

DATA SCIENCE WITH R PROGRAMMING

COURSE CONTENTS

R Programming

- ❖ Course Introduction
- ❖ How to get help in the course
- ❖ How to install the Software
- ❖ Development Environment Overview
- ❖ Introduction to R Basics
- ❖ Arithmetic in R
- ❖ Variables
- ❖ R Basic Data Types
- ❖ Vectors
 - Vector Basics
 - Vector Operations
 - Vector Indexing and Slicing
 - Vector Exercise
- ❖ Comparison Operators
- ❖ Matrices
 - Introduction to R matrices
 - Creating a Matrix
 - Matrix Arithmetic
 - Matrix Operations
 - Matrix Selection and Indexing
 - Factor and Categorical Matrices
 - Matrix Exercise
- ❖ Data Frames
 - Introduction to Data Frames
 - Data Frame Basics
 - Data Frame Indexing and Selection
 - Data Frame Operations
 - Data Frame Training Exercise
- ❖ Lists
- ❖ Data Input and Output with R
 - CSV Files
 - Excel Files
 - SQL with R
 - Web Scraping with R
- ❖ R Programming Basics
 - Logical Operators
 - If, else, and else if Statements
 - While Loops
 - For Loops
 - Functions
- ❖ Advanced R Programming
 - Built-in R Features
 - Apply Functions
 - Math Functions with R
 - Regular Expressions
 - Dates and Timestamps
- ❖ Data Manipulation with R
 - Dplyr Functions
 - Pipe Operator
- ❖ Charting in R
- ❖ Data Visualization With R
 - Overview of ggplot2
 - Histograms
 - Scatterplots
 - Barplots
 - Boxplots
 - 2 Variable Plotting
 - Coordinates and Faceting
 - Themes
- ❖ Interactive Visualizations with Plotly
 - Overview of Plotly and Interactive Visualizations
 - Resources for Plotly and ggplot2
- ❖ Capstone Data Project

Business Statistics

Introduction to Analytics

- ❖ Analytics Industry Overview
- ❖ Application of Analytics & Challenges to Analytics

Data Understanding

- ❖ Data Types
- ❖ Summarizing Techniques
- ❖ Five Number Summary
- ❖ Histograms & Ogives
- ❖ Box Plots
- ❖ Scatter Diagram
- ❖ Frequency Tables and Distribution
- ❖ Cumulative Distributions

Measure of Central Tendency, Dispersion and its importance

- ❖ Understanding Range
- ❖ Inter Quartile Range

IT TRAINING & SERVICES

Regd. Office # 207, II floor, HUDA Maithrivanam, Ameerpet, Hyderabad 500 038.

Ph # +91 40 40310000, 23743392.

E-mail: enq@peerstech.com URL: www.peerstech.com

- ❖ Variance
- ❖ Standard Deviation

Probability and Probability Distribution

- ❖ Introduction to Probability
- ❖ Types of Probability
- ❖ Probability Rules
- ❖ Probability Distribution
- ❖ Random Variables
 - Discrete Random Variable
 - Continuous Random Variable
- ❖ Discrete Distributions
 - Binomial Distribution
 - Negative Binomial Distribution
 - Geometrical Distribution
 - Poisson Distribution
- ❖ Continuous Distribution
 - Normal Distribution
 - Standard Normal Distribution
 - Z scores

Sampling and Sampling Distribution

- ❖ Introduction to Sampling
- ❖ Random Sampling & Non Random Sampling
- ❖ Sampling Techniques
 - Stratified Sampling Method
 - Cluster Sampling Method
- ❖ Sampling Distribution
- ❖ Central Limit Theorem
- ❖ Standard Error Concept

Statistical Inference

- ❖ Estimation
 - Introduction
 - Point Estimates and Interval Estimates
 - Calculating Interval Estimates using 'Z' table
 - Introduction to 't' distribution
 - Degrees Of Freedom
 - Calculating Interval Estimates using 't' table
 - Confidence Intervals with t & z distributions
 - Determining Sample Size in Estimation

Hypotheses Testing

- ❖ Introduction

- ❖ Testing Procedure
- ❖ Testing Hypotheses
- ❖ One Sample Test & Two Sample Tests
 - Z test
 - t test
 - One Tail & Two Tail Test
 - Dependent & Independent Samples
- ❖ Concept of p-value
- ❖ ANOVA
 - Introduction
 - F distribution
 - One way ANOVA
 - Two Way ANOVA
- ❖ CHISQ Test
- ❖ Some Non Parametric Tests
- ❖ Man-Whitney U Test
 - Wilcoxon Test
 - Kruskal Wallis Test

Simple Regression & Correlation

- ❖ Introduction
- ❖ Dependent and Independent Variables
- ❖ Correlation Analysis
 - Spearman Correlation
 - Pearson Correlation
- ❖ Estimation in Regression
- ❖ Least Squared Method
- ❖ Standard Error Of Line
- ❖ Finding Regression Equation
- ❖ Hypotheses Testing for estimates
- ❖ Limitations & Errors in Simple Regression Analysis
- ❖ Multiple Regression analysis:
 - Introduction
 - ❖ Multicollinearity
 - ❖ Fitting the model
 - ❖ Regression Assumptions
 - ❖ Residual Analysis for Regression Assumptions
 - ❖ Transformation Of Variables

Logistic Regression

- ❖ Understanding Logistics Regression
- ❖ Difference between linear and logistics regression
- ❖ Odds Ratio
- ❖ Logit Model
- ❖ Building Models
- ❖ ROC concept

IT TRAINING & SERVICES

Regd. Office # 207, II floor, HUDA Maithrivanam, Ameerpet, Hyderabad 500 038.

Ph # +91 40 40310000, 23743392.

E-mail: enq@peerstech.com URL: www.peerstech.com

- ❖ Model Fitting
- ❖ Evaluation of goodness of fit
- ❖ Model Suitability

Cluster Analysis

- ❖ Introduction to Cluster Analysis
- ❖ Clustering Algorithm
 - Hierarchical Clustering Procedure
 - Agglomerative Clustering Technique
 - Non Hierarchical Procedure
 - K-means
- ❖ Evaluation of Clustering Results
- ❖ Application

Factor Analysis

- ❖ Definition and examples
- ❖ Factor Analysis
- ❖ Communality
- ❖ Rotation Of Factors
- ❖ Implementation
- ❖ Evaluation

MACHINE LEARNING USING R

- ❖ Introduction to Machine Learning
- ❖ Linear Regression
- ❖ Project : Linear Regression
- ❖ Logistic Regression
- ❖ Project : Logistic Regression
- ❖ K Nearest Neighbors
- ❖ Project : K Nearest Neighbors
- ❖ Decision Trees and Random Forests
- ❖ Project : Decision Trees and Random Forests
- ❖ Support Vector Machines
- ❖ Project : Support Vector Machines
- ❖ K – Means Clustering
- ❖ Project : K – Means Clustering
- ❖ Natural Language Processing
- ❖ Project : Natural Language Processing
- ❖ Neural Nets
- ❖ Project : Neural Nets

IT TRAINING & SERVICES

Regd. Office # 207, II floor, HUDA Maithrivanam, Ameerpet, Hyderabad 500 038.

Ph # +91 40 40310000, 23743392.

E-mail: enq@peerstech.com URL: www.peerstech.com