23BS22T3 - STATISTICAL METHODS FOR DATA SCIENCE

(Common to CSE-DS and AI&DS)

Course Category:	Basic Sciences	Credits:	3
Course Type:	Theory	Lecture-Tutorial-Practical:	3-0-0
Prerequisite:	Intermediate Mathematics	Sessional Evaluation: Univ. Exam Evaluation: Total Marks:	30 70 100
Objectives:	• Statistical methods used in data science help to explore data and find patterns and trends, descriptive statistics summarize data using indexes like the mean and median. • It helps to understand the data by providing a clear overview and checking its quality		

	Upon successful completion of the course, the students will be able to:	
Course Outcomes	CO1	Understand the basic concepts of Statistics.
	CO2	Analyze the data and draw conclusion about collection of data under study using Point estimation
	CO3	Analyze data and draw conclusion about collection of data under study using Interval estimation.
	CO4	Analyze to test various hypotheses included in theory and types of errors for large samples.
	CO5	Apply the different testing tools like t-test, F-test, chi-square test to analyze the relevant real life problems.
Course Content	**UNIT-I** **Basic Concepts:** Random variables (discrete and continuous), probability density functions, properties, mathematical expectation. Probability distributions: Binomial, Poisson and Normal-their properties. Population, sample, parameter and statistic, characteristics of a good estimator, Consistency – Invariance property of Consistent estimator, sufficient condition for consistency; Unbiasedness Sufficiency. **UNIT-II** **Point Estimation:** Point Estimation- Estimator, Estimate, Methods of point estimation – Maximum likelihood method (the asymptotic properties of ML estimators are not included), Large sample properties of ML estimator (without proof)- applications, Method of moments, method of least squares, method of minimum chi-square and modified minimum chi-square-Asymptotic Maximum Likelihood Estimation and applications.. **UNIT-III** **Interval Estimation:** Confidence limits and confidence coefficient; Duality between	

acceptance region of a test and a confidence interval; Construction of confidence intervals for population proportion (small and large samples) and between two population proportions(large samples) Confidence intervals for mean and variance of a normal population, Difference between the mean and ratio of two normal populations

UNIT-IV

Testing of hypotheses: Types of errors, power of a test, most powerful tests; Neyman-Pearson Fundamental Lemma and its applications; Notion of Uniformly most powerful tests; Likelihood Ratio tests: Description and property of LR tests - Application to standard distributions.

UNIT-V

Small sample tests: Student's t-test, test for a population mean, equality of two population means, paired t-test, F- test for equality of two population variances, Chi-square test for goodness of fit and test for independence of attributes, $\chi 2$ test for testing variance of a normal distribution.

Text Books & References Books	**TEXT BOOKS:** 1. Miller and Freunds, Probability and Statistics for Engineers,7/e, Pearson, 2008. 2. Manoj Kumar Srivastava and Namita Srivastava, Statistical Inference – Testing of Hypotheses, Prentice Hall of India, 2014 **REFERENCE BOOKS:** 1. S.C. Gupta and V.K. Kapoor, Fundamentals of Mathematical Statistics, 11/e, Sultan Chand & Sons Publications, 2012. 2. S. Ross, a First Course in Probability, Pearson Education India, 2002. 3. W. Feller, an Introduction to Probability Theory and its Applications, 1/e, Wiley, 1968. 4. Robert V Hogg, Elliot A Tannis and Dale L.Zimmerman, Probability and Statistical Inference, 9th edition, Pearson publishers,2013.
E-Resources	1. https://onlinecourses.nptel.ac.in/noc21_ma74/preview 2. https://onlinecourses.nptel.ac.in/noc22_mg31/preview