



**OFFICE OF THE DEPUTY PRINCIPAL  
ACADEMICS, STUDENT AFFAIRS AND RESEARCH**

---

**UNIVERSITY EXAMINATIONS  
2020 /2021 ACADEMIC YEAR  
SECOND YEAR SECONDSEMESTER REGULAR EXAMINATION**

**FOR THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS WITH  
COMPUTING)**

**COURSE CODE: STA 219  
COURSE TITLE: CATEGORICAL DAT ANALYSIS**

**DATE: 26/7/2021**

**TIME: 1300-1600HRS**

---

**INSTRUCTION TO CANDIDATES**

- **SEE INSIDE**

**THIS PAPER CONSISTS OF 4 PRINTED PAGES**

**PLEASE TURN OVER**

**REGULAR – MAIN EXAM**

**STA 219: CATEGORICAL DATA ANALYSIS**  
**STREAM: ASC**

**DURATION: 3 Hours**

**INSTRUCTION TO CANDIDATES**

Answer **ALL** questions from section A and any **THREE** from section B.

**SECTION A [31 Marks]. Answer ALL questions.**

**QUESTION ONE (15 Marks)**

a) When the 2012 General Social Survey asked subjects whether they would be willing to accept cuts in their standard of living to protect the environment, 544 of 1170 subjects said “yes.”

- (i) Estimate the population proportion who would say “yes.” (1 mark)
- (ii) Conduct a significance test to determine whether a majority or minority of the population would say “yes.” Report and interpret the  $P$ -value. (4 marks)
- (iii) Construct and interpret a 99% confidence interval for the population proportion who would say “yes” (2 marks)

b) A study was conducted to investigate the effectiveness of bicycle safety helmets in preventing head injury. The data consist of a random sample of 793 persons who were involved in bicycle accidents during a one-year period

Wearing Helmet	Head Injury		Total
	Yes	No	
Yes	17	218	235
No	130	428	558
Total	147	646	793

- (i) Compare the proportion of head injuries by helmet use (2 marks)
- (ii) Calculate the relative risk for this data and interpret your results (2 marks)
- (iii) Compute the odds ratio for this data and interpret your results (2 marks)
- (iv) Compute the 95% confidence interval for the odds ratio (2 marks)

**QUESTION TWO (16 Marks)**

25 students were asked whether they were vegetarians. 6 answered “yes”. Let  $\pi$  denote the population proportion who would say “yes.” Consider  $H_0: \pi = 0.50$  and  $H_1: \pi \neq 0.50$

- a) Conduct the Wald, Score (6 marks)
- b) and Likelihood Ratio tests (6 marks)
- c) Compute the 95% confidence intervals for the Wald and Score tests (4 marks)

**SECTION B (39 MARKS)**

**QUESTION THREE (13 Marks)**

An experiment of a multinomial distribution with 4 categories results in the outcome (32, 20, 23, 25). Let  $H_0: p_1 = p_2 = p_3 = p_4 = 0.25$ . Perform both the Wald test and the likelihood ratio test on each of the parameters. (13 marks)

**QUESTION FOUR (13 Marks)**

a) A study was conducted to identify reasons for the exceptionally high rate of lung cancer among male residents of coastal town. The primary risk factor under investigation was employment in shipyards, and data are tabulated separately for three levels of smoking.

Smoking	Shipbuilding	Case	Controls
No	Yes	13	61
	No	67	286
Moderate	Yes	71	51
	No	400	350
Heavy	Yes	15	5
	No	124	87

Use the Mantel-Haenszel Method to compute the odds ratio of cancer for those employed in the shipbuilding industry. (7 marks)

b) The table below shows the results of a survey in which each subject of a sample of 1169 adults was asked to indicate their level of the agreement with the statement "Under eight year old should be taught using mother tongue"

Highest level of Education	Level of agreement				
	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
Primary	70	102	65	94	13
Secondary	38	76	136	101	121
Tertiary	66	98	74	20	95

Calculate the generalized odds ratio and comment on your result. (6 marks)

**QUESTION FIVE (13 Marks)**

A study on educational aspirations of high school students measured aspirations based on family as summarized in the table below:

		Family Income		
		Low	Middle	High
Education Level	Some high school	9	11	9
	High school graduate	44	52	41
	Some college	13	23	12
	College graduate	10	22	27

- Test independence of aspirations and family income using  $X^2$ . (3 marks)
- Test independence of aspirations and family income using  $G^2$ . (3 marks)
- Explain the deficiency of this test for these data. (2 marks)
- Conduct a more powerful test. Interpret results. (5 marks)



**QUESTION SIX (13 Marks)**

- a) Explain in detail the Fisher's exact test (7 marks)
- b) The table below contains results of a study comparing radiation therapy with surgery in treating cancer of the larynx. Use Fisher's exact test to test  $H_0: OR = 1$  against  $H_a: OR > 1$ . Interpret results.

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

(6 marks)

**QUESTION SEVEN (13 Marks)**

- a) Assume that the number of cases of tetanus reported in the United States during a single month in 2005 has a Poisson distribution with parameter  $\mu$ . The number of cases reported in January and February are 1 and 3 respectively. Find likelihood function and write some simple R codes to make the plot for the likelihood function over the space of potential values for  $\mu$ . (3 marks)
- b) Discuss three assumptions of logistic regression (3 marks)
- c) For three-way contingency tables:
- When any pair of variables is conditionally independent, explain why there is homogeneous association. (3 marks)
  - When there is not homogeneous association, explain why no pair of variables can be conditionally independent. (3 marks)

\*\*\*\*\*