PARVATHANENI BRAHMAYYA SIDDHARTHA COLLEGE OF ARTS & SCIENCE
Autonomous
Siddhartha Nagar, Vijayawada-520010
Re-accredited at 'A+'by the NAAC

Offered to: M.Sc. (Computational Data Science)

CourseName	Applied Data Analysis			**L**	**T**	**P**	**C**	**CIA**	**SEE**	**TM**
CourseCode	22DS4E4			4	0	0	4	30	70	100
Year of Introduction: 2022	**Year of Offering:** 2022	**Year of Revision**: Nil		**Percentage of Revision:**						
L-Lecture, T-Tutorial, **P**-Practical, **C**-Credits, **CIA**-Internal Marks, **SEE**-External Marks, **TM**-Total Marks										

Course Description and Purpose: Applied Data Analysis is a course that illustrates *concepts ofR-Programming*, *Data Structures*, *Descriptive Statistical Analysis*, *Basic Graphs*, *Analysis of ANOVA*, *Multivariate Analysis, Files & Databases.*
Course Objectives: : This course will help enable the students to understand and familiar with *R-Programming*, *Data Structures*, *Descriptive Statistical Analysis*, *Basic Graphs*, *Analysis of ANOVA*, *Multivariate Analysis, Files & Databases.*

Course Outcomes:
On successful completion of this course, the students able to:
CO1: How to use R environment, R packages, Data Types and control Flow Statements. .
CO2: Explain basic implementation techniques in R, including regression and time series analysis
CO3: Apply Descriptive statistics and hypothesis tests in R including parametric and non-parametric tests.
CO4: Analyze variance within datasets using various ANOVA models
CO5: Create and interpret Basic, and Advanced Graphs, Database connection using MYSQL in R.

CO-PO MATRIX								
COURSE CODE	**CO-PO**	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**
22DS4E4	**CO1**	H				H		
	CO2	H		H				
	CO3	H		H				L
	CO4	M		M				
	CO5	M					M	L

UNIT-I (12 Hours)

Introduction to R: Why use R?, R Environment, Working with R Packages, Understanding Datasets, Data Types, Data Structures (Operations on Data Structures), Missing Values, Sorting Data, Merging Datasets, Subsetting Datasets, Control Flow Statements, Aggregation and Restructurings.

UNIT-I (12 Hours)

Descriptive Statistics: Introduction to Descriptive Statistics (Measures of Central Tendency, Measures of Dispersion of Variability, Measures of Shapes (Skewness and Kurtosis)), Introduction to Sampling(Sampling Types), Hypothesis Testing with R(One Sample Test, One Sample Sign Test, Two Samples Test), Parametric Test(Correlations, Z-Test, T-Test), Non Parametric Tests (Wilcoxon Signed- Rank Test, Chi Square Test).

UNIT-III (12 Hours)

Basic Graphs: Bar Plots, Pie Charts, Histograms, Line, Dot Plots, Kernel Density Plots and Dot Plots.

The Advanced Graphics: The ggplot2 Package.

Analysis of Variance: Fitting ANOVA Models, One-way ANOVA, One-way ANCOVA, Two-way factorial ANOVA, Repeated measures ANOVA, Multivariate Analysis of Variance (MANOVA)

UNIT-IV (12 Hours)

Basic Multivariate Analysis: Regression (Simple Linear Regression, Multiple Linear Regression, Logistic Regression), Time Series Analysis (Creating Time Series, Components of Time Series Analysis, Seasonal Decomposition, Exponential Models), Forecasting (Simple Moving Averages, Weighted Moving Averages, Single Exponential Smoothing)

UNIT-V (12 Hours)

Connecting R to External Interfaces: CSV Files (Reading From a CSV File, Writing to a CSV File), Microsoft Excel (Reading from XLSX File, Writing to XLSX File), Databases (Connecting R to MYSQL ,Creating Tables, Inserting Rows, Updating Rows, Deleting Rows, Querying Rows, Querying Tables, Dropping Tables), XML Files (Reading From XML Files, JSON Files, Reading From JSON Files), Binary Files (Writing to Binary Files, Reading From Binary Files).

Prescribed Text Book			
	Author	Title	Publisher
1	Dr. Rob Kabacoff	R in Action : Data Analysis and Graphics with R. [UNIT-I ,UNIT-II ,UNIT-III]	Manning Publications Co, Edition 2011.
2	Dr.Jeeva Jose	A Beginners Guide For Data Analysis Using R Programming. (UNIT IV and UNIT V) UNI IV: Chapter-11 11.3 [11.3.1 to 11.3.3] 11.5,11.6 [11.6.1 to 11.6.3] UNIT V: Chapter-6 [6.1 to 6.6]	Khanna Book Publishing Co.(P) Ltd, Edition 2019.

Reference Text Books			
	Author	Title	Publisher
1	Dr. Dhaval Maheta	Data Analysis using R	Notion Press, September 2021
2	Michael J.Crawley	The R Book	Wiley, Edition: 2007
3	Ken Black John	Business Statistics for Contemporary Decision Making	Wiley & Sons, Inc., Edition 2013

PARVATHANENI BRAHMAYYA SIDDHARTHA COLLEGE OF ARTS & SCIENCE
Autonomous
Siddhartha Nagar, Vijayawada-520010
Re-accredited at 'A+'by the NAAC

M.Sc.(Computational Data Science) **Semester :IV**

Course Code: 22DS4E4Course Name: Applied Data Analysis

Time: 3 Hours **Max Marks: 70**

SECTION-A

Answer the following questions. (5×4=20Marks)

1. (a)Whatarethedifferent*DataTypes*usedinR.(CO1,L1)
   (or)
   (b) Define*Subsetting*and *Merging*.(CO1,L1)
2. (a) How to remove missing values in R?(CO1,L1)
   (or)
   (b) What is meant by *Random Sampling* and*Non Random Sampling*? (CO3,L1)
3. (a) What is *Correlation*? Explain its types (CO3,L1)
   (or)
   (b) Whatismeantbyt-test and f-test? Give one example using R (CO3.L1)
4. (a) Whatis thepurposeofANOVA?(CO4,L1)
   (or)
   (b) Define *Logistic Regression. Give one* example using R. (CO2,L1)
5. (a) Define*TimeSeriesAnalysis* and its components.(CO2,L1)
   (or)
   (b) Whatisthesyntaxusedtoread*XMLFiles*.(CO6,L1)

SECTION-B

Answer the following questions. (5×10=50Marks)

6. (a)Outlinethedifferent*DataStructures*usedinR.(CO1,L2)
   (or)
   (b) Explain *Control Flow Statements* in R.(CO1,L2)
7. (a) Explain the different statistical measures used in *Descriptive Statistics*. (CO3,L5)
   (or)
   (b) Explain *Non Parametric Test* and *Wilcoxon Signed-Rank Test* in R (CO3,L5)
8. (a) List*VariousTypesofCharts*inR. (CO6,L4)
   (or)
   (b) Analyze *One-way ANOVA* and *Two-way factorial ANOVA*. (CO4,L4)
9. (a) Distinguish *Simple*and*MultipleRegression* inR withExample.(CO2,L4)
   (or)
   (b) Classify various components used in *Time Series Analysis in R* with example. (CO3,L4)
10. (a) Explain procedure to *connecttoadatabase*inRusingMYSQL with an example. (CO6,L5)
    (or)
    (b) Explain procedure to import*csvfile*and*binaryfile* inR with an example. (CO6,L5)