CETPA INFOTECH PVT. LTD.

CURRICULUM OF R Programming

Module 1: Essential to R programming

1: An Introduction to R

- History of S and R
- Introduction to R
- The R environment
- What is Statistical Programming?
- Why use a command line?
- Your first R session

2: Introduction to the R language

- Starting and quitting R
- Recording your work
- Basic features of R
- Calculating with R
- Named storage
- Functions
- Exact or approximate?
- R is case-sensitive
- Listing the objects in the workspace
- Vectors
- Extracting elements from vectors
- Vector arithmetic
- Simple patterned vectors
- Missing values and other special values
- Character vectors
- Factors
- More on extracting elements from vectors
- Matrices and arrays
- Data frames
- Dates and times
- Built-in functions and online help
- Built-in examples
- Finding help when you don't know the function name
- Built-in graphics functions
- Additional elementary built-in functions
- Logical vectors and relational operators
- Boolean algebra
- Logical operations in R
- Relational operators
- Data input and output
- Changing directories
- dump() and source()
- Redirecting R output
- Saving and retrieving image files
- Data frames and the read.table function

3: Programming statistical graphics

- High-level plots
- Bar charts and dot charts
- Pie charts
- Histograms
- Box plots
- Scatterplots
- QQ plots
- Choosing a high-level graphic
- Low-level graphics functions
- The plotting region and margins
- Adding to plots
- Setting graphical parameters

4: Programming with R

- Flow control
- The for() loop
- The if() statement
- The while() loop
- Newton's method for root finding
- The repeat loop, and the break and next statements
- Managing complexity through functions
- What are functions?
- Scope of variables
- Miscellaneous programming tips
- Using fix()
- Documentation using#

- Some general programming guidelines
- Top-down design
- Debugging and maintenance
- Recognizing that a bug exists
- Make the bug reproducible
- Identify the cause of the bug
- Fixing errors and testing
- Look for similar errors elsewhere
- The browser() and debug()functions
- Efficient programming
- Learn your tools
- Use efficient algorithms
- Measure the time your program takes
- Be willing to use different tools
- Optimize with care

5: Simulation

- Monte Carlo simulation
- Generation of pseudorandom numbers
- Simulation of other random variables
- Bernoulli random variables
- Binomial random variables
- Poisson random variables
- Exponential random numbers
- Normal random variables
- Monte Carlo integration
- Advanced simulation methods
- Rejection sampling
- Importance sampling
-
- 6: Computational linear algebra
-
- Vectors and matrices in R
- Constructing matrix objects
- Accessing matrix elements; row and column names
-
- Matrix properties
- Triangular matrices
- Matrix arithmetic
- Matrix multiplication and inversion
- Matrix inversion
- The LU decomposition
- Matrix inversion in R
- Solving linear systems
- Eigenvalues and eigenvectors
- Advanced topics
- The singular value decomposition of a matrix
- The Choleski decomposition of a positive definite matrix
- The QR decomposition of a matrix
- The condition number of a matrix
- Outer products
- Kronecker products
- apply()

7: Numerical optimization

- The golden section search method
- Newton–Raphson
- The Nelder–Mead simplex method
- Built-in functions
- Linear programming
- Solving linear programming problems in R
- Maximization and other kinds of constraints
- Special situations
- Unrestricted variables
- Integer programming
- Alternatives to lp()

Quadratic programming

Module 2: Data Manipulation Techniques using R programming

1: Data in R

- Modes and Classes
- Data Storage in R
- Testing for Modes and Classes
- Structure of R Objects
- Conversion of Objects
- Missing Values
- Working with Missing Values

2: Reading and Writing Data

- Reading Vectors and Matrices
- Data Frames: read.table
- Comma- and Tab-Delimited Input Files
- Fixed-Width Input Files
- Extracting Data from R Objects
- Connections
- Reading Large Data Files
- Generating Data
- Sequences
- Random Numbers
- Permutations
- Random Permutations
- Enumerating All Permutations
- Working with Sequences
- Spreadsheets
- The RODBC Package on Windows
- The gdata Package (All Platforms)
- Saving and Loading R Data Objects
- Working with Binary Files
- Writing R Objects to Files in ASCII Format
- The write Function
- The write.table function
- Reading Data from Other Programs 3: R and Databases
-
-
- A Brief Guide to SQL
- Navigation Commands
- Basics of SQL
- Aggregation
- Joining Two Databases
- Subqueries Modifying Database Records
- ODBC
- Using the RODBC Package
- The DBI Package
- Accessing a MySQL Database
- Performing Queries
- Normalized Tables
- Getting Data into MySQL
- More Complex Aggregations

4: Dates

- as.Date
- The chron Package
- POSIX Classes
- Working with Dates
- Time Intervals
- Time Sequences

5: Factors

- Using Factors
- Numeric Factors
- Manipulating Factors
- Creating Factors from Continuous Variables
- Factors Based on Dates and Times
- Interactions

- 6: Subscripting

- Basics of Subscripting
- Numeric Subscripts
- Character Subscripts
- Logical Subscripts
- Subscripting Matrices and Arrays
- Specialized Functions for Matrices
- Lists
- Subscripting Data Frames

7: Character Manipulation

- Basics of Character Data
- Displaying and Concatenating Character
- Working with Parts of Character Values
- Regular Expressions in R
- Basics of Regular Expressions
- Breaking Apart Character Values
- Using Regular Expressions in R
- Substitutions and Tagging

8: Data Aggregation

- Table
- Road Map for Aggregation
- Mapping a Function to a Vector or List
- Mapping a function to a matrix or array
- Mapping a Function Based on Groups
- There shape Package
- Loops in R

9: Reshaping Data

- Modifying Data Frame Variables
- Recoding Variables
- The recode Function
- Reshaping Data Frames
- The reshape Package
- Combining Data Frames
- Under the Hood of merge

Module 3: Statistical Applications using R programming

1: Basics

- First steps
- An overgrown calculator
- Assignments
- Vectorized arithmetic
- Procedures
- Graphics
- R language essentials
- Expressions and objects
- Functions and arguments
- Vectors
- Quoting and escape sequences
- Missing values
- Functions that create vectors
- Matrices and arrays
- Factors
- Lists
- Data frames
- Indexing
- Conditional selection
- Indexing of data frames
- Grouped data and data frames
- Implicit loops
- Sorting

2: The R environment

- Session management
- The workspace
- Textual output
- 3 Scripting
- Getting help
- Packages
- Built-in data
- attach and detach
- subset, transform, and within
- The graphics subsystem
- Plot layout
- Building a plot from pieces
- Using par
- Combining plots
- R programming
- Flow control
- Classes and generic functions
- Data entry
- Reading from a text file
- Further details on read.table
- The data editor
- Interfacing to other program

3: Probability and distributions

- Random sampling
- Probability calculations and combinatorics
- Discrete distributions
- Continuous distributions
- The built-in distributions in R
- Densities
- Cumulative distribution functions
- Quantiles
- Random numbers

4: Descriptive statistics and graphics

- Summary statistics for a single group
- Graphical display of distributions
- Histograms
- Empirical cumulative distribution
- Q–Q plots
- Boxplots
- Summary statistics by groups
- Graphics for grouped data
- Histograms
- Parallel boxplots
- Stripcharts
- Tables
- Generating tables
- Marginal tables and relative frequency
- Graphical display of tables
- Barplots
- Dotcharts
- Piecharts

5: One- and two-sample tests

One-sample t test
Wilcoxon signed-rank test
Two-sample t test
Comparison of variances
Two-sample Wilcoxon test
The paired t test
The matched-pairs Wilcoxon test

6: Regression and correlation

- Simple linear regression
- Residuals and fitted values
- Prediction and confidence bands
- Correlation
- Pearson correlation
- Spearman's ρ
- Kendall's τ

7: Analysis of variance and the Kruskal–Wallis test

- One-way analysis of variance
- Pairwise comparisons and multiple testing
- Relaxing the variance assumption
- Graphical presentation
- Bartlett's test
- Kruskal–Wallis test
- Two-way analysis of variance
- Graphics for repeated measurements
- The Friedman test
- The ANOVA table in regression analysis

8: Tabular data

- Single proportions
- Two independent proportions
- k proportions, test for trend
- r × c tables

- 9: Power and the computation of sample size

- The principles of power calculations
- Power of one-sample and paired t tests
-

- Power of two-sample t test
- Approximate methods
- Power of comparisons of proportions
- Two-sample problems
- One-sample problems and paired tests
- Comparison of proportions
- 10: Advanced data handling
-
- Recoding variables
- The cut function
- Manipulating factor levels
- Working with dates
- Recoding multiple variables
- Conditional calculations
- Combining and restructuring data frames
- Appending frames
- Merging data frames
- Reshaping data frames
- Per-group and per-case procedures
- Time splitting
-
- 11: Multiple Regression
-
- Plotting multivariate data
- Model specification and output
- Model search
-
- 12: Linear models
-
- Polynomial regression
- Regression through the origin
- Design matrices and dummy variables
- Linearity over groups
- Interactions
- Two-way ANOVA with replication

Analysis of covariance

- Graphical description Comparison of
-

- regression lines Diagnostics

13: Logistic regression

- Generalized linear models
- Logistic regression on tabular data
- The analysis of deviance table
- Connection to test for trend
- Likelihood profiling
- Presentation as odds-ratio estimates
- Logistic regression using raw data
- Prediction
- Model checking

14: Survival analysis

-
- Essential concepts
- Survival objects
- Kaplan–Meier estimates
- The log-rank test
- The Cox proportional hazards model
-

15: Rates and Poisson regression

-
- Basic ideas
- The Poisson distribution
- Survival analysis with constant hazard
- Fitting Poisson models
- Computing rates
- Models with piecewise constant intensities
- 16: Nonlinear curve fitting
-
- Basic usage
- Finding starting values
- Self-starting models
- Profiling
- Finer control of the fitting algorithm
-