



# UNIVERSITY OF EMBU

2016/2017 ACADEMIC YEAR

SECOND SEMESTER EXAMINATION

FIRST YEAR EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE  
STATISTICS

STA 122: COMPUTATIONAL METHODS AND DATA ANALYSIS I

DATE: APRIL 10, 2017

TIME: 8:30-10:30 AM

INSTRUCTIONS:

Answer Question ONE and ANY Other TWO Questions.

## QUESTION ONE (30 MARKS)

- a) Define the following terms
- i. Factors [1 Mark]
  - ii. Function [1 Mark]
  - iii. Matrix [1Mark]
  - iv. Data frame [1Mark]
  - v. Vectors [1Mark]
- b) List the advantages of using R in data analysis and as a platform for statistical computing. [4 Marks]
- c) Name the four components of the flow control in Statistical computing [4 Marks]
- d) State what the following command does [5 Marks]
- ```
x <- 3 if(x>2) y <-2*x else y<-3*x
```
- e) Write an R code to compute [5 Marks]

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x - \bar{x})^2$$

**QUESTION TWO (20 MARKS)**

- a) Name the four components of the flow control in Statistical computing [4 Marks]
- 
- b) State a step by step process of plotting the pie chart of a continuous numeric variable called eruptions duration in faithful dataset of R [4 Marks]
- c) Define the term vector and comment on the following two lines below
- (i) `c(3,4,10,18,38)`
  - (ii) `c("TRUE", "TRUE", "FALSE", FALSE, "TRUE")` [3 Marks]
- d) Suppose we have the data 2, 4, 8, 2, 13,3,7,21,19. Write an R code to read the data into R command line and plot the histogram. [3 Marks]
- e) The sample mean of the values of a variable X is given by the formula
- $$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$
- write a computer programme to implement the above formula [3 Marks]
- f) State the main three components of a function structure [3 Marks]

**QUESTION THREE (20 MARKS)**

- a) Discuss a step by step procedure of importing data from [9 Marks]
- (i) SPSS
  - (ii) EXCEL
  - (iii)Text file
- b) Define the term exploratory data analysis and state its attributes [5 Marks]
- c) Write an R code to create a frequency distribution table
- (i) Draw a bar chart [3 Marks]
  - (ii) Pie chart [3 Marks]

#### QUESTION FOUR (20 MARKS)

a) Consider the following data of low temperatures for 50 days.

|    |    |    |    |    |
|----|----|----|----|----|
| 57 | 39 | 52 | 52 | 43 |
| 50 | 53 | 42 | 58 | 55 |
| 58 | 50 | 53 | 50 | 49 |
| 45 | 49 | 51 | 44 | 54 |
| 49 | 57 | 55 | 64 | 45 |
| 50 | 45 | 51 | 54 | 58 |
| 53 | 49 | 52 | 51 | 41 |
| 52 | 40 | 44 | 49 | 45 |
| 43 | 47 | 47 | 43 | 51 |
| 55 | 55 | 46 | 54 | 41 |

(i) write an r code to capture the following data [2 Marks]

(ii) Write an r code to create a frequency distribution table [5 Marks]

(iii) Write an R program to calculate the cumulative frequency curve of the above data. [5 Marks]

b) Write an R program that will sort a vector of integers in ascending order. [8 Marks]

#### QUESTION FIVE (20 MARKS)

a) Consider the following data for a group of patients in a study.

| <b>Id</b> | <b>patient initials</b> | <b>date of interview</b> | <b>age</b> | <b>Ethnic</b> | <b>marital status</b> | <b>religion</b> |
|-----------|-------------------------|--------------------------|------------|---------------|-----------------------|-----------------|
| AA0001    | DK                      | 17.06.2003               | 38         | Other         | Married               | Protestant      |
| AA0002    | SM                      | 17.06.2003               | 33         | Kikuyu        | single                | Other           |
| AA0003    | PL                      | 25.06.2003               | 40         | Luo           | divorced              | Other           |
| AA0004    | SN                      | 26.06.2003               | 28         | Kikuyu        | Married               | Protestant      |
| AA0005    | RM                      | 26.06.2003               | 43         | Kamba         | Married               | Other           |
| AA0006    | SC                      | 30.06.2003               | 30         | Kikuyu        | widowed               | Protestant      |
| AA0007    | WN                      | 01.07.2003               | 32         | Kikuyu        | single                | Protestant      |
| AA0008    | SK                      | 03.07.2003               | 26         | Kikuyu        | single                | Protestant      |

|        |    |            |    |        |           |            |
|--------|----|------------|----|--------|-----------|------------|
| AA0009 | JN | 07.07.2003 | 42 | Kikuyu | Married   | Protestant |
| AA0010 | PR | 09.07.2003 | 40 | Kikuyu | widowed   | Catholic   |
| AA0011 | BO | 09.07.2003 | 29 | Luo    | separated | Protestant |
| AA0012 | KM | 11.07.2003 | 30 | Kikuyu | widowed   | Protestant |
| AA0013 | MM | 14.07.2003 | 38 | Kikuyu | Married   | Protestant |
| AA0014 | BO | 16.07.2003 | 35 | Luo    | single    | Protestant |
| AA0015 | JO | 18.07.2003 | 27 | Luo    | Married   | Catholic   |
| AA0016 | SN | 22.07.2003 | 35 | Kikuyu | single    | Protestant |
| AA0017 | GO | 22.07.2003 | 26 | Luo    | Married   | Protestant |
| AA0018 | PM | 23.07.2003 | 32 | Kamba  | Married   | Catholic   |
| AA0019 | RM | 23.07.2003 | 31 | Kamba  | separated | Protestant |
| AA0020 | FG | 30.07.2003 | 39 | Kikuyu | widowed   | Catholic   |

(i) Assuming the data above is named carol, describe the procedure of creating a database coral in excel and importation into R software for analysis [6 Marks]

(ii) Write a step by step procedure of creating a frequency distribution table of the variable age in the above data in R [6 Marks]

b) Write a program that calculates the sum of cubes of positive integers from 1 to  $n$  for a given value of  $n$ , i.e.,

$$\sum_{i=1}^n i^3$$

Check your code against the direct formula

$$\sum_{i=1}^n i^3 = \left(\frac{n(n+1)}{2}\right)^2$$

for different values of  $n$ , such as  $n = 3$ ,  $n = 30$ , and  $n = 300$ .

[8 Marks]

---END---