pig               192.000  180.0  0.94      C
```

b) Give the R command to produce the following output:

Output:

```
>                    body  brain ratio group
Rabbit              2.500   12.1  4.84     C
Goat               27.660  115.0  4.16     C
Grey wolf          36.330  119.5  3.29     C
Sheep              55.500  175.0  3.15     C
Donkey            187.100  419.0  2.24     C
Gorilla           207.000  406.0  1.96     C
Asian elephant   2547.000 4603.0  1.81     C
Kangaroo           35.000   56.0  1.60     C
Jaguar            100.000  157.0  1.57     C
Giraffe           529.000  680.0  1.29     C
Horse             521.000  655.0  1.26     C
Pig               192.000  180.0  0.94     C
Cow               465.000  423.0  0.91     C
African elephant 6654.000 5712.0  0.86     C
Triceratops      9400.000   70.0  0.01     C
Brachiosaurus   87000.000  154.5  0.00     C
Dipliodocus     11700.000   50.0  0.00     C
Mouse               0.023    0.4 17.39     B
Potar monkey       10.000  115.0 11.50     B
Chimpanzee         52.160  440.0  8.44     B
Golden hamster      0.120    1.0  8.33     B
Cat                 3.300   25.6  7.76     B
Rat                 0.280    1.9  6.79     B
Mountain beaver     1.350    8.1  6.00     B
Guinea pig          1.040    5.5  5.29     B
Rhesus monkey       6.800  179.0 26.32     A
Mole                0.122    3.0 24.59     A
Human              62.000 1320.0 21.29     A
>
```

(4 marks)

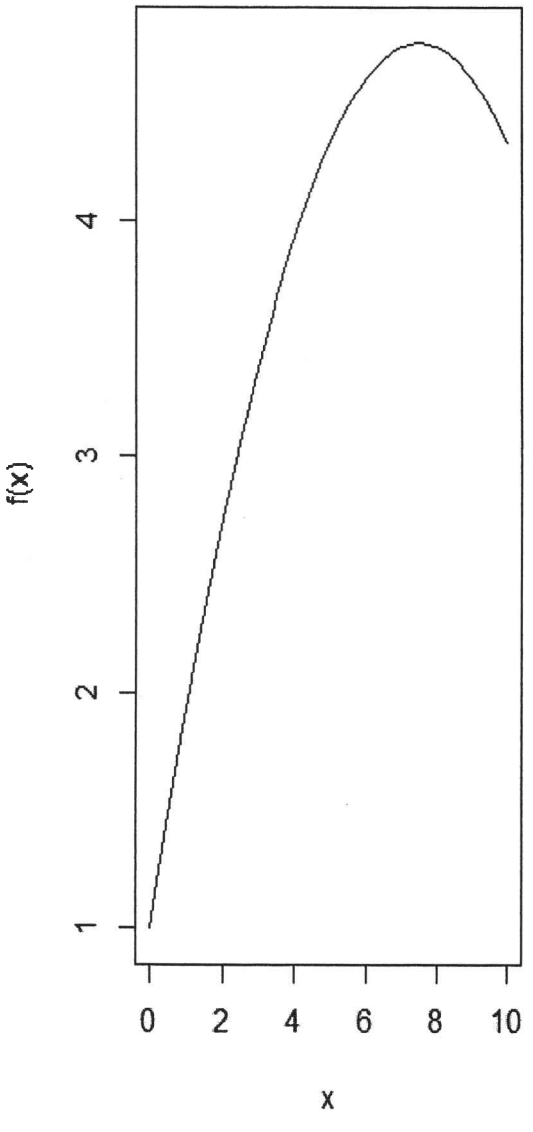
c) Write the R function to obtain the following descriptive statistics:

```
$body
     mean median skewness
A   22.97    6.8     0.37
B    8.53    1.2     1.74
C 7038.77  207.0     3.26

$brain
    mean median skewness
A 500.67  179.0     0.36
B  74.69    6.8     1.65
C 822.77  175.0     2.16

$ratio
   mean median skewness
A 24.07  24.59    -0.20
B  8.94   8.04     1.13
C  1.76   1.57     0.66
```

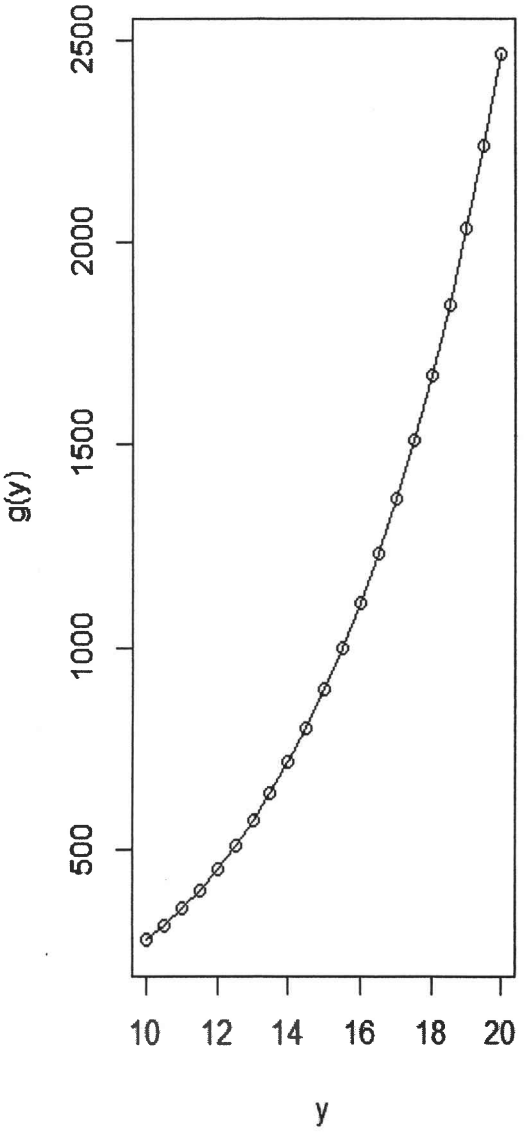
(6 marks)

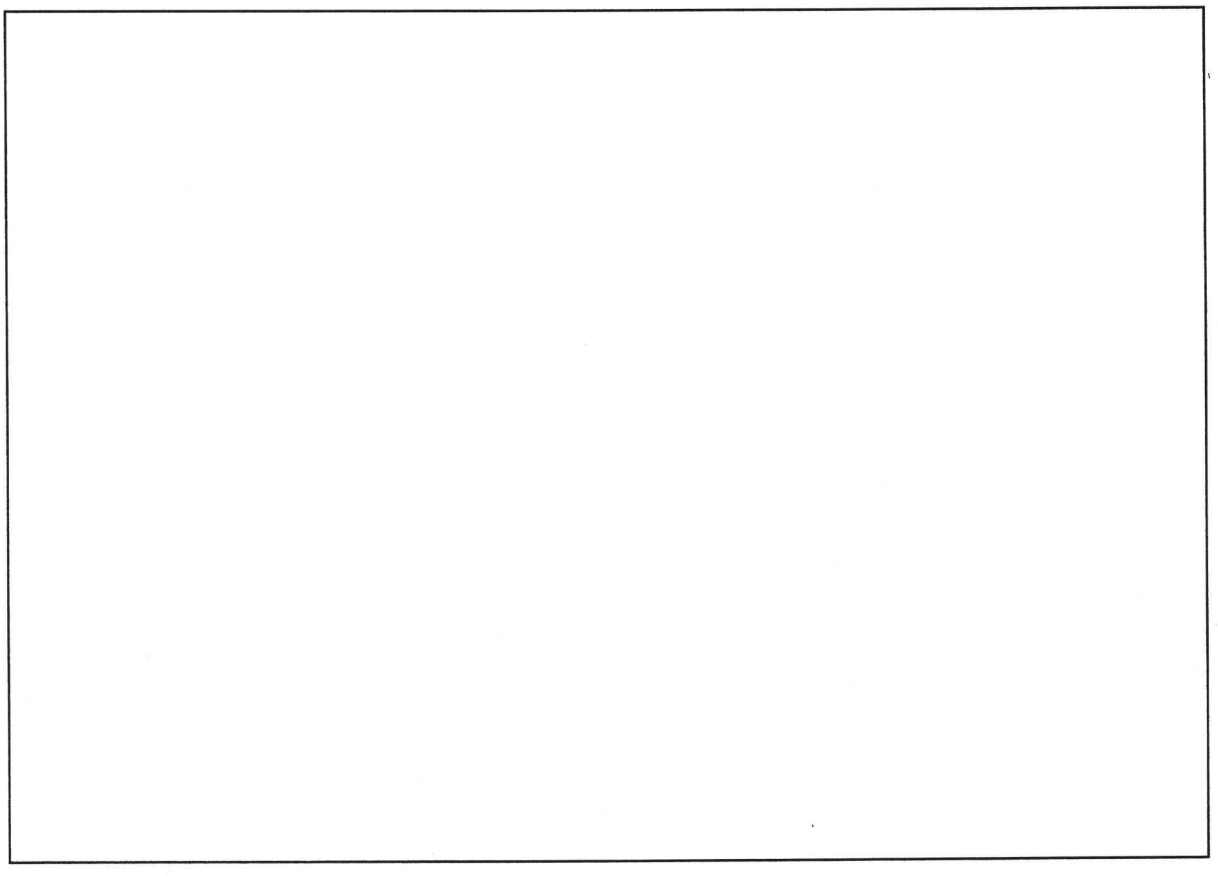
d) Write the R functions to produce each of the following output:

$$f(x) = 1 + x - x^{\frac{2}{15}}$$



$$g(y) = 10 + 1.2\beta^{\sqrt{y}}, \beta = 5.5$$





(6 marks)

PART B

STUDENT ID NO.	:	
PROGRAMME CODE	:	
GROUP	:	
LECTURER'S NAME	:	
COURSE CODE	:	STA705

PART B

QUESTION 1

The following table depicts a sample of EMPLOYEE data:

id	gender	bdate	educ	jobcat	salary	salbegin	jobtime	prevexp	minority
1	m	2/3/1952	15	3	57000	27000	98	144	0
2	m	5/23/1958	16	1	40200	18750	98	36	1
3	f	7/26/1929	12	1	21450	12000	98	381	0

with the following variable values:

gender: m-Male, f-Female
jobcat: 1-Clerk, 2-Custodial, 3-Managerial
minority: 1-Yes, 0-No

a) You are given a program that creates a SAS data set of EMPLOYEE using formatted input. This program contains eight (8) errors. Spot six (6) errors only and rewrite the program.

```
data employee;
   input id 2. +2 gender $ 1. +2 bdate mmddyyyy10. +2
   educ 2. +2 jobcat 1. +2 salary dollar10.2 +2
   salbegin dollar10.2 +2 jobtime 1. +2 prevexp 3. +2
   minority 1.;
   datalines;
1 m 2/3/1952 15 3 $57,000.00 $27,000.00 98 144 0
02 m 5/23/1958 16 1 $40,200.00 $18,750.00 98 36 0
3 f 4/15/1947 8 1 $21,900.00 $13,200.00 98 190 0
...
10 f 02/13/1946 20 1 $24,000.00 $13,500.00 98 244
0
;
run;
proc print; run;
```

(6 marks)

b) Write down SAS statements required to create a permanent data set called EMPLOYEE in any user defined library of your choice.

(1 mark)

c) Using EMPLOYEE data, write a SAS program to generate the following permanent format, variable label, and value label. Include in your program a print statement that will print output as shown in Figure 1.

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Label
3	bdate	Num	8	MMDDYY10.	Date of Birth
4	educ	Num	8		Educational Level (years)
2	gender	Char	1	$SEX.	Gender
1	id	Num	8		Employee Code
5	jobcat	Num	8	JOB.	Employment Category
8	jobtime	Num	8		Months since Hire
10	minority	Num	8	YES.	Minority Classification
9	prevexp	Num	8		Previous Experience (months)
6	salary	Num	8	DOLLAR11.2	Current Salary
7	salbegin	Num	8	DOLLAR11.2	Beginning Salary

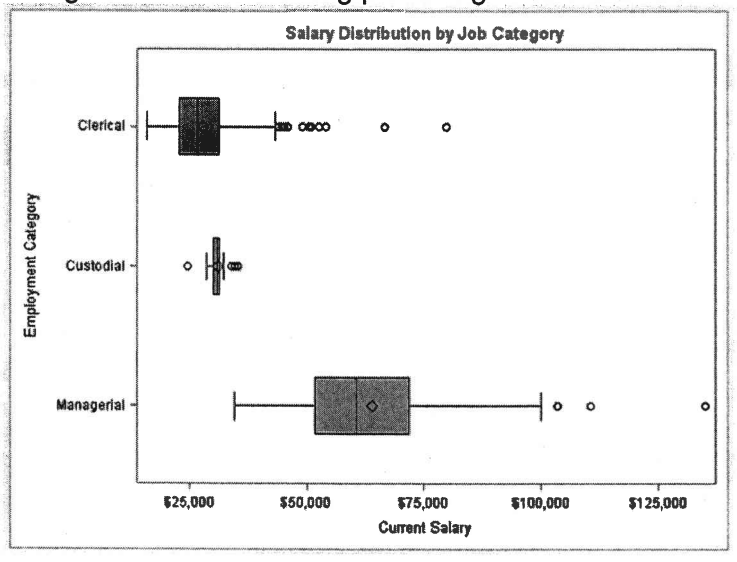
Current Salary for Employee Gender and Beginning Salary Category

Obs	id	gender	bdate	educ	jobcat	salary	salbegin	jobtime	prevexp	minority	salcat	salbegincat
170	170	Male	06/13/1964	12	Clerical	$26,550.00	$15,000.00	86	38	Yes	Low	Low
171	171	Male	01/21/1930	12	Clerical	$26,700.00	$13,500.00	86	367	Yes	Low	Low
172	172	Female	06/13/1953	15	Clerical	$29,850.00	$15,000.00	86	79	Yes	Low	Low
173	173	Male	01/15/1950	20	Managerial	$69,250.00	$42,480.00	85	134	No	Med	Low
174	174	Male	01/07/1935	8	Custodial	$31,950.00	$15,000.00	85	438	No	Low	Low
175	175	Male	01/08/1938	8	Clerical	$26,250.00	$15,600.00	85	171	No	Low	Low

Figure 1

(8 marks)

d) Write a program to generate the following plot using PROC SGPLOT.



(4 marks)

e) In order to test whether job category (jobcat) has a specified multinomial distribution, a Chi-square goodness of fit test is to be conducted. It is assumed that the distribution of jobcat is 20% Clerical, 30% Custodial and 50% Managerial. Write a program using PROC FREQ to carry out the test. The output of the test is as shown below.

The FREQ Procedure

Employment Category			
jobcat	Frequency	Percent	Test Percent
Clerical	363	76.58	20.00
Custodial	27	5.70	30.00
Managerial	84	17.72	50.00

Chi-Square Test for Specified Proportions	
Chi-Square	950.8671
DF	2
Pr > ChiSq	<.0001

(6 marks)

QUESTION 2

a) Based on EMPLOYEE data in question 1, write a SAS program to create a new variable "Beginning Salary Category", SALBEGINCAT, using the following criteria:

SALBEGINCAT = 'LOW' if salbegin less than $50000
SALBEGINCAT = 'MEDIUM' if salbegin is between $50000 and $75000
SALBEGINCAT = 'HIGH' if salbegin is more than $75000

(5 marks)

b) Using PROC MEANS, write a SAS program to generate the following output:

Current Salary for Employee Gender and Beginning Salary Category

The MEANS Procedure

Analysis Variable : salary Current Salary							
Gender	Beginning Salary Category	N Obs	N	Mean	Std Dev	Minimum	Maximum
Female	Low	216	216	26031.921	7558.021	15750.000	58125.000
Male	High	1	1	135000.000	.	135000.000	135000.000
	Low	255	255	40732.275	18222.502	19650.000	110625.000
	Medium	2	2	85125.000	25986.174	66750.000	103500.000

(5 marks)

c) Write a SAS program to produce the following output by using appropriate SAS function to generate the new variables yr, mth, and dy

Obs	bdate	yr	mth	dy
1	02/03/1952	1952	2	3
2	05/23/1958	1958	5	23
3	07/26/1929	1929	7	26
4	04/15/1947	1947	4	15
5	02/09/1955	1955	2	9

(5 marks)

d) Your first task is to simulate four data sets for One-Way ANOVA, separately. The parameters of each data set are described in the following table:

	Data 1	Data 2	Data 3	Data 4
Treatment	1	2	3	4
Mean	100	90	80	95
SD	30	25	30	40
n	20	30	10	15

The second task is to merge all the four data sets and name the combined temporary data as ALL.

Hint: The data sets for One-Way ANOVA must fulfill the normality assumption.

(10 marks)

END OF QUESTION PAPER