MACHAKOS UNIVERSITY COLLEGE

University Examinations for 2016/2017 academic year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS
SECOND YEAR EXAMINATION FOR DEGREE IN BACHELOR OF STATISTICS AND PROGRAMMING

SST 200: INTRODUCTION TO COMPUTER INTERACTIVE STATISTICS

INSTRUCTIONS:

ATTEMPT QUESTION ONE AND ANY TWO QUESTIONS

SECTION A

QUESTION ONE [3OMKS]

a) Define the following concepts of modern of statistics (4mks)
   i. Estimation
   ii. Inferential statistics
   iii. Hypothesis testing
   iv. Descriptive statistics

b) Let $X = \begin{pmatrix} 10 \\ 20 \\ 5 \end{pmatrix}$ and $Y = \begin{pmatrix} -2 \\ 15 \\ -6 \end{pmatrix}$. Write an R program to evaluate $X^2 + 2X + {X}/{9} + 3$
   (4mks)

c) Write the R output of the following
   i. ourmatrix<-matrix(c(2,4,3,1,-1,6),nrow=2,ncol=3,byrow=TRUE) (2mks)
   ii. our.matrix<-matrix(c(2,4,3,1,-1,6),nrow=2,ncol=3,byrow=FALSE) (2mks)

iii. seq(1,10) (1mk)

iv. s1<-seq(1,10,by=0.5) (2mks)

v. s2<-rep(c(1,4),c(10,15) (2mks)

vi. children=factor(c(1,0,1,0,0,0),leels=c(0,1),labels=c("boy","girl")) (2mks)

d) Give four reasons for using R-program (4mks)

e) The data y<-c(33,44,29,16,25,45,33,19,54,22,21,49,11,24,56) contain sales of milk in litres for 5 days in three different shops(the first 3 values are for shops 1,2 and 3 on Monday, e.t.c). Produce a statistical summary of the sales for each day of the week and also for each shop using R program   (5mks)

f) Generate Binomial random vector of x of n=1000, size=1, probability =0.5 and hence plot   (2mks)

SECTION B

QUESTION TWO [20MKS]

a) Define the term Data frame (2mks)

b) Given a data frame called our.data with the following entries.

	Year	Mean_weight	Gender	Mean_height
1	1980	71.5	M	179.3
2	1988	72.1	M	179.9
3	1996	73.7	F	180.5
4	1998	74.3	F	180.1
5	2000	75.2	M	180.3
6	2002	74.7	M	180.4

i. Read the dataset in R   (3mks)

ii. Display the Mean_weight values only from the data frame (2mks)

iii. Select the data for Males only   (2mks)

iv. Select the data that displays the Mean_height and Mean_weight only from the data set above (2mks)

c) Suppose $\bar{X} = 30, S^2 = 4, n = 100$, Determine the 95% confidence interval for the true mean. (4mks)

d) Given the following datasets 144,145,170,129,165,77,134,174,88, 100,146,172,128,121,171,117,123,120. Write R program that reads and display the data hence display its summary statistics (5mks)

QUESTION THREE [20MKS]

a) Given the following two matrices

$A = \begin{pmatrix} 0 & 6 & -3 \\ 5 & 2 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 3 & -1 \\ 4 & 1 & 6 \end{pmatrix}$. Write the R program that does the following;

   i.   Reads and display the two matrices A and B (2mks)

   ii.   Adds the two matrices   (1mk)

   iii.   Transpose of the expected output in ii. Above (2mks)

b) The table below shows the weights and heights of the seven students in Machakos university

Weight	40	60	72	57	90	95	72
Height	1.55	1.75	1.80	1.65	1.90	1.74	1.91

   i.   Read in the data in R hence find the standard deviation of weights (4mks)

   ii.   Calculate the corresponding BMI's (3mks)

   iii.   Plot the weights versus heights, clearly labeling the x and y axes with the main tittle as 'WEIGHT VS HEIGHT' (4mks)

   iv.   Plot a boxplot of weight on height (1mk)

c) Consider the following system of linear equation, solve for x1 and x2 using R (3mks)

$$3x1 + 4x2 = 4$$
$$x1 + x2 = 2$$

QUESTION FOUR [20MKS]

a) The following dataset was saved in desktop as ourdata.csv

100,146,172,128,144,145,170,129,165,77,134,174,88,121,171,117,123,120,109,153,152,
96,87,98,111,145,87,114,124,74,105,157,128,115,46,53,128,65,126,132,187,59,104,158,
89,137,185,155,120,135,106,104,144.

Write R program that;

   i)   Reads file containing the dataset and hence display the data (3mks)

   ii)   Calculates the confidence interval of the above data (4mks)

b) State the procedure for hypothesis testing (4mks)

c) Determine the output displayed in the following R program by systematically displaying the matrix at each stage (4mks)

```
Matrix1<-matrix(c(5,6,2,-3), nrow=2,ncol=2,byrow=TRUE)
Matrix1
Matrix2<-matrix(c(2,6,2,3,-1,6), nrow=3,ncol=2,byrow=FALSE)
Matrix2
Summatrix<-Matrix1+Matrix2
```

d) Write a program in R that takes in 4 variables, compute the average and returns the same (5mks)

QUESTION FIVE [20MKS]

a) Consider a survey that has data on 200 females and 300 males. If the first 200 values are from females and the next 300 values are from males, write R program that represent this vector   (5mks)

b) Five people are asked to rate the performance of a product on a scale of 1-5, with 1 representing very poor performance and 5 representing very good performance. Given the following datasets 1,3,4,2,2,1. Represent the following in R programme and give the expected output   (5mks)

c) State and explain any two types of errors used in hypothesis testing (4mks)

d) Given the following vector, $A = \begin{pmatrix} 40 \\ 100 \\ 90 \end{pmatrix}$. Write R program that find the mean and minimum of the vector A   (4mks)

e) The following R program was run. The program could not run. Identify the mistake and explain why it couldn't run   (2mks)

Matrix1<-matrix(c(2,4,3,1,-1,6),nrow=2,ncol=3,byrow=FALSE))

MATRIX1