Semester	**Seven**
Course	**Major**
Paper Code	
Paper Title	**Categorical Data Analysis & Large Sample theory - II**
No. of Credits	**6**
Theory/Composite/ Practical	**Theory**
Minimum No. of preparatory hours per week a student has to devote	**4** **Module 1: 2 periods/week** **Module 2: 2 periods/week**
Module	**2**
Syllabus	**Module 1: Categorical Data Analysis** **Unit 1:** ***Association in 3-way contingency tables:*** Partial and marginal tables. Conditional versus marginal association and independence. Conditional and marginal odds ratio. Homogeneous association. Simpson's paradox. [8L] **Unit 2:** ***Introduction to Generalized linear Model (GLM):*** Components of a GLM. Fitting a GLM using iterative weighted least squares and maximum likelihood method. Deviance and generalized Pearson chi square. Pearson, Anscombe and deviance residuals. [5L] **Unit 3:** ***GLM for binary and Count data:*** Linear Probability models (LPM). Logistic and Probit regression models: model fitting & interpretation, Confusion matrix, ROC & AUC. Fitting of Poisson regression model. [10L] **Unit 4:** ***Log linear model for contingency table :*** Log linear model of independence and saturated model for two-way tables. Model assumptions and parameter interpretation. Log linear and logistic connection. [3L] **Module 2: Large Sample Theory- II** **UNIT 1:** ***Standard Errors of Statistics:*** Delta method***,*** derivation of large sample standard errors of sample moments, standard deviation, coefficient of variation, moment measures of skewness and kurtosis, correlation coefficient, odds ratio and their uses in large sample tests and interval estimation under normality assumption. Asymptotic distribution of sample quantiles. [12L] **UNIT 2**: ***Variance Stabilization:*** Transformation of Statistics. Derivation and uses of $\sin^{-1}$, square root, logarithmic and Fisher's Z transformations. [6L] **U** **3:**

	Pearsonian χ^2: Large Sample distribution of Pearsonian χ^2 statistic and its uses in tests for goodness of fit, independence and homogeneity. Yates' correction in a 2x2 contingency table. [8L]
Learning Outcomes	1. Understand the large sample behavior of various statistics in the context of estimation and hypothesis testing. 2. Apply the delta method to derive large sample standard errors of different statistics. 3. Analyze the asymptotic distribution of sample quantiles and use them in inferential contexts. 4. Implement variance-stabilizing transformations and facilitate inference. 5. Evaluate the large sample distribution of Pearsonian χ^2 statistic and apply it to tests of goodness-of-fit, independence, and homogeneity. 6. Evaluate different measures of association in three-way contingency tables. 7. Understand the concept of generalized linear model. 8. Remember and apply different measures for goodness-of-fit of a model. 9. Apply logit and probit regression models to binary data. 10. Apply Poisson regression to count data.
Reading/ Reference List	1. McCullagh, P and Nelder, J. A. (1995): Generalized Linear Models. Chapman and Hall. 2. Simonoff, J. F. (2010): Analyzing Categorical Data. Springer. 3. Fienberg, S. E. (2007): The Analysis of Cross Classified data. 2nd Edition, Springer. 4. Agresti, A. (2007): An Introduction to Categorical data analysis. Wiley. 5.Goon A.M., Gupta M.K.: Das Gupta. B. (2005), Outline of Statistics, Vol. I & II, World Press, Calcutta. 6.Rohatgi V. K. and Saleh, A.K. Md. E. (2009): An Introduction to Probability and Statistics. 2nd Edn. (Reprint) John Wiley and Sons. 7.Miller, I. and Miller, M. (2002) : John E. Freund's Mathematical Statistics (6th addition, low price edition), Prentice Hall of India. 8.P. Mukhopadhyay (2006): Mathematical Statistics. 3rd Edn, Books and Allied Limited. 9.C.R. Rao (1983): Linear Statistical Inference and its Application. 3rd Edn, Wiley Eastern Limited. 10. R.V. Hogg and A.T. Craig (2002): Introduction to Mathematical statistics. 5th Edn, Pearson Education.
Evaluation	CIA: 30 End-Sem: 70 Total: 100
	Module 1 (35 marks) | **Module 2 (35 marks)**

Paper Structure for Semester Exam	Short questions (5 marks each): 4 out of 6 Long questions (15 marks each): 1 out of 2	Short questions (5 marks each): 4 out of 6 Long questions (15 marks each): 1 out of 2