

COURSE TITLE	STATISTICS ESSENTIALS FOR AI & DATA SCIENCE USING R
COURSE CODE	01AS0502
COURSE CREDITS	4

Objective:

- 1 The objective of this subject is to build a strong foundation in statistics and R concepts relevant to Artificial Intelligence and Data Science, with a focus on probability, time series and statistical analysis using R. The course aims to enable students to remember, understand, and apply to applyze statistical methods for data analysis puppose. It emphasizes that the practical implementation of statistical techniques, development of analytical and logical thinking skills, and the ability to interpret results effectively. Students will strengthen their computational skills in handling and analyzing real-life data sets, preparing them for advanced data science subjects and real-world problem-solving contexts.

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

- 1 Learn R programming for statistical analysis and structured data handling. (Understand, Remember)
- 2 Perform descriptive and inferential statistical analysis using R. (Apply, Anlysize)
- 3 Apply probability and statistics in real life problems and interpret regression models for predictive analytics. (Apply, Analyze)
- 4 Demonstrate time series models and forecast time series data using statistical techniques in R. (Remember and Understand)
- 5 Apply and demonstrate PCA (Principle Component Analysis) and multivariate analysis techniques for dimensionality reduction. (Apply and Evaluate)

Pre-requisite of course:Basic statistics (mean, variance, and probability) and logical reasoning for programming.

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
3	0	2	50	30	20	25	25

Contents : Unit	Topics	Contact Hours
1	Basics of R Programming & Data Handling Introduction to R and RStudio environment, R 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	8

Contents : Unit	Topics	Contact Hours
2	Essential Statistical Analysis & Visualization using R Apply descriptive Statistics: Measures of central tendency, Measures of dispersion, Skewness and kurtosis, Summary statistics using R, Apply probability distributions (ImplementationFocus): Normal distribution, Binomial distribution, Poisson distribution, Distribution functions: dnorm(), pnorm(), rnorm(), etc., Hypothesis Testing in R: One-sample and two-sample t-tests, Chi-square test, One-way ANOVA, Visualization (2D & 3D): Histogram, boxplot, density plot, Scatter plot and correlation heatmap, 3D scatter plots (plotly), 3D surface visualization, Stem-and-Leaf Plot.	10
3	Regression Modeling & Interpretation Simple Linear Regression, Multiple Linear Regression, Model assumptions and diagnostics, Residual analysis, Logistic Regression, Model evaluation metrics (R^2 , RMSE, Confusion Matrix), Visualization of regression plane (3D), Interpretation of coefficients and model output	9
4	Time Series Analysis & Forecasting using R Introduction to Time Series Data, Components: Trend, Seasonality, Random Component, Time series visualization, Moving averages, Autocorrelation (ACF) & Partial Autocorrelation (PACF), Exponential smoothing, stationarity and unit root testing, Augmented Dickey Fuller (ADF), Kwiatkowski Phillips Schmidt Shin(KPSS), Basic ARIMA, SARIMA, and Autoregressive Model (AR(p)) models, Forecasting and interpretation of time series models	10
5	Multivariate Analysis & Dimensionality Reduction Techniques Covariance and correlation matrix, Principal Component Analysis (PCA), Scree plot and variance explanation, PCA visualization (2D and 3D), Introduction to clustering techniques, Interpretation of multivariate datasets	8
Total Hours		45

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	Practical 1 Installation and configuration of R and RStudio	2
2	Practical 2 Basic R operations and data structures	2
3	Practical 3 Importing and cleaning real datasets and of descriptive statistics	2
4	Practical 4 Visualization using ggplot2. 2D and 3D with plotly	2
5	Practical 5 Functions for probability distributions in R	2

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
6	Practical 6 Implementation of probability distributions	2
7	Practical 7 Hypothesis testing using R with data set	2
8	Practical 8 Linear regression modelling with data set	2
9	Practical 9 Logistic regression implementation with data set	2
10	Practical 10 Time series decomposition (trend, seasonality)	2
11	Practical 11 ARIMA modelling using pre-defined dataset	2
12	Practical 12 PCA implementation using Real data	2
Total Hours		24

Textbook :

- 1 Introduction to Statistics and Data Analysis: With Exercises, Solutions and Applications in R, Christian Heumann, Michael Schomaker, and Shalabh, Springer, 2016
- 2 An Introduction to Statistical Learning with Applications in R, G. James, D. Witten, T. Hastie and R. Tibshirani, Springer, 2021

References:

- 1 Introductory Statistics with R, Introductory Statistics with R, Peter Dalgaard, Springer, 2008
- 2 R in Action, R in Action, Robert Kabacoff, Manning Publications, 2022

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
15.00	25.00	30.00	20.00	10.00	0.00

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.

Instructional Method:

- 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 .

Supplementary Resources:

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