MAASAI MARA UNIVERSITY

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

SCHOOL OF SCIENCE

BACHELOR OF SCIENCE IN APPLIED STATISTICS WITH COMPUTING

COURSE CODE: STA 2215-1

COURSE TITLE: PRINCIPLES OF STATISTICAL INFERENCE

DATE: 22/4/24 **TIME: 1100-1300HRS**

INSTRUCTIONS TO CANDIDATES

1. Answer **ALL** questions from section A and any **TWO** from section B.
2. Use of sketch diagrams where necessary and brief illustrations are encouraged.
3. Read the instructions on the answer booklet keenly and adhere to them.

*This paper consists of **four** printed pages. Please turn over.*

SECTION A

QUESTION ONE (20 MARKS)

a) Briefly explain the four types of categorical data **[4mks]**

b) When X is a binomial random variable with parameters n and p, Show that the sample proportion $\hat{p} = \frac{x}{n}$ is an unbiased estimator of ?. **[3Mks]**

c) A study was taken to establish whether there is a difference in the mean sales between the male marketers and female ones. The monthly sales in KES 100,000 for the marketers grouped by gender are shown below. Test the appropriate hypotheses using $\alpha = 0.05$ significance level. **[5Mks]**

Male: 27.4, 25.4, 28.5, 31.1, 30.4, 31.5, 23.4, 27.5, 30.2, 25.8, 26.6, 24.6, 24.3, 26.3 and 26.5

Female: 27.8, 31.2, 29.2, 25.6, 26.8, 32.9, 28.3, 30.3, 25.5, 28.8, 28.8 and 26.8

d) Let $x_1, x_2, ...x_n$ be a random sample from a gamma distribution with parameters α and β having $E[x] = \alpha\beta$ and $Var[x] = \alpha\beta^2$.

(i) Obtain the moment estimators of α and β **[3mks]**

(ii) The data below shows the survival time X in weeks of a randomly selected male mouse exposed to 240 rads of gamma radiation. Assuming it has a gamma distribution, compute the estimates of α and β.

152, 115, 109, 94, 88, 137, 152, 77, 160, 165, 125, 40, 128, 123, 136, 101, 62, 153, 83 and 69. **[5mks]**

SECTION B (30 MARKS)

Answer any TWO Questions

QUESTION TWO (15 MARKS)

a) A company packages a particular product in cans of three different sizes, each one using a different production line. Most cans conform to specifications, but a quality control engineer has identified the following reasons for non-conformance: (1) blemish on can; (2) crack in can; (3) improper pull tab location; (4) pull tab missing; (5) other. A sample of nonconforming units is selected from each of the three lines, and each unit is categorized according to reason for nonconformity, resulting in the following contingency table data:

Production Line	Reason for Nonconformity					
	Blemish	**Crack**	**Location**	**Missing**	**Other**	**Sample size**
1	34	65	17	21	13	150
2	23	52	25	19	6	125
3	32	28	16	14	10	100
Total	89	145	58	54	29	375

Does the data suggest that the proportions falling in the various non-conformance categories are not the same for the three lines? Use a level of significance of 0.05

[8mks]

b) A study was made on the profitability of certain small ventures depending on amount invested. The data were recorded as follows in KES 10000;

Invested amount (x)	2.1	1.6	1.9	1.7	1.4	1.2	1.3	1.1	2.3	1.4
Profit (y)	10.6	7.7	8.6	7.6	7.8	5.9	7.2	5.4	9.6	5.6

(i) Determine the regression equation **[3mks]**

(ii) Evaluate the Pearson product moment correlation coefficient **[2mks]**

(iii) Compute the Spearman rank correlation coefficient **[2mks]**

QUESTION THREE (15 MARKS)

A local bank has three branch offices. The bank has a liberal sick leave policy, and a vice-president was concerned about employees taking advantage of this policy. She thought that the tendency to take advantage depended on the branch at which the employee worked. To see whether there were differences in the time employees took for sick leave, she asked each branch manager to sample employees randomly and record the number of days of sick leave taken during 2015. Twenty employees were chosen, and the data are listed below:

A	**B**	**C**	**D**
13	13	11	13
13	15	12	7
12	15	12	12
16	17	14	8
25	25	22	10

Does the data indicate a difference in branches? Use $\alpha = 0.05$ **[15mks]**

QUESTION FOUR (15 MARKS)

a) Let y be the sales at a fast-food outlet (KES 1000), x_1 be the population within a 2 kilometer radius (1000's of people) and x_2 be number (in hundreds) of competing outlets within a 2 kilometer radius. Fit a multiple linear regression model and test the significance of both the fitted model and the two independent variables. **[12mks]**

Sales (y)	Population (x_1)	Competition (x_2)
101	81.7	19.9
142	103.8	18.7
117	96.5	26.1
104	95.2	24.5
109	92.9	21.6
132	99.1	23.3
107	85.4	28.2
118	90.5	21.4
103	95.6	25.5
120	83.4	19.9
131	106.7	21.6
123	92.4	22.9

c) A plastics manufacturer has developed a new type of plastic trash can and proposes to sell them with an unconditional 6-year warranty. To see whether this is economically feasible, 20 prototype cans are subjected to an accelerated life test to simulate 6 years of use. The proposed warranty will be modified only if the sample data strongly suggests that fewer than 90% of such cans would survive the 6-year period. During the test 12 cans survive the test. Should the manufacturer implement the unconditional 6-year warranty? Test at $\alpha = 0.05$. **[3mks]**

END//