



MACHAKOS UNIVERSITY

University Examinations 2021/2022 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

THIRD YEAR FIRST SEMESTER EXAMINATION FOR

BACHELOR OF SCIENCE (STATISTICS AND PROGRAMMING)

SST 302: PROGRAMMING LANGUAGE FOR STATISTICS II

DATE: 15/12/2021

TIME: 2:00 – 4:00 PM

INSTRUCTION:

Answer Question One and Any Other Two Questions

QUESTION ONE (30 MARKS)

- a) In a large city, telephone calls to 911 come on the average of two every three minutes. If one assumes an approximate Poisson process:
- What is the probability of five or more calls arriving in a nine-minute period? (3 marks)
 - Write r code to compute the above problem. (2 marks)
- b) Let $A = \begin{pmatrix} 10 \\ 20 \\ 5 \end{pmatrix}$ and $B = \begin{pmatrix} -3 \\ 15 \\ -6 \end{pmatrix}$. Write R code to evaluate $A^2 + 2A + B/9 + 3$ (3 marks)
- c) Write r code to sketch the curve $y = 2x^2 + 2x + 3$. (4 marks)
- d) In a certain restaurant, 3 coffees and a cup cake cost a total of 7 dollars. 2 coffees and 4 cup cakes cost a total of 8 dollars.
- Formulate and solve the system of linear equations. (4 marks)
 - Write r code that you would use to solve the problem above. (3 marks)
- e) Given the equation $y = (2x^2 + 3x + 3)^3$.
- Determine dy/dx . (2 marks)
 - Write r code to solve the equation. (3 marks)

- f) Given the equation $2x^2 + 3$
- $\int_0^3 2x^2 + 3 \, dx$ (2 marks)
 - Write r code to solve (i) above. (4 marks)

QUESTION TWO (20 MARKS)

- If the random variable X has a chi-square distribution with five degrees of freedom, calculate:
 - $P(X > 6.5)$ (2 marks)
 - $P(X < 11.8)$ (2 marks)
 - Write r code you would use to compute the above probabilities. (2 marks)
- In a one-year mortality investigation, 45 of the 250 ninety-year-olds present at the start of the investigation died before the end of the year. Assuming that the number of deaths has a binomial distribution with parameters $n = 250$ and q , find a symmetrical 90% confidence interval for the unknown mortality rate q . (6 marks)
 - Write r code that you would use to achieve this. (2 marks)
- Consider the data below on weekend sales of ice cream in Nairobi.

Temperature (X)	25	16	28	20	22	23	16	18
Sales (in 100 shillings) (Y)	125	79	140	103	111	115	80	91

Assume that the data is in a notepad file named *icedata* stored in directory E and folder myDocs in your computer. Write R code that does the following:

- Reads in the data into R. (2 marks)
- Does a scatter plot of the data with correct labels for the X and Y axes and has the title 'SCATTER PLOT OF ICE CREAM SALES'. (2 marks)
- Fits a simple linear regression model. (2 marks)

QUESTION THREE (20 MARKS)

- Briefly explain the difference between point and interval estimation. (2 marks)
- Explain briefly the importance of simulation. (2 marks)
- Explain briefly why it is important to set the random number seed. (2 marks)
- Suppose you are given the linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$, where $x \sim N(0,3)$ and $\varepsilon \sim N(0,2)$. Further you are given that $\beta_0 = 0.5$ and $\beta_1 = 2$.

- i. Write reproducible r code that you would use to simulate the above model taking $n=100$. (6 marks)
 - ii. Write r code to summarize the variable y. (2 marks)
 - iii. Write r code to plot the above data, clearly labelling the y and x axes and include a title for your plot. (2 marks)
- e) Write reproducible r code you would use to sample ten numbers from a sequence of numbers ranging from 1 to 100. In the event that you wish to sample with replacement, write code to do the same. (4 marks)

QUESTION FOUR (20 MARKS)

- a) Given the matrices $A = \begin{bmatrix} 2 & 4 \\ 5 & 8 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 2 \\ 6 & 7 \end{bmatrix}$, write R code to
- i. Read in and display the matrices in R. (2 marks)
 - ii. Obtain the sum of the matrices and display the result in R. (2 marks)
- b) The data `sales<-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 three values are for shops 1,2 and 3 on Monday, the next three values are for the three shops on Tuesday and so on. Produce a statistical summary of the sales for each day of the week and for each shop using R. (5 marks)
- c) A study was conducted to investigate the relationship between the price of a plot of land and the distance from the nearest tarmac road. A random sample of 12 plots was taken and the data relating the distance from the nearest tarmac road in hundred metres and the price of the plot in million Kenya shillings is as shown in the table below.
- | | | | | | | | | | | | | |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Distance | 48 | 78 | 35 | 72 | 10 | 60 | 40 | 22 | 84 | 25 | 12 | 48 |
| Price | 1.2 | 0.4 | 1.4 | 0.8 | 1.8 | 0.6 | 1.5 | 1.2 | 0.8 | 1.5 | 1.6 | 1.2 |
- i. Assume the data is stored as file *plot.txt* in directory E and folder RDocs in your computer. Read in the file into R with the headings. (2 marks)
 - ii. Write R code to calculate the mean and standard deviation of the two variables. (2 marks)
 - iii. Write R code to compute the Pearson's product moment co-efficient of correlation between the distance from the tarmac and the price of the plot. (2 marks)
- d) Six 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. The following data on

their responses 1,3,4,2,2,1 was obtained. Write R code to represent this data and give the expected output. (5 marks)

QUESTION FIVE (20 MARKS)

- a) Outline the procedure used for hypothesis testing. (4 marks)
- b) Differentiate between Type I and Type II errors as used in hypothesis testing. (2 marks)
- c) Explain briefly what is meant by the term power of a test as used in hypothesis testing. (2 marks)
- d)
 - i. Determine a 95% confidence interval for the average height of 10 year old children assuming that heights have a $N(\mu, \sigma^2)$ distribution based on a random sample of 5 children whose heights are 124cm, 122cm, 130cm, 125cm and 132cm. (6 marks)
 - ii. Write r code that you would use to read in the data and compute the confidence interval. (4 marks)
- e) In order to increase the efficiency with which employees in a certain organisation can carry out a task, 5 employees are sent on a training course. The time in seconds to carry out the task both before and after the training course is given below for the 5 employees:

Employee	1	2	3	4	5
Before	42	51	37	43	45
After	38	37	32	40	48

Assuming the following data is already saved in the workspace as *employeedata*, write R code to test whether the training course has had the desired effect. (2 marks)