DEPARTMENT OF MANAGEMENT SCIENCES

INDIAN INSTITUTE OF TECHNOLOGY KANPUR

Proposal for New Course

Course Title: Introductory Programming for Data Sciences and Quantitative Research

Course Number: DMSXXX

Credits: 3-0-0-0 [5]

Prerequisite: The course is only prescribed for Ph.D. students doing quantitative research in management, economics, mathematics, and the engineering domains.

Proposing department: Prof. Abhinava Tripathi, Department of Management Sciences.

Other interested faculty: Prof. BV Phani, Prof. Suman Saurabh, Prof. Harshal Mulay

COURSE DESCRIPTION

This doctoral-level course introduces programming foundations for data-driven and quantitative research using R, a leading language for statistical computing, visualization, and reproducible analysis. Designed for PhD scholars pursuing empirical research across management, economics, mathematics, engineering, and social sciences, the course emphasizes the integration of programming logic, statistical reasoning, and applied econometric modeling. The course begins with an introduction to R and RStudio, data structures, and essential programming constructs. Students learn to manipulate, transform, and visualize complex datasets, develop skills in data wrangling, conditional programming, and advanced visualization with ggplot2. Subsequent modules bridge programming with quantitative methods, covering linear regression, logistic regression, panel data, and quantile regression, each reinforced through case-based applications drawn from real-world research problems. Beyond traditional econometrics, the course introduces auxiliary quantitative techniques including Principal Component Analysis (PCA), clustering methods, Markov regime-switching regression, and basic text analytics, enabling students to expand their methodological toolkit for interdisciplinary research. By the end of the course, PhD scholars will possess the computational proficiency and methodological depth required to implement reproducible workflows, perform robust empirical analyses, and independently develop data-driven research pipelines suitable for high-quality publications.

1. Course Objectives
The course aims to

- To equip PhD scholars with programming literacy in R for data-intensive research.
- To develop the ability to structure, clean, and visualize large datasets efficiently.
- To integrate statistical and econometric modeling techniques within research workflows.
- To foster reproducible and transparent quantitative analysis practices in empirical research.
- To introduce advanced modeling frameworks (PCA, clustering, regime-switching, text analytics) applicable to interdisciplinary studies.

2. Course Learning Outcomes
By the end of the course, the student should be able to

- Design and execute end-to-end data analysis pipelines using R for academic research.
- Implement and interpret statistical and econometric models with empirical rigor.
- Perform reproducible data cleaning, transformation, and visualization for publication-quality outputs.
- Employ advanced data analytics methods (PCA, clustering, Markov, text mining) in complex datasets.
- Integrate programming-based quantitative approaches into dissertation-level research and scholarly writing.

3. Course Contents

S. No.	Topic	No of Lectures
1	**Introduction to Programming with R** • What is R and R studio • Data structures in R • Basic mathematical, logical, and string operations • Basic data visualization	2
2	**Data wrangling with R** • Loops and conditional statements • Reading, writing, and other basic data processing • Sort, merge, filter, select, and perform other advanced operations • Advanced data visualization with GGPLOT	2
3	**Introduction to Regression with R** • Basic statistical operations with R (normal distribution, confidence intervals, hypothesis testing) • Basics of Regression • R implementation of regression • Regression case study with R	2
4	**Logistic Regression with R** • Basics of classification algorithms • R implementation of logistic regression • Logistic Regression case study with R	2
5	**Panel data regression** • Basics of Panel data • R implementation of Panel data modeling • Panel data case study with R	2
6	**Quantile Regression** • Basics of quantile regression • R implementation of quantile regression • Quantile regression case study with R	2
7	**Auxiliary quantitative approaches** • PCA and Clustering analysis with R • Markov regime switch regression with R • Basic text analytics with R	2

4. Prerequisite: The course is only applicable to PhD students (management, economics, mathematics, and engineering).

5. Short Summary: This course provides a comprehensive overview R implementation of various quantitative methods used in data sciences.
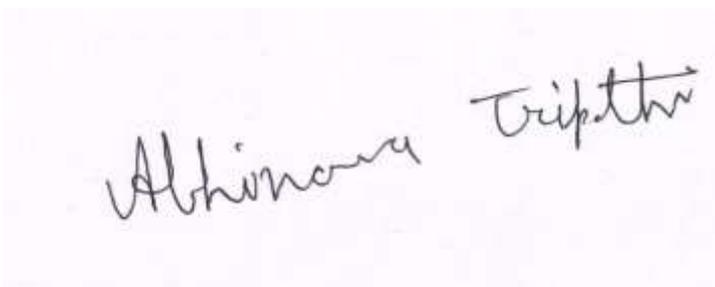
Proposer: Prof. Abhinava Tripathi (abhinavat@iitk.ac.in)
Dated:06/10/2025

DPGC Convener: Dated:

The course is approved / not approved

Chairperson, SPGC
Date:

Supplementary Readings
1. Hadley Wickham, "R for Data Science," O'Reilly, 1st Edition
2. Robert I. Kabacoff, "R in Action: Data Analysis and Graphics with R," Manning, 2nd Edition
3. Chris Chapman and Elea McDonnell Feit, "R for Marketing Research and Analytics," Springer, 2nd Edition
4. John Fox and Sanford Weisberg, "An R Companion to Applied Regression," Sage, 3rd Edition
5. Hadley Wickham, "ggplot2 Elegant Graphics for Data Analysis," Springer, 2nd Edition
6. Basic Econometrics by Gujarati, 4th Edition
7. Statistics for Business and Economics by Anderson and Sweeney 14th edition onwards
8. Introductory Econometrics for Finance, Chris Brooks, 3rd Edition