**Training on Statistical Data Analysis using R**

**About the Course**

R is an open-source programming language that provides a wide variety of statistical and graphical techniques.  R has “become the de-facto standard for writing statistical software among statisticians. This Training on Statistical Data Analysis using R will give you a solid foundation in creating statistical analysis solutions using the R language, and how to carry out a range of commonly used analytical processes.

**Target Participants**

This Training on Statistical Data Analysis using R is intended for Data Scientists, Data Analysts, Business Intelligence Analysts and any other professional who want to explore the vast range of analytical and graphical capabilities of R.

**Course Duration**

* **Classroom-based:**5 Days

**What you will learn**

**By the end of this training the participants will be able to learn:**

* An introduction to R, basic data types, and R/RStudio installation
* Overview of base R concepts and specific data wrangling packages in R
* Connecting to databases, executing database queries in R
* How to use R for graphical summary
* R programming
* How to carry out a range of analyses using R

**Course Outline**

1. **Introduction to Statistical Analysis**
   1. Explain the basic steps of the research process
   2. Explain differences between populations and samples
   3. Explain differences between experimental and non-experimental research designs
   4. Explain differences between independent and dependent variables
2. **Introduction to R software for statistical computing**
   1. Overview of the R Studio IDE
   2. Installing, loading and updating R packages
   3. Creating objects in R
   4. Data types
   5. Data structures
   6. Sorting vectors and data frames
   7. Directory management commands
   8. Direct data entry in R (for small data sets)
   9. Importing data from other software
   10. Decision structures (if, if-else, if-else if-else)
   11. Repetitive structures (for and while loops)
   12. Other important programming functions (break, next, warn, stop)
3. **Data Wrangling and Cleaning in R**
   1. Working with variables
   2. Transform continuous variables to categorical variables
   3. Add new variables to data frames
   4. Handling missing values
   5. Sub-setting data frames
   6. Appending and merging data frames
   7. Spit data framesStack and unstack data frames
4. **Explanatory Data Analysis (EDA) in R**
   1. Creating tables of frequencies and proportions
   2. Cross tabulations of categorical variables
   3. Descriptive statistics for continuous variables
5. **Data Visualization using R base package**
   1. Introduction to graphs and charts in R
   2. Customizing graph attributes (titles, axes, text, legends)
   3. Graphs for categorical variables
   4. Graphs for continuous variables
   5. Graphs to investigate relationship between variables
6. **Mean Comparison Tests in R**
   1. One Sample T Test
   2. Independent Samples T Test
   3. Paired Samples T Test
   4. One-way analysis of variance (ANOVA)
7. **Tests of Associations in R**
   1. Chi-Square test of independence
   2. Pearson’s Correlation
   3. Spearman’s Rank-Order Correlation
8. **Predictive Regression Models using R**
   1. Linear Regression
   2. Multiple Linear Regression
   3. Binary Logistic Regression
   4. Ordinal Logistic Regression

**Training Approach**

This Training on Statistical Data Analysis using R is delivered by our seasoned trainers who have vast experience as expert professionals using R programming language. The course is taught through a mix of practical activities, theory, group works and case studies.

Training manuals and additional reference materials are provided to the participants.

**Prerequisites**

Basic knowledge of Statistics ideal.

**Certification**

Upon successful completion of this course, participants will be issued with a certificate.

**Tailor-Made Course**

We can also do this as a tailor-made course to meet organization-wide needs. A training needs assessment will be done on the training participants to collect data on the existing skills, knowledge gaps, training expectations, and tailor-made needs.