Semester	**FOUR**
Paper Number	**HSTSE4021P**
Paper Title	**Statistical Data Analysis Using R**
No. of Credits	**2**
Theory/Composite	**Practical**
No. of periods assigned	**2+2**
Module	Single
Course description/objective	*At the end of the course, a student is expected to:* o Create, Access and Save Files, access help pages and load/install new packages. o Use R as a calculator to compute basic mathematical functions. o Draw diagrams and add titles and legends to them. o Compute statistical measures using in-built functions. o Learn programme structures and implement them to write one's own code. o Read and Write Data from external file sources.
Syllabus	**UNIT 1:** ***Introduction :*** History and Overview of R, the CRAN, Installing the R Software, The R-Console and the R-Script. Saving and Accessing Files. Libraries in R. Loading and Installing Packages in R. The *quit()* and the *history()* commands. [2L] ***R as a calculator :*** Basic Mathematical Functions. The Base Library. Defining variables, calling variables, Unary and Binary Operators on Variables. [3L] ***Modes of Data Storage :*** Vectors, Matrices, Data Frames, Lists. The *c(), edit()* and *scan()* commands. Defining Attributes. Creating Patterned Data – the *rep()* and *seq()* commands. Extracting rows and columns in data frames and lists. Assigning names to columns of data frames and matrices and rows of lists. The *$* operator. The *attach()/detach()* command. Conditional selections and subsetting of objects. The *length()* command. Merging multiple vectors or columns of different data frames into one - The *cbind(), rbind()* and *merge()* commands. Inter-Conversions of the various modes of storages. [6L] **UNIT 2:** ***Diagrammatic representations of Non-Frequency Data :*** the *plot()* command. Line Diagram, Bar (Horizontal and Vertical) diagrams, Multiple Bar diagrams, Multiple Line diagrams, Pie and Subdivided Charts. Adding legends, Title, labels, limits on the axis. The *'graphics'* package and the *'ggplot2'* package. The *par()* parameter and its arguments. [6L] ***Diagrammatic representations of Frequency Data :*** Frequency Distributions, the *table()* command. Column Diagrams and Histograms. Box Plots - the *summary()* command. Cumulative Frequency Diagrams. Juxtaposing frequency curves over

	histograms. [8L] **UNIT 3:** ***Univariate Statistics:*** Descriptive Measures of Central Tendency, Dispersion, Skewness and Kurtosis. The *'moments'* package and its functions. [4L] ***Bivariate Statistics:*** Scatterplot, Various forms of correlations. Regression Theory – the *lm()* command, polynomial regression. Residual Plots. [2L] ***Linear Algebra:*** Algebra of Matrices. The *'Matrix'* package. Obtaining Determinants, Trace, Rank and Inverse of a Matrix. Obtaining row reduced forms of matrices, obtaining an orthonormal basis. Eigen Values and Eigen Vectors. Solving a system of equations. Diagonalisation of Matrices. [2L] **UNIT 4:** ***Programming in R:*** Control Statements: if, if else. Loop Structures: for, while, repeat. User defined functions – Passing arguments, calling functions and returning values. [7L] ***Statistical Simulations:*** Drawing Random Samples from different finite and infinite probability distributions – the *set.seed()* command. Illustrations through statistical problems (probability estimates by long-run relative frequencies, Bias and MSE's of estimates, coverage of Confidence Intervals, calculating empirical level and power of tests). Optimisation of Functions – the *optim()* function and its various arguments. [8L] ***File Handling:*** Importing and Exporting Data from/to other softwares. [4L]
List of Practical	1. Computing Basic Mathematical Functions using R as a calculator. 2. Storing Data in various modes - vectors, matrices, data frames and lists. 3. Representing Non-Frequency Data by diagrams. 4. Obtain Frequency distributions from raw discrete and continuous data. 5. Representing Frequency Data by diagrams. 6. Univariate Statistical Measures in R. 7. Bivariate Statistical Measures in R. 8. Arithmetical Operations on Matrices and computing determinants, rank, inverse, characteristic roots and vectors of matrices. 9. Control Structures and Loops in R. 10. Applications of control structures and loops to write programme codes of various statistical problems. 11. Export and Import Data from/to other software 12. Install and load new packages and libraries in R.

Reading/ Reference list	1. Dalgaard, P : Introductory Statistics with R, Springer Pubications, 2nd edition, 2008. 2. Maindonald, J. & Braun, J. : Data Analysis and Graphics Using R , Cambridge University Press, Cambridge, 2nd edition, 2007. 3. Faraway, J. J. : Linear Models with R ,Chapman& Hall/CRC Texts in Statistical Science.
Evaluation	CIA: 20 End Sem: 80 Total: 100