LOYOLA COLLEGE (AUTONOMOUS) CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION – **STATISTICS**

SECOND SEMESTER – **APRIL 2025**

PST2MC03 – CATEGORICAL DATA ANALYSIS

Date: 29-04-2025 Dept. No. Max. : 100 Marks

Time: 01:00 PM - 04:00 PM

SECTION A – K1 (CO1)

	Answer ALL the questions	**(5 x 1 = 5)**
1	**Define the following**	
a)	Ordinal data.	
b)	Cohort Study.	
c)	Relative Risk.	
d)	Odds ratio.	
e)	Conditional models for Matched Pairs.	

SECTION A – K2 (CO1)

	Answer ALL the questions	**(5 x 1 = 5)**
2	**Choose the correct answer**	
a)	The Wald test is used to test the significance of: a) The entire model b) Individual regression coefficients c) Residual errors d) Multicollinearity	
b)	The Likelihood Ratio Test compares: a) Two different datasets b) Two nested models c) Two independent models d) Two sets of coefficients	
c)	Stepwise selection methods are used in: a) Model selection b) Data visualization c) Data transformation d) Neural networks	
d)	Which method is best suited for analyzing the strength and direction of association between two ordinal variables? a) Linear regression b) Spearman's rank correlation c) ANOVA d) t-test	
e)	Cohen's Kappa is used to measure agreement between: a) Two raters b) More than two raters c) Two independent groups d) A dependent and independent variable	

SECTION B – K3 (CO2)

	Answer any THREE of the following	**(3 x 10 = 30)**
3	Derive the p.m.f. of Multinomial distribution from Poisson distribution.	
4	Explain the importance and testing procedure of Residuals in contingency tables.	
5	Estimate M.L.E of parameters of multinomial distribution.	
6	Explain the components of Generalized Linear Models	
7	Two judges evaluate 100 research papers as either "Accept" or "Reject." Their ratings are given in the table below:	

	Judge B: Accept	Judge B: Reject	Total
Judge A: Accept	30	10	40
Judge A: Reject	15	45	60
Total	45	55	100

(a) Calculate the **Observed Agreement** .
(b) Calculate the **Expected Agreement** under chance agreement.
(c) Compute Cohen's Kappa (κ) and interpret the result.

SECTION C – K4 (CO3)

Answer any TWO of the following **(2 x 12.5 = 25)**

8 Derive the log likelihood function of Logistic Regression model. Also state any four assumptions of Logistic regression.

9 A company produces light bulbs, and each bulb either works (success) or doesn't work (failure). A random sample of 20 light bulbs is tested. The company claims that 70% of the bulbs work. Test this claim against the alternative hypothesis that the proportion of working bulbs is not 70% using LRT. Use a significance level of alpha = 0.05.

10 A clinical trial records the response (Success = 1, Failure = 0) of 100 patients before and after receiving a new drug. The results are:

Response Before	Response After	Count
Recovery	Recovery	50
Recovery	No recovery	15
No recovery	Recovery	25
No recovery	No recovery	10

Estimate the **marginal proportion** of success before and after treatment.

11 Explain Deviance, AIC, BIC, Pseudo R^2 measure in multivariate analysis.

SECTION D – K5 (CO4)

Answer any ONE of the following **(1 x 15 = 15)**

12 A factory produces widgets, and you want to test if the defect rate is p=0.1 (10%). Suppose you observe 50 widgets and find that 8 are defective. Conduct a score test at a 5% significance level to determine if p=0.1 is plausible.

13 Compute Concordant pairs and Discordant pairs for the following cross-classification table on job satisfaction (Y) by income (X).

Income ($)	Job Satisfaction			
	Extremely Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied
<15000	1	3	10	6
15000 - 25000	2	3	10	7
25000 - 40000	1	6	14	12
>40000	0	1	9	11

SECTION E – K6 (CO5)

Answer any ONE of the following **(1 x 20 = 20)**

14 Derive the test statistics for Likelihood Ratio test for binomial distribution and hence estimate the confidence interval.

15 Estimate the sensitivity and specificity for the following data using KS Statistics:

Predicted Probability	0.4	0.6	0.3	0.9	0.7	0.6	0.2	0.5	0.41	0.5	0.11	0.24	0.67
Y	0	1	0	1	1	1	0	0	0	1	0	0	1

$$$$$$$$$$$$