

UNIVERSITY EXAMINATIONS

SECOND SEMESTER 2023/2024 ACADEMIC YEAR

**THIRD YEAR EXAMINATION FOR THE DEGREE OF
BACHELOR OF SCIENCE (STATISTICS)**

STAT 324: CATEGORICAL DATA ANALYSIS

STREAM: R

TIME: 2 HRS

DAY: THURSDAY [2.30 – 4.30 P.M]

DATE: 11/04/2024

THIS QUESTION PAPER CONSISTS OF FIVE (5) PAGES

PLEASE DO NOT OPEN UNTIL THE INVIGILATOR SAYS SO

INSTRUCTIONS: Answer question ONE and any other TWO questions

QUESTION ONE (30 MARKS)

- a) Which scale of measurement is most appropriate for the following variables –nominal, or ordinal?
(4 Marks)
- i) Capital cities (Nairobi, Kampala, Kigali).
 - ii) Patient temperature (low, medium, high)
 - iii) Mode of transportation (Bicycle, Bus, Motorbike).
 - iv) Favorite type of music (classical, country, folk, jazz, rock).
- b) A meta-analysis compared aspirin with placebo on incidence of heart attack and of stroke, separately for men and for women. For the Women’s Health Study, heart attacks were reported for 198 of 19,934 taking aspirin and for 150 of 19,942 taking placebo.
- i) Construct the 2×2 table that cross classifies the treatment (placebo ,aspirin) with whether a heart attack was reported (yes, no). (2 Marks)
 - ii) Estimate the Relative risk and its 95% confidence interval. (4 Marks)
 - iii) Estimate the odd ratio and its 95% confidence interval. (6 Marks)
- c) A sample of 200 women suffering from excessive menstrual bleeding have been taking an analgesic designed to diminish the effects. A new analgesic is claimed to provide greater relief. After trying the new analgesic, 40 women out of 100 reported greater relief with the standard analgesic, and 60 out of 100 reported greater relief with the new one. Test at 5% significance level the hypothesis that the probability of greater relief with the standard analgesic is the same as the probability of greater relief with the new analgesic. (6 Marks)
- d) Describe the Binomial sampling schemes, it assumptions and probability mass function.
(4 Marks)



QUESTION TWO (20 MARKS)

- a) Describe the Fisher exact test and its assumptions. **(5 Marks)**
- b) The following shows the results of a retrospective study comparing radiation therapy with surgery in treating cancer of larynx. The response

	Cancer controlled	Cancer not controlled
Surgery	3	7
Ration therapy	13	4

- i) Report and interpret the p value for Fisher exact test with $H_a : \theta < 1$. **(7 Marks)**
- ii) Report and interpret the p value for Fisher exact test with $H_a : q^1 \ 1$ **(8 Marks)**

QUESTION THREE (20 MARKS)

- a) The table below shows the response of respondent asked to rate their agreement to the statement job security is good with respect to the work they were doing.

	Not too happy	Pretty happy	Very happy
Not at all true	15	25	5
Not too true	21	47	21
Somewhat true	64	248	100
Very true	73	474	311

- i) Calculate the number of concordant pairs. **(3 Marks)**
- ii) Calculate the number of discordant pairs. **(3 Marks)**
- iii) Calculate the gamma statistics and interpret. **(3 Marks)**



b) The table below shows the infant malfunction and .alcohol consumption.

Alcohol consumption	Infant malfunction	
	Absent	Present
0	17066	48
<1	14464	38
1-2	788	5
3-5	126	1
≥ 6	37	1

Determine whether there is association between infant malfunction and alcohol consumption using Cochran-Mantel-Haenszel Test at 5% significance level. **(11 Marks)**

QUESTION FOUR (20 MARKS)

a) Consider the table below.

	Not too or Pretty happy	Very happy
Not at all true	40	5
Very true	547	311

Find the Cumulative Odds ratio and its 95% confidence interval. **(5 Marks)**

b) The contingency table gives the Income Level and Participation in School Programs by Parents of Elementary School Students.

Program participation	Income level		
	Low (L)	Middle (M)	High (H)
Never (N)	28	18	16
Occasional (O)	22	65	14
Regular (R)	17	74	3

i) Prepare the joint probability table based on these sample data. **(3 Marks)**

ii) Find the marginal probabilities. **(2 Marks)**



iii) Find $P(M / R)$

$P(L \text{ and } R)$

(1 Mark)

$P(H \text{ or } R)$

(2 Marks)

iv) Test whether the Program participation is related to Income level at 10% significance level.

(7 Marks)

QUESTION FIVE (20 MARKS)

a) The table below classifies $n=800$ boys according to socioeconomic status (S), whether they participated in boy scouts (B), and whether they have been labeled as a juvenile delinquent (D):

Socioeconomic status	Boy scout	Delinquent	
		Yes	No
Low	Yes	11	43
	No	42	169
Medium	Yes	14	104
	No	20	132
High	Yes	8	196
	No	2	59

i) State the response and explanatory variables in the data.

(2 Marks)

ii) Determine whether mutual independence model fit the data at 5% significance level.

(9 Marks)

Test the hypothesis whether Boy Scout and Socioeconomic status are jointly independent of Delinquent at 5% significance level.

(9 Marks)

