Kasarani Campus
Off Thika Road
Tel. 2042692 / 3
P. O. Box 49274, 00100
NAIROBI
Westlands Campus
Pamstech House
Woodvale Grove
Tel. 4442212
Fax: 4444175

KIRIRI WOMENS' UNIVERSITY OF SCIENCE AND TECHNOLOGY
UNIVERSITY EXAMINATION, 2016/2017 ACADEMIC YEAR
SECOND YEAR, SECOND SEMESTER EXAMINATION
FOR THE DEGREE OF BACHELOR OF SCIENCE
(MATHEMATICS)

Date: 9th August, 2016.
Time: 11.00am – 1.00pm

KMA 200 - COMPUTER INTERACTIVE STATISTICS

INSTRUCTIONS TO CANDIDATES

ANSWER QUESTION ONE (COMPULSORY) AND ANY OTHER TWO QUESTIONS

QUESTION ONE (30 MARKS)

a) The functions $cumsum(x)$ and $cumprod(x)$ return a vector whose elements are the cumulative sums and products respectively. Determine the output from the following R program

```
vals1 <- c(-1,5,0,2,5,0,6,-6,5,8,2,6,-1,4,5)
vals2 <- c(5,4,4,2,4,1,3,2,3,3)
cumsum(vals1)
cumsum(vals2[c(2,6,1,10)])
cumprod(vals2)
cumprod(vals1[c(7:10,8,2,11,13,10,15,3)])
```

(4 Marks)

b) Consider the scores for students who sat for a mathematics test

$$62, 58, 68, 58, 52, 82, 48, 62, 65, 68$$

Write a well commented R program that performs the following tasks;

i) Reads and prints the following scores

(3 Marks)

ii) Computes the Arithmetic Mean, Harmonic Mean and Geometric Mean of the score.

(5 Marks)

Hence determine the output of part (ii) above

(6 marks)

c) It is expected that 10% of production from a continuous process would be defective. Find the probability that in a sample of 10 units chosen at random;

i) Exactly 2 will be defective

(3 Marks)

ii) At least 2 will be defective

(3 Marks)

Write a well commented R program that computes the values of part (i) and (ii) above

(2 Marks)

d) A study on the IQ for second year students in KWUST was carried out and the scores mean was 89 with a standard deviation of 10. Assuming that the IQ scores were normally distributed. Write a well commented R program that generates a normal random vector of n = 1000 and summarizes the simulation using a histogram superimposed with a density curve.

(4 Marks)

QUESTION TWO (20 MARKS)

Assuming that the data in Table 1 is normally distributed, write a well commented program in R that;

Expenses in KES												
51	43	58	52	50	49	52	49	49	54	52	42	53
51	50	47	52	53	44	53	45	54	44	53	48	54
43	55	52	51	52	49	52	46	49	50	44	55	50
49	54	49	47	53	48	55	50	49	53	52	52	44
51	52	48	47	5 6	52	50	58	52	48	52	54	54

Table 1

i) Reads and prints the data file *expense.csv* from drive C in the folder *KMA200*

(3 Marks)

ii) Generates the frequency distribution table of the data

(2 Marks)

iii) Determines the mode of the data

(3 Marks)

iv) Computes the measures of distribution that is kurtosis (k) the measure of degree of peakedness and skewness (sk) measure of symmetry of the data using the formulae below

$$k = \frac{(Q_3 - Q_1)}{2(P_{90} - P_{10})}, \qquad sk = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

Where

$Q_i, i = 1,2,3$: The Quartile i^{th} value

$P_i, i = 1,...,100$: The percentile i^{th} value

(10 Marks)

v) Generates a box plot of the data

(2 Marks)

QUESTION THREE (20 MARKS)

Table 2 gives laboratory and lecture scores out of 10.

Laboratory	8	3	9	2	7	10	4	6	1	5
Lecture	9	5	10	1	8	7	3	4	2	6

Table 2

Write a well comment R program that computes the following;

i) Pearson's Product Moment Correlation Coefficient

(4 Marks)

ii) Spearman Rank Correlation Coefficient

(2 Marks)

Hence or otherwise determine the values of the part (a) and (b) above, interpret the output

(14 Marks)

QUESTION FOUR (20 MARKS)

In an experiment the following values in Table 3 were obtained for two variables X and Y

X	39	65	62	90	82	75	25	98	36	78
Y	47	53	58	86	62	68	60	91	51	84

Table 3

a) Write a well commented R program that performs the following tasks;

i) Estimates and outputs the model coefficients of the simple linear model below

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

Where

i is the ith observation

(5 Marks)

ii) Predicts and outputs the Y values for X values 33, 40, 70, 100

(2 Marks)

iii) Generates a scatter diagram for the data and superimposed with the least squares regression line on the scatter diagram.

(4 Marks)

b) Determine the model coefficient estimates using the least squares methods. Hence of otherwise, predict the values of Y for X values 33, 40, 70, 100

(9 Marks)

QUESTION FIVE (20 MARKS)

The four branches of a financial firm made the sales (in millions KShs) in Table 4 in eight months of a financial year.

A	1.13	1.71	1.39	1.68	1.74	1.19	1.15	1.36
B	1.34	1.47	1.74	1.53	1.74	1.19	1.15	1.36
C	1.15	1.33	1.01	1.28	1.20	1.59	1.59	1.68
D	1.03	1.68	1.78	1.51	1.23	1.91	1.34	1.82

Table 4

Write a well commented program in R that does the following;

a) Output a vector for the mean and variance of the branches.

(4 Marks)

b) Output a vector for the coefficient of variation of the branches.

(4 Marks)

c) Compute the following parameter estimate;

$$CS = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2+(n_3-1)S_3^2+(n_4-1)S_4^2}{n_1+n_2+n_3+n_4-4}}$$

i)

$$CM = \frac{n_1\overline{X}_1+n_2\overline{X}_2+n_3\overline{X}_3+n_4\overline{X}_4}{n_1+n_2+n_3+n_4}$$

ii)

(4 Marks)

Where;

$n_i =$ Sample size for ith branch, for $i = 1,2,3,4$

$\overline{X}_i =$ Sample mean for ith branch, for $i = 1,2,3,4$

$S_i^2 =$ Sample variance for ith branch, for $i = 1,2,3,4$

d) Hence determine the output of the estimates in part(c)

(8 Marks)