

**DS502PC: INTRODUCTION TO DATA SCIENCE****B.Tech. III Year I Sem.**

L	T	P	C
3	1	0	4

**Course Objectives:**

- Learn concepts, techniques and tools they need to deal with various facets of data science practice, including data collection and integration
- Understand the basic types of data and basic statistics
- Identify the importance of data reduction and data visualization techniques

**Course Outcomes:**

- Understand basic terms of statistical modeling and data science
- Implementation of R programming concepts
- utilize R elements for data visualization and prediction

**UNIT- I****Introduction**

Definition of Data Science- Big Data and Data Science hype – and getting past the hype - Datafication - Current landscape of perspectives - Statistical Inference - Populations and samples - Statistical modeling, probability distributions, fitting a model – Over fitting.

**Basics of R:** Introduction, R-Environment Setup, Programming with R, Basic Data Types.

**UNIT- II Data Types & Statistical Description**

**Types of Data:** Attributes and Measurement, Attribute, The Type of an Attribute, The Different Types of Attributes, Describing Attributes by the Number of Values, Asymmetric Attributes, Binary Attribute, Nominal Attributes, Ordinal Attributes, Numeric Attributes, Discrete versus Continuous Attributes.

Basic Statistical Descriptions of Data: Measuring the Central Tendency: Mean, Median, and Mode, Measuring the Dispersion of Data: Range, Quartiles, Variance, Standard Deviation, and Interquartile Range, Graphic Displays of Basic Statistical Descriptions of Data.

**UNIT- III**

**Vectors:** Creating and Naming Vectors, Vector Arithmetic, Vector sub setting,

**Matrices:** Creating and Naming Matrices, Matrix Sub setting, Arrays, Class.

**Factors and Data Frames:** Introduction to Factors: Factor Levels, Summarizing a Factor, Ordered Factors, Comparing Ordered Factors, Introduction to Data Frame, subsetting of Data Frames, Extending Data Frames, Sorting Data Frames.

**Lists:** Introduction, creating a List: Creating a Named List, Accessing List Elements, Manipulating List Elements, Merging Lists, Converting Lists to Vectors

**UNIT- IV**

**Conditionals and Control Flow:** Relational Operators, Relational Operators and Vectors, Logical Operators, Logical Operators and Vectors, Conditional Statements.

**Iterative Programming in R:** Introduction, While Loop, For Loop, Looping Over List.

**Functions in R:** Introduction, writing a Function in R, Nested Functions, Function Scoping, Recursion, Loading an R Package, Mathematical Functions in R.

**UNIT- V**

**Charts and Graphs:** Introduction, Pie Chart: Chart Legend, Bar Chart, Box Plot, Histogram, Line Graph: Multiple Lines in Line Graph, Scatter Plot.

**Regression:** Linear Regression Analysis, Multiple Linear regression

**TEXT BOOKS:**

1. Doing Data Science, Straight Talk from The Frontline. Cathy O'Neil and Rachel Schutt, O'Reilly, 2014.
2. K G Srinivas, G M Siddesh, "Statistical programming in R", Oxford Publications.

**REFERENCE BOOKS:**

1. Jiawei Han, Micheline Kamber and Jian Pei. Data Mining: Concepts and Techniques, 3rd ed. The Morgan Kaufmann Series in Data Management Systems.
2. Introduction to Data Mining, Pang-Ning Tan, Vipin Kumar, Michael Steinbach, Pearson Education.
3. Brian S. Everitt, "A Handbook of Statistical Analysis Using R", Second Edition, 4 LLC, 2014.
4. Dalgaard, Peter, "Introductory statistics with R", Springer Science & Business Media, 2008.
5. Paul Teetor, "R Cookbook", O'Reilly, 2011.