KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS

2020/2021 ACADEMIC YEAR

FOURTH YEAR SECOND SEMESTER

MAIN EXAMINATION

FOR THE DEGREE OF BACHELOR OF SCIENCE

COURSE CODE: STA 452

COURSE TITLE: CATEGORICAL DATA ANALYSIS

DATE: 5/10/2021   **TIME**: 2:00 pm – 4:00 pm

INSTRUCTIONS TO CANDIDATES
Answer Question One and Any other TWO Questions

TIME: 2 Hours

QUESTION 1:

(a) Using suitable examples, illustrate how you would determine the strength of Association between Categorical variables by means of odds and odds ratio

(6 marks)

(b) (i) Define the following

        I)   Concordant pairs   (2 marks)

        II) Discordant pairs   (2 marks)

(ii) Explain how you would use the Concordant and Discordant pairs to measure the

Strength of Association between two Categorical variables   (6 marks)

(c) In a certain General Social Survey, the following data pertaining to job satisfaction and income was recorded

INCOME	JOB SATISFACTION			
(Us $)	Very dissatisfied	Little dissatisfied	Moderately satisfied	Very satisfied
< 6	20	24	80	82
6 - 15	22	38	104	125
15 – 25	13	28	81	113
>25	7	18	54	92

Obtain,

    i)   The total number of concordant pairs, C   (3 marks)

    ii) The total number of Discordant pairs, D   (3 marks)

From the values of C and D, comment on the status of income and job satisfaction

(3 marks)

Calculate the value of gamma, $\hat{\gamma}$ and use it to conclude about the nature of association between the categorical variables

(5 marks)

QUESTION 2:

(a) Determine, on the basis of sample data shown in the following table, whether the true proportion of shoppers favouring detergent A over detergent B is the same in all counties of Western Kenya:

	Number favouring Detergent A	Number favouring Detergent B	Total
Busia	232	168	400
Bungoma	260	240	500
Kakamega	197	203	400

Use the 0.05 level of significance   (8 marks)

(b) The data in the Table below is of records of accidents in 2015 compiled by the ministry of Transport in the Republic of Kenya.

Safety equipment in use	Injury	
	Fatal Fatal	Non
None	1601	162,527
Seat Belt	510	412,368

Work out

(i)the odds favouring non-fatal injury for seat belt and none seat belt users
(4 marks)

(ii) the odds favouring Fatal injury for seat belt and none seat belt users
(4 marks)

(iii) the odds ratio   (2 marks)

Comment on the association between the safety equipment in use and the nature of injury

(2 marks)

QUESTION 3:

In a certain cohort study, it was noted that per year, the proportion who died from lung Cancer was 0.0014 for cigarette smokers and 0.0001 for non-smokers. The proportion who died from coronary heart disease was 0.00669 for smokers and 0.00413 for non-smokers.

(a) Describe the association of smoking with each of lung cancer and heart disease using the relative risk. Interpret   (6 marks)

(b) Describe the associations using the difference of proportions. Interpret. (5 marks)

(c) By taking lung Cancer to be a more severe disease caused by cigarette smoking, use the concept of Concordant and Discordant pairs to determine the nature of the Association between Causes of Death and Status of smoking   (6 marks)

(d) Which response is more strongly related to cigarette smoking in terms of the reduction in the number of deaths that would occur with elimination of cigarettes   (3 marks)

QUESTION 4:

Consider the following data on the job status of students working within a university. The students are categorized as Freshmen or Sophomores and were required to confirm whether they had a job or not.

	Job Status	
	No	Yes
Freshmen	25	12
Sophomores	11	14

a) For Freshmen, obtain the odds favouring not having a job   (3 marks)

b) For Sophomores, obtain the odds favouring not having a job   (3 marks)

c) Calculate the odds ratio, $\hat{\theta}$ (5 marks)

d) Use the fact that $\dfrac{\ln(\hat{\theta}) - 0}{SE\left[\ln(\hat{\theta})\right]}$ has a distribution that is approximately standard normal for large values of n to ascertain whether there is an evidence of a strong association between job status and class category   (9 marks)

QUESTION 5:

Opinion about promotions, too dependent on published work by persons interested in teaching or research was taken and displayed as below

Interest	Promotion dependent on published work		Total
	Agree	Disagree	
Teaching	90	10	100
Research	70	30	100
Total	160	40	200

(a) Test the null hypothesis that promotion dependent on published work is independent of the field of interest at 5% level of significance. (10 marks)

(b) Discuss the strength of association between promotion dependent on published work and ones field of interest at 1% level of significance (10 marks)