20MA504T					Probability & Statistics for Data Science					
Teaching Scheme					Examination Scheme					
L	T	P	C	Hrs/Week	Theory			Practical		Total Marks
					MS	ES	IA	LW	LE/Viva	
3	0	0	3	3	25	50	25	---	---	100

COURSE OBJECTIVES

- Identify the independent and dependent variables in a research problem.
- To equip students with consequently requisite quantitative skills that they can employ and build on in flexible ways
- To identify the concept of estimation of parameter.
- To apply the hypothesis test and t, F and Chi square test to real world problems

Unit 1 PROBABILITY THEORY & RANDOM VARIABLES **10 Hrs.**

Basics: sample space, outcomes, probability, Events: mutually exclusive, independent, Calculating probability: sets, counting, tree diagram, Conditional probability, Law of total probability, Bayes' theorem

Random variables: Overview and Discrete RVs, Discrete and Continuous RVs, Mean, Moments, Variance pmf, pdf, cdf, Discrete RVs: Bernoulli, Binomial, Geometric, Indicator, Uniform (a, b), Exponential(λ), Normal (μ, σ2), and its several properties

Unit 2 PROBABILITY DISTRIBUTIONS & MARKOV CHAINS **10 Hrs.**

Joint distributions & conditioning, Joint probability distribution, Linearity and product of expectation, Conditional expectation, Probability Inequalities: Weak Law of Large Numbers, Central Limit Theorem, Markov chains, Stochastic processes, Setting up Markov chains, Balance equations

UNIT 3 PARAMETRIC & NON-PARAMETRIC INFERENCES **10 Hrs.**

Basics of inference, Empirical PMF, Sample mean, bias, se, MSE, Empirical Distribution Function (or eCDF)

Kernel Density Estimation (KDE), Statistical Functionals, Plug-in estimator, Confidence intervals-Percentiles, quantiles, Normal-based confidence intervals, DKW inequality, Parametric inference: Consistency, Asymptotic Normality, Basics of parametric inference, Method of Moments Estimator (MME), Properties of MME, Basics of MLE, Maximum Likelihood Estimator (MLE), Properties of MLE

UNIT 4 HYPOTHESES TESTING & REGRESSION **10 Hrs.**

Basics of hypothesis testing , Wald Test, Type I and Type II errors , Z-test, t-test, ANOVA, Kolmogorov-Smirnov test (KS test), p-values, permutation test, Pearson correlation coefficient, Chi-square test for independence, Bayesian reasoning & inference, Simple Linear Regression , Multiple Linear Regression

COURSE OUTCOMES

On completion of the course, student will be able to

CO1 - Understand theoretical foundations of probability theory and mathematical statistics

CO2 - Understand the concepts of various parameter estimation methods, like method of moments, maximum likelihood estimation and confidence intervals.

CO3 - Apply the central limit theorem to sampling distribution.

CO4 - Identify the appropriate hypothesis testing procedure based on type of outcome variable and number of samples

CO5 - Analyze hypotheses tests of means, proportions and variances using both one-and two-sample data sets.

CO6 - Implement basic simulation methods using statistical software to investigate sampling distributions.

TEXT/REFERENCE BOOKS

1. Wasserman, Larry, “All of Statistics: A Concise Course in Statistical Inference” Springer, 2004.
2. S.M. Ross, Introduction to Probability Models, Academic Press
3. Miller & Freund’ Probability and statistics for engineers, ninth edition, Richard a. Johnson, Pearson.
4. Devore. J.L., “Probability and Statistics for Engineering and the Sciences”, Cengage Learning, New Delhi, 8th Edition, 2012.
5. S.M. Ross, Stochastic Processes, Wiley