Academic Council Meeting No. and Date : 03 / February 14, 2022

Agenda Number : 6 Resolution Number : 12 / 5.2

Vidya Prasarak Mandal's
B. N. Bandodkar College of Science (Autonomous), Thane

Certificate Course

in

Basics of R software

With effect from

Academic Year 2022-2023

PREAMBLE

Basics of R Software (course) is introduced as a certificate course conducted by Department of Statistics VPM's B.N. Bandodkar College of Science, Thane. In choice based credit system the course would be of 45 hrs. duration. There would be single paper based on syllabus consisting of 4 units. This course will offer credit of 2 on successful completion of the course.

Any scientific task without the knowledge of software is difficult to imagine and complete in the current scenario. Different software packages are available to analyze the data. It is felt that there is need to well equip the students, teachers and others who are using Statistics and related topics. R is free software that is capable of handling mathematical and statistical manipulations. R is an **open-source software environment for statistical computing** that is rapidly becoming the tool of choice for data analysis. It's developed by a large international community of scientists and programmers and is **at the forefront of new developmentsin statistical computing.** It has its own programming language as well as built-in functions to perform any specialized task. We intend to give the basics of R software in this course.Many students, teachers and others face the problem of using appropriate Statistical tools, calculations, understanding and interpreting the numerical answers. In order to understand the basic concepts of Statistics using R software, we have organized this Certificate Course.

Examination will be conducted in in two parts; internal and practical. Internal examination will be taken in a form of assignments. For each unit one assignment will be given which will help in strengthening the knowledge of the student. Practical examination will focus on application of the knowledge gained in the course. Teaching time allotted will be 20 periods for theory and 25 periods for practical.

COURSE COVERAGE:

1. Master the use of the R interactive environment.
2. Understand basic regular expressions in R.
3. Understand base R graphics.
4. Use R for descriptive statistics.
5. Use R for inferential statistics.
6. Multiple regression in R.

OBJECTIVES OF THE COURSE:

1. To familiarize them withbasic regular expressions in R.
2. To impart knowledge about creating charts, plots and vectors.
3. To impart knowledge aboutbasic data analysis procedures.
4. To understand the multiple regression models.
5. To understand the procedure to perform all conventional statistical models and analysis tests.
6. To develophands-on experience on solving problems related to real-life data.

LEARNING OUTCOMES: After completion of this course participant would be able to

1. Edit and save changes in existing data.
2. Summarise and graphically display data.
3. Perform all conventional statistical analysis tests.
4. Read existing datasets into R or create new ones.
5. Handle any form of data.

ELIGIBILITY:

The certificate course offers a comprehensive content designed for researchers, educators and students of diverse backgrounds who have limited knowledge in statistical data analyses, but are enthusiastic to earn competency in learning the computational science of data using R.

It's suitable for **undergraduates, graduates and researchers** from any field that uses statistical computing.

Certificate Course Basics of R Software

CourseCode	Course Title	No. of lectures	Credits
BNBCCBR1T1	Basics of R Software	**20**	**2**
BNBCCBR1P1	Practicals in R	**25**	

Syllabus of Basics of R Software

Theory

UNIT	SUB-TOPICS	No. of lectures
I	**Fundamentals of R** **Introduction (Introduction to R, Installation of R, Starting & Ending in R, Getting help in R)** What is Statistics ?(with Examples), What is Data? **Basic operations**→ + , - , *,÷, ^, sqrt. **Numerical Functions**→log10, log, sort, max, min, unique, range, length, var, prod, sum, dim, etc. **Data Types** →Vector, List, Matrices, Array and data frames **Variable types and Factor**→ Logical, Numerical, Integer, Complex, Character. R-Commands to Input data a) **Assignment statement:** • =, <- or -> b) **Creating vectors:** • c( ) → Concatenate Function • scan( ) → Scan Function c) **Generating sequences** • : → Sequence operator • seq( ) → (from=a ,to=b,by=c) • seq(length=d,from=a,by=c) d) **Replicating object or elements**: • rep( ) → Replicate function **e) Matrix Operations:** • matrix( ) [With examples] **Data manipulation** → Selecting random N rows, Removing duplicate rows, Renaming variables, Creating a new variable, Selection of random fraction of rows, Appending of rows and columns, Simulation of variables. • Data frames [With examples] Accessing value /data from data frames. *** Inbuilt data sets and Resident data** *** Data Processing** → Data Import and Export , Setting working directory, Checking structure of data: str( ), class( ), changing types of variables (for ex. as.factor, as.numeric)	5

II	**Descriptive Statistics** **a) Data Visualization :- (Graphs and Diagrams)** Simple bar diagram, Sub divided bar diagram, Multiple bar diagram, Pie diagram, Box plot for one and more variables, Histogram Frequency polygon, Frequency curve, Stem and leaf, Scatter plot. **b) Measures of central tendency :-** • Mathematical averages : Mean (A.M., G.M., H.M.) • Positional averages : Median, Quartiles, Deciles, Percentiles • Mode c) **Calculations of Measures of Central Tendency using R** (psych package) Examples solved using R Ungrouped data Ungrouped data with NA values **Measures of Dispersion** • Range, Coefficient of Range • Quartile Deviation, Coefficient of Quartile Deviation • Mean Deviation about a (Mean, Median), Coefficient of Mean Deviation • Variance, Standard Deviation, Coefficient of Variation **Measures of Skewness** • Absolute Measures of Skewness • Karl Pearson's Measures of Skewness • Bowley's Measures of Skewness **Relative or Coefficient of Skewness** • Karl Pearson's Coefficient of Skewness • Bowley's Coefficient of Skewness • Measure based on Moments **Kurtosis** :- Leptokurtic, Mesokurtic, Platykurtic (Examples)	**5**
III	**Probability and Probability Distributions** **Probability (Introduction)** P(A)=n(A)/n(S) = Total no. Of cases favourable to Event A/Total no. Of cases *** Permutations, Combinations** *** Probability** **Probability Distributions** • Binomial Distribution, Poisson Distribution, Uniform Distribution • Exponential Distribution • Normal Distribution • Student's t Distribution, F Distribution • Chi Square Distribution (With Examples) *** Testing of Hypothesis** *** Statistical Tests using R** • t-test (One Sample t test, Two Sample t test, Paired t test) • Chi-Square Test for Association, Chi-Square Test for Variance • F-test for Equality of Variances	**5**

IV	**<u>Correlation Analysis, Linear Regression Analysis and Curve Fitting</u>** **Introduction :-** **Linear Regression Model** with One Explanatory variable • **Data Pre-processing** :- Detection and Treatment of Missing values and Outliers, Model building • Interpretation of Output produced by lm command in R. **Multiple Linear Regression Model** • Data Pre processing **Curve Fitting** • Quadratic Curve • Exponential Curve • Power Curve • Logarithmic Curve	**5**
	Total Theory Periods	**20**

Practical

UNIT	SUB-TOPICS	No. of lectures
I	**Fundamentals of R** a) Assignment statement: b) Creating vectors: c) Generating sequences d) Replicating object or elements: e) Matrix Operations: Data manipulation Inbuilt data sets and Resident data	**6**
II	**Descriptive Statistics** a) Data Visualization (Graphs and Diagrams) b) Measures of central tendency c) Calculations of Measures of Central Tendency using R d) Measures of Dispersion e) Measures of Skewness f) Relative or Coefficient of Skewness g) Kurtosis :- Leptokurtic, Mesokurtic, Platykurtic (Examples)	**6**
III	**Probability and Probability Distributions** a) Probability (Introduction) b) Permutations, Combinations c) Probability Distributions d) Statistical Tests using R	**6**
IV	**Correlation Analysis, Linear Regression Analysis and Curve Fitting** **Introduction :-** **Linear Regression Model** with One Explanatory variable • **Data Pre-processing** :- Detection and Treatment of Missing values and Outliers, Model building • Interpretation of Output produced by lm command in R. **Multiple Linear Regression Model** • Data Pre processing **Curve Fitting** • Quadratic Curve • Exponential Curve • Power Curve • Logarithmic Curve	**7**
	Total Practical Periods	**25**

Suggested Readings:

1. Crawley, M. J. (2006). Statistics - An introduction using R. John Wiley,London
2. Purohit, S.G.; Gore, S.D. and Deshmukh, S.R. (2015). Statistics using R,second edition.Narosa Publishing House, New Delhi.
3. Shahababa , B. (2011). Biostatistics with R, Springer, New York
4. Verzani, J. (2005). Using R for Introductory Statistics, Chapman and Hall/CRC Press,New York
5. Asha Jindal (Ed.)(2018), Analysing and Visualising Data with R software-APracticalManual, ShailjaPrakashan, K.C.College.

Evaluation Scheme

Assignments and Practical Examination

Total number of assignments:
(04 each carrying) **05 marks *4 assignments**
Total marks: **20 marks**

Internal Assignments have to be submitted in the softcopy format in the department.

Practical Examination: Suggested Format of Question paper
Duration: $2^{1/2}$ Hours **Total Marks: 80**

All questions are compulsory

Q. 1	Based on Unit I	20
Q. 2	Based on Unit II	20
Q. 3	Based on Unit III	20
Q. 4	Based on Unit IV	20

Total of Internal Assignments **20 Marks**
Total of Practical Examination **80 Marks**
Grand Total **100 Marks**

~ * ~ * ~ * ~ * ~ * ~