## APPENDIX-II
## SYLLABI FOR OPEN ELECTIVES

### 22OE301: R-PROGRAMMING

| Course Name | R-Programming | | | L | T | P | C | CIA | SEE | TM |
|---|---|---|---|---|---|---|---|---|---|---|
| Course Code | 22OE301 | | | 3 | 0 | 0 | 3 | 30 | 70 | 100 |
| Year of Introduction: 2020 | Year of Offering: 2023 | Year of Revision: No Revision | | Percentage of Revision: Nil | | | | | | |
| **L**-Lecture, T-Tutorial, **P**-Practical, **C**-Credits, **CIA**-Internal Marks, **SEE**-External Marks, **TM**-Total Marks | | | | | | | | | | |

**Course Description and Purpose:** This course provides a comprehensive understanding of R programming, descriptive statistics, bi-variate analysis, regression, analysis of variance, hypothesis testing, parametric and non-parametric tests, and connecting R to external interfaces for practical data analysis and interpretation. This course equip learners with essential skills in R programming and statistical analysis, enabling them to manipulate data, conduct hypothesis tests, perform regression analysis, and connect R to external interfaces for robust decision-making in various fields.

**Course Objectives:** This course will help the students to learn about Introduction to R Programming, Descriptive Statistics and Bi-variate Analysis using R, Regression Using R, Analysis of Variance Using R, Testing of Hypothesis Using R, Parametric Tests, Non Parametric Tests and Connecting R to External Interfaces.

**Specific objectives include:**
- To provide an understanding on Introduction to R Programming ..
- To learn Descriptive Statistics and Bi-variate Analysis using R
- To gain knowledge on Regression Using R and Analysis of Variance Using R.
- To lean about Testing of Hypothesis Using R, Parametric Tests and Non Parametric Tests.
- To provide understanding on Connecting R to External Interfaces.

**Course Outcomes:**
**CO1:** By mastering R programming fundamentals, data manipulation techniques, and statistical functions, students will gain the ability to explore, analyze, and visualize diverse datasets effectively, making informed decisions through a variety of data representation methods.
**CO2:** Students will develop expertise in descriptive statistics, mastering measures of central tendency, dispersion, and correlation, enabling them to analyze diverse data sets, discern patterns, and make informed decisions in both categorical and numerical contexts.
**CO3:** Students will master regression techniques (simple linear, multiple linear, and logistic) and ANOVA methods (one-way, two-way, multivariate, and repeated measures), enabling them to model complex relationships and conduct in-depth analysis of variance in diverse datasets using R programming.
**CO4:** Students will acquire expertise in hypothesis testing methodologies, mastering the steps and types of hypothesis testing, as well as parametric tests (t-test, z-test, f-test) and non-parametric tests (Mann-Whitney U test, Kruskal-Wallis test, Chi-Square test), empowering them to assess and draw valid conclusions from a wide array of data sets using R programming.
**CO5:** Students will gain the ability to seamlessly connect R to external interfaces, including CSV files, Microsoft Excel, databases (MySQL), XML files, and JSON files, enabling them to efficiently import, export, manipulate, and analyze data from diverse sources, enhancing their data integration and analysis capabilities.

## UNIT-I (12 Hours)

**Introduction to R Programming:** Why use R?, R Features, R Environment, Working with R Packages, Define Dataset, Data Types using R, Data Structures using R, Missing Values, Sorting Data, Merging Datasets, Subsetting Datasets, Operators in R, Important Statistical Functions in R, Exploratory Data Analysis - Bar Chart, Pie Chart, Histogram, Line Plot, Box Plot, Scatter Plot and Density Plot.

## UNIT-II (12 Hours)

**Descriptive Statistics and Bi-variate Analysis using R:** Introduction to Descriptive Statistics Measures of Central Tendency, Measures of Dispersion of Variability, Measures of Shapes-Skewness and Kurtosis, Correlation Meaning - Types of Correlation-Measures of Correlation - Scatter Diagram, Karl Pearson's Coefficient of Correlation, Spearman's Rank Correlation Coefficient, Bi Serial Correlation, Bi-variate Analysis of Categorical Variables and Numerical Variables.

## UNIT-III (12 Hours)

**Regression Using R:** Estimation the Method of Least Square, Introduction to regression, Types of Regression Models-Simple Linear Regression, Multiple Linear Regression, Logistic Regression and its implementation using R Programming

**Analysis of Variance Using R**: Definition of ANOVA, Types of ANOVA **-** One way ANOVA**-**Two way ANOVA**-**Multivariate Analysis of Variance (MANOVA) and Repeated Measure ANOVA

## UNIT-IV (12 Hours)

**Testing of Hypothesis Using R:** Definition of Hypothesis Testing, Steps in Testing of Hypothesis, Types of Hypothesis Testing - Null Hypothesis, Alternative Hypothesis and Statistical Hypothesis.

**Parametric Tests:** t-test, z-test and f-test, Differences between t-test and z-test.

**Non Parametric Tests:** The Mann Whitney U Test, Kruskal Wallis Test and Chi Square Test.

## 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)

**Reference Text Books:**


1. Sharma, J. K., Business Statistics (UNIT-I,UNIT-III), New Delhi: Pearson Education, 2013
2. Anderson,D.,Sweeney,D.,Williams,T., Camm, J., & Cochran, J., Statistics for Business and Economics, Cengage Learning, 2013, New Delhi
3. Dr. Rob Kabacoff, R in Action: Data Analysis and Graphics with R (UNIT-IV), Manning Publications CO, Edition 2011.
4. Dr.Jeeva Jose, A Beginners Guide for Data Analysis Using R Programming. (UNIT-II, UNIT-V, UNIT-III), Khanna Book Publishing Co.(P) Ltd, Edition 2019.
5. Michael J. Crawley, John Wiley & Sons, Statistics: An Introduction using R, Weily, 2015.
6. Aczel,A.D.& Sounderpandian, J, Complete Business Statistics, Tata McGraw Hill, 2011, New Delhi.
7. Davis, G., & Pecar, B., Business Statistics using Excel, New Delhi: Oxford University Press, 2014.

**PARVATHANENI BRAHMAYYA SIDDHARTHA COLLEGE OF ARTS & SCIENCE**

(An Autonomous College in the jurisdiction of Krishna University)

**Course Name:** R-Programming (Open Elective), Third Semester

**Course Code:** 22OE301

**(w.e.f admitted batch 2022-23)**

**SECTION-A**

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

**Answer ALL questions** **(5×4 = 20 Marks)**

1. (a) What are the different *Data Types* used in R ? (CO1,L1)

(or)

(b) What are the important statistical functions used in R.?(CO1, L1)

2. (a) Explain about Skewness as a measure of shape in Descriptive Statistics (CO2,L2)

(or)

(b) Explain about Types of Correlations used in R (CO2, L2)

3. a) Explain about Logistic Regression used in R CO3,L5)

(or)

b) Explain about MANOVA used in R (CO3,L5)

4. (a) What are the different steps used in hypothesis testing? (CO4.L1)

(or)

(b) What is meant by f-test with example ? (CO4,L1)

5. a) How we can insert data into R using MYSQL ? (CO5,L2)

(or)

b) Explain process of reading data from XML file with example (CO5, L2)

**SECTION-B**

**Answer Five Questions Choosing One Question from Each Unit.**

**All Questions Carry Equal Marks.** **(5×10 = 50 Marks)**

6. (a) Outline the different *Data Structures* used in R. (CO1,L2)

(or)

(b) Explain about different Operators in R. (CO1,L2)

7. (a) Explain measure of central tendency and measure of dispersion used in *Descriptive Statistic*s (CO2,L2)

(or)

(b) Explain Bi-variate Analysis of Categorical Variables and Numerical Variables using R with examples **.** (CO2,L2)

8. (a) Apply Simple Linear Regression and Multiple Linear Regression using R. (CO3,L3)

(or)

(b) Construct One Way ANOVA and Two Way ANOVA using R. (CO3, L3)

9. (a) Explain different Types of Hypothesis Testing used in R. (CO4,L2)

(or)

(b) Explain about z-test and its types used in R with examples. (CO4,L2)

10. (a) Explain process of Reading and writing data from CSV and Excel files with examples. (CO5,L5)

(or)

(b) How do you connect to a database in R using MYSQL ? Give one example (CO5, L5)