Skill Based (SB) Courses

Skill Based/Specialized Laboratory (SB) Course 3a: Data Analysis Using R

Structure 1: PG Curricular Structure with only Course Work

Course Title & Code	Credits	Credit Distribution of the Course			Eligibility Criteria	Prerequisite of the course (if any)
		Lecture (00 Hours)	Tutorial (00 Hours)	Practical (60 hours)		
SB 3a: Data Analysis Using R	**2**	**0**	**0**	**2**	**NIL**	**NIL**

Course Objectives:

- To enhance the programming skills and working knowledge of R software.
- Equip students for data management and optimization using R.

Course Learning Outcomes: After successful completion of this course, the students will be able to:

- Understand R programming for data analysis
- Accessing R packages and use different functions in R
- Visualize and summarize the data using statistical functions and graphs
- Develop computational skills for the generation of random samples
- Perform MLEs and application problems based on the fitting of a suitable distribution.
- Design and analyse incomplete block designs.

Unit I: (14 Hours)

Introduction to R, R data structures, kernel density plots, ggplots2 package. Panel display, surface plots, contour plots, plots in 2-D and 3-D. Exploratory data analysis of the empirical distribution function and its properties, quantile function, confidence interval of quantiles of order p, tolerance and convergence. Kernel function for density estimation.

Unit II: (14 Hours)

The inverse transformation for the generation of random variables from discrete, continuous and exponentiated distributions. One-dimensional optimization, Maximum Likelihood Estimation through ***nlm,*** and ***optim*** packages. Non-parametric tests: Kruskal-Wallis test, Wilcoxon, Mann-Whitney test.

Unit III: (14 Hours)

Hypothesis testing of one and paired sample t-test. ANOVA: Fixed, Random and Mixed effect models. Incomplete and Confounded Block Designs: Balanced Incomplete Block (BIB) Designs, Confounded 2^K and 2^{k-p} Designs.

Unit IV: (14 Hours)

Cluster analysis: Hierarchical Cluster analysis, K-Means Cluster Analysis, Factor analysis Principal Components Analysis, Discriminant Analysis.

Essential Readings:

1. Davies, T.M. (2016). *The Book of R: A First Course in Programming and Statistics*, No Starch Press.
2. Lawson, J. (2015). *Design and Analysis of Experiments with R*, Chapman and Hall/CRC Press.
3. Rizzo, M.L. (2019). *Statistical Computing with R*, Chapman & Hall/CRC Press.

Suggested Readings:

1. Crawley, M.J. (2023). *The R Book*, John Wiley & Sons.
2. Gardener, M. (2017). *Beginning R: The statistical programming language*, John Wiley & Sons.
3. Kabacoff, R.I. (2015). *R in Action: Data Analysis and Graphics in R*, Manning Publications.