[This question paper contains 2 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 7460 **L**

Unique Paper Code : 2373010020

Name of the Paper : Generalized Linear Models

Name of the Course : **B.Sc. (Honours) Statistics, NEP-UGCF**

Semester : VIII

Duration : 3 Hours Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Attempt **FIVE** questions in all.
3. Use of a non-programmable scientific calculator is allowed.
4. Use of Statistical Tables is allowed.

1. (a) Explain the method of estimation of parameters in a nonlinear regression model. What are the main challenges in statistical inference for such models compared to linear regression?
   (b) In logistic regression, derive the likelihood function for a binary response variable. Write the log-likelihood and the score function for one parameter.
   (c) What are graphical and decomposable models in the context of contingency tables? Give one example of a graphical model that is not decomposable. (6x3)

2. (a) Discuss the method of Maximum Likelihood Estimation (MLE) for nonlinear models when errors are normally distributed. How does it differ from Nonlinear Least Squares (NLS)
   (b) A logistic regression model is fitted to predict customer churn (1=churn, 0=stay) using monthly charges (X_1) and contract type ($X_2 = 1$ for month-to-month, 0 for long-term).
   The fitted model is:
   $$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = -3.5 + 0.08X_1 + 1.5X_2$$
   $SE(\hat{\beta}_1) = 0.02$, $SE(\hat{\beta}_2) = 0.30$.
   (i) Test whether monthly charges significantly affect churn.
   (ii) Compute and interpret the odds ratio for contract type.
   (iii) Find the probability of churn for a customer with monthly charges =50 and month-to-month contract. (6,12)

3. (a) Define the binomial GLM. Show how logistic regression is a special case. Explain the concept of logit transformation.
   (b) The population of bacteria (in thousands) is measured over time (in hours):

Time (t)	0	1	2	3	4	5	6
Population (P)	1.0	1.8	3.2	5.7	10.2	18.3	32.9

It follows the exponential growth model $P = P_0 e^{rt} + \varepsilon$.
(i) Transform the model to linear form and estimate P_0 and r.

P.T.O.

(ii) Predict the population at $t = 8$ hours.
(iii) Estimate the doubling time.

(6,12)

4. (a) What is meant by the Odds Ratio? Explain how it relates to the slope parameter in logistic regression.

(b) A case-control study examines the relationship between a genetic marker (G=1 if present, 0 if absent) and disease (D=1 if case, 0 if control). Data are stratified by age group (Young/ Old):

Age Group	G	D = 1	D = 0
Young	1	25	15
Young	0	10	30
Old	1	40	20
Old	0	20	40

Test for homogeneity of odds ratios across age groups (Breslow-Day test concept). Then fit a logistic regression model with main effects and interaction term to check if age modifies the effect of the genetic marker.

(6,12)

5. (a) Explain the concept of confounding and interaction in logistic regression. How can they be detected and tested?

(b) What is overdispersion? Explain its impact on the standard errors of parameter estimates in both logistic and Poisson regression models.

(9, 9)

6. (a) Describe the negative binomial regression model as an alternative to Poisson regression. Write its mean-variance relationship and explain when it is preferred.

(b) The number of accidents at 20 intersections is modeled with traffic volume (V in 1000 cars/day) and road type (R=1 for highway, 0 for local). Fitted Poisson model:

$$\ln(\hat{\mu}) = -1.0 + 0.20 \times V + 0.50 \times R$$

The estimated dispersion parameter $\hat{\phi} = 1.8$ (quasi-Poisson).
(i) Interpret the coefficient of V.
(ii) Adjust the standard errors if raw $SE(\hat{\beta}_V) = 0.05$ and $SE(\hat{\beta}_R) = 0.20$.
(iii) Recompute 95% confidence intervals for β_V and β_R after adjustment.

(9,9)

7. (a) What are log-linear models in the context of contingency tables? Describe how the parameters in such models are interpreted.

(b) A study produced the following a 2x2 table:

	Disease	No Disease
Smokers	25	35
Non-Smokers	15	45

(i) Calculate the Odds Ratio (OR).
(ii) Compute the Relative Risk (RR).
(iii) Comment on the association between Smokers and Disease.

(9,9)

(500)