MAASAI MARA UNIVERSITY

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

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

COURSE CODE: STA 2219-1

COURSE TITLE: CATEGORICAL DATA ANALYSIS

DATE: **TIME:**

INSTRUCTIONS TO CANDIDATES

INSTRUCTIONS

i) Answer Question **One** and **any Two** Questions

ii) Show all the workings clearly

iii) Do not write on the question paper

Question One –(20 Marks)

a) Define the following categorical variables and give an example for each
i) Ordinal variable. **(3 Marks)**
ii) Nominal variable. **(3 Marks)**

b) When commercial aircraft are inspected, wing cracks are reported as nonexistent, detectable, or critical. The history of a particular fleet indicates that 70% of the planes inspected have no wing cracks, 25% have detectable wing cracks, and 5% have critical wing cracks.five planes are randomly selected. Find the probability that
i) one has a critical crack,two have detectable cracks, and two have no cracks. **(3 Marks)**
ii)at least one plane has critical cracks. **(4 Marks)**

c) A 20-year study of British male physicians noted that the proportion who died from lung can-Cer was 0.00140 per year for cigarette smokers and 0.00010 per year for nonsmokers. The Proportion who died from heart disease was 0.00669 for smokers and 0.00413 for nonsmok-ers.
i) Describe the association of smoking with lung cancer and with heart disease, using the diffe-rence of proportions, the relative risk, and the odds ratio. Interpret. **(4 Marks)**
ii) Which response (lung cancer or heart disease) is more strongly related to cigarette smoking, in terms of the reduction in deaths that could occur with an absence of smoking? **(3 Marks)**

Question two –(15 Marks)

a) A survey of voter sentiment was conducted in four areas to compare the fraction of voters favoring candidate A. Random samples of 200 voters were polled in each of the four areas, and the results are shown in the table below.

	Area			
Opinion	1	2	3	4
Favor A	59	48	76	53
Do not favor A	141	152	124	147

Do the data present sufficient evidence to indicate that the fractions of voters favoring candi-Date A differ in the four areas? Take $\alpha = 0.05$. **(7 Marks)**

b) A coin is flipped twice. Let X= number of heads obtained, when the probability of a head for a flip equals π.
i) Assuming π=0.50, specify the probabilities for the possible values for X, and find the distributions mean and standard deviation. **(3 Marks)**
ii) Find the binomial probabilities for X when π equals 0.60 and 0.40. **(2 Marks)**
iii) Suppose you observe *y*=1, calculate the maximum likelihood estimator of π.
(3 Marks)

Question Three –(15 Marks)

a) When the 2000 General Social Survey asked subjects whether they would be willing to accept cuts in their social standard of living to protect the environment 344 of 1170 subjects said "yes."

i) Estimate the population proportion who would say "yes." **(2 Marks)**

ii) Contruct and interpret a 99% confidence interval for the population proportion who would say "yes." **(4 Marks)**

b) For a period of n=50 weeks, the number of accidents at an intersection was checked. The data is
Given in the table below:

Number of accidents, x	0	1	2	3 or more
Frequency, f	32	12	6	0

At $\alpha = 0.05$, test the hypothesis that the random variable X has a poisson distribution.

(9 Marks)

Question Four – (15 Marks)

The table below gives the total seasonal rainfall at Sacramento, California, for the 90-year period 1854-1944.

Rainfall (mm)	f
7.5 – 10.5	12
10.5 – 13.5	10
13.5 – 16.5	15
16.5 – 19.5	19
19.5 – 22.5	12
22.5 – 25.5	14
25.5 – 37.5	8

At α =0.05 level of significance can the data be considered as satisfying a normal distribution?
(9 Marks)

b) For diagnostic testing, let X= true status(1= disease, 2= no disease) and Y = diagnosis(1= positive, 2= negative). Let $\pi_i = P(Y=1 | X=i), i=1,2.$

i) Explain why sensitivity = π_1 and specificity = $1-\pi_2$. **(2 Marks)**

ii) Let γ denote the probability that a subject has the disease. Given that the diagnosis is positive, use Baye's theorem to show that the probability a subject truly has the disease is
$\pi_1\gamma / \left[\pi_1\gamma + \pi_2(1-\gamma)\right]$. **(4 Marks)**