COURSE TITLE	**STATISTICS FOR DATA SCIENCE**
COURSE CODE	**01AS1103**
COURSE CREDITS	**5**

Objective:

1 This course aims to build a strong foundation in statistical concepts and their practical implementation using R programming. It enables students to understand data types, organize and visualize data, apply statistical measures, and perform data analysis for real-world AI and data science applications.

Course Outcomes: After completion of this course, student will be able to:

1 Understand fundamental statistical concepts, data types, and data representation techniques

2 Apply statistical measures and R programming techniques for data handling and analysis.

3 Analyze datasets using statistical methods and visualization tools to interpret patterns and insights

4 Develop R-based solutions and simple programs for real-world data analysis tasks.

Pre-requisite of course: Foundational arithmetic and basic logical reasoning.

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
4	0	2	50	30	20	25	25

Contents : Unit	Topics	Contact Hours
1	**Basics of statistics.** Concept of Statistical population, Attributes and Variables (Discrete and Continuous), Different types of scales - Nominal, Ordinal, Ratio and Interval, Presentation of data: Classification, Tabulation, Diagrammatic & Graphical Representation of Grouped data, Frequency distributions, Cumulative frequency distributions and their graphical representations, Histogram, Frequency polygon and frequency curve, Ogives with their utility	12
2	**Measure of Central tendency** Apply the Measures of Central tendency (mean, median, mode, etc..) and Dispersion(Range, Standard deviation, Variance, Quartile Deviation, Mean deviation, etc.) with their properties, Merits and Demerits, Measures of Skewness and Kurtosis and their significance, Measures based on quartiles, Deciles, Percentiles	10

DR. TARANNUM ALIMAHMAD BLOCH
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Contents : Unit	Topics	Contact Hours
3	**Basics of R Programming & Data Handling** Introduction to R and RStudio environment, Data types: vectors, matrices, arrays, lists, data frames, Basic arithmetic and logical operations, Statements (looping concepts) and functions, Data import/export (CSV, Excel), Data cleaning and transformation using dplyr, Basic plotting in R, Handling of missing and duplicated cases/observations, Writing own functions	10
4	**Essential Statistical Analysis & Visualization using R** Apply the R programming for Histogram, boxplot, density plot, Scatter plot and 3D scatter plots (plotly), 3D surface visualization, Stem-and-Leaf Plot, Matplot, Plot options; Multiple plots in a single graphic window, Adjusting graphical parameters, Analyse descriptive Statistics: Measures of central tendency, Measures of dispersion, Skewness and kurtosis, Summary statistics using R	13
Total Hours		**45**

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Practical 1** Study and implementation of data types in R (vectors, matrices, lists, data frames)	2
2	**Practical 2** Data import/export using CSV and Excel files in R .	2
3	**Practical 3** Data cleaning: handling missing and duplicate values .	2
4	**Practical 4** Implementation of descriptive statistics (mean, median, mode) .	2
5	**Practical 5** Computation of variance, standard deviation, and quartiles .	2
6	**Practical 6** Creation of frequency distribution tables .	2
7	**Practical 7** Plotting histograms and frequency polygons .	2
8	**Practical 8** Visualization using boxplot and density plots .	2
9	**Practical 9** Scatter plot and correlation analysis .	2
10	**Practical 10** Multiple plots and graphical parameter adjustments in R .	2
11	**Practical 11** Writing user-defined functions in R .	2

DR. TARANNUM ALIMAHMAD BLOCH
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
12	**Practical 12** Mini project: Data analysis and visualization using real dataset.	2
Total Hours		**24**

Textbook :

1 The Art of R Programming: A Tour of Statistical Software Design, Norman Matloff, No Starch Press, 2011

2 R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, Hadley Wickham & Garrett Grolemund, O'Reilly Media, 2023

3 Applied Statistics and Probability for Engineers, Douglas C. Montgomery & George C. Runger, Wiley, 2018

References:

1 Introductory Statistics with R, Introductory Statistics with R, Peter Dalgaard, Springer, 2008

2 Think Stats, Think Stats, Allen B. Downey, O'Reilly Media, 2014

3 Hands-On Programming with R, Hands-On Programming with R, Garrett Grolemund, O'Reilly Media, 2014

4 Statistical Inference via Data Science, Statistical Inference via Data Science, Chester Ismay & Albert Y. Kim, CRC Press, 2019

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery					
Remember / Knowledge	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Higher order Thinking / Creative**
10.00	25.00	30.00	20.00	10.00	5.00

Instructional Method:

1 The course delivery method will depend upon the requirement of content and need of students. The teacher in addition to conventional teaching method by black board, may also use any of tools such as demonstration, role play, Quiz, brainstorming, MOOCs etc.

2 The internal evaluation will be done on the basis of continuous evaluation of students in the laboratory and class-room.

3 Practical examination will be conducted at the end of semester for evaluation of performance of students in laboratory.

4 Students will use supplementary resources such as online videos, NPTEL videos, e-courses.

DR. TARANNUM ALIMAHMAD BLOCH
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Supplementary Resources:

1 http://mathworld.wolfram.com
2 https://pll.harvard.edu/course/statistics-and-r