Semester: III						
STATISTICS FOR DATA SCIENCE **Category: PROFESSIONAL CORE COURSE** **(Theory)**						
Course Code	**:**	**AI235AT**		**CIE**	**:**	**100 Marks**
Credits: L:T:P	**:**	**03:01:00**		**SEE**	**:**	**100 Marks**
Total Hours	**:**	**45L+30T**		**SEE Duration**	**:**	**3Hours**

Unit-I	**08 Hrs**
Descriptive Statistics - Describing data sets - Frequency tables and graphs, relative frequency tables and graphs, Grouped data, histograms, Summarising data sets - Sample Mean, sample median, sample mode, sample variance and sample standard deviation, percentiles and box-plots	
Unit – II	**08 Hrs**
Sampling and Sampling Distributions - Types of sampling, Sample Mean, Sample Variance, Sampling distributions from a normal population, sampling from a finite population, Normal Distribution, approximating binomial, Poisson distributions using normal distribution	
Unit –III	**08 Hrs**
Correlation, Covariance and Independent Random Variables: Joint behavior of random variables, Correlation, Covariance, variance-covariance matrix, Independent random variables, Sums of independent random variables, Law of Large Numbers, Central Limit Theorem	
Unit –IV	**10 Hrs**
Large Sample Estimation – Statistical Inference, Types of Estimators, Point estimation - Point estimation of a population parameter, Interval Estimation - Constructing a confidence interval, Large-Sample Confidence Interval for a Population Mean Interpreting the confidence interval, Large sample confidence interval for a population proportion, Estimating the difference between two population means, Estimating the difference between two binomial distributions, One-sided confidence bounds, Choosing the Sample size.	
Unit –V	**11 Hrs**
Hypothesis Testing - Testing of hypothesis about population parameters, Statistical Test of hypothesis, A large-sample test about the population mean - Essentials of the test, calculating the *p*-value, two types of errors, power of a statistical test, A large-sample test of hypothesis for the difference between two population means - Hypothesis testing and confidence intervals, Hypothesis testing for the binomial, Some comments on testing of hypothesis	

Course Outcomes: After completing the course, the students will be able to:-	
CO1	Apply the knowledge of statistics in providing solutions to some common real-life and business problems.
CO2	Visualize data better, make logical inferences about the data in real-world scientific/business use cases, and present the analysis results.
CO3	Make inferences about a population from samples through various statistical techniques.
CO4	Use statistical tools to illustrate the principles of data distribution, data sampling, and data visualization.
CO5	Appraise the knowledge of statistics in data science to build a successful career as an AI&ML engineer, work in teams, and communicate their ideas effectively.

Reference Books	
1.	Sheldon M. Ross, Introduction to Probability and Statistics for Engineers and Scientists, 5th Edition, Academic Press, 2014
2.	David Freedman, Robert Pisani and Roger Purves, Statistics, 4th Edition, Nortan & Company, 2007
3.	Richard A. Johnson, Miller & Freund's - Probability And Statistics For Engineers, 9th Edition, Pearson, 2018
4.	William Mendenhall, R J Beaver, B M Beaver, Introduction to Probability and Statistics - Cengage Learning, 2019.
5.	The R Book, Michael J. Crawley, Second Edition, John Wiley Publications, 2013

RUBRIC FOR THE CONTINUOUS INTERNAL EVALUATION (THEORY)		
#	**COMPONENTS**	**MARKS**
1.	**QUIZZES:** Quizzes will be conducted in online/offline mode. TWO QUIZZES will be conducted & Each Quiz will be evaluated for 10 Marks. **THE SUM OF TWO QUIZZES WILL BE THE FINAL QUIZ MARKS.**	**20**
2.	**TESTS:** Students will be evaluated in test, descriptive questions with different complexity levels (Revised Bloom's Taxonomy Levels: Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating). Two tests will be conducted. Each test will be evaluated for 50 Marks, adding upto 100 Marks. **FINAL TEST MARKS WILL BE REDUCED TO 40 MARKS.**	**40**
3.	**EXPERIENTIAL LEARNING:** Students will be evaluated for their creativity and practical implementation of the problem. Case study based teaching learning (10), Program specific requirements (10), Video based seminar/presentation/demonstration (10) Designing & Modeling (10) **Phase 2 will be done in the exhibition mode (Demo/Prototype/any outcome). ADDING UPTO 40 MARKS**.	**40**
	MAXIMUM MARKS FOR THE CIE THEORY	**100**

RUBRIC FOR SEMESTER END EXAMINATION (THEORY)		
Q. NO.	**CONTENTS**	**MARKS**
	PART A	
1	Objective type questions covering entire syllabus	20
	PART B (Maximum of TWO Sub-divisions only)	
2	Unit 1 : (Compulsory)	16
3 & 4	Unit 2 : Question 3 or 4	16
5 & 6	Unit 3 : Question 5 or 6	16
7 & 8	Unit 4 : Question 7 or 8	16
9 & 10	Unit 5: Question 9 or 10	16
	TOTAL	**100**