LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION - **STATISTICS**

SECOND SEMESTER – **APRIL 2013**

ST 2817 - CATEGORICAL DATA ANALYSIS

Date : 04/05/2013 Dept. No. Max. : 100 Marks

Time : 9:00 - 12:00

SECTION – A

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

1. Define a 'nominal variable' with an example.
2. Mention any one distribution relevant for categorical data and the situation in which it arises.
3. State the Wald's statistic for testing a specified value of Poisson parameter.
4. Define 'Relative Risk' and explain with an example how it is more insightful than 'Difference of Proportions'.
5. Define 'Sample Odds Ratio' for 2 x 2 contingency tables. Show that it is unaffected even with disproportionately large or small samples for marginal categories of any of the variables.
6. State the asymptotic distribution of the sample logit.
7. State the Pearson and Standardized Residuals for two-way tables.
8. Describe the 'Partitioned Chi-Square' approach for 2 X J contingency tables.
9. Give the intuitive motivation for the 'log' link function in modeling Poisson Count data.
10. Define 'Deviance' for a Generalized Linear Model.

SECTION – B

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

11. Suppose that $P(Y_i = 1) = \pi$, i = 1,2,…,n. Let $Y = \sum_{i=1}^{n} Y_i$

    (a) When $\{Y_i\}$ have pairwise correlation $\rho > 0$, show that $Var(Y) > n\,\pi(1 - \pi)$.

    (b) Suppose that heterogeneity exists: $P(Y_i = 1 \mid \pi) = \pi$ for all i, Y_i's are independent, but π is a r.v. with density function g(.) on [0 , 1] having mean ρ & positive variance. Show that $Var(Y) > n\,\rho\,(1 - \rho)$. (3 + 5)

12. Give the general frameworks of the (a) Likelihood Ratio Test; and (b) the single parameter and multiparameter versions of Score Test. (3 + 5)

13. Discuss with examples: Case-Control, Cohort and Cross-Sectional Studies.

14. The following table is based on medical records of people in a village:

		Stomach Ailments	
Oil Used	Yes	No	
Ordinary Oil		40	4050
Refined Oil		3	1030

Identify the response variable and find difference of proportions, relative risk and oddsratio. Why are the relative risk andodds ratio almost same? Interpret the results. [Cont'd]

15. The following table reports the results of a survey on educational level versus religious beliefs:

Education level	Religious beliefs: Fundamentalist	Moderate	Liberal
Less than high school	59	46	36
High school – Junior college	190	216	147
Entered College level	46	84	86

Compute 'Goodman and Kruskal' Gamma coefficient and interpret the result.

16. Apply the Delta Method of getting the asymptotic standard error of log odds ratio and point out the amendments needed in the case of zero frequency cells.

17. The following are the computed logit scores obtained in building a binary logit model involving 20 data points:

DV	1	1	0	1	0	0	1	1	1	0
Logit score	21.25	28.58	0.59	1.68	–3.73	–5.97	14.74	17.55	3.78	3.95
DV	1	1	0	1	0	0	1	1	1	0
Logit score	18.23	22.55	24.31	4.31	–2.94	5.96	12.15	25.57	26.68	2.86

Construct the 'Gains Table' and compute the KS statistic for the model.

18. Develop the 'Cumulative Logits' for an ordinal response variable. Explain how the probabilities for the different response categories are estimated.

SECTION – C

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

19. (a) Derive the 'Score Test' and the 'Score Confidence Interval' for the Binomial parameter.
(b) In a survey, a question requiring yes / no (support / oppose) response was addressed to a sample of 1800 persons to find out the support for a new proposal. 810 said 'yes'. Let π be the population proportion who would reply 'yes' (support). Find the p-value for testing H_o: $\pi = 0.5$ using the score test and construct a 95% CI for π. Interpret the results.(12+ 8)

20. (a) Present the 'Proportional Reduction in Variation' formulation for measuring association between nominal variables. Develop the 'Uncertainty Coefficient' (U) and 'Concentration Coefficient' (τ).

(b) A study was conducted to see how the award of death penalty in murder cases is associated with the race of accused. Here the response variable was Y = Death penalty verdict, with categories (Yes, No) and explanatory variable was X = Race of accused, with categories (white, black). The other covariate was Z = Race of victim, with categories (white, black). Data on 674 murder cases are summarized in the following 2 x 2 x 2 table:

[Cont'd]

Race of Victim	Race of Accused	Death Penalty Verdict: Yes	No
White	White	53	414
	Black	11	37
Black	White	0	16
	Black	4	139
ALL	White	53	430

	Black	15	176

Using the above data, show that conditional odds ratios are capable of revealing certainrelationships that are hidden by marginal odds ratio. (12+ 8)

21. a) Briefly sketch the Pearson's Chi-square and Likelihood Ratio Chi-square tests for independence in two-way tables.
(b)Using the data in Q. No. (15), test the hypothesis of independence between education level and religious belief, by Pearson's X^2 Statistic and the Likelihood Ratio G^2 Statistic. (6 + 14)

22. (a) Develop the 'Baseline Category Logits' for nominal response variables. Explain how the probabilities for the different response categories are estimated. (b) A study on food preference of invitees for a party was conducted by considering Gender, Nativity (Tamilnadu / Other South Indian States / North India / NorthEast India), Age (Youth / Middle aged / Old) as predictors. The food varieties available were Madras Meals / Andhra Meals / Tandoori. With Madras Meals as the baseline category, a Multinomial logit model was built and the two logit equations' coefficients are summarized below. Gender was found insignificant.

Logit	Intercept	Southern	Northern	North-east	Youth	Middle-aged
$\text{Log}(p_{\text{Andhra}}/\ p_{\text{Madras}})$	–1.55	1.46	–1.66	0.94	1.12	2.79
$\text{Log}\ (p_{\text{Tandoor}}/p_{\text{Madras}})$	–2.09	–0.63	0.70	0.65	3.28	1.52

Find the probabilities for different food preferences for invitees of the following categories:

(i) An old person from a southern state(other than Tamilnadu)

(ii) A youth from Tamilnadu

(iii) A middle-aged person from North-east

Interpret these probabilities. (8 +12)

* * * * * * *