



MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY

P.O. Box 972-60200 – Meru-Kenya.
Tel: +254(0) 799 529 958, +254(0) 799 529 959, +254 (0)712 524 293
Website: www.must.ac.ke Email: info@mucst.ac.ke

UNIVERSITY EXAMINATIONS 2021/2022

FOURTH YEAR, SECOND SEMESTER SPECIAL SUPPLEMENTARY EXAMINATION
FOR DEGREE OF BACHELOR OF SCIENCE IN STATISTICS

SMS 3458: CATEGORICAL DATA ANALYSIS

DATE: JANUARY 2023

TIME: 2 HOURS

INSTRUCTIONS: Answer Question ONE and any other TWO questions.

QUESTION ONE (30 MARKS)

- Define categorical data. (1 Mark)
- Distinguish between ordinal and nominal variables. (2 Marks)
- A clinical trial aspirin use to prevent heart attack was done, approximately equal numbers of subjects were assigned to placebo and aspirin use groups.

	Heart Attack			
Aspirin use		Yes	No	Total
	Placebo	189	10,845	11,034
	Aspirin	104	10,933	11,037
	Total	293	21,778	22,071

Compute

- Risk estimates and interpret (3 Marks)
- Relative risk estimate and interpret (3 Marks)
- Odd ratio estimate and interpret (3 Marks)
- Compute the standard error for the difference in proportions, hence the 95% confidence interval. (3 Marks)

d) Discuss the components of generalized linear model. (6 Marks)

e) The data shown in the following table refers to the effect of passive smoking on lung Cancer. It summarizes results of case-control studies from three Counties among non-smoking women married to smokers.

County	Spouse Smoked	Cases	Controls
Murang'a	No	21	82
	Yes	73	188
Nyeri	No	5	16
	Yes	19	38
Kirinyaga	No	71	249
	Yes	137	363

a) Describe the association in the partial tables. (6 Marks)

b) Describe the association in the marginal table. Is there evidence of Simpson's Paradox? (3 Marks)

QUESTION TWO (20 MARKS)

Data below record exposure to video display terminals and spontaneous abortion (SAB), depending on the month of gestation.

Month of Gestation	Unexposed SAB/Pregnancies	Exposed SAB/Pregnancies
1	10/512	1/366
2	38/502	30/365
3	15/462	12/335
4	7/449	5/323
5	2/442	4/318
6	4/440	1/314
7	2/436	1/313

i. Perform a stratified analysis of this association considering strata defined by month of gestation. (10 Marks)

ii. Compute the partial and marginal associations in the strata obtained above. (10 Marks)

QUESTION THREE (20 MARKS)

- a) From the data given in question two, perform Cochran Mantel Haenszel to test for association between Exposure to video display terminals and spontaneous abortions (take $\alpha = 0.05$).

$$H_0: OR_{\text{common}} = 1$$

$$H_1: OR_{\text{common}} \neq 1 \quad (10 \text{ Marks})$$

- b) Test for goodness of fit the normal probability distribution for the following data compromised of $n=14,736$ blood pressure reading given that the mean and variance of the data is $\bar{x} = 80.68$ and $S^2 = 12^2$ respectively.

Class interval	Observed count, O_i
1 < 50	57
2 ≥ 50 to < 60	330
3 ≥ 60 to < 70	2132
4 ≥ 70 to < 80	4584
5 ≥ 80 to > 90	4604
6 ≥ 90 to < 100	2119
7 ≥ 100 to < 110	659
8 ≥ 1100	251

(10 Marks)

QUESTION FOUR (20 MARKS)

Consider the following data

	Disease	Healthy	Totals
Exposed	2	8	10
Not exposed	10	290	300
Totals	12	298	310

Compute

- Risk difference and the 95% C.I for the risk difference. (4 Marks)
- Relative risk and the 95% C.I for the relative risk (4 Marks)
- Odds ratio and the 95% C.k for the odds ratio (4 Marks)

- iv. Test for the association between the exposure and disease status. Take $\alpha = 0.05$

QUESTION FIVE (20 MARKS)

- a) Define the following
- i. Pearson's Chi-square test (2 Marks)
 - ii. Likelihood ratio (2 Marks)
 - iii. Yate's correction (2 Marks)
- b) Highlight assumptions under chi-square test (2 Marks)
- c) Explain how chi-square test is done in SPSS (6 Marks)
- d) What does the following statistics measure.
- i. Phi (2 Marks)
 - ii. Cramer's V (2 Marks)
 - iii. Lambda (2 Marks)