**Course Contents**

Session-1

**Basic Functionalities & Data Cleaning**

In this session we will learn installation of R, user interface, basic R programming, data frames and matrices, cleaning and reading data in R from different file types and exploratory analysis. This section will give you a very introduction to the world of R. You will get the exposure how to use these open source software to manipulate and handle large data sets.

Session-2 & 3

**Data Summarization & Visualization**

Data visualization is one of the core skills in data science. In order to start building useful models, we need to understand the underlying dataset. Always we need to explore the variables in depth before moving on to building a model or doing something else with the data. Effective data visualization is the most important tool of getting started doing something using the dataset. Plots: Bar (Different kinds), scatter, histogram, boxplot, violin plot, hexplot, line plot, area plot, heatmap, surface plot, Choropleth, mapping etc.

Session-4

**Hypothesis Testing**

This session will introduce you to one of the most important topics in statistics: Hypothesis testing. Topics on hypothesis testing in both R ranges from one & two population mean test, non-parametric test, chi-square test of association, ANOVA to MANOVA. The knowledge of hypothesis testing will help you how to formulate hypotheses and test it in using real world data.

Session-5

**Linear & Logistic Regression**

Regression is one the most important tools in statistics that is used widely in real world decision making. Building linear regression model, checking the assumptions of linear regression model, model validation, variable selection, logistic regression, and variable selection for logistic regression are the main topics to be discussed in this session. You will learn how to apply the knowledge of regression modeling in real data and build a robust regression model using data.

Session-6

**Time Series Regression**

Time series analysis deals with series of data points ordered in time. The characteristics of time series include auto-correlation, non-stationarity and seasonal or cyclic variability – which are required to address while modeling this type of data. The main objective of time series analysis is forecast future values based on past values with certain confidence limit. In this session we will learn how to prepare data for time series analysis including daily, weekly, monthly and yearly data, decompose a series into different components, meet assumptions of fitting time series model, find best time series model on different scenario and details about fitting auto-regressive integrated moving average (ARIMA) model.

Session-7

**Panel Regression**

Panel (data) analysis is a statistical method, widely used in social science, epidemiology, and econometrics to analyze two-dimensional (typically cross sectional and longitudinal) panel data. The data are usually collected over time and over the same individuals and then a regression is run over these two dimensions. Multidimensional analysis is an econometric method in which data are collected over more than two dimensions (typically, time, individuals, and some third dimension). There are fixed effect and random effect models of panel data.

Session-8

**Multivariate Analysis**

The main objective of multivariate analysis is to study complex set of data in which multiple responses were collected from each of the individuals. Multivariate techniques are usually useful to reduce dimensionality of the data, finding patterns and grouping of the data. In this session we will learn about principal component analysis (PCA) – a popular data reduction method. We will also learn factor analysis which describes variability among observed variables in terms of a potentially lower number of unobserved variables called factors. Later we will learn how to group a set of observation through clustering.

Session-9

**Structural Equation Modelling (SEM)**

Structural equation modeling (SEM) includes a diverse set of mathematical models, computer algorithms, and statistical methods that fit networks of constructs to data. SEM includes confirmatory factor analysis, confirmatory composite analysis, path analysis, partial least squares path modeling, and latent growth modeling. Use of SEM is commonly justified in the social sciences because of its ability to impute relationships between unobserved constructs (latent variables) and observable variables.

Session-10

**Introduction to Machine Learning in R**

In modern era, machine learning poses high demand in terms of computational advancement and technological feasibility. Prediction and classification of object with high accuracy are of interest in many fields of science. This session will discuss different types of classifiers including binary classifier, random forest classifier and boosting techniques. Neural network framework, which was developed from idea of human learning mechanism using neurons, will also discuss and practice here.

**Course Delivery Mode:** Online (Zoom)