Indian Statistical Institute
Chennai Centre
M-Stat (NB Stream)Semester II : 2016-18
Categorical Data Analysis

End-Semester Examination

April 24, 2017] [9:30 am - 12:30 noon

- Answer **All** the questions.
- The paper is of **65** marks. The Maximum you can score is **60**.

1. (a) The data in the table below are the observations on independent response variables Y_i which follow Poisson(λ_i). The linear predictors of $E(Y_i)$ are $\eta_i = \beta_1 + \beta_2 x_i$, $i = 1, \cdots, 7$ with the link function $g(\mu_i) = \mu_i^{\frac{1}{2}}$.

y_i	120	42	20	2	39	51	100
x_i	-3	-2	-1	0	1	2	3

   i. Express the likelihood function $L(\beta_1, \beta_2; y_1, \cdots, y_7)$ as a function of β_1 and β_2 and derive the likelihood equations for (β_1, β_2).

   ii. Suppose the numerical solution of the likelihood equations gave the maximum likelihood estimates of the parameters $\hat{\beta}_1 = 0.55215$ and $\hat{\beta}_2 = 3.64430$ for which the deviance is equal to 43.94.
   State with justification whether the assumed model gives adequate fit to the data.

   iii. Suppose we wish to the test of hypotheses

$$H_o : \beta_2 = 0 \quad \text{against} \quad H_1 : \beta_2 \neq 0.$$

   Is it possible to test the hypotheses using the given information? Justify. If yes, carry out the formal test of hypotheses using the data and the deviance given in (ii) above.

   (b) For a nominal-scale response variable Y with J categories, specify the Baseline-Category logit model with a single covariate x. Define $\pi_{ij} = P[Y_i = j|x_i]$, $j = 1, \cdots, J$ based on n observations Y_i, $i = 1, \cdots, n$, which is expressible in terms of the model parameters you have specified. Derive the likelihood function in terms of π_{ij} and obtain log-likelihood equations of the model parameters.

[9 + 6 = 15]

2. (a) Suppose X, Y, Z are three categorical variables with I, J, K categories respectively under a log-linear model where Z is jointly independent of X and Y.

   i. Clearly specify the model in terms of the model parameters.

   ii. Assume full multinomial sampling distribution of the random cell frequencies N_{ijk}'s (state it in the context), and derive the maximum likelihood estimators of the model parameters you have specified in (i) above.

(b) The cross-tabulated data on three categorical variables: Gender (male, female), Political party affiliation (Democrat, Republican, Independent) and Opinion about current laws legalizing abortion (oppose, support) is presented in the table below.

		Political party affiliation		
Gender	Opinion	Democrats	Republican	Independent
Male	Oppose	132	1776	127
	Support	42	6	12
Female	Oppose	56	79	40
	Support	172	54	105

The interest is to verify whether given the Political party affiliation, gender and opinion about current laws legalizing abortion are independent of each other.

i. What is the common odds ratio in the context? Compute and interpret it.

ii. Clearly state your hypotheses in terms of the odds ratios. Carry out Cochran-Mantel-Haenszel test with continuity correction to test for the hypotheses and interpret your result (in terms of the p-value).

[9 + 6 = 15]

3. (a) Suppose K players are participating in a tournament which is carried out at n locations. At every location, every player plays a match with every other player (i.e. each pair plays total n matches) and every match results in either a win or a loss. The available cross-tabulated records are n_{ij}, where n_{ij} is the number of times player i won over player j. Assume that the win or loss within a pair is independent of win or loss within any other pair at any location, irrespective of whether there is a common player in the two pairs.

i. Briefly state the Bradley-Terry model, specify the parameters (β's in standard notation) and explain how you will use the model in the above set up.

ii. Derive the log-likelihood function in terms of the parameters (β's) for fitting the Bradley-Terry Model for the above set up. Clearly state the assumptions you have made.

(b) The table below displays the diagnosis of carcinoma of 106 patients based on two different pathological evaluations.

	Evaluation B		
Evaluation A	Malignant	Benign	Other
Malignant	38	21	11
Benign	33	14	12
Other	15	18	21

i. Estimate the kappa measure of agreement and interpret it.

ii. Estimate the weighted kappa measure of agreement, selecting appropriate weights (to be specified). **[10 + 10 = 20]**

4. (a) Draw path diagram for the following system of equations in variables U_1, U_2, U_3 and V_1, V_2 where ω represents error term, either measurement error or residual error, depending on the situation.

$$\begin{aligned} U_1 &= a_1V_2 + b_1U_2 + \omega_1 \\ U_2 &= a_2V_1 + a_3V_2 + \omega_2 \\ U_3 &= a_4V_1 + a_5V_2 + b_2U_1 + b_3U_2 + \omega_3 \end{aligned}$$

(b) Consider the path diagram of a Structural Equation Model(SEM) given below to answer the following.

i. Identify exogeneous and endogeneous variables.

ii. Give path diagram of any one measurement model and any one structural model from the SEM.

iii. Express the SEM in terms of the system of equations with standard notations, to be specified. (The error components are not shown in the diagram but must be appropriately incorporated in your equations.)

iv. Find the correlation between the nodes a) U_3 and U_9, b) U_{11} and U_9 and c) A_4 and A_5.

[4 + 11 = 15]