

Semester	<b>FIVE</b>
Paper Number	<b>HSTDS5012T &amp; HSTDS5012P</b>
Paper Title	<b>Categorical Data Analysis</b>
No. of Credits	<b>6</b>
Theory/Composite	<b>Composite</b>
No. of periods assigned	Th: 4 Pr: 3
Module	single
Course description/objective	<p><i>At the end of the course a student should know</i></p> <ul style="list-style-type: none"> <li>○ The difference between ordinal and nominal scales of measurement.</li> <li>○ Concept of contingency tables.</li> <li>○ Different measures of association for a kx1 contingency table.</li> <li>○ Different measures of association in three way tables.</li> <li>○ How to model cell frequencies in a contingency table.</li> <li>○ Logit and Probit regression models with reference to binary data</li> </ul>
Syllabus	<p><b>UNIT 1:</b>  <i>Association in two way tables:</i> Introduction to Categorical Data, 2x2 contingency table, types of observational studies, notion of independence &amp; association, ideas of complete and absolute association. Yules measures of association and colligation, Cramer's measure of association, Extension to kx1 contingency table: Pearson's chi-square, Kendall's <math>\tau</math> &amp; <math>\tau_b</math>, Goodman-Kruskal's <math>\gamma</math>. Difference of proportions, relative risk, odds ratio, log odds ratio. [20L]</p> <p><b>UNIT 2:</b>  <i>Association in three way tables:</i> Partial tables, marginal tables, conditional associations, conditional versus marginal association, Simpson's paradox. Conditional and marginal odds ratio, Conditional independence versus marginal independence. Homogeneous association. [10L]</p> <p><b>UNIT 3:</b>  <i>Generalized linear Model:</i> Components of a generalized linear model, Random component, systematic component, Link function  [6L]  <i>Generalized linear model for binary data:</i> Logistic and probit regression model, Multiple logistic regression. Model fitting by using score function. [8L]</p> <p><b>UNIT 4:</b>  <i>Model for contingency table:</i> Log linear model of independence for two way table, Interpretation of the parameters in independence model, saturated model for two way table. The log-</p>

	linear and logistic connection. [8L]	
List of Practical	<ol style="list-style-type: none"> <li>1. Measures of association for 2x2 contingency table.</li> <li>2. Relative risk, odds ratio</li> <li>3. Measures of association for kx1 contingency table.</li> <li>4. Fitting a logit model</li> <li>5. Fitting a probit model</li> <li>6. Fitting of multiple logistic regression.</li> </ol>	
Reading/Reference Lists	<ol style="list-style-type: none"> <li>1. Goon A.M., Gupta M.K., Dasgupta, B. (2005), Fundamentals of Statistics, Vol II, World Press, Calcutta.</li> <li>2. McCullagh, P &amp; Nelder, J.A.(1995), Generalized Linear Models. Chapman and Hall.</li> <li>3. Simonoff, J.F.(2010): Analyzing Categorical Data. Springer</li> <li>4. Fienberg, S.E. (2007): The Analysis of Cross Classified data, 2<sup>nd</sup> Edn, Springer</li> <li>5. Michael S. Lewis Beck (1993): Basic Statistics. Sage Publication.</li> <li>6. Agresti, A. (2007): An Introduction to Categorical data analysis. Wiley.</li> </ol>	
Evaluation	<b>Theory</b>	<b>Practical</b>
	CIA: 10	Continuous assessment: 40
	End-Sem: 50	
	Total: 60	
Paper Structure for End Sem Theory	Short questions (5 marks each)	Long questions (15 marks each)
	4 out of 6	2 out of 3