



(Knowledge for Development)

KIBABII UNIVERSITY
UNIVERSITY EXAMINATIONS
2017/2018 ACADEMIC YEAR
FIRST YEAR SECOND SEMESTER
SPECIAL/ SUPPLEMENTARY EXAMINATION
FOR THE DEGREE OF MASTER OF SCIENCE IN
STATISTICS

COURSE CODE: STA 851

COURSE TITLE: CATEGORICAL DATA ANALYSIS

DATE: ¹²
~~08~~/10/18

TIME: 8 AM -11 AM

INSTRUCTIONS TO CANDIDATES

Answer Question One and Any Other TWO Questions

TIME: 3 Hours

This Paper Consists of 4 Printed Pages. Please Turn Over.

QUESTION ONE (30 MARKS)

- (a) Explain the following terms us used in categorical data analysis
- (i) Response variable
 - (ii) Nominal variable
 - (iii) Relative risk
 - (iv) Odds ratio
- (b) Explain in details the main probability distribution models in categorical data analysis

The following table was taken from the General Social Survey:

Belief in Afterlife			
Race	Yes	No or Undecided	Total
White	31	12	43
Black	9	4	13

Find the relative risk and the 95% approximate confidence interval for the true value of the relative risk.
($z_{.95} = 1.645$; $z_{.975} = 1.96$)

QUESTION TWO (20 Marks)

In the United States, the estimated annual probability that a woman over the age of 35 dies of lung cancer equals 0.001304 for current smokers and 0.000121 for nonsmokers ²M. Pagano and K. Gauvreau, *Principles of Biostatistics*, Duxbury Press, Pacific Grove, CA. 1993, p. 134..

- a. Find and interpret the difference of proportions and the relative risk. Which measure is more informative for these data? Why?
- b. Find and interpret the odds ratio. Explain why the relative risk and odds ratio take similar values.

QUESTION THREE (20 MKS)

The Table below shows the results of a retrospective study comparing radiation therapy with surgery in treating cancer of the larynx. The response

TABLE 3.13 Data for Problem 3.13

	Cancer Controlled	Cancer Not Controlled
Surgery	21	2
Radiation therapy	15	3

Source: Reprinted with permission from W. M. Mendenhall, R. R. Million, D. E. Sharkey, and N. J. Cassisi, *Internat. J.*

SAS OUTPUT

```

                Fisher's Exact Test
-----
Cell (1,1) Frequency (F)                21
Left-sided Pr <= F                       0.8947
Right-sided Pr >= F                      0.3808
Table Probability (P)                    0.2755
Two-sided Pr<= P                         0.6384
-----

Odds Ratio                               2.1000

Asymptotic Conf Limits:   95% Lower Conf Limit  0.3116
                          95% Upper Conf Limit 14.1523
Exact Conf Limits:       95% Lower Conf Limit  0.2089
                          95% Upper Conf Limit 27.5522
-----
```

indicates whether the cancer was controlled for at least two years following treatment.

- a. Report and interpret the P -value for Fisher's exact test with $\hat{\lambda}_i$. H_a :
1, and $\hat{\lambda}_{ii}$. H_a : 1. Explain how the P -values are calculated.
- b. Interpret the confidence intervals for . Explain the difference between them and how they were calculated.
- c. Find and interpret the one-sided mid- P -value. Give advantages and disadvantages of this type of P -value.

QUESTION FOUR (20 MKS)

- (a) Describe the purpose of the link function of a GLM. What is the identity link? Explain why it is not often used with binomial or Poisson responses.
- (b) Explain how logistic regression model is applied in categorical data analysis

QUESTION FIVE (20 MKS)

Suppose we have the following three-way contingency table:

Victim's Race	Defendant's Race	Death Penalty		Total
		Yes	No	
White	White	2	13	15
	Black	1	5	6
	Total	3	18	21
Black	White	0	5	5
	Black	1	15	16
	Total	1	20	21

- (a). Compute and interpret the sample conditional odds ratios, adding 0.5 to each cell to reduce the impact of the 0 cell count.
- (b). Test the hypothesis that death penalty is independent of defendant's race, controlling for victim's race using the Cochran-Mantel-Haenszel test with $\alpha = 5\%$. ($\chi^2(.95) = 3.8415, \chi^2(.95)$)