



(University of Choice)

**MASINDE MULIRO UNIVERSITY OF SCIENCE AND TECHNOLOGY
(MMUST)
MAIN CAMPUS**

**UNIVERSITY EXAMINATIONS
MAIN EXAM**

2021/2022 ACADEMIC YEAR

**SECOND YEAR SECOND SEMESTER EXAMINATION
FOR THE DEGREE OF BACHELOR OF SCIENCE IN EPIDEMIOLOGY AND
BIOSTATISTICS (BSc EPIMED)**

**COURSE CODE: HEM 225
COURSE TITLE: CATEGORICAL DATA ANALYSIS**

DATE: 21/04/2022

TIME: 8.00-10.00 AM

INSTRUCTIONS TO CANDIDATES:

Answer all Questions from section A and any other two questions from section B
TIME: 2 Hours

MMUST observes ZERO tolerance to examination cheating
Paper Consists of 4 Printed Pages. Please Turn Over



SECTION A (40 MKS) COMPULSORY

1. State and explain 2 types of data that a statistician can use to describe data. (4 marks)
2. Given a random sample of size n from a population whose pdf is;

$$f(x, \alpha, \beta) = \begin{cases} \frac{1}{\sqrt{2\pi\beta}} e^{-\frac{1}{2\beta}(x-\alpha)^2} & -\infty < x < \infty, \beta > 0 \\ 0, elsewhere \end{cases}$$

Obtain the MLE of α and β (6 marks)

3. In a sociological investigation, three men and three women are asked if they watch football regularly. All the men said yes while all the women said no. Is this difference statistically significant at $\alpha = 0.01$ (8 marks)
4. Using simple algebra show that ;

$$\sum_{i=1}^r \sum_{j=1}^c \log E_{ij} = c \sum_{i=1}^r \log n_i + r \sum_{j=1}^c \log n_j - rc \log n$$

Can be simplified to;

$$\log E_{ij} = U + U_1(i) + U_2(j)$$

where

$$U = \frac{\sum_{i=1}^r \sum_{j=1}^c \log E_{ij}}{rc}$$

$$U_{1(i)} = \frac{\sum_{j=1}^c \log E_{ij}}{c} - \frac{\sum_{i=1}^r \sum_{j=1}^c \log E_{ij}}{rc}$$

$$U_{2(j)} = \frac{\sum_{i=1}^r \log E_{ij}}{r} - \frac{\sum_{i=1}^r \sum_{j=1}^c \log E_{ij}}{rc}$$

(6 marks)

5. The following results were obtained in a study to identify whether disease was associated with exposure to non treated water.

untreated water	case	control
Yes	55	19
No	128	164

- a. Calculate the odds ratio and discuss your results. (3 marks)
- b. Calculate the relative risk and discuss your results. (3 marks)
- c. Compute the 95% confidence interval for the odds ratio. (4 marks)
- d. Determine the Chi-square statistic, test the hypothesis and interpret your results. Compare your result with your result in (b) above. (6 marks)

SECTION B (answer any two questions)

QUESTION ONE (15 MKS)

- a. In a broad general sense, psychiatric patients can be classified as psychotics or neurotics. A psychiatric white studying the symptoms of a random sample of 40 from each of these populations found that where else 16 patients in the neurotic group has suicidal feelings, only 8 in the psychiatric group suffered in this way. Test if there is an association between the two psychiatric groups in the presence or absence of suicidal feelings. The data was as shown below. (8 marks)

	Type of patient	
	psychotics	Neurotics
Suicidal feelings	8	16
No suicidal feelings	32	24

- b. The table below shows data collected during an investigation into attempted suicides and show suicidal intent and a depression rating score for a sample of 91 cases. Compute Kendal's tau statistic and interpret your results. (7 marks)

	Depression Rating (X)				
	Did not want to lie	10	14	8	2
Suicidal intent (Y)	Unsure	2	4	7	2
	Wanted to lie	5	9	11	17

QUESTION TWO (15 MKS)

- a. State and describe three types of distributions for categorical data hence give their probability mass functions respectively. (6 marks)
- b. The following output was obtained when effect of temperature (temp) on the probability of thermal distress (td) was modeled. Study the table hence answer the questions that follow;

```

Call:
glm(formula = td ~ temp, family = binomial(link = logit), data = mydata)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0810 -0.7682 -0.3726  0.4246  2.2272
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 15.6211    7.6339   2.046  0.0407 *
temp        -0.2402    0.1116  -2.151  0.0315 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
    Null deviance: 28.267  on 22  degrees of freedom
Residual deviance: 20.007  on 21  degrees of freedom
AIC: 24.007
Number of Fisher Scoring iterations: 5

```

- i. Write down the fitted model (1 marks)
- ii. Estimate the probability of thermal distress at 31°F. (4 marks)
- iii. At what temperature does the estimated probability equal 0.5? (3 marks)
- iv. Interpret the value of AIC in the output. (1 marks)

QUESTION THREE (15 MKS)

- a. Study the following 2x2x2 contingency table and answer the following questions;

Clinic	Treatment	Response	
		success	Failure
1	A	18	12
	B	12	8
2	A	2	8
	B	8	32

- i. Determine the conditional odds ratio. Interpret your results (6 marks)
 - ii. Determine the marginal odds ratio. Interpret your results (1 marks)
- b. State and describe 4 scales of measurement of data (8 marks)