## Course Objectives

The objective of this course is to teach the R programming language to MBA students and marketing professionals at IIFT. The course aims to equip the student with the programming knowledge as well as insight needed to solve real-world data analysis problems that they are likely to encounter in the course of employment. It does this by using several realistic examples of data analysis problems that are solved with R. These examples form the basis of exposition into the techniques of data-analysis with R.  
  
Pedagogy

Programming is very much a ‘learn-by-doing’ discipline.

To help students learn, the entire course will be offered as series of web based editable Jupyter Notebooks which contain all the material taught in each session. The notebooks will contain exercises the students will need to complete as part of their participation in the course.  
Every problem identified in session 1 below is solved in a dedicated Jupyter notebook.

Students will be able to download the notebooks and the changes they made to them (each student will have their own editable copy of the online notebooks)  
  
In addition all materials (including presentations) used in class will be made available as part of the course site and can be downloaded under an Open source license.  
  
Use of Large Language Models for the purposes of learning is STRONGLY ENCOURAGED. This reflects how AI is being used in the corporate sector and in the experience of this instructor really helps with learning when used appropriately.   
  
Group assignments are created with the expectation that students will use an LLM to solve the problems assigned to them.

## Assessment

20% of student grade will come from the group assignment

30% of student grade will come from a written mid-term examination

50% of the student grade will come from a written final examination

### Group Assignment

In class assessment will take place in the form of a group assignment that each student group will need to complete. Students will be grouped into teams of 5-6 students. The group assignment may be complete after the teaching sessions have been completed, but before the end of the semester.

### Written Examinations

Students will be given two written examinations (mid-term and final) which shall test their ability to convert a problem description into an R program.

Answers will be graded for logical correctness. Minor syntax errors, which would be automatically corrected in a real-world environment, and minor errors in function, module package names will not result in loss of marks.

## Module 1/Day 1: Introduction to R (4 hrs)

### Session 1: Why this Course? Why R? (1hrs)

* Course Objectives
* Course Mechanics (Labs, Assignments, Exams, Getting Help, LLM Use Policy, Honor Code)
* What is R? Why Use R?
* Workflow in an R Project
* Overview of Running Examples.

We will use a series of running examples from real-world problems show how R can be used to solve them. This an overview of the examples that we will use throughout the course

Example 1: Microeconomics: Estimating Demand Elasticity

Example 2: Macroeconomics: Modelling GDP growth using ARIMA/S-ARIMA

Example 3: Finance (derivatives): Pricing Options using Black-Scholes

Example 4: Finance (insurance): Ruin-Theory Simulation

Example 5: Digital Marketing: A/B Testing

Example 6: IT: Code Quality Analysis

Example 7: AI/ML: Image Classification using CNNs

### Session 2: Overview of The R Programming Language (3 hrs)

* Getting Setup: Using R in Jupyter Notebooks, Using R in R-Studio, Using R on the command line
* Simple Manipulations (numbers and vectors)
* Objects
* Order and Unordered Factors
* Arrays and Matrices
* Lists and Data Frames
* Reading from Files
* Probability Distributions
* Grouping, Loops and Conditional Execution
* Functions
* Graphical Procedures
* Statistical Models in R
* Standard Packages
* OS facilities

## Module 2/Day 2: Exploratory Data Analysis (4hrs)

### Session 3: Visualizing Data (2 hours)

* Choice of Visualization Libraries: ggplot2, plotly
* Basic Charts
* Statistical and Scientific Charts
* Financial charts

### Session 4: Data Manipulation (1hrs)

* The tidyverse and dplyr package
* The pipe operator
* Loading, Selecting and Filtering Data
* Arranging and Mutating Data
* Joining Data

### Session 5: Advanced Visualizations (1hrs)

* 3D Plots
* Plots for Machine Learning
* Custom Controls
* Animations

## Module 3/Day 3: Statistical Analysis (4hrs)

### Session 6: Basic Statistical Analysis (2hrs)

* Descriptive Statistics (mean(), median(), sd(), var(), min(), max(), summary(), quantile())
* Probability Distributions (normal, t, chi-squared, F, binomial, Poisson) and Sampling from a distribution
* Statistical Inference and Hypothesis Testing
* Linear Models and Regression Analysis
* Generalized Linear Models (Logistic and Poisson Regression)

### Session 7: Advanced Statistical Analysis (2hrs)

* Resampling
* Linear Model Selection and Regularization
* Multivariate Data Analysis
* Monte Carlo Methods
* Non-Linear Models

## Module 4/Day 4: Machine Learning and Neural Networks (4hrs)

### Session 8: Tree Based Methods and Support Vector Machines (2hrs)

* Basics of Decision Trees (rpart, party, and tree)
* Classification Trees and Regression Trees
* Support Vector Machines

### Session 9: Supervised and Unsupervised Learning and Neural Networks (2hrs)

* Supervised and Unsupervised Learning
* Neural Networks – Using MNIST data
* Kaggles Cats and Dogs Data Set
* Building a simple CNN in R

### Module 5/Day 5: Software Engineering Considerations (4hrs)

### Session 10: Importing and Cleaning Data (2hrs)

* Data Sources
* Spreadsheets and CSV Files
* Databases
* Web APIs
* Scraping
* Parquet Files and Apache Arrow
* Data Cleanup Techniques

### Session 11: Applying R to New Problems (2hrs)

A demonstration of all the techniques taught so far as applied to a new problem.

* Identifying Buyer Segments for a Software Product

## References

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2023) *An Introduction to Statistical Learning: With Applications in R*. 3rd edn. Berlin/Heidelberg: Springer Nature.

Wickham, H., Çetinkaya-Rundel, M. and Grolemund, G. (2023) *R for Data Science*, 2nd edn. Sebastopol, CA: O'Reilly Media, Inc. Available at: [https://r4ds.hadley.nz/]

Chollet, F. and Allaire, J.J. (2018) *Deep Learning with R*. Manning Publications Co. Available at: [https://www.manning.com/books/deep-learning-with-r]

Hair, J.F., Babin, B.J., Anderson, R.E., and Black, W.C. (2018) *Multivariate Data Analysis*. 8th edn. NOIDA, India: Cengage (India) Pvt. Ltd. Available at: [https://www.amazon.in/Multivariate-Analysis-Joseph-Anderson-William/dp/9353501350/]