INDIAN STATISTICAL INSTITUTE

Semestral exam. (Semester I: 2025-2026)

Course Name: M. Stat. 1st year

Subject Name: Categorical Data Analysis

Date: 19.11. 2025: Total Marks: 60: Maximum Marks: 50: Duration: 2 hr 30 min

Notes: You may answer as many questions as you like. Maximum you can score is 50. Calculators are allowed, but statistical tables are not.

1. Suppose there are two urns, labelled as urn A and urn B, each having 4 yellow, 5 blue and 6 orange balls initially. Let Y_1 and Y_2 be two discrete random variables. Draw a ball from the urn A, and Y_1 represents the variable corresponding to the colour of the drawn ball. Consequently, we add 3 additional balls of the same colour of the drawn ball to the urn B. Now, draw one ball from this modified urn B, and Y_2 represents the colour of the drawn ball. Find the marginal distribution of Y_2. Find Goodman and Kruskal's τ for Y_2 on Y_1. [2+6=8]

2. Find the sample values of the measure of association Goodman-Kruskal's τ for the following three tables, where r and s are the last two digits of your class roll number. (Here the rows correspond to the categories of X variable, and the columns correspond to the categories of the Y variable.)

1	3	10	6	7
2	3	10	7	6
1	6	14	$10r+s$	5
0	1	9	11	3

1	6	14	$10r+s$	5
0	1	9	11	3
1	3	10	6	7
2	3	10	7	6

$10r+s$	1	5	6	14
11	0	3	1	9
6	1	7	3	10
7	2	6	3	10

[6]

3. (a) Give geometric interpretation of all possible 2×2 contingency tables in a multinomial set up.

   (b) Discuss the interpretation of all 2×2 tables with *log odds ratio* $= 1$.

   (c) From (b), give the geometric interpretation of all 2×2 tables with *log odds ratio* $= -1$.

   [3+3+2=8]

4. Consider the following logistic regression model with a binary outcome Y, and two binary covariates x and z:

$$logit\ (P(Y = 1|x, z)) = \beta_0 + \beta_x\ x + \beta_z\ z + \beta_{xz}\ x \times z.$$

   Interpret all model parameters in terms of different types of odds ratios. [5]

5. Obtain the conditional maximum likelihood estimator for binary matched pair data for a logit model. [6]

6. Describe the proportional odds model by clearly mentioning the setup. Why the model is named so? Write down the expression of the likelihood. [3+3+2=8]

7. There are 161 patients, and they are diagnosed (disease: present or absent) before and after using the drug. Test marginal homogeneity for the following table using McNemar test at 5% level of significance:

	After: present	After: absent	Total
Before: present	59	6	65
Before: absent	16	80	96
Total	75	86	161

   [4]

8. (a) Draw path diagram for the following causal relationship:

$$\begin{aligned} X_1\ X_2 &\longrightarrow Y_1 \\ X_2\ X_3\ Y_1\ Y_3 &\longrightarrow Y_2 \\ Y_1\ Y_2 &\longrightarrow Y_3 \end{aligned}$$

   (b) Use the iterative proportional fitting procedure to find the expected values for the data given below, assuming that a person's level of schooling is independent of their parents' wealth.

	Univ.	High School	None	Total
Wealthy	20	15	10	45
Middle class	40	25	20	85
Poor	8	11	23	45
Total	68	51	53	175

[1+5=9]

9. Suppose that observations on three possibily correlated random variables (X, Y, Z) are recorded for some individuals where X is continuous, Y is count, and Z is ordinal categorical. How would you write a joint model for (X, Y, Z)?

[6]