



UNIVERSITY OF EMBU

2017/2018 ACADEMIC YEAR

SECOND SEMESTER EXAMINATIONS

FIRST YEAR EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE IN
STATISTICS

STA 122 COMPUTATIONAL METHODS AND DATA ANALYSIS I

DATE: APRIL 5, 2018

TIME: 11:00 AM – 1:00 AM

INSTRUCTIONS:

Answer Question ONE and ANY Other TWO Questions.

QUESTION ONE (30 MARKS)

- a) Define the following terms
- i) Algorithm (1 mark)
 - ii) Debugging (1 mark)
 - iii) Program (1 mark)
 - iv) Data frame (1 mark)
 - v) Vectors (1 mark)
- b) Name the four components of the flow control in Statistical computing (4 marks)
- c) State what the following command does (3 marks)
- $$x <- 3 \text{ if}(x > 2) \text{ y } <- 2 * x \text{ else } \text{ y} <- 3 * x$$
- d) State and describe five types of data representation in a computer (5 marks)
- e) Write an algorithm for computing the variance given by the equation below (4 marks)

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

- f) Describe a general algorithm for determining the largest representable number without producing overflow. (5 marks)

- g) Consider the continuous function

$$f(x) = \begin{cases} x^2 + 2x + 3 & \text{if } x < 0 \\ 2x - 0.5x^2 & \text{if } 0 \leq x < 2 \\ x^2 + 4x - 7 & \text{if } 2 \leq x \end{cases}$$

Write a function tmpFn which takes a single argument xVec. The function should return the vector of values of the function $f(x)$ evaluated at the values in xVec. Hence plot the function $f(x)$ for $-3 < x < 3$. (4 marks)

QUESTION TWO (20 MARKS)

- a) 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)
- b) Discuss the three sources of errors in computational methods (6 marks)
- c) Find the smaller root of the equation $x^2 - 32x + 1 = 0$ correct to four significant figures (3 marks)
- d) Find the number of terms of the exponential series such that their sum gives the value of e^x correct to six decimal places at $x = 1$. (3 marks)
- e) Describe the four features of an Algorithm (4 marks)

QUESTION THREE (20 MARKS)

- a) Highlight the consequences of violating exploratory data analysis assumptions (5 marks)
- b) Discuss a step by step procedure of importing data from (4 marks)
- i) SPSS
- ii) EXCEL
- into R.
- c) Define the term exploratory data analysis and state its attributes (5 marks)

- d) With examples, Explain what the following programming terms mean in R (6 marks)
- (i) *if()*
 - (ii) *while()*
 - (iii) *ifelse()*

QUESTION FOUR (20 MARKS)

- a) Describe major data representation in the computer (6 marks)
- b) Find the smaller root of the equation $x^2 - 400x + 1 = 0$ using four digit arithmetic. (4 marks)
- c) Define the term normalized floating point and give an example (4 marks)
- e) Clearly describe three philosophies of data analysis (6 marks)

QUESTION FIVE (20 MARKS)

- a) Identify and describe five errors encountered in numerical computation (5 marks)
- b) State the four major assumptions of exploratory data analysis (5 marks)
- c) Suppose $x_0 = 1$ and $x_1 = 2$ and

$$x_j = x_{j-1} - \frac{1}{x_{j-1}} \text{ for } j = 1, 2, \dots$$

Write a function *testLoop* which takes the single argument n and returns the first $n-1$ values of the sequence $\{x_j\}_{j \geq 0}$. that means the values of $x_0, x_1, x_2, \dots, x_{n-2}$. (5 marks)

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

$$\sum_i^n i^3$$

Check your code against the direct formula

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

for different values of n , such as $n = 3, n = 30$, and $n = 300$. (5 marks)

---END---

