MAASAI MARA UNIVERSITY

REGULAR UNIVERSITY EXAMINATIONS
2023/2024ACADEMIC YEAR
SECOND YEAR FIRST SEMESTER

SCHOOL OF PURE, APPLIED, AND HEALTH SCIENCES
BACHELOR OF APPLIED STATISTICS WITH COMPUTING

COURSE CODE: STA 2220-1
COURSE TITLE: STATISTICAL DATA ANALYSIS II

DATE: 16/4/2024 **TIME: 1100-1300 HRS**

INSTRUCTIONS TO CANDIDATES

- **Question One is Compulsory**
- **Answer Any Other TWO QUESTIONS**

This paper consists of 4 printed pages. Please turn over.

Question One (20 Marks)

a) Give advantages and disadvantages of face-to-face interviews **[2marks]**
b) outline the procedure for importing data from excel files into R **[3marks]**
c) outline the procedure of creating data frames in R **[3marks]**
d) Discuss the concept and meaning of inferential statistics **[2marks]**
e) Explain the meaning of power of a test **[1marks]**
f) Classify each of the following variables as nominal or ordinal
   i. Sex; male, female **[1mark]**
   ii. Color of hair: Blonde, Red, Neutral, Dark **[1mark]**
   iii. Degree of certification: low, medium, high **[1mark]**
   iv. Yearly income: below 10000, 10001-20000,20001-30000, **[1mark]**
g) Each subject in a sample of 100 men and 100 women is asked to indicate which of the facts is responsible for increase in teenage crime: A- increasing gap in income between the rich and the poor; B- the increase in unemployment; C- Bad television programs.
   A cross tabulation of the response by gender is given below.

	Factors Responsible		
Gender	A	B	C
Male	60	81	75
Female	75	87	86

   i. Is it valid to apply the chi-squired test of independence to this 2×3 contingency table? **[1marks]**
   ii. Test hypothesis that responses are independent of gender using both the chi-squire test and the likelihood test **[4marks]**

Question Two (15 Marks)

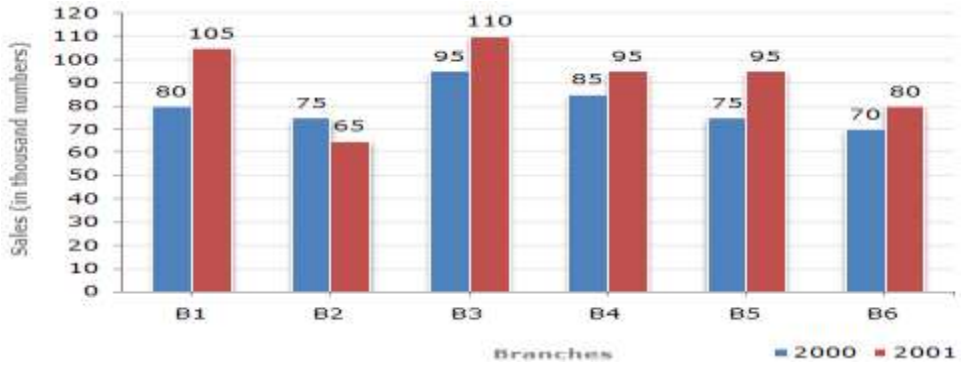
a) Given a data set; 13,26,18,11,9,36,15,12,10.

i) Write a command that will assign the data set to an object "x" in R **[2marks]**

ii) Write R commands and their corresponding outputs to evaluate; length, mean, range, and summary for the random variable "x" **[3marks]**

b) The bar graph given below shows the sales of books (in thousand number) from six branches of a publishing company during two consecutive years 2000 and 2001.

Sales of Books (in thousand numbers) from Six Branches - B1, B2, B3, B4, B5 and B6 of a publishing Company in 2000 and 2001.



---

i. What is the ratio of the total sales of branch B2 for both years to the total sales of branch B4 for both years? **[2marks]**

ii. Total sales of branch B6 for both the years is what percent of the total sales of branches B3 for both the years? **[2marks]**

iii. What percent of the average sales of branches B1, B2 and B3 in 2001 is the average sales of branches B1, B3 and B6 in 2000? **[2marks]**

iv. What are the average sales of all the branches (in thousand numbers) for the year 2000? **[2marks]**

v. Total sales of branches B1, B3 and B5 together for both the years (in thousand numbers) is? **[2marks]**

Question three (15marks)

i) Give an R command that can solve the following problems

a. $\int_{-1.96}^{1.96} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ **(3marks)**

b. $\int_{0}^{\infty} \frac{1}{(x+1)\sqrt{x}} dx$ **(3marks)**

ii) Give R codes for differentiating with respect to x the following functions;

a. Sin (x-3) **(3marks)**

b. $x^3 - 2x^2 + 3,$ **(3 marks)**

c. e^{-x^2} **(3marks)**

Question four (20marks)

a) Discuss the procedure for hypothesis testing. **[3marks]**

b) Distinguish the following terms as they apply in data analysis

i. Simple linear and multiple linear regression **[2marks]**

ii. Response variable and independent variable **[2marks]**

iii. Data cleaning and data screening **[2marks]**

iv. Multicollinearity and outliers **[2marks]**

b) In a comparison of the cleaning action of four detergents, 20 pieces of white cloth were first soiled with India ink. The cloths were then washed under controlled conditions with 5 pieces washed by each of the detergents. Unfortunately, three pieces of cloth were lost in the course of the experiment. Whiteness readings, made on the 17 remaining pieces of cloth are as shown below.

Detergent

A	B	C	D
77	74	73	76
81	66	78	85
61	58	57	77
76	70	69	64
69	20	63	72

Assuming all whiteness readings to be normally distributed with common variance, test the hypothesis of no difference between the four brands as regards mean whiteness readings after washing. **[4marks]**

//END//