COURSE TITLE	**STATISTICS FOR DATA SCIENCE**
COURSE CODE	**05DS0202**
COURSE CREDITS	**5**

Course Outcomes: After completion of this course, student will be able to:

1. To compute descriptive statistics including diagrammatic representation and interpretation.
2. To calculate probability and classify the probability usage in different area.
3. To test sampling characteristic from a population based on statistical measures & construct a sample for experiment.
4. To verify the decision statement by using parametric methods.
5. To determine simple linear regression analysis and correlation to understand time series analysis and its application to predict value of unknown variable.

Pre-requisite of course: Basic Knowledge of Mathematical Fundamentals

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
4	1	0	50	30	20	0	0

Contents : Unit	Topics	Contact Hours
1	**Descriptive Statistics** Introductions to statistics : Scale of data measurement, Mean, Median, Mode, Percentiles , Quartiles , Five Number Summary , Range, interquartile range, Mean absolute Deviation, Population Variance and Population Standard deviation, Sample Variance and sample Standard deviation, Frequency distribution, Relative Frequency, Percentage Frequency, Cumulative Frequency , Data Graphs: Histogram, ogive, Frequency Polygon , Bar Graphs, Stem and leaf, Class Test 1	12
2	**Probability** Introduction : Definition of Probability , Random Sample, Sample Space , Events, Elementary Events, Independent Events , Complementary Events, Mutually Exclusive Events, Experiment , Formula for finding probability , Union probability, Joint Probability , Examples , Addition Theorem of Probability , Examples, Conditional Probability , Examples , Probability Matrices , Examples , Activity : MCQ Quiz	12

Contents : Unit	Topics	Contact Hours
3	**Sampling & Estimation of single population** Introduction :sampling techniques Random sampling (simple random sampling, stratified sampling, systematic sampling, cluster sampling, Non-Random sampling :convenience sampling, judgment sampling , Quota Sampling, Snowball Sampling , Estimation of population mean using z statistics(sigma known – finite sample size), Examples, Estimation of population mean using z statistics (sigma known – infinite sample size), Examples , Estimation of population mean using t statistics (sigma unknown), Estimation of population proportion, Estimation of population variance, Examples , Activity : Lecture by Students	12
4	**Testing of Hypothesis for single Population** Introduction , Types of Hypothesis , Rejection region, Non rejection region, Type-I and Type-II error, Testing hypothesis about a population mean using z – statistics (sigma known), Examples , Testing hypothesis about a population mean using t – statistics (sigma unknown), Examples , Mix Examples , Class Test 2	12
5	**Correlation & Regression** Karl Pearson Correlation Coefficient , Examples , Introduction to simple regression analysis, Determine equation of regression Line , Prediction of dependent variable by regression line , Residual Analysis, Sum of Square due to error , Sum of square of regression, standard error estimation, Total sum of square, coefficient of determination, Activity : Formula Based Quiz	12
Total Hours		**60**

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Descriptive Statistics** Examples of Mean, Median and Mode, Examples of Five Number Summary , Examples of Variance and Standard Deviation , Examples of Freqquency Distribution	4
2	**Probability** Examples of Probability , Examples of Addition Thorem of Probability , Examples of Conditional Probability , Examples of Probability Matrices	4
3	**Sampling & Estimation of single population** Estimation of population mean using z statistics(sigma known – finite sample size), Estimation of population mean using z statistics (sigma known – infinite sample size), Estimation of population mean using t statistics (sigma unknown), Estimation of population proportion, Estimation of population variance	4
4	**Testing of Hypothesis for single Population** Testing hypothesis about a population mean using z – statistics (sigma known), Testing hypothesis about a population mean using t – statistics (sigma unknown)	4

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
5	**Correlation & Regression** Karl Pearson Correlation Coefficient , Determine equation of regression Line , Residual Analysis, Sum of Square due to error , Total sum of square, coefficient of determination	4
Total Hours		**20**

Textbook :

1 Business Statistics for Contemporary Decision Making , Ken Black, Wiley Publication, 2010

References:

1 Statistics for business and economics, Statistics for business and economics, 1. Anderson, Sweeney, Williams, Thompson Publication, 2000

2 Statistical Methods, Statistical Methods, 2. S P Gupta, S Chand, 2002

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery and evaluation

Remember / Knowledge	Understand	Apply	Analyze	Evaluate	Higher order Thinking
20.00	30.00	25.00	15.00	10.00	0.00

Instructional Method:

1 Board Work

2 PPT

Supplementary Resources:

1 https://www.tutorialspoint.com/statistics/index.html

2 https://www.w3schools.com/statistics/index.php