



MAASAI MARA UNIVERSITY

**REGULAR UNIVERSITY EXAMINATIONS
2018/2019 ACADEMIC YEAR
THIRD YEAR SECOND SEMESTER
EXAMINATION**

**SCHOOL OF SCIENCE AND INFORMATION
SCIENCE
BACHELOR OF SCIENCE IN APPLIED
STATISTICS WITH COMPUTING**

**COURSE CODE: STA 3221
COURSE TITLE: STATISTICAL
MODELLING**

**DATE: 25TH APRIL, 2019
13:00HRS**

TIME: 11:00 -

INSTRUCTIONS:

**ANSWER QUESTION ONE AND ANY OTHER TWO
QUESTIONS.**

SECTION ONE (30 MARKS)

a) i) State four principle steps followed when using monte-Carlo method in simulations of data

(4Marks)

ii) Assuming a distribution $f(x)$ is to be sampled using markov chain Monte Carlo-Metropolis Algorithm approach. What are the five steps that must be used

(5marks)

b) Differentiate between stochastic and deterministic models citing examples in each

(4marks)

c) Discuss the following modelling approaches :

(6marks)

i) Fixed effect models, random effect models and mixed effects model

ii) Linear and non-linear models

iii) Multivariate models and univariate modelling

d) In building a simulation model for the petrol filling station with a single pump served by a single service man. Assuming the arrival of cars and service time are random variables. Identify the following as used in simulation (6 Marks)

i) States

ii) Events

iii) Entities

iv) Queue

v) Random realizations

vi) Distributions

- d) Write the probability density function (P.D.F.) for the GLM using Gamma distribution and show the distribution is a member of the exponential family. Show the expected value and variance of the distribution (5marks)

QUESTION TWO (20 MARKS)

A survey was conducted by researchers to collect data on demographic characteristics and attitudes of residents. In 2007 the survey had two attitude items measured on a 5-point Likert scale.

Item 1: A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.

Item 2: Working women should have paid maternity leave.

Responses to these items are tabulated below;

Item 1	Item2					
	Strongly Agree	Agree	Neither	Disagree	Strongly Disagree	
Strongly Agree	90	96	22	17	2	227
Agree	102	199	48	30	5	384
Neither	7	2	5	8	0	22
Disagree	42	100	20	36	7	205
Strongly Disagree	9	18	7	10	2	46
	250	415	102	101	16	884

What's the nature of the dependency between the two items and any assumptions you made?

QUESTION THREE (20 MARKS)

The following data were collected after a food poisoning outbreak. It is suspected that the potato salad, the crab salad or both were the cause. The contingency table below shows the results of a random survey of 304 diners: whether they were sick (food-poisoned) and the food that they ate.

	Potato Salad		No Potato Salad	
	Crab Salad	No Crab Salad	Crab Salad	No Crab Salad
Not Sick	80	24	31	23
I Sick	120	22	4	0

(a) What is a generalized linear model? What is the saturated model in the context of generalized linear models?

(4 Marks)

(b) A log-linear generalized linear model with a Poisson distribution was fitted to the data. The computer output

below shows the analysis of deviance table for these data.

left-hand column of that row and all the rows above it.

	Deviance	Change in Deviance
Intercept	295.253	
Sick	294.779	0.474
Potato	169.664	125.115
Crab	73.871	95.793
potato: crab	63.196	10.676
sick:potato	6.482	56.714
sick:crab	2.743	3.739
sick:potato:crab	4.123e-10	2.743

Each row of the table refers to a model containing the terms given in the

Find a suitable model for these data and give an interpretation.
What can be concluded about the likely cause of the outbreak?

(6 Marks)

(c) How is a Pearson residual defined in this model? .

(2 Marks)

(c) Calculate the Pearson residuals for your fitted model.
Do they indicate an adequate fit to the data?

(4 Marks)

(a) Another researcher suggests that a logistic regression model with sickness as response would be more appropriate for these data than the log-linear model. Describe briefly the different aims of these two approaches, and discuss whether your colleague's suggestion is a good one.

QUESTION FOUR (20 MARKS)

In surgery, it is desirable to give enough anesthetic so that patients do not move when an incision is made. It is also desirable not to use much more anaesthetic than necessary. In an experiment, patients are given different concentrations of anaesthetic. The response variable is whether or not they move at the time of incision 15 minutes after receiving the drug.

	Concentration					
	0.8	1.0	1.2	1.4	1.6	2.5
Move	6	4	2	2	0	0
No move	1	1	4	4	4	2
Total	7	5	6	6	4	2

- a) Suggest an appropriate model to explain the impact of anaesthetic on the response variable.(6 Marks)
- b) Write an R program which reads these data into R data set called ana. The program should then produce a contingency table and a glm analysis(8marks)

- c) From the glm analysis below, what can you conclude between concentrations of anaesthetic and movement of patients?(6maks)

	coef.est	coef.se
(Intercept)	-6.469	2.418
cone	5.567	2.044

k =
n = 30, 2

residual deviance = 27.8, null deviance = 41.5 (difference = 13.7):

QUESTION FIVE (20 MARKS)

- Use appropriate algorithm explain the use the Monte Carlo technic to prove that Ordinal Least Square (OLS) estimators of General Linear Regression (GLR) model are BLUEs
- Highlight four instances I that will prompt a researcher to simulate data instead of real data.