(University of Choice)

MASINDEMULIROUNIVERSITY OF SCIENCE AND TECHNOLOGY (MMUST)

UNIVERSITY SPECIAL/ SUPPLEMENTARY EXAMINATIONS 2017/2018 ACADEMIC YEAR

FIRST YEAR SECOND SEMESTER REGULAR EXAMINATIONS FOR THE DOCTOR OF PHYLOSOPHY OF BIOMEDICAL SCIENCES, MA DOCTOR OF PHYLOSOPHY OF MEDICAL LABORATORY SCIENCE AND TECHNOLOGY, DOCTOR OF PHYLOSOPHY OF SCIENCE IN HUMAN NUTRITION AND DIETETICS, DOCTOR OF PHYLOSOPHY OF SCIENCE IN OPTIOMETRY AND VISION SCIENCE AND DOCTOR OF PHYLOSOPHY OF PUBLIC HEALTH .

COURSE CODE: **BML 912**

COURSE TITLE: **ADVANCED BIOSTATISTICS AND COMPUTING**

DATE: **TIME:**

INSTRUCTIONS TO CANDIDATES
Answer Question one and any other three questions

Time: 3 hours

Question One (Compulsory) (25 MARKS)

a. Differentiate between the following terms. (6 marks)
    i.   A n estimate and an estimater
    ii.   Linear and non linear correlation
    iii.   Relative risk and odds ratio

b. Study the table below hence attempt the following questions;

	Death by suicide	
Habitat	Yes	No
Inner city	102	59898
Surburbs	121	189879

    i.   Obtain the relative risk. (4 marks)
    ii.   Compute the 95% CI for the relative risk. Interpret your results. (5 marks)
c. Outline the procedure for downloading optional packages in R. (10 marks)

Question Two (25 MARKS)

a.

$$P = \begin{bmatrix} 2 & 7 & 4 \\ 6 & 1 & 3 \\ 8 & 4 & 5 \end{bmatrix} \quad and \quad Q = \begin{bmatrix} 6 & 7 & 1 \\ 2 & 9 & 2 \\ 3 & 3 & 5 \end{bmatrix}$$

Using the three matrices in (a) above, write a code to perform the following:
i. Determines PQ(1 marks)
ii. Determines P+Q(1 marks)
iii. Calculates the determinant of P (1 marks)
iv. Computes the inverse of PQ (1 marks)
v. Augments matrix P to Q column wise (1 marks)
vi. Evaluates $P^2+2P+P/Q+3$ (2 marks)

b. Consider the following data called*ExamData*whichwe assume to be saved as a text file in a folder called *StatisticalData* in the C directory.

Weigt	51	59	64	76	93	106	125	149	171	199
age	2	4	6	8	10	12	14	16	18	20

Write a single well commented program in R that does the following:
    (i)   Reads the above data into R (2 marks)
    (ii)   Sorts the weights into decreasing order (1 marks)
    (iii)   Computes the minimum and the maximum values of both weight and age. (4 marks)
    (iv)   Plots the scatter diagram of weight against age where; (7 marks)
               X-axis is labeled as 'Age'
               Y-axis is labeled as 'Weight'

The main title of the graph is 'A graph of weight against age'
The sub title is 'Question one graph'
(v). Superimpose a straight line of best fit to the graph in (iv) above. (2 marks)
(vi). Summarize each of the above variables Weight and Age in ahistogram (2 marks)

Question Three (25 MARKS)
The data below shows the household expenditure, income and savings. Fit the least squares regression model to determine the amount of expenditure given that income is $10 and savings is $2.

Income	12	15	8	5
Savings	3	1	0	2
expenditure	4	3	2	1

(25 marks)

Question Four (25 MARKS)

a. Define the term discriminant analysis and hence state the underlying assumptions in studying it. (5 marks)

b. Given that $\overline{X}_1 = (20 \quad 2 \quad 12)^T$ comes from population 1 and $\overline{X}_2 = (17 \quad 9 \quad 2)^T$ comes from population2 and both populations have a common covariance matrix,

$$S = \begin{pmatrix} 20 & -4 & 15 \\ -4 & 16 & 0 \\ 15 & 0 & 4 \end{pmatrix}$$

Use discriminant analysis to classify X=(24 1 3) (20 marks)

Question Five (25 MARKS)

a. Consider two samples of data X and Y such that $n_1 = n_2 = 9$.

$\overline{X} = 35.22, \quad \sum X^2 = 11361, \quad \overline{Y} = 31.56, \quad \sum Y^2 = 9122$

Estimate the true mean difference $\mu_1 = \mu_2$ at 5% level of significance. (15 marks)

b. The following results were obtained in a study to identify whether disease was associated with exposure to non-treated water

Untreated water	case	control
Yes	55	19
No	128	164

i.   Calculate the odds ratio and discuss your results (4 marks)

ii.   Calculate the 95% confidence interval for the OR. Interpret your results. (6 marks)