

INTERNATIONAL MANAGEMENT INSTITUTE, BHUBANESWAR

POST GRADUATE DIPLOMA IN MANAGEMENT

IS: 603 -DATA SCIENCE USING R

CREDIT: 1.5 credits

SESSION DURATION: 60 Minutes

TERM: IV
YEAR: 2020-2021
BATCH: 2019-21

FACULTY: Dr. Santanu Das

Telephone: 8658006959

E-Mail: santanu@imibh.edu.in

Office hours: 9.30 am to 5.30 pm

Consulting hours: Can be decided based on prior appointment.

Course Objective: Of late, R has become very popular software to undertake basic as well as advanced data analysis. Being freely available, it is used even by the corporate. The basic objective of this course is to expose the students to the R environment, so that they can use it in future to make business decisions.

Learning Outcomes: At the end of this course, students will develop a basic understanding about how R works and learn basic commands. This course will enable them to undertake independent research in their areas of interest as well as applying in their corporate life.

LO1: Introduction and basic understanding of the R environment

LO2: Understanding data analytical tools and their applications

LO3: Applying the software to undertake analysis and interpretation.

Suggested Readings: R for Everyone – Jared Lander, Pearson Education (First Ed, 2014)

Pedagogy: The pedagogy will involve using a given dataset on the R Studio for analysis. All the sessions will be held in this software.

Evaluation Criteria:

Components	Learning Outcomes	Weightage (%)
End Term Examination	LO1, LO2, LO3	40%
Project (in a group of 2 each)	LO3	30%
Assignments	LO3	10%
Report Writing**	LO2, LO3	20%
Total		100%

** The students will be given a business situation which will involve data analysis. (S)he will then write a report in a way which can be presented to the board of the company for decision making effectively.

Plagiarism

We are committed to upholding the highest standards of academic integrity and honesty. Plagiarism is the use of or presentation of ideas, works that are not one's own and which are not common knowledge, without granting credit to the originator. You may refer the already available content just for your reference and to get the basic ideas. Only 20% of such content is acceptable, above that comes under the definition of Plagiarism which is unacceptable in IMI and will be treated seriously. All such cases will be referred to the appropriate body of the Institute for suitable disciplinary action.

Session Plan:

Session No.	Topics	Learning Outcome	Readings
1-2	Getting Started with R <ul style="list-style-type: none"> • Comments, indents • Opening a dataset • Simple mathematical operations • Creating matrices • Subset data and declaring new variables • Types of data (Data frames, Vectors, Lists) 	LO1	Class exercise and handouts
3-4	Data Preprocessing and cleaning <ul style="list-style-type: none"> • Missing data handling • Removing rows and columns • Data transformation (Min-Max, Z, Decimal scaling) 	LO1, LO2	Class exercise and handouts
5-6	Data Visualization <ul style="list-style-type: none"> • Level graphs • Multiple graph • Histogram and kernels • qq plots • gg plots • Scatter plots 	LO1, LO2, LO3	Class exercise and handouts
7-8	Regression Techniques <ul style="list-style-type: none"> • Univariate • Multivariate (linear, logit, probit) 	LO1, LO2, LO3	Class exercise and handouts
9-11	Data Manipulation <ul style="list-style-type: none"> • Pipe operator • Filtering 	LO2, LO3	Class exercise and handouts

	<ul style="list-style-type: none"> • Mutation • Summarising • Looping 		
12-13	Handling outliers in the data	LO2, LO3	Class exercise and handouts
14-15	Data Wrangling <ul style="list-style-type: none"> • Reshaping Data • Joining Tables • String Processing 	LO1, LO2, LO3	Class exercise and handouts
16-17	Project Presentation		