

Semester	SEVEN
Course	Major
Paper Code	C4ST230711T
Paper Title	Categorical Data Analysis & Asymptotic Methods
No. of Credits	6
Theory / Practical /Composite	Theory
Classes per week	4 Module I : 2 periods/week Module II : 2 periods/week
Module	2

Course Outcomes for Module I

<p>1. Remember key concepts related to association and independence in contingency tables, components of generalized linear models, and basic log-linear model structures.</p>
<p>2. Understand conditional versus marginal association, odds ratios, Simpson’s paradox, and the theoretical foundations of GLMs for categorical and count data.</p>
<p>3. Apply appropriate generalized linear models (logistic, probit, Poisson, and log-linear models) to categorical data and compute relevant measures such as odds ratios, deviance, and residuals.</p>
<p>4. Analyze Examine associations in multi-way contingency tables and assess model adequacy using residuals, deviance statistics, confusion matrices, ROC curves, and AUC.</p>
<p>5. Evaluate alternative models for categorical data, assess goodness of fit, and judge the suitability of different modeling approaches for binary and count responses.</p>
<p>6. Create suitable GLM or log-linear models for real-life categorical data problems and interpret the results in a meaningful statistical context.</p>

Course Outcomes for Module II

<p>1. Remember concepts of large sample distributions, standard errors of statistics, variance-stabilizing transformations, and Pearsonian χ^2 statistics.</p>
<p>2. Understand the principles of the delta method, asymptotic distributions of estimators and sample quantiles, and the rationale behind variance-stabilizing transformations.</p>
<p>3. Apply large sample standard errors, apply the delta method, and use appropriate transformations for interval estimation and hypothesis testing.</p>
<p>4. Analyze the asymptotic behavior of common statistics and analyze the impact of transformations on variance stabilization and normality assumptions.</p>
<p>5. Evaluate the suitability of asymptotic approximations, transformations, and Pearsonian χ^2 tests for goodness of fit, independence, and homogeneity.</p>
<p>6. Create large sample inference procedures by selecting suitable transformations and test statistics for practical statistical problems involving categorical and continuous data.</p>

Syllabus

Unit/Module	Content	No. of lectures	CO mapping	Cognitive levels
Module I				
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.	8	CO1 CO2	K1 K2
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.	5	CO2 CO3	K1 K2
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.	10	CO3 CO4	K3 K4
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.	3	CO6	K6
Module 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.	12	CO1 CO2 CO3	K1 K2 K3
Unit 2	Variance Stabilization: Transformation of Statistics. Derivation and uses of sin-1, square root, logarithmic and Fisher's Z transformations.	6	CO1 CO2 CO4	K1 K2 K4

Unit 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.	8	CO1 CO5 CO6	K1 K5 K6
---------------	--	---	-------------------	----------------

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	
Paper Structure for Semester Exam	Module 1 (35 marks)	Module 2 (35 marks)
	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

CO	CO Description for Module I	Cognitive levels
CO1	Remember key concepts related to association and independence in contingency tables, components of generalized linear models, and basic log-linear model structures.	K1
CO2	Understand conditional versus marginal association, odds ratios, Simpson's paradox, and the theoretical foundations of GLMs for categorical and count data.	K2
CO3	Apply appropriate generalized linear models (logistic, probit, Poisson, and log-linear models) to categorical data and compute relevant measures such as odds ratios, deviance, and residuals.	K3
CO4	Analyze Examine associations in multi-way contingency tables and assess model adequacy using residuals, deviance statistics, confusion matrices, ROC curves, and AUC.	K4
CO5	Evaluate alternative models for categorical data, assess goodness of fit, and judge the suitability of different modeling approaches for binary and count responses.	K5
CO6	Create suitable GLM or log-linear models for real-life categorical data problems and interpret the results in a meaningful statistical context.	K6

CO	CO Description for Module II	Cognitive levels
CO1	Remember concepts of large sample distributions, standard errors of statistics, variance-stabilizing transformations, and Pearsonian χ^2 statistics.	K1
CO2	Understand the principles of the delta method, asymptotic distributions of estimators and sample quantiles, and the rationale behind variance-stabilizing transformations.	K2
CO3	Apply large sample standard errors, apply the delta method, and use appropriate transformations for interval estimation and hypothesis testing.	K3
CO4	Analyze the asymptotic behavior of common statistics and analyze the impact of transformations on variance stabilization and normality assumptions.	K4
CO5	Evaluate the suitability of asymptotic approximations, transformations, and Pearsonian χ^2 tests for goodness of fit,	K5

	independence, and homogeneity.	
CO6	Create large sample inference procedures by selecting suitable transformations and test statistics for practical statistical problems involving categorical and continuous data.	K6