



MACHAKOS UNIVERSITY

University Examinations for 2017/2018

SCHOOL OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF COMPUTING AND INFORMATION TECHNOLOGY

FOURTH YEAR SECOND SEMESTER EXAMINATION FOR

BACHELOR OF SCIENCE (STATISTICS AND PROGRAMMING)

BACHELOR OF SCIENCE (MATHEMATICS AND COMPUTER SCIENCE)

SST 406/SMA 468: STATISTICAL COMPUTING

DATE: 18/12/2017

TIME: 8.30-10.30 AM

INSTRUCTIONS: Answer Question One and Any Other Two Questions

QUESTION ONE-COMPULSORY (30 MARKS)

The following data representing the weight of a sample of 10 new born babies from Machakos County was collected during a study to investigate the relationship between baby weight and rate of respiratory infections. Use the data set to answer the questions that follow:

Value of weight (Kg)

2.0, 0.75, 4.0, 2.3, 5.0, 3.8, 2.4, 3.0, 3.5, 4.1

- a) Create a program in R-Programming with functions to find the mean, mode, median and print the results for the above data set (10 marks)
- b) Explain the use of the trim parameter when finding the mean of a data set (2 marks)
- c) Show the resultant vector for the above data set when a trim of 0.2 is applied (2 marks)
- d) Discuss the use of NA option when calculating the mean for the above data set (2 marks)

- e) Using an example describe a variable (2 marks)
- f) Giving reasons, state whether the following variable names qualify to be valid variable names in R-Programming
 - i **patient_name** (1 mark)
 - ii **patient_5weigth** (1 mark)
- g) Using appropriate examples, discuss four types of operators commonly used in R programming (10 marks)

QUESTION TWO (20 MARKS)

- a) Distinguish between user defined functions and in-built functions in R-Programming (2 marks)
- b) Using an example, describe what a comment is and why comments are important when writing code in R-Programming (3 marks)
- c) Create a function to print the square of numbers in a sequence. Write code to call the function and pass 4 arguments (5 marks)
- d) Using appropriate examples describe the in-built functions used to generate binomial distributions in R-Programming (10 marks)

QUESTION THREE (20 MARKS)

- a) Outline the steps you would follow in establishing a regression model in R Programming (5 marks)
- b) The following sample data representing the weight and height of a sample of 10 residents from Machakos County was collected during a study determine the relationship between height and weight among residents.
Values of weight: 74, 56, 87, 100, 45, 50, 58, 62, 70, 78
Values of height: 134,164,131,112,124,156,177,165,134,188
 - i. Describe the syntax for creating the relationship model and obtaining the coefficients for the above data sets (5 marks)
 - ii. Describe the steps and syntax to predict the height for a person with weight 75 Kg (5 marks)
 - iii. Describe the steps to visualize the regression graphically (5 marks)

QUESTION FOUR (20 MARKS)

- a) Explain five frequently used data types in R Programming Language (5 marks)
- b) Distinguish between the following types of statistical analysis and tests commonly carried out in R-programming:
 - i. R-analysis of covariance and time-series analysis (2 marks)
 - ii. R-none-linear least square test and R-Chi Square test (3 marks)
- c) Describe the syntax of a time series regression model in R-Programming explaining each of the parameters that go into the regression function (10 marks)

QUESTION FIVE (20 MARKS)

- a) Using appropriate examples, discuss three common control statements used to change program execution from its normal sequence in R Programming (6 marks)
- b) Distinguish between the following types of R-Regression models
 - iii. R -Linear regression and R-multiple regression (2 marks)
 - iv. R-Logistic regression and R-Poisson regression (2 marks)
- c) Using appropriate examples describe the in-built functions used to generate normal distributions in R-Programming (10 marks)