

Statistical Learning with R

Dibyajyoti Bora and Sahana Bhattacharjee

Kalyani Publishers, 2025, India, ISBN: 978-93-6440-364-1

Brijesh P Singh

Statistical Learning with R by Dibyajyoti Bora and Sahana Bhattacharjee is a cornerstone text for those entering the field of Statistics and data science. The book provides a practical and accessible introduction to statistical learning techniques, emphasising their implementation in R. The book covers a wide array of topics including logistic/linear regression, classification, clustering, non-parametric methods, time series analysis, and survival analysis in addition to descriptive statistics. Each chapter includes theoretical explanations, real-world examples, and R-based tutorials for hands-on learning.

The R software and its code play a key role in data analysis and statistical modelling. The book is an accessible and comprehensive guide for beginners in statistical computing and data analysis using R. The book is structured across 23 well-organised chapters, starting from the foundational concepts of R programming and extending to advanced statistical methodologies. The progression of topics is instructionally effective, making it suitable for self-learners, undergraduate students and early-stage researchers in statistics and data science.

Chapters 1 to 6 introduce readers to R programming environment, particularly R Studio, and cover essential concepts such as mathematical operations, construction of different types of vectors (numeric, character, logical, categorical), sequences, lists, matrices, determinants, and data frames. These chapters serve as a solid grounding in R, presented in simple language that avoids overwhelming the reader.

Chapters 7 and 8 shift focus to data summarisation through tabular formats (one-way, two-way, and multi-way tables), and graphical representation techniques including bar diagrams, box plots, pie charts, histograms, frequency polygons, and ogives. These visual tools are important for effective data communication, and the inclusion of formatting options in R enhances the practical utility of these chapters. Chapter 9 explores descriptive statistics, such as measures of central tendency (mean, median, mode) and dispersion (range, variance, standard deviation), with clear R implementation, helping readers to perform basic statistical analysis with confidence.

Chapter 10 introduces probability distributions including all discrete distributions such as Bernoulli, Binomial, Poisson, Hyper-geometric, Geometric, Negative Binomial as well as continuous distributions such as Normal, Uniform, Exponential, Weibull, Multinomial distributions and many more. The inclusion of R code to demonstrate each distribution makes theoretical concepts more approachable and application oriented.

Chapters 11 to 13 delve into fundamentals of statistical inference. Key topics covered include hypothesis testing, types of errors, test statistics for both small and large samples, and application of Student's t and F distributions in hypothesis testing scenarios, again supported with ample R code.

Chapter 14 covers correlation and regression analysis including scatter plots, correlation coefficient, and regression equations. The examples are clearly explained, making complex concepts understandable for beginners. Chapter 15 is devoted to ANOVA (Analysis of Variance) and the Design of Experiments. Topics such as one-way and two-way ANOVA, randomised block design, Latin square design, factorial experiments, split-plot designs, and incomplete block designs are covered comprehensively. The examples provided with R syntax ensure a practical understanding of these experimental methods.

Chapters 16 and 17 focus on non-parametric tests and Chi-square test including one-sample, two-sample, and k -sample tests. These chapters are especially useful for scenarios where data assumptions are not met, and the corresponding R code enhances comprehension. Chapter 18 introduces time series analysis, covering data transformation, trend, and seasonal variation analysis, and forecasting techniques using ARMA, ARIMA and SARIMA models. The application of R for time series forecasting is particularly well-explained and practically oriented.

Chapter 19 discusses survival analysis, including survival and hazard functions, different types of censoring, and both parametric and non-parametric approaches. The inclusion of the Cox proportional hazard model shows the intent of authors to extend the reach of the book to health sciences and related fields. Chapters 20 and 21 provide concise but effective coverage of logistic regression, principal component analysis (PCA), and factor analysis which are the key tools in multivariate statistical analysis under the environment of R. Finally, Chapters 22 and 23 offer a hands-on guide to writing user-defined functions and creating loops in R (for, while, and repeat), rounding off the understanding of the reader with essential programming constructs.

I would like to mention here the strengths of the book in terms of following points:

1. **Accessibility:** The book assumes only a basic understanding of statistical tools, making it approachable for readers without advanced mathematical backgrounds. The authors prioritise intuition over rigor, using clear explanations and visualisations to demystify complex concepts.
2. **Practical Focus:** Each chapter concludes with R labs that guide readers through implementing techniques using real datasets. These tutorials are praised for their clarity and relevance, bridging theory and application.
3. **Engaging Visuals:** Colour graphics enhance understanding, and datasets used in variety of examples make the material relatable across disciplines.

Apart from the above strengths, the book has some limitations. It does not cover inflated, truncated, size and length biased distributions and various kinds of censoring schemes. Other kinds of correlations should have also been discussed to enhance the readability of the book.

Overall, *Statistical Learning with R* is a commendable introductory text that balances theoretical explanation with hands-on coding in R. The book stands out in its clarity of presentation and breadth of topics. By incorporating practical examples and structured R scripts for each concept, authors have successfully created a learning resource that is both informative and appealing. It is particularly useful for beginners in R and statistical analysis and can also serve as a supplementary text in undergraduate courses in statistics, data science, and related disciplines. The book is highly recommended for students, early-career researchers and professionals who wish to build a solid foundation in statistical learning using R. The best feature of the book is its choice of examples, which are outstanding and should appeal especially to the readers. Its accessibility and comprehensive scope make it a valuable addition to any academic or personal library.

Brijesh Pratap Singh
Banaras Hindu University
Varanasi