LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

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

SECONDSEMESTER – APRIL 2018

17/16PST2MC04- CATEGORICAL DATA ANALYSIS

Date: 23-04-2018 Dept. No. Max. : 100 Marks

Time: 09:00-12:00

SECTION – A

Answer ALL the following questions (10 x 2 = 20 marks)

1) Illustrate with an example how a variable in ratio scale can be converted to ordinal scale.
2) Present the general formulation of Wilk's LR test.
3) Explain a Cohort study with an example.
4) State any two properties of odds ratio.
5) Define 'Sensitivity' and 'Specificity' of a diagnostic procedure.
6) Define Goodman-Kruskal Gamma coefficient and the sample version, explaining the notations.
7) Present the Delta Method of constructing Wald CI for a parametric function $g(\theta)$.
8) Under usual notations, state the mean and variance functions for the random component of a GLM under the 'exponential dispersion family' setting.
9) State a situation in which 'extreme value models' are more suited than logit or probit models for a binary response variable.
10) Define 'Cumulative Logits' for ordinal response variables.

SECTION – B

Answer any FIVE questions (5 x 8 = 40 marks)

11) Let Y_i be independent Bernoulli r.v.'s with $P(Y_i = 1 \mid \pi) = \pi$, $i = 1,2,\ldots,n$ where π is a r.v. with p.d.f. g(.) on (0 ,1) with mean ρ and positive variance. Show that $Y = \sum_{i=1}^{n} Y_i$ has over-dispersion relative to B(n, ρ). If π is a uniform r.v., find Var(Y).
12) Explain Poisson and Independent-Multinomial Sampling scheme in the context of construction of Contingency tables and give examples for both.
13) A survey was carried out among employees of a large company to know their opinion on the recreation facilities introduced recently:

		Opinion on Recreation Facilities		
		Unnecessary	Neutral	Welcome
Employee Experience Level	Freshers (< 1 year)	117	292	745
	Medium Experience	624	636	758
	Senior Employees	597	208	204

Taking the scores for Y (opinion) as 0, 1, 2 in the order given and scores for X (Experience) as 0.5, 3, 7.5 in the order given, carry out the 'linear trend analysis' to test the hypothesis of independence against the one-sided alternative that 'Higher experience is associated with negative opinion'. **[Cont'd]**

14) Give the general formulation for 'Proportional Reduction in Variation'. Derive an expression for 'Uncertainty Coefficient' with motivation from 'entropy'.

15) (a) Test whether X and Y are independent from the following contingency table (at 5% significance level) by constructing the Wald Asymptotic confidence interval for the log odds ratio:

X \ Y	Success (1)	Failure (0)
1	480	240
0	120	96

16) The following table classifies a sample of women by their 'current marital status' and by their opinion on the 'source of marital problems':

	Opinion on Source of Marital Problems		
Marital Status	**In-Laws' Attitudes**	**Husband's Attitude**	**External Factors**
Divorced	356	270	215
Living Separately	285	321	218
Living with Husband	265	504	527

Carry out the 'Chi-Square Residual Analysis' using 'Standardized Residuals'. For each group of women, identify which source(s) they point out as the major reason and the least-likely reason for marital problems. [Pearson X^2 Statistic not needed]

17) Derive the mean and variance functions for the random component of a GLM under the 'exponential dispersion family' setting and derive the likelihood equations for estimating the vector of parameters **β** in the systematic component.

18) Explain Adjacent-Category Logits and bring out the relationship with Baseline-Category Logits. Explain the method of estimating the probabilities for the outcome categories from an 'Adjacent-Category Logit Model'.

SECTION – C

Answer any TWO questions (2 x 20 = 40 marks)

19) (a)Bring out the similarity of the conditional likelihood function of Poisson distribution (by conditioning on the sample total) to the likelihood of Multinomial distribution.
(b) Derive the Wald, LR and Score test statistics for the Poisson parameter. If the sample mean of a sample of 100 observations from a Poisson distribution is found to be 1.15, carry out all three tests to test the hypothesis that the population mean equals 1. (5+15)

20) (a) Define Homogeneous association. For a 2x2x2 table, establish the symmetric nature of homogeneous association by showing that equal XY conditional odds ratios is equivalent to equal XZ conditional odds ratios and equal YZ conditional odds ratios.
(b)In studying the dependence / independence between school-board and timely completion of outreach activities, the following data were reported. The college-major group (Arts/ Science/Engineering) is the covariate considered:

[Cont'd]

	State Board		CBSE	
Major Subject	**Completed**	**Not Completed**	**Completed**	**Not Completed**
Arts	105	21	45	9
Science	192	24	96	12
Engineering	40	45	48	54

Compute the marginal and conditional odds ratios to relate school-board to completion of outreach activities. What do these measures indicate? (10+10)

21) (a) Derive the Likelihood Ratio G^2 Statistic for testing independence in a contingency table. Using this technique, test the hypothesis of independence of "women's marital status" and "opinion on marital problems" using the data in Q.No. (16)

(b) Apply the 'Partitioning of Chi-Square' technique to find the combination(s) of opinions associated with combinations of marital-status groups. (8+12)

22) (a) Describe the components of a GLM explaining the notations used. Identify these components for a model with a count variable as response.

(b) A binary logit model was built with 15 records and the response variable values and the scores for the linear predictor (systematic component) are given below:

Y	1	0	0	1	0	0	0	1	0	1	1	1	1	0	1
Score	-0.453	-0.541	1.284	1.685	0.217	-0.713	-0.403	1.144	-0.129	1.435	1.277	1.719	-0.231	-0.474	-0.082

Compute the probability scores for each record and the Kolmogorov-Smirnov Statistic for the model and comment on the model performance. Obtain the optimal cut-point for prediction of 1's and 0's.

(8+12)

* * * * * *