



MACHAKOS UNIVERSITY

University Examinations 2021/2022 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

FOURTH YEAR SECOND SEMESTER EXAMINATION FOR
BACHELOR OF SCIENCE (STATISTIC AND PROGRAMMING)

BACHELOR OF SCIENCE (MATHEMATICS)

BACHELOR OF ARTS

SMA 468: STATISTICAL COMPUTING

DATE: 15/12/2021

TIME: 2:00 – 4:00 PM

INSTRUCTION:

Answer Question One and Any Other Two Questions

QUESTION ONE (30 MARKS)

- a) Differentiate the following R-functions as they apply in data analysis
- i. `data.frame()` and `cbind()`
 - ii. `seq()` and `rep()`
 - iii. `is.()` and `as.()`
 - iv. `head()` and `tail()`, (8 marks)
- b) Given $A = \{2, -1, 3, 1, 4, 0, 5, 1, 1\}$, write down the R commands for;
- i. Constructing a 3×3 matrix A, such that its first column elements are 2, -1, 3, second column; 1, 4, 0 and third column 5, 1, 1
 - ii. $A_1 = A \times A$
 - iii. The determinant of A_1
 - iv. The inverse of A_1 (5 marks)

- c) Write the R script for creating the following
- A sample of 50 numbers starting at 10 which are incremented by 5.
 - A random sample of 20 numbers from 70 to 120 with replacement
 - The first ten multiples of 9 (6 marks)
- d) Suppose the mean checkout time of a supermarket cashier is three minutes. Write the R script for determining the probability of a customer checkout being completed by the cashier in less than two minutes. (3 marks)
- e) The probability distribution of a continuous variable x is defined by

$$f(x) = \begin{cases} \frac{1}{40}(2x+3), & 0 \leq x \leq 5 \\ 0 & \text{elsewhere} \end{cases}$$

Write the R script for determining;

- The expected value of x
- Variance of x (8 marks)

QUESTION TWO (20 MARKS)

- a) Given the vectors $A <- c(11, 34, 29, 31, 9)$ and $B <- c(2, 5, 7, 3, 1)$. Write down the output after executing the following R functions;
- `order(B)`
 - `tail(A, 3)`
 - `rep(c(B), 2)`
 - `append(A, B)` (4 marks)
- b) Past experience has shown that the average number of school buses passing the university gate per hour is 3. Write R script to determine the probability that, in a given hour 5 buses will pass by the university gate. (4 marks)
- c) Two groups of patients were placed on diets with high and low protein content, and their blood pressure monitored. The results are summarized in table 3.

Table 3: Patients Blood Pressure

Protein status	Protein							
High Protein	134	146	104	119	124	83	113	129
Low Protein	70	118	101	85	107	102	-	-

The test statistics output at $\alpha=5\%$, was as summarized in table 4.

Table 4: Sample t-test output

	<i>t</i>	<i>df</i>	<i>Sig.(2-tailed)</i>	<i>Mean difference</i>	<i>95% CI of the difference</i>	
					<i>Lower</i>	<i>Upper</i>
<i>Attendance</i>	<i>2.19</i>	<i>12</i>	<i>0.049</i>	<i>1.500</i>	<i>0.107</i>	<i>43.559</i>

- State the hypothesis for the above scenario
- Write the R script that executed the given output
- Based on the sample t-test output make statistical conclusion
- Interpret the 95% CI lower and upper difference values (12 marks)

QUESTION THREE (20 MARKS)

- Highlight three basic data types in R; (3 marks)
- Given $A=\{23,34,56,77,32,11,34,54,61,49,45,26\}$ and $B=\{15, 23, 24, 37, 19, 46, 54, 22, 34, 38,29,33\}$

Write the R Scripts for constructing/executing

- Scatter plot and a line of best fit (2 marks)
 - Box plot for each in a common axis (2 marks)
 - Average, variance and the standard deviation for each (3 marks)
- c) Write R script for determining;
- A binomial distribution table for head showing up after tossing a coin three times. (2 marks)
 - Probability of no head in the first toss (2 marks)
 - Expected value (2 marks)
 - Variance (2 marks)
 - Cumulative distribution (2 marks)

QUESTION FOUR (20 MARKS)

- a) The heights of 100 students in a class is assumed to be normally distribute with a mean height of 130cm and standard deviation of 15cm. Write the R script for determining;
- The probability that the height of a student randomly picked is between 135cm to 150cm
 - The 95% confidence interval (8 marks)
- b) Three types of machines' output was noted and recorded for one week and results were summarized in table 5 and the ANOVA analysis results in table 6.

Table 5: Machines output

Output	23	46	77	29	21	43	54	65	65	47	44	52	69	76	38
Machine type	1	1	2	1	3	3	3	2	2	1	1	2	2	2	1

Table 6: ANOVA Table

Source of Variance	df	Sum sq	Mean sq	F value	Pr(>F)
Group	2	3032	x	12.39	0.0012
Residue	z	y	122.4		
Total	14	4501			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Using the above information;

- State the hypothesis testing for the above scenario
- Determine the values of x , y and z
- Write the R script that generated the above output
- With justifications state whether the machines' output differed significantly at $\alpha=5\%$. (12 marks)

QUESTION FIVE (20 MARKS)

- a) Briefly discuss four types of optimality criteria and their efficiencies (8 marks)
- b) Tumaini chain stores run several outlets in and out of the city (CBD). The management wishes to establish if; distance from the CBD in kilometers, amount of money spent on advertisement and salesmanship in Kenya shillings ("000") , and the gender of the cashier influences the sales in Kenya shillings ("000"). A random sample of their records is summarized below.

Distance from CBD(x1)	5	4	1	9	7	3	4	5
Promotions costs (x2)	16.7	16.8	18.2	16.3	17.3	18.2	15.9	17.2
Cashiers' Gender (x3)	F	M	M	M	F	F	F	M
Sales (y)	174.4	164.4	244.2	154.6	181.6	207.5	152.8	163.2

Using the data above data, write the R script to answer the following

- Fitting a simple linear regression model using distance to predict sales
- Fitting a multiple linear regression model using distance, advertisement and cashier's gender to predict sales
- By grouping all the independent variable into a design matrix (X)

Write the R script of determining the best fit and the regression coefficients.

(12 marks)