

Course Code	Course Title	L	T	P	C
OLMCA506	Statistics for Data Science	3	1	0	4
Pre-requisite	NIL	Syllabus version			
Course Objectives		1.0			
1. To understand the descriptive measures of statistics.					
2. To understand the Probability Theory and Distributions.					
3. To understand the analysis and applications of various types of correlation and regression.					
4. To understand the method of estimations.					
5. To understand the hypothesis testing, test of significance and ANOVA.					
Course Outcome					
1. Students are able to analyze the data using the techniques of Measures of central tendency, Dispersion, Skewness and Kurtosis.					
2. Students will be able to evaluate and apply the properties to obtain the solutions of various statistical problems by using the knowledge of probability theory and distributions.					
3. Students will be able to evaluate and infer the concept of correlation for relating two or more related variables and regression methods for forecasting the data.					
4. Students will be able to evaluate and apply the properties to obtain the solutions of various statistical problems by using the knowledge of estimation theory.					
5. Students will be able to evaluate and apply the properties to obtain the solutions of various statistical problems by using the knowledge of hypothesis testing, test of significance and ANOVA.					
Module:1	Descriptive Statistics				
Representation of Data - collection, tabulation and diagrammatic representation of different types of univariate and bivariate data- Measures of Central Tendency-Arithmetic Mean, Geometric Mean, Harmonic Mean, Median, Mode - Measures of Dispersion -Range, Variance and Standard Deviation, Mean Deviation, Coefficient of Variation - Skewness and Kurtosis.					
Module:2	Probability Theory				
Random Experiment – Sample Space – Events – Axiomatic Definition of Probability – Addition Theorem – Multiplication Theorem – Baye's Theorem - Applications. Continuous and Discrete Random Variables – Distribution Function: Probability Mass Functions and Probability Density Functions.					
Module:3	Probability Distributions				
Discrete Probability Distributions: Rectangular Distribution, Binomial Distribution and Poisson Distribution - Continuous Probability Distributions: Uniform Distribution, Exponential and Normal Distribution - Central Limit Theorem.					
Module:4	Correlation and Regression Analysis				
Correlation: Scatter plot, Karl Pearson coefficient of correlation, Spearman's rank correlation coefficient, multiple and partial correlations. Regression: Simple and Multiple linear regression – Variable Selection.					
Module:5	Estimation Theory				
Method of moments, Method of maximum likelihood estimation, Interval estimation and its connection with hypothesis testing. Interval estimation for parameters of Normal, Binomial and Poisson distributions					
Module:6	Hypothesis Testing				
Null and alternative hypotheses, Acceptance and critical regions, Type I and Type II errors, Level of significance and Power of a test, Most powerful tests, Tests for mean and variance					

of a normal distribution, and Randomized test.			
Module:7		Test of Significance and ANOVA	
Test of significance for mean(s) and variance(s) based on Normal distribution, t, F and Chi-square distribution. ANOVA – Fixed effect model for one-way classified data, two-way classified data with one observation per cell.			
		Total Lecture Hours	120 hours
Text Book(s)			
1.	Jay L. Devore, Kenneth N. Berk and Matthew A. Carlton. Modern Mathematical Statistics with Applications. 3rd Edition, Springer New York, 2021.		
2.	Walpole R.E, Myers R.H, and Myers S.L, Probability and Statistics for Engineers and Scientists, Pearson, New Delhi, 2017.		
Reference Books			
1.	Gupta S.C and Kapoor V.K, Fundamentals of Mathematical Statistics, 12th edition, Sultan Chand & Sons, New Delhi, 2020		
2.	Gupta S.C and Kapoor V.K, Fundamentals of Applied Statistics, 4th revised edition, Sultan Chand & Sons, New Delhi, 2019.		
3.	Peter Bruce, Andrew Bruce and Peter Gedeck, Practical Statistics for Data Scientists, 2nd Edition, O'Reilly Media, Inc., 2020.		
4.	Rohatgi V.K and Saleh E, An Introduction to Probability and Statistics, 3rd edition, John Wiley & Sons Inc., New Jersey, 2015.		
Mode of Evaluation: Quiz/Assignment/Mid-Term, FAT			
Recommended by Board of Studies		27.07.2022	
Approved by Academic Council		N0. 67	Date 08-08-2022