



MACHAKOS UNIVERSITY

University Examinations 2019/2020 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

FOURTH YEAR SPECIAL /SUPPLEMENTARY EXAMINATION FOR

BACHELOR OF SCIENCE IN MATHEMATICS

BACHELOR OF SCIENCE IN STATISTIC & PROGRAMMING

BACHELOR OF ARTS

SMA 468: STATISTICAL COMPUTING

DATE: 21/1/2021

TIME: 2.00-4.00 PM

INSTRUCTION: Answer Question *ONE* which is compulsory and any other *TWO* Questions

QUESTION ONE (30 MARKS)

- a) Given $A = \{2, -1, 3, 1, 4, 0, 5, 1, 1\}$, write down the R commands for
- 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
 - $A_1 = A \times A$
 - The determinant of A_1
 - The inverse of A_1 (8 marks)
- b) Write the R script for the following
- Creating 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
 - A sample of (6 marks)
- c) The probability distribution of a continuous variable x is defined by

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

Write the R script for determining;

- i. The expected value of x
- ii. Variance of x (8 marks)

d) Makongeni eight upper primary school pupils visiting the school clinic had their age and weight taken as summarized in the table below.

Student	A	B	C	D	E	F	G	H
Age (years)	9	10	11	12	13	13	14	14
Weight (kg)	25	23	24	26	36	35	35	34

Write the R script for executing the following

- i. Scatter diagram for age and weight
- ii. A smooth line of best fit on the scatter diagram
- iii. The correlation between age and weight
- iv. The regression equation weight as a function of age (8 marks)

QUESTION TWO (20 MARKS)

a) The weight and the gender of 12 children born on 1st January in a certain hospital in Nairobi was recorded as below

Child	A	B	C	D	E	F	G	H	I	J	K	L
Weight	2.9	4.3	5.2	3.5	3.7	2.5	4.0	3.9	4.3	5.1	3.2	4.1
Gender	F	F	M	M	F	F	M	F	F	M	M	F

Write down the R script for determining

- i. Average weight of the 12 children
- ii. Middle weight
- iii. The box plot per gender in a single axis
- iv. Histogram of the weight distribution (8 marks)

b) Two groups of patients were placed on diets with high and low protein content, and their blood pressure monitored. The results are summarized below.

Protein status	Protein							
High Protein	13	146	104	11	124	83	113	129
	4			9				
Low Protein	70	118	101	85	107	102	-	-

The test statistics output at $\alpha=5\%$. Is as summarized below

Table 1: Sample t-test output

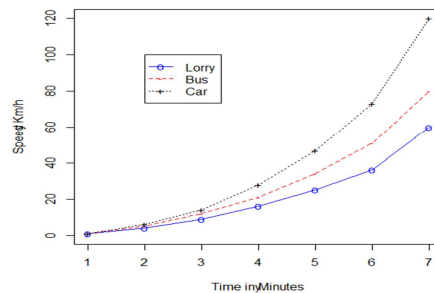
	<i>t</i>	<i>df</i>	<i>Sig.</i> (2-tailed)	<i>Mean difference</i>	<i>95% CI of the difference</i>	
					<i>Lower</i>	<i>Upper</i>
<i>Attendance</i>	<i>2.1</i>	<i>1</i>	<i>0.049</i>	<i>1.500</i>	<i>0.107</i>	<i>43.559</i>
	<i>9</i>	<i>2</i>				

- 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)

- Using R script write a distribution table for head showing up after tossing a coin three times and hence the R script for determining
 - Probability of no head in the first toss
 - Expected value
 - Variance
 - Cumulative distribution (10 marks)
- The data below shows the speed gained by three types of vehicles in a given time
 Time in Minutes {1,2,3,4,5,6,7}
 Lorry {10,40,50,66,70,75,80}
 Bus {10, 50, 72, 85, 94, 101, 120}
 Car {10, 60, 84, 95, 105, 110, 115}

The speed gain is represented in the linear graphs below.



Write the R script for the output (10 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 picked randomly is 135 to150
 - The 95% confidence interval (8 marks)
- b) Three types of machines' output was noted and recorded for one week and results were summarized in the ANOVA table below.

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

Table 2: 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 output values and with justifications;

- 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
- By justifying state whether the machines output differed significantly at $\alpha=5\%$

(12 marks)

QUESTION FIVE (20 MARKS)

- a) Past experience has shown that the average number of motorbikes passing along the university gate per minute is 3. Write R script to determine the probability that, in a given minute 5 motorbikes will pass by the university gate. (4 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, amount spent to advertise and do salesmanship, and the gender of the cashier influences the amount of sales. By using the data below advice the Tumaini management.

Distance(x1)	5	4	1	9	7	3	4	5
Advertisement(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, write the R script to answer the following

- i. Fit a simple linear regression model using distance to explain sales
- ii. Fit a multiple linear regression model using distance and advertisement to explain sales
- iii. Fit a multiple linear regression model using distance, advertisement and cashier gender to explain sales
- iv. By combining all the independent variable into a design matrix determine the best fit and interpret the betas.

(16 marks)