

Branch Name:	Artificial Intelligence & Data Science
Program Code:	108
Course Name:	Probability and Statistical data analysis
Course Code:	1ET1080501
Pre-requisite Course:	Basic knowledge of mathematics, including calculus and algebra. Familiarity with programming concepts and a programming language (preferably Python).

Course Objectives:

1. This course provides a comprehensive introduction to probability theory and statistical data analysis.
2. To develop students' understanding of fundamental probability concepts and equip them with essential statistical techniques to analyze and interpret data.
3. Emphasis will be placed on the application of probability and statistics in AI and data science domains.
4. To impart basic managerial skills for collection & analysis of statistical data & taking prompt decision.

Teaching and Examination Scheme:

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)				
Lecture (L)	Tutorial (T)	Practical (P)	Credit	Theory (Marks)		Practical (Marks)		Total (Marks)
				University Assessment	Continuous Assessment	University Assessment	Continuous Assessment	
03	-	02	04	70	30	-	50	150

Subject Contents:

Sr. No	Topic	Total Hours	Weightage (%)
1	Introduction to Probability Theory - Basic concepts of probability: sample space, events, and probability axioms - Combinatorics: permutations and combinations - Conditional probability and independence - Random variables and probability distributions	05	10
2	Probability Distributions - Discrete probability distributions: Bernoulli, binomial, geometric, and Poisson distributions - Continuous probability distributions: uniform, normal (Gaussian), exponential, and gamma distributions - Multivariate probability distributions	07	20

3	Statistical Inference - Estimation: point estimation and interval estimation - Hypothesis testing: null and alternative hypotheses, p-values, type I and type II errors - Sampling distributions and the Central Limit Theorem	05	10
4	Parametric and Non-Parametric Tests - Parametric tests: t-tests, analysis of variance (ANOVA) - Non-parametric tests: Mann-Whitney U test, Kruskal-Wallis test, chi-square test - Goodness-of-fit tests	07	15
5	Regression Analysis - Simple linear regression: estimation, interpretation, and inference - Multiple linear regression: model building, variable selection, and diagnostics - Logistic regression for classification problems	07	15
6	Exploratory Data Analysis (EDA) - Data visualization techniques: histograms, scatter plots, box plots, etc. - Summary statistics: measures of central tendency, variability, and correlation - Outlier detection and handling missing data	07	15
7	Probability and Statistics in AI and Data Science - Bayesian statistics and probabilistic modeling - Decision theory and Bayesian inference - Statistical concepts in machine learning: cross-validation, regularization, bias-variance tradeoff	07	15

Text Books:

- "Probability and Statistics for Engineering and the Sciences" by Jay L. Devore
- "Introduction to Probability" by Dimitri P. Bertsekas and John N. Tsitsiklis
- "Statistical Inference" by George Casella and Roger L. Berger
- "Introduction to Linear Regression Analysis" by Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining

Reference Books:

- Quantitative Techniques for Managerial Decision by Srivastava, New Age
- Probability & Statistics for Engineers by Rao, SCITECH
- Statistics for Management by Lewis, Pearson
- Quantitative Techniques in Management by Vohra, TMH
- Applied Statistics & Probability for Engineers by Sharma, Willey
- Introduction to Probability & Statistical Application by P.A. Meyer

List of Practical:

- Problems related to various types of Data Collection and Analysis.
- Problems related to Probability Distribution.
- Problems related to Correlation and Regression Analysis.
- Problems related to Statistical Tests and Testing of Hypothesis
- Problems related to Testing of Hypothesis.
- Case study.

List of Open Source Software/learning website:

1. Khan Academy: The Khan Academy website provides free online courses and tutorials on various topics, including probability, statistics, and data analysis. You can visit their website at <https://www.khanacademy.org/>.
2. Coursera: Coursera offers online courses from top universities and institutions worldwide. They have a variety of courses related to probability, statistics, and data analysis. You can explore their offerings at <https://www.coursera.org/>.
3. edX: edX is another platform that offers online courses from leading universities and organizations. They provide courses on probability, statistics, and data analysis. Visit their website at <https://www.edx.org/> for more information.
4. OpenIntro: OpenIntro is a project that offers free textbooks and resources on statistics and data analysis. Their materials are widely used in educational institutions. You can access their resources at <https://www.openintro.org/>.

Course Learning Outcomes (CLO): On completion of this course, the students will be able to:

CLO	Description	Bloom's Taxonomy Level
CLO1	Understand the fundamental concepts of probability theory and its application in AI and data science.	2 Understanding
CLO2	Apply statistical techniques to analyze and interpret data in the context of AI and data science problems.	3 Applying
CLO3	Utilize statistical software and programming tools for data analysis.	2 Understanding 3 Applying
CLO4	Evaluate and interpret statistical results to draw meaningful conclusions.	3 Applying
CLO5	Apply statistical reasoning and decision-making in AI and data science applications.	3 Applying

Mapping of CLOs with POs & PSOs

Course Learning Outcomes	Program Outcomes (POs)												Program Specific Outcomes(PSOs)			
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3	PSO4
CLO1	L	L			L							L				
CLO2	M	M	L		M				M			M				
CLO3	H	M			M				M			L				
CLO4	L	M	L													
CLO5	M	M	L		M				M			L				

H: High, M: Medium, L: Low