



INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

STATISTICAL PROGRAMMING LABORATORY								
III Semester: CSE (Data Science)								
Course Code	Category	Hours / Week			Credits	Maximum Marks		
		L	T	P	C	CIA	SEE	Total
ACDD02	Core	0	0	2	1	40	60	100
Contact Classes: NIL	Tutorial Classes: NIL	Practical Classes:45			Total Classes: 45			
Prerequisite: Basics of Mathematics.								

I. COURSE OVERVIEW:

This course emphasizes the practical applications of probability and statistics, it equips students with essential tools for tackling real-world problems in the field of data science. The course serves as a valuable introduction for data science students, showcasing the relevance and importance of probability theory and mathematical statistics in their analytical toolkit for making informed decisions and deriving meaningful insights from data.

II. COURSES OBJECTIVES:

The students will try to learn:

- I. The different types of statistical models using R in an efficient way.
- II. Statistical programming, computation, graphics, and modeling, writing functions.
- III. Manipulate data within R and to create graphs and charts used analyze data for data driven applications.

III. COURSE OUTCOMES:

At the end of the course students should be able to:

- CO1 Classify data in R and execute basic mathematical operations on datasets.
- CO2 Apply various statistical distributions using R, through hypothesis testing, and graphical representation in data analysis projects.
- CO3 Construct data visualizations using R plots for effective representation and communication of analytical insights.
- CO4 Interpret linear regression models in R for comprehensive data analysis.
- CO5 Apply R knowledge to real-life data analytics, addressing practical challenges and deriving meaningful insights.
- CO6 Analyze datasets, formulate testable hypotheses, and identify suitable statistical tests for robust data interpretation.

IV. COURSE CONTENT:

WEEK 1: INTRODUCTION TO R PROGRAMMING

1. Data structures in R
2. Concept of vectorization
3. Loops in R
4. Functions in R
5. Apply-family functions

WEEK 2: APPLY-TYPE FUNCTIONS

1. Defining user defined classes and operations, Models and methods in R
2. Customizing the user's environment
3. Utilizing Conditional statements, Loops and iterations.

WEEK 3: DATA IMPORT AND MANIPULATION

1. Important libraries: dplyr, purrr, tidyr, ggplot2
2. Data reading, writing, and manipulations.
3. Exploring and cleaning datasets
4. Subsetting and filtering data
5. Data visualizations in R

WEEK 4: EXPLORATORY DATA ANALYSIS (EDA)

1. Descriptive statistics: mean, median, mode
2. Distribution plots and summary statistics
3. Correlation analysis
4. Characteristics of univariate data sets
5. Characteristics of bivariate data sets
6. Descriptive visualizations

WEEK 5: PROBABILITY DISTRIBUTIONS

1. Probability of events
2. Random variables and distribution functions
3. Visualizing data distributions

WEEK 6: BINOMIAL DISTRIBUTION

1. Study of binomial distribution. Plots of density and distribution functions.
2. Normal approximation to the Binomial distribution.

WEEK 7: DATA VISUALIZATION

1. Creating basic plots (scatter plots, histograms, bar charts)
2. Customizing plot aesthetics
3. Multi-plot layouts and visualization best practices

WEEK 8: INFERENCE STATISTICS & TESTING

1. Study of Confidence Intervals for Means of Large and Small Samples.
2. Conducting Large and Small Sample Tests.
3. Performing F- Test

WEEK 9: HYPOTHESIS TESTING

1. One-sample and two-sample t-tests
2. Chi-square tests
3. ANOVA for multiple groups

WEEK 10: INFERENCE STATISTICS

1. Parameter estimation
2. Confidence intervals
3. Bootstrap Resampling for Confidence intervals
4. Analysis of Covariance (ANCOVA)

WEEK 11: REGRESSION

1. Simple linear regression
2. Multiple linear regression
3. Assessing model fit and making predictions

WEEK 12: TIME SERIES ANALYSIS

1. Handling time series data in R
2. Time series decomposition
3. Forecasting using methods like ARIMA or exponential smoothing

WEEK 13: ADVANCED DATA VISUALIZATION

1. Creating interactive visualizations using tools like Shiny
2. Dashboards for data presentation
3. Customizing visualizations for storytelling

WEEK 14: TEXT ANALYTICS WITH R:

1. Introduction to text mining and natural language processing
2. Preprocessing text data
3. Text classification or sentiment analysis

V. TEXT BOOKS:

1. Maria Dolores Ugarte, Ana F. Militino, Alan T. Arnholt “Probability and Statistics with R” 2nd edition on, CRC Press, 2016.
2. P. Dalgaard. Introductory Statistics with R, 2nd edition. (Springer 2008).

VI. REFERENCE BOOKS:

1. Michael Akritas, “Probability & Statistics with R for Engineers and Scientists”, 2nd edition on, CRC Press, 2016.

VII. ELECTRONICS RESOURCES:

1. <http://nptel.ac.in/courses/106104135/48>
2. <http://nptel.ac.in/courses/110106064>

VIII. MATERIALS ONLINE (Include full stack)

1. <https://akanksha.iare.ac.in/>