

Descriptive and Inferential Statistics Using R

**Hands-On Examples with
Practice Exercises and Solutions**

Dr. Preeti Dabas



Sultan Chand & Sons

Descriptive and Inferential Statistics Using R

Descriptive and Inferential Statistics Using R

Dr. Preeti Dabas

Assistant Professor (Economics)

SRM University, Delhi-NCR, Sonapat



SULTAN CHAND & SONS[®]

Educational Publishers

New Delhi

SULTAN CHAND & SONS®

Educational Publishers

23, Daryaganj, New Delhi-110002

Phones : 011-23281876, 23266105, 41625022 (*Showroom & Shop*)

011-23247051, 40234454 (*Office*)

E-mail : sultanchand74@yahoo.com; info@sultanchandandsons.com

Fax : 011-23266357; Website : www.sultanchandandsons.com

First Edition: 2024

ISBN: 978-93-91820-80-0 (TC-1307)

Price: ₹ 450.00

EVERY GENUINE COPY OF THIS BOOK HAS A HOLOGRAM



In our endeavour to protect you against counterfeit/fake books, we have pasted a copper hologram over the cover of this book. The hologram displays the full visual image, unique 3D multi-level, multi-colour effects of our logo from different angles when tilted or properly illuminated under a single light source, such as 3D depth effect, kinetic effect, pearl effect, gradient effect, trailing effect, emboss effect, glitter effect, randomly sparking tiny dots, micro text, laser numbering, etc.

A fake hologram does not display all these effects.

Always ask the bookseller to put his stamp on the first page of this book.

All Rights Reserved: No part of this book, including its style and presentation, may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopying, recording or otherwise without the prior written consent of the Publishers. Exclusive publication, promotion and distribution rights reserved with the Publishers.

Warning: The doing of an unauthorised act in relation to a copyright work may result in both civil claim for damages and criminal prosecution.

Special Note: Photocopy or xeroxing of educational books without the written permission of Publishers is illegal and against Copyright Act. Buying and selling of pirated books is a criminal offence. Publication of key to this is strictly prohibited.

General: While every effort has been made to present authentic information and avoid errors, the author and the publishers are not responsible for the consequences of any action taken on the basis of this book.

Limits of Liability/Disclaimer of Warranty: The publisher and the author make no representation or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation warranties of fitness for a particular purpose. No warranty may be created or extended by sales or promotional materials. The advice and strategies contained herein may not be suitable for every situation. This work is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If professional assistance is required, the services of a competent professional person should be sought. Neither the publisher nor the author shall be liable for damage arising herefrom.

Disclaimer: The publisher have taken all care to ensure highest standard of quality as regards typesetting, proofreading, accuracy of textual material, printing and binding. However, they accept no responsibility for any loss occasioned as a result of any misprint or mistake found in this publication.

Author's Acknowledgement: The writing of a Textbook always involves creation of a huge debt towards innumerable author's and publications. We owe our gratitude to all of them. We acknowledge our indebtedness in extensive footnotes throughout the book. If, for any reason, any acknowledgement has been left out we beg to be excused. We assure to carry out correction in the subsequent edition, as and when it is known.

Foreword



Prof. S.P. Gupta
Management Consultant

Ex. Head and Dean,
Faculty of Management Studies
University of Delhi, Delhi-7

“Statistics is the grammar of Science.”

Karl Pearson (English Mathematician & Biostatistician, 1857-1936)

It is with great pleasure and enthusiasm that I contribute the foreword to Preeti Dabas’s insightful and comprehensive book, *‘Descriptive and Inferential Statistics using R’*. In a world inundated with data, it is imperative that we possess a solid understanding of appropriate statistical tests to extract meaningful insights and answer crucial research questions.

The book serves as a valuable guide, taking readers on a journey of statistical thinking and analysis. It unfolds in four distinct parts, establishing a robust foundation for statistical analysis. Part 0 introduces essential concepts and RStudio functions, Part I covers univariate analysis, Part II explores bivariate analysis, and Part III delves into multivariate analysis. The author’s guidance equips readers to confidently navigate diverse research scenarios and draw meaningful conclusions, making this book an invaluable resource for both learners and practitioners seeking statistical proficiency. By integrating RStudio illustrations and interpretations, the author bridges the gap between theory and practice, empowering readers to apply statistical analyses with confidence.

This structured approach provides readers with a comprehensive understanding of Statistics from both descriptive and inferential perspectives and to perform the same in R. They will learn how to effectively represent data, explore its characteristics, and draw meaningful insights. Additionally, they will be equipped with the knowledge and techniques required to make informed decisions, test hypotheses, analyse relationships, and assess the magnitude of effects.

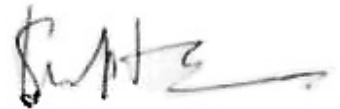
What sets this book apart is its emphasis on practical application and clarity of explanation. The author successfully strikes a balance between accessibility and depth, making it suitable for individuals at various stages of their academic or research journey. Whether pursuing higher studies, embarking on research endeavours, or seeking self-guided learning, readers will find this book to be a valuable companion.

I must emphasise that this book is not merely a compilation of statistical formulae or a technical manual for using R software. Instead, it is a holistic resource that instils statistical thinking and enables readers to apply appropriate tests, analyse data, and interpret results accurately. While the book primarily focuses on data analysis and hypothesis testing, it also provides supplementary resources, including R scripts, datasets, practice questions along with answers, and a glossary of key terms and R functions used in the book.

This book is a testament to the author's expertise and passion for Statistics. Through her clear explanations, illustrative examples, and meticulous guidance, she equips readers with the necessary skills to navigate the intricate world of statistical analysis. I have no doubt that this book will be a valuable asset for any individual seeking a comprehensive understanding of Descriptive and Inferential Statistics in research. I wholeheartedly endorse this work and congratulate the author on her remarkable achievement.

May Statistics guide your journey!

Best Regards.



Prof. S.P. Gupta

Foreword



Dr. Neeraj Kaushik

Associate Professor

Head, Department of Business Administration
National Institute of Technology, Kurukshetra

“Some books are to be tasted, others to be swallowed, and some few to be chewed and digested”.

Francis Bacon (English Philosopher, 1561-1626)

Why one more book on Statistics? Well, Statistics has become an integral part of our lives, influencing everything from the decisions we make to the policies we create. In today’s data-driven world, the ability to analyse and interpret data is a fundamental skill that empowers individuals and organisations alike. In this era of abundant data, understanding Statistics is not just a valuable asset; it is a necessity.

It is with great pleasure that I introduce you to ‘*Descriptive and Inferential Statistics using R*’, a comprehensive guide that equips readers with the necessary tools to navigate the realm of Statistics and harness the power of the statistical programming language, R. This book provides a practical and inclusive approach to mastering statistical analysis. From univariate to bivariate to multivariate analysis, the content delves into various statistical methods and their interpretation.

Each research problem is discussed in a methodical manner, starting with the essential components of descriptive statistics which include tabular representation, graphical representation, measures of central tendency, and measures of dispersion, bringing life to the raw data followed by inferential statistics covering hypothesis testing, regression analysis, post-hoc tests, and effect size. What sets this book apart is its practical approach, where the concepts are illustrated using R, ensuring a seamless integration of theory and application. The reader will not only learn how to interpret the output generated in R but also acquire the skills to effectively report and communicate your results.

With access to readymade sample R scripts and hypothetical datasets through GitHub, hands-on examples using R, and a detailed guide for interpreting and reporting results, this book equips readers with the skills necessary to effectively analyse and interpret data. The incorporation of practice questions that span across disciplines, with thorough answers, addressing anticipated queries, and a convenient summarised table of descriptive and inferential statistics further enhance the learning experience.

Whether you are a student, researcher, or professional, this book serves as an invaluable resource for honing statistical proficiency and leveraging the power of R for data analysis and decision-making.

The author's dedication to providing a clear and organised framework for learning Statistics is evident throughout the book. Readers can confidently navigate through each chapter, knowing that they are building a solid foundation and acquiring the skills necessary to tackle real-world data analysis challenges. I applaud the author for her foresight in recognising the importance of Statistics in multiple disciplines and for creating this invaluable resource.

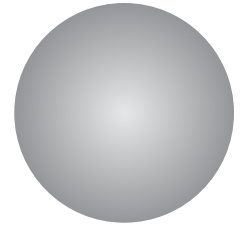
'Descriptive and Inferential Statistics using R' is not just a book; it is a gateway to statistical literacy and analytical prowess, empowering readers to extract valuable insights from data and make informed decisions. As you delve into its pages, you will gain the skills and knowledge to navigate the intricacies of statistical analysis, empowering you to uncover meaningful insights and make informed decisions. Embrace the possibilities that lie ahead as you discover the beauty of Statistics and harness the potential of R.

Happy Learning!

A handwritten signature in black ink that reads "Neeraj". The signature is stylized with a long horizontal line underneath the name.

Dr. Neeraj Kaushik

Preface



“If you torture the data long enough, it will confess to anything.”

Ronald Coase (British Economist & Nobel Prize Winner, 1991)

Well said! Hence, the need for understanding appropriate statistical tests to answer the research questions. The beauty lies in the fact that only one particular test can be applied in a given scenario depending upon the number and type of variable(s) involved. This book is an attempt to help the readers to develop statistical thinking by explaining how to deal with a research problem in a detailed and comprehensive manner which includes understanding the variable type, choosing the appropriate descriptive and inferential statistics, performing the same in R, and finally interpreting and reporting the results. In this book, all the research scenarios are divided under three broad heads: one variable scenario, two variables scenario and more than two variables scenario.

Organisation of the book

The book is divided into four parts in order to present the contents in a systematic way.

Part **zero** serves as a foundation for building the base of key terminologies and concepts essential for conducting data analysis. It comprises three chapters, each aimed to familiarise readers with terminologies used frequently throughout the book and to foster a sense of ease for analysing and interpreting the data. Chapter **one** gives a brief introduction to Statistics at the outset. It then builds an understanding of different terms used while entering and analysing the data followed by four levels of measurement: nominal, ordinal, interval and ratio. It further delves into the terms used while interpreting the data followed by differentiating between absolute change, relative change, percentage change, percentage point and growth rate. Furthermore, it also gives the reader an insight of the term ‘significance’ which is heavily used while testing hypotheses. Chapter **two** deals with understanding the meaning and formulation of hypothesis testing and the related concepts: null and alternative hypotheses, one and two-tailed tests, errors associated with it, level of significance, confidence level, confidence interval, p -value, power of the test and test statistic in detail. It thoroughly discusses the key assumptions of statistical tests, with an emphasis on the assumption of normality. It also differentiates between statistical significance and practical significance along with elucidating the meaning of effect size. It furthermore discusses decision rules for rejecting or retaining the null hypothesis, leading to a conclusion about the population based on the sample. Chapter **three** takes you on the journey of navigating through RStudio. The main aim of this chapter is to assist readers in downloading and installing the software, and to get familiar with the software and perform key functions in RStudio which include: how to import data, how to code, how to install and load packages, how to apply mathematical and statistical operations, how to plot graphs, how to export the results and last but not the least how to cite the software in research work.

Part **I** deals with the univariate analysis, that is, analysis of research problems having only one variable (one dependent variable and no independent variable). Chapter **four** covers all such possible scenarios.

Part **II** deals with the bivariate analysis, that is, analysis of research problems having two variables (one dependent variable and one independent variable). It consists of **five** chapters. Chapters **five, six, seven and eight** cover the research problems having two variables when dependent variable is binary, nominal, ordinal and metric respectively. Chapter **nine** covers the research problems when the two variables are paired.

Part **III** deals with the multivariate analysis, that is, analysis of research problems having more than two variables (one dependent variable and more than one independent variable), which is often going to be the case in real-life scenarios. It consists of four chapters. Chapters **ten, eleven, twelve and thirteen** cover the research problems having more than two variables when dependent variable is binary, nominal, ordinal and metric respectively.

The sequence of discussion for a given research problem would be as follows:

Given a research problem, we will first identify the variable(s) and then get an impression of our sample and present the same in tabular form followed by its graphical representation. Then we will try to answer the research question by applying appropriate measures of central tendency and dispersion (if applicable) followed by statistical test to know what the sample tells about the population followed by the post-hoc analysis in case our independent variable has more than two categories and then we shall compute the effect size. In addition, we will also regress dependent variable on independent variable(s) to see the magnitude of the impact. We will then learn to perform the same in RStudio followed by the key interpretations. Lastly, we shall conclude each scenario by learning how to report the results in research work.

For whom this book is?

This book caters to individuals seeking a comprehensive and lucid understanding of the appropriate application of descriptive and inferential statistics in various research problems. It is particularly useful for those pursuing higher studies, budding researchers, and self-learners across diverse disciplines. This book is good to go for those who fear or always wanted to skip the formulae but have an inquisitive mind and are curious to get an answer to their research questions.

What is included and what is not?

Please bear in mind that this book is neither a book on Statistics nor on Research Methodology nor on R software. Rather, it focuses on applications of Statistics in research using R software. The primary objective of this book is to provide readers with a seamless and user-friendly approach to navigating through various statistical tests.

The scope of this book is limited to data analysis only and does not include text analysis. Furthermore, it covers analysis on cross-sectional data and excludes that on time-series data. It includes preliminary concepts related to hypothesis testing in detail followed by all possible tests but derivation of formulae is not touched upon. Research problems having more than one dependent variable are not discussed. Topics related to research methods like sampling design, sampling techniques and determination of sample size are out of scope of this book.

This book also devotes a chapter exclusively for R software so as to make the reader familiar and comfortable with its interface. Having said that, it is expected from the reader to explore more about R through its tutorial since this book stresses more on choosing the appropriate statistical test and

analysing and interpreting data. An illustration for every research problem is given separately where the analysis is done using R software.

Pre-requisites for this book

It is assumed by the author that the reader is familiar with the elementary concepts of Statistics. The reader should have R software, which is a freeware, downloaded and installed in the system.¹

How to go for this book?

If anyone is starting the research journey from the scratch (given that the reader has basic knowledge of Statistics), then one should follow the same sequence as the book follows. However, if one already has or after having the background knowledge of the preliminary concepts and getting familiar with the software, can jump to any of the required chapter(s) of interest.

What you will get for your money?

- Hands-on experience with R which is a freeware
- Sample R scripts for the research scenarios discussed in this book (on GitHub)
- Hypothetical datasets used in this book (on GitHub)
- A guide to interpreting and reporting the results
- Practice questions at the end of each chapter along with their answers and explanations (wherever required)
- The book by default answers some of the questions that will eventually come in mind of an inquisitive reader
- A summarised table of all the descriptive and inferential statistics applied in different research scenarios
- Glossary of key terms
- Glossary of R functions and packages used in the book.

It is hoped that the content of the book will be beneficial and fruitful to the readers.

For any critical comments or feedback, you are welcome to reach out to the author at preetiidabas871@gmail.com.

Hope you are as excited to go ahead with this book as I am.

Best Wishes!

Preeti Dabas

Disclaimer: The datasets used in this book are purely hypothetical in nature. Hence, the results obtained or conclusions drawn based on them should not be taken literally.

¹. Chapter 3 helps to download and install the same.

Acknowledgement

Writing this book has been a transformative journey, and I am deeply grateful to everyone who has contributed to its completion. Their unwavering support and encouragement have been instrumental in shaping this work.

First and foremost, I extend my heartfelt gratitude to *Dr. Neeraj Kaushik*, Associate Professor at NIT Kurukshetra, for sharing his profound knowledge through workshops and insightful discussions on the Google group. His teachings have had a significant influence on my understanding, and I am indebted to him.

I am also deeply grateful to *Prof. DV Singh* for his invaluable support and the wealth of experiences he has generously shared throughout. His profound insights and unwavering encouragement have been pivotal in shaping my academic journey.

I am also blessed to have had the support of my teachers, who played a crucial role in my academic development and offered their guidance throughout my educational journey. I would like to give a special mention to *Dr. Anita Dagar*, whose dedication and expertise laid the foundation for my understanding of this subject.

My heartfelt appreciation goes to my late grandparents for their blessings, guiding me on this path. I am profoundly grateful to my parents and parents-in-law for providing me with the precious gift of time and their unwavering support. Their understanding and encouragement made this endeavor possible, and I am forever thankful for their love and blessings. Special thanks to my husband, *Siddharth*, whose patient listening and understanding made the journey smoother, and my daughter, *Vrinda*, who brought an abundance of luck and has illuminated my life with boundless joy. I extend my sincere gratitude to my brother-cum-mentor, *Nirmal*, and my bhabhi, *Dipti*, for their constant love and support. This book is a testament to their unconditional love and encouragement.

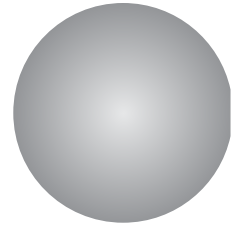
I extend my special thanks to my colleagues-cum-friends, *Dr. Minesh* and *Aashish*, for offering thoughtful suggestions to improve its quality. I would like to express my gratitude to my friends *Sakshi* and *Riya* for spending quality time and taking regular updates on the progress of work, and *Kajal* for always being there to help whenever required. I am deeply thankful for their presence in my life. My sincere thanks also goes to the creators and designers of the freeware R, which played a significant role in the development of this book.

Last but not the least, I express my heartfelt gratitude to *Sultan Chand & Sons Publishers* for believing in my work and agreeing to publish it.

Finally, I extend my thanks to all those I encountered during this journey, each of whom contributed in their unique way to support me and help me bring this book to fruition. To all these wonderful individuals, I am deeply grateful, and I will always cherish the invaluable support they have provided me. This book would not have been possible without them.

Preeti Dabas

Snapshot of the Book



| <i>S.No.</i> | <i>Chapter Name</i> | <i>Page No.</i> |
|---------------------------------|--|-----------------|
| Part 0: Getting Started | | |
| 1 | Background Concepts | 3 - 14 |
| 2 | Hypothesis Testing | 15 - 39 |
| 3 | Handshake with RStudio | 41 - 64 |
| Part I: Univariate Analysis | | |
| 4 | Univariate Analysis: Only One Variable | 69 - 107 |
| Part II: Bivariate Analysis | | |
| 5 | Bivariate Analysis: With Binary Dependent Variable | 113 - 138 |
| 6 | Bivariate Analysis: With Nominal Dependent Variable | 139 - 169 |
| 7 | Bivariate Analysis: With Ordinal Dependent Variable | 171 - 199 |
| 8 | Bivariate Analysis: With Metric Dependent Variable | 201 - 227 |
| 9 | Bivariate Analysis: With Two Paired Variables | 229 - 259 |
| Part III: Multivariate Analysis | | |
| 10 | Multivariate Analysis: With Binary Dependent Variable | 265 - 287 |
| 11 | Multivariate Analysis: With Nominal Dependent Variable | 289 - 313 |
| 12 | Multivariate Analysis: With Ordinal Dependent Variable | 315 - 336 |
| 13 | Multivariate Analysis: With Metric Dependent Variable | 337 - 363 |

Contents

| | |
|-----------------------------|-------|
| <i>Foreword</i> | v |
| <i>Foreword</i> | vii |
| <i>Preface</i> | ix |
| <i>Acknowledgement</i> | xiii |
| <i>Snapshot of the Book</i> | xiv |
| <i>Contents</i> | xv |
| <i>List of Tables</i> | xix |
| <i>List of Figures</i> | xxi |
| <i>Abbreviations</i> | xxv |
| <i>Note to Readers</i> | xxvi |
| <i>About the Author</i> | xxvii |

Part 0: Getting Started

| | |
|--|----------------|
| 1. Background Concepts | 3 - 14 |
| 1.1 Introduction to Statistics | 3 |
| 1.2 Key Terms Related to Data Entry | 4 |
| 1.3 Levels of Measurement: Nominal, Ordinal, Interval and Ratio | 5 |
| 1.3.1 Non-Metric/ Categorical/ Qualitative Variables | 6 |
| 1.3.2 Metric/ Continuous/ Quantitative Variables | 7 |
| 1.3.3 Deciding the Level of Measurement of a Variable | 8 |
| 1.4 Understanding the Differences among Differences | 9 |
| 1.5 Meaning of the Word ‘Significance’ | 10 |
| 1.5.1 Same Absolute Difference but Different in Relative Terms | 10 |
| 1.5.2 Same in Relative Terms but Different in Absolute Differences | 11 |
| <i>Key Takeaways from this Chapter</i> | 11 |
| <i>Test Yourself</i> | 12 |
| 2. Hypothesis Testing | 15 - 39 |
| 2.1 Hypothesis Testing: Meaning, Types and the Associated Errors | 15 |
| 2.1.1 Meaning and Types of Hypothesis | 15 |
| 2.1.2 Errors Associated with Hypothesis Testing | 16 |
| 2.2 Key Concepts Related to Hypothesis Testing | 18 |

| | | |
|--|---|-----------------|
| 2.2.1 | Probability Value (p -Value) | 19 |
| 2.2.2 | Level of Significance | 19 |
| 2.2.3 | Connecting the Dots: p -Value and Level of Significance | 19 |
| 2.2.4 | Does the Insufficiency of Evidence to Reject Null Hypothesis Imply its Acceptance? | 20 |
| 2.2.5 | Confidence Level | 21 |
| 2.2.6 | Confidence Interval | 21 |
| 2.2.7 | Power of a Test | 22 |
| 2.3 | Difference between Statistical Significance and Practical Significance | 23 |
| 2.4 | Hypothesis Testing: Formulation and Decision Rules | 23 |
| 2.4.1 | Formulation of Hypothesis | 24 |
| 2.4.2 | Decision Rules in Hypothesis Testing | 25 |
| 2.5 | Key Assumptions of Statistical Tests | 26 |
| 2.6 | Concept of Normality | 28 |
| 2.6.1 | The Importance of Assessing Data Normality | 29 |
| 2.6.2 | Methods for Testing Normality of Data | 29 |
| 2.6.3 | Dealing with Non-Normality | 34 |
| | <i>Key Takeaways from this Chapter</i> | 36 |
| | <i>Test Yourself</i> | 38 |
| 3. | Handshake with RStudio | 41 - 64 |
| 3.1 | Downloading and Installing the Software | 41 |
| 3.2 | Interface of RStudio | 43 |
| 3.2.1 | 'Script Editor' Pane | 44 |
| 3.2.2 | 'Console' Pane | 45 |
| 3.2.3 | 'Files/ Plots/ Packages/ Help' Pane | 46 |
| 3.2.4 | 'Environment/ History/ Connections/ Tutorials' Pane | 51 |
| 3.3 | Let's Code | 52 |
| 3.4 | Citing the Software in Research Work | 56 |
| | <i>Key Takeaways from this Chapter</i> | 57 |
| | <i>Test Yourself</i> | 58 |
| | <i>Appendix A: Cheat Sheet for Basic Operations</i> | 60 |
| <hr/> Part I: Univariate Analysis <hr/> | | |
| 4. | Univariate Analysis: Only One Variable | 69 - 107 |
| 4.1 | When the Only Variable is a Binary Variable | 70 |
| 4.2 | When the Only Variable is a Nominal Variable | 77 |
| 4.3 | When the Only Variable is an Ordinal Variable | 87 |
| 4.4 | When the Only Variable is a Metric Variable | 94 |
| | <i>Key Takeaways from this Chapter</i> | 103 |
| | <i>Test Yourself</i> | 104 |
| | <i>Appendix B: Cheat Sheet for Univariate Analysis</i> | 107 |

Part II: Bivariate Analysis

| | |
|---|------------------|
| 5. Bivariate Analysis: With Binary Dependent Variable | 113 - 138 |
| 5.1 Binary = f (Binary) | 114 |
| 5.2 Binary = f (Nominal/ Ordinal) | 121 |
| 5.3 Binary = f (Metric) | 130 |
| <i>Key Takeaways from this Chapter</i> | 135 |
| <i>Test Yourself</i> | 136 |
| 6. Bivariate Analysis: With Nominal Dependent Variable | 139 - 169 |
| 6.1 Nominal = f (Binary) | 140 |
| 6.2 Nominal = f (Nominal/ Ordinal) | 148 |
| 6.3 Nominal = f (Metric) | 158 |
| <i>Key Takeaways from this Chapter</i> | 166 |
| <i>Test Yourself</i> | 167 |
| 7. Bivariate Analysis: With Ordinal Dependent Variable | 171 - 199 |
| 7.1 Ordinal = f (Binary) | 172 |
| 7.2 Ordinal = f (Nominal/ Ordinal) | 182 |
| 7.3 Ordinal = f (Metric) | 192 |
| <i>Key Takeaways from this Chapter</i> | 197 |
| <i>Test Yourself</i> | 198 |
| 8. Bivariate Analysis: With Metric Dependent Variable | 201 - 227 |
| 8.1 Metric = f (Binary) | 202 |
| 8.2 Metric = f (Nominal/ Ordinal) | 211 |
| 8.3 Metric = f (Metric) | 220 |
| <i>Key Takeaways from this Chapter</i> | 224 |
| <i>Test Yourself</i> | 225 |
| 9. Bivariate Analysis: With Two Paired Variables | 229 - 259 |
| 9.1 Two Paired Binary Variables | 230 |
| 9.2 Two Paired Nominal Variables | 236 |
| 9.3 Two Paired Ordinal Variables | 242 |
| 9.4 Two Paired Metric Variables | 249 |
| <i>Key Takeaways from this Chapter</i> | 255 |
| <i>Test Yourself</i> | 256 |
| <i>Appendix C: Cheat Sheet for Bivariate Analysis</i> | 258 |

Part III: Multivariate Analysis

| | |
|--|------------------|
| 10. Multivariate Analysis: With Binary Dependent Variable | 265 - 287 |
| 10.1 Binary = f (Categorical, Categorical) | 266 |
| 10.2 Binary = f (Categorical, Continuous) | 275 |
| 10.3 Binary = f (Continuous, Continuous) | 281 |

| | |
|---|------------------|
| <i>Key Takeaways from this Chapter</i> | 286 |
| <i>Test Yourself</i> | 286 |
| 11. Multivariate Analysis: With Nominal Dependent Variable | 289 - 313 |
| 11.1 Nominal = f (Categorical, Categorical) | 290 |
| 11.2 Nominal = f (Categorical, Continuous) | 299 |
| 11.3 Nominal = f (Continuous, Continuous) | 305 |
| <i>Key Takeaways from this Chapter</i> | 311 |
| <i>Test Yourself</i> | 311 |
| 12. Multivariate Analysis: With Ordinal Dependent Variable | 315 - 336 |
| 12.1 Ordinal = f (Categorical, Categorical) | 316 |
| 12.2 Ordinal = f (Categorical, Continuous) | 324 |
| 12.3 Ordinal = f (Continuous, Continuous) | 329 |
| <i>Key Takeaways from this Chapter</i> | 334 |
| <i>Test Yourself</i> | 334 |
| 13. Multivariate Analysis: With Metric Dependent Variable | 337 - 363 |
| 13.1 Metric = f (Categorical, Categorical) | 338 |
| 13.2 Metric = f (Categorical, Continuous) | 347 |
| 13.3 Metric = f (Continuous, Continuous) | 353 |
| <i>Key Takeaways from this Chapter</i> | 360 |
| <i>Test Yourself</i> | 360 |
| <i>Summary of Descriptive Statistics</i> | 365 - 367 |
| <i>Summary of Inferential Statistics</i> | 369 - 371 |
| <i>Glossary of Key Terms</i> | 373 - 380 |
| <i>Glossary of R Functions</i> | 381 - 398 |
| <i>Appendix D: Probability, Odds, log Odds and Odds Ratio</i> | 399 - 406 |
| <i>Appendix E: Growth Models</i> | 407 - 408 |
| <i>Appendix F: Common Errors and Debugging Tips</i> | 409 - 410 |
| <i>Answers to 'Test Yourself'</i> | 411 - 435 |
| <i>References</i> | 437 - 440 |
| <i>Index</i> | 441 - 443 |

e-Contents



Datasets



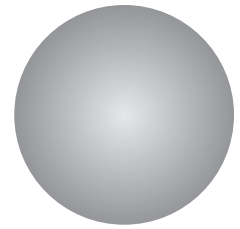
R Scripts

List of Tables

| | |
|---|-----|
| 2.1: Four Possibilities while Drawing Inferences about Population on the Basis of Sample | 16 |
| 2.2: Decision Outcomes in Legal Trials | 17 |
| 2.3: Decision Outcomes in Medical Diagnostics | 18 |
| 2.4: Decision Outcomes in Hypothesis Testing | 18 |
| 2.5: Hypothesis Testing: Type I and Type II Errors, Level of Confidence and Power of a Test | 22 |
| 2.6: Box-Cox Transformation | 36 |
| 3.1: Keyboard Shortcuts | 52 |
| 4.1: Does Increasing Sample Size Decreases p -Value | 76 |
| 4.2: Calculating Odds and \log Odds in Univariate Ordinal Variable Scenario | 93 |
| 4.3: Tabular Representation for Univariate Metric Variable | 99 |
| 5.1: Derivation of Expected Values from Observed Values for Binary = f (Binary) | 115 |
| 5.2: Tabular Representation for Binary = f (Binary) | 120 |
| 5.3: Derivation of Expected Values from Observed Values for Binary = f (Nominal) | 123 |
| 5.4: Tabular Representation for Binary = f (Nominal) | 128 |
| 6.1: Derivation of Expected Values from Observed Values for Nominal = f (Binary) | 141 |
| 6.2: Tabular Representation for Nominal = f (Binary) | 146 |
| 6.3: Derivation of Expected Values from Observed Values for Nominal = f (Nominal) | 149 |
| 6.4: Tabular Representation for Nominal = f (Nominal) | 156 |
| 7.1: Tabular Representation for the Scenario When Ordinal = f (Binary) | 181 |
| 7.2: Tabular Representation for Ordinal = f (Nominal) | 191 |
| 8.1: Tabular Representation for Metric = f (Binary) | 208 |
| 8.2: Tabular Representation for Metric = f (Nominal) | 218 |
| 8.3: Tabular Representation for Metric = f (Metric) | 223 |
| 9.1: Tabular Representation for Two Paired Binary Variables | 234 |
| 9.2: Tabular Representation for Two Paired Nominal Variables | 241 |
| 9.3: Tabular Representation for Two Paired Ordinal Variables | 248 |
| 9.4: Tabular Representation for Two Paired Metric Variables | 253 |
| 10.1: Maths Behind Regression for Binary = f (Categorical, Categorical) | 271 |
| 10.2: Tabular Representation for Binary = f (Categorical, Categorical) | 272 |
| 10.3: Maths Behind the Odds for Binary = f (Categorical, Categorical) | 274 |

| | |
|--|-----|
| 11.1: Tabular Representation for Nominal = $f(\text{Categorical}, \text{Categorical})$ | 297 |
| 12.1: Tabular Representation for Ordinal = $f(\text{Categorical}, \text{Categorical})$ | 322 |
| 13.1: Tabular Representation for Metric = $f(\text{Categorical}, \text{Categorical})$ | 343 |
| D.1: Probability-Odds- <i>log</i> Odds Conversion | 402 |

List of Figures



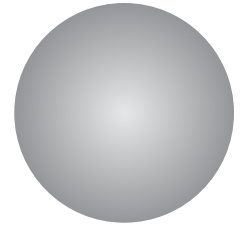
| | |
|--|----|
| 1.1: Placing Respondents and Variables in a Spreadsheet while Entering Data | 5 |
| 1.2: Four Levels of Measurement | 6 |
| 1.3: Decision Tree for the Level of Measurement of a Variable | 8 |
| 2.1: COVID Report: Usage of False Positive and False Negative | 16 |
| 2.2: Plot of the Claimed Value with respect to Lower and Upper Limits of Confidence Interval | 25 |
| 2.3: Plot of Normal Distribution | 28 |
| 2.4: Histogram: Normal and Non-Normal Distributions | 30 |
| 2.5: Box Plots: Normal and Non-Normal Distributions | 31 |
| 2.6: Normality Check through QQ Plot | 31 |
| 3.1: Downloading RStudio | 41 |
| 3.2: Choosing Appropriate Installer | 42 |
| 3.3: Downloading R | 42 |
| 3.4: Interface of RStudio | 43 |
| 3.5: Switching 'Console Pane' to Left/Right | 43 |
| 3.6: Customising Pane Layout | 44 |
| 3.7: Sample of R Script | 45 |
| 3.8: Output Display on 'Console' | 45 |
| 3.9: Working in 'Console' - I | 46 |
| 3.10: Working in 'Console' - II | 46 |
| 3.11: 'Plots' Pane | 47 |
| 3.12: 'Package' Pane | 47 |
| 3.13: Installing a Package through 'Console' - I | 48 |
| 3.14: Installing a Package through 'Console' - II | 48 |
| 3.15: Installing a Package through 'Console' - III | 48 |
| 3.16: Installing a Package through 'Console' - IV | 49 |
| 3.17: Installing a Package through 'Package' Pane - I | 49 |
| 3.18: Installing a Package through 'Package' Pane - II | 49 |
| 3.19: 'Help' Pane | 50 |
| 3.20: Cheat Sheets through 'Help' Menu | 51 |
| 3.21: Executing the Command from 'History' Pane | 52 |
| 3.22: Keyboard Shortcuts Help | 52 |
| A.1: Cheat Sheet: Source Editor and Tab Panes | 60 |
| A.2: Cheat Sheet: Package Development | 60 |

| | |
|--|-----|
| A.3: Cheat Sheet: Visual Editor | 61 |
| A.4: Cheat Sheet: Keyboard Shortcuts | 61 |
| A.5: Cheat Sheet: Different Functions - I | 62 |
| A.6: Cheat Sheet: Different Functions - II | 63 |
| A.7: Cheat Sheet: Different Functions of 'Mosaic' Package | 64 |
| 4.1: A Brief Summary of Possible Scenarios Under Univariate Analysis | 70 |
| 4.2: Graphical Representation for Univariate Binary Variable | 76 |
| 4.3: Graphical Representation for Univariate Nominal Variable | 86 |
| 4.4: Graphical Representation for Pairwise Comparison | 87 |
| 4.5: Graphical Representation for Univariate Ordinal Variable | 94 |
| 4.6: Understanding the Box Plot | 95 |
| 4.7: Normality Check of a Metric Variable using QQ Plot | 100 |
| 4.8: Graphical Representation for Univariate is a Metric Variable | 100 |
| 4.9: Plot of One Sample t -Test Result for Univariate Metric Variable | 102 |
| B.1: Univariate Graphical Representation through 'Mosaic' Package | 107 |
| 5.1: A Brief Summary of Possible Scenarios Under Bivariate Analysis with Binary Dependent Variable | 113 |
| 5.2: Graphical Representation for Binary = f (Binary) | 120 |
| 5.3: Plot of χ^2 Test Results for Binary = f (Binary) | 121 |
| 5.4: Graphical Representation for Binary = f (Nominal) | 129 |
| 5.5: Plot of χ^2 Test Results for Binary = f (Nominal) | 129 |
| 5.6: Graphical Representation for Binary = f (Metric) | 135 |
| 6.1: A Brief Summary of Possible Scenarios Under Bivariate Analysis with Nominal Dependent Variable | 139 |
| 6.2: Graphical Representation for Nominal = f (Binary) | 147 |
| 6.3: Plot of χ^2 Test Results for Nominal = f (Binary) | 147 |
| 6.4: Graphical Representation for Nominal = f (Nominal) | 156 |
| 6.5: Plot of χ^2 Test Results for Nominal = f (Nominal) | 157 |
| 6.6: Graphical Representation for Nominal = f (Metric) | 165 |
| 6.7: Graphical Representation for effects of IDV (Metric) on the Probabilities of Different Categories of the DV (Nominal) | 166 |
| 7.1: A Brief Summary of Possible Scenarios Under Bivariate Analysis with Ordinal Dependent Variable | 171 |
| 7.2: Graphical Representation for Ordinal = f (Binary) | 181 |
| 7.3: Graphical Representation for Ordinal = f (Nominal) | 191 |
| 7.4: Graphical Representation for Ordinal = f (Metric) | 196 |
| 8.1: A Brief Summary of Possible Scenarios Under Bivariate Analysis with Metric Dependent Variable | 201 |
| 8.2: Three Possibilities Under Metric = f (Binary) | 203 |
| 8.3: Graphical Representation for Metric = f (Binary) | 209 |
| 8.4: Plot of Two Sample t -Test Results for Metric = f (Binary) | 210 |
| 8.5: Three Possibilities Under Metric = f (Nominal) | 212 |
| 8.6: Graphical Representation for Metric = f (Nominal) | 219 |
| 8.7: Graphical Representation for Metric = f (Metric) | 223 |

| | | |
|-------|---|-----|
| 9.1: | A Brief Summary of Possible Scenarios Under Bivariate Analysis with Paired Variables | 230 |
| 9.2: | Graphical Representation for Two Paired Binary Variables | 235 |
| 9.3: | Graphical Representation for Two Paired Nominal Variables | 241 |
| 9.4: | Graphical Representation for Two Paired Ordinal Variables | 248 |
| 9.5: | Normality Check of Metric Variables using QQ Plot | 253 |
| 9.6: | Graphical Representation for Two Paired Metric Variables | 254 |
| 9.7: | Plot of Paired t -Test Results for Two Paired Metric Variables | 255 |
| C.1: | Cheat Sheet: Bivariate Graphical Representation through 'Mosaic' Package - I | 258 |
| C.2: | Cheat Sheet: Bivariate Graphical Representation through 'Mosaic' Package - II | 259 |
| 10.1: | A Brief Summary of Possible Scenarios Under Multivariate Analysis with Binary Dependent Variable | 265 |
| 10.2: | Graphical Representation for Binary = f (Categorical, Categorical) | 272 |
| 10.3: | Graphical Representation of the Regression Results for Binary = f (Categorical, Categorical) | 274 |
| 10.4: | Graphical Representation of the Regression Results for Binary = f (Categorical, Continuous) | 279 |
| 10.5: | Graphical Representation of Regression Results for Binary = f (Continuous, Continuous) | 284 |
| 11.1: | A Brief Summary of Possible Scenarios Under Multivariate Analysis with Nominal Dependent Variable | 289 |
| 11.2: | Graphical Representation for Nominal = f (Categorical, Categorical) | 297 |
| 11.3: | Graphical Representation of Regression Results for Nominal = f (Categorical, Categorical) | 298 |
| 11.4: | Graphical Representation of Regression Results for Nominal = f (Categorical, Continuous) | 304 |
| 11.5: | Graphical Representation of the Regression Results for Nominal = f (Continuous, Continuous) | 309 |
| 12.1: | A Brief Summary of Possible Scenarios Under Multivariate Analysis with Ordinal Dependent Variable | 315 |
| 12.2: | Graphical Representation for Ordinal = f (Categorical, Categorical) | 322 |
| 12.3: | Graphical Representation of Regression Results for Ordinal = f (Categorical, Categorical) | 323 |
| 12.4: | Graphical Representation of Regression Results for Ordinal = f (Categorical, Continuous) | 328 |
| 12.5: | Graphical Representation of Regression Results for Ordinal = f (Continuous, Continuous) | 332 |
| 13.1: | A Brief Summary of Possible Scenarios Under Multivariate Analysis with Metric Dependent Variable | 337 |
| 13.2: | Graphical Representation for Metric = f (Categorical, Categorical) | 344 |
| 13.3: | Graphical Representation of Regression Results for Metric = f (Categorical, Categorical) | 345 |
| 13.4: | Graphical Representation of the Interaction Plots for Metric = f (Categorical, Categorical) | 346 |
| 13.5: | Graphical Representation for Metric = f (Categorical, Continuous) | 350 |
| 13.6: | Graphical Representation of Regression Results for Metric = f (Categorical, Continuous) | 352 |
| 13.7: | Graphical Representation for Metric = f (Continuous, Continuous) | 358 |

| | |
|--|-----|
| 13.8: Graphical Representation of Regression Results for Metric = f (Continuous, Continuous) | 359 |
| D.1: Probability on Scale | 401 |
| D.2: Odds on Scale | 401 |
| D.3: <i>Log</i> of Odds on Scale | 401 |
| D.4: Probability on Scale (Example) | 401 |
| D.5: Odds on Scale (Example) | 402 |
| D.6: <i>Log</i> of Odds on Scale (Example) | 402 |
| D.7: Relationship Between Probability and Odds | 403 |
| D.8: Relationship Between Odds and <i>log</i> Odds | 403 |
| D.9: Linear vs Logistic Regression | 405 |

Abbreviations



| | |
|-----------------|--------------------------------------|
| χ^2 | : Chi-Square |
| % | : Percentage |
| AIC | : Akaike Information Criterion |
| ANCOVA | : Analysis of Covariance |
| ANOVA | : Analysis of Variance |
| CF | : Cumulative Frequency |
| CI | : Confidence Interval |
| CRAN | : Comprehensive R Archive Network |
| CV | : Coefficient of Variation |
| DF | : Degree(s) of Freedom |
| DV | : Dependent Variable |
| GLM | : Generalised Linear Model |
| H_0 | : Null Hypothesis |
| H_1 | : Alternative Hypothesis |
| IDE | : Integrated Development Environment |
| IDV | : Independent Variable |
| IQR | : Inter-Quartile Range |
| KS | : Kolmogorov-Smirnov |
| LL | : Lower Limit |
| LOS | : Level of Significance |
| OR | : Odds Ratio |
| <i>p</i> -Value | : Probability Value |
| PP | : Percent-Percent |
| QQ | : Quantile-Quantile |
| R^2 | : <i>R</i> -Squared |
| SD | : Standard Deviation |
| SE | : Standard Error |
| UL | : Upper Limit |

Note to Readers

The book follows the following conventions throughout:

- R would mean RStudio.
- Bold letters are defined in Glossary appended at the end of the book.
- # indicates the author is addressing the reader while explaining R syntax. These comments are bits of text not executed by the software. For example,
this is a comment
- > represents R syntax in illustrations.
- As you explore R syntax, be aware that R is space insensitive, that is, R does not differentiate between spaces and may interpret “Variable=value” and “Variable = value” interchangeably.
- Extra spaces from the software results have been removed throughout.
- The titles of tables and graphs in the “Reporting the Results” sections are kept according to the given scenarios. However, in real research result reporting, the headings should be in terms of the variables. For example,
“Figure : Graphical Representation for Metric = f (Metric)” should be written as “Figure : Scatterplot of Relationship between Profits and Advertisement Spendings”.
- Additionally, in the graphical representation under “Reporting the Results”, the data is presented using more than one graph. However, present the data using only one of them in real research result reporting. This will ensure clarity and avoid redundancy in the presentation of results.

About the Author

Preeti Dabas began her academic journey at Motilal Nehru School of Sports, Rai (Haryana), where she completed her schooling in 2012. She furthered her education with a Bachelor's degree in Economics Honours from Miranda House, University of Delhi, in 2015. She completed her Master's degree in Economics in 2017 and earned Doctoral degree on the topic "*An Analysis of Burden of Health Care Cost on Households in Haryana: Empirical Evidences from Rohtak City*" under the guidance of *Prof. Anita Dagar* from the Department of Economics, Maharshi Dayanand University, Rohtak in February 2022.

She cleared January 2017 UGC-NET exam and was recipient of UGC Junior Research Fellowship. She also secured an all-India 77th rank in GATE Economics exam, 2021.

Her areas of interest encompass Microeconomics, Statistics, Econometrics, Data Analysis and Research Methodology. More than 15 research paper publications in reputed journals stand to her credit. She has also presented nine research papers in different national and international conferences. She also won '*Best Presenter Award*' and '*Young Researcher Award*' in International Conferences in 2020 and 2021 respectively.

She is skilled in R, MS-Excel and Jamovi. She has also completed an online course on "Bayesian Statistics – from Concepts to Data Analysis (with Honors)" from University of California, New York.

Presently, she is working as Assistant Professor (Economics) with SRM University, Delhi-NCR, Sonapat since August 2019. She also shared her knowledge and insights as a Visiting Professor at the esteemed Centre for Advanced Studies in Policy Research Foundation, Pune, from February 2021 to January 2022.

She has been the Chief Editor of '*Economy Briefs*', a quarterly magazine by the Department of Economics, SRM University, Delhi-NCR, Sonapat, since March 2021. Her editorial skills ensure the publication of insightful economic analysis and scholarly works.

Salient Features

- Hands-on experience with R, a powerful and freely available software.
- A comprehensive guide to interpreting and reporting results.
- Practice questions at the end of each chapter with answers and detailed explanations where necessary.
- A convenient summarised table presenting descriptive and inferential statistics applied in various research scenarios, for quick reference.
- Glossaries of key statistical terms and R functions.
- Anticipated questions that often arise in the mind of an inquisitive reader are addressed throughout the book.
- Availability of dummy datasets used in the book on GitHub.

Book Reviews

“*Descriptive and Inferential Statistics using R*” is an insightful guide that navigates with clarity through the hands-on applications of statistical analysis. The book begins with foundational concepts and seamlessly progresses from univariate to bivariate to multivariate analysis. With practical examples and clear explanations, the book guides readers through hypothesis testing using RStudio and interpretations of software results. An indispensable resource that merges theory and application, it is a valuable addition to the realm of statistics. I am sure the book will serve as a practical guide for students of economics, statistics and policy studies.

Prof. Ram Singh

Professor, Delhi School of Economics

The book titled “*Descriptive and Inferential Statistics Using R*” by Dr. Preeti Dabas is a gateway to statistical proficiency across disciplines. With its engaging approach, practical examples, and Rscripts, this book not only simplifies the intricacies of statistical analysis but also empowers eager learners to confidently navigate the world of data. A must-read for those seeking to unlock the potential of data-driven decision-making.

Prof. (Retd.) C.S. Sharma

Shri Ram College of Commerce, University of Delhi
Visiting Faculty at Ambedkar University, Delhi

About the Author

Dr. Preeti Dabas completed her Bachelor’s degree in Economics Honours from Miranda House, University of Delhi, in 2015. She completed her Master’s degree in Economics in 2017 and earned Doctoral degree under the guidance of Dr. Anita Dagar in February 2022 from the Department of Economics, Maharshi Dayanand University, Rohtak.



She was recipient of UGC Junior Research Fellowship and also secured an all-India 77th rank in GATE Economics exam, 2021. More than 15 research paper publications in reputed journals stand to her credit. She has also presented nine research papers in different national and international conferences. She also won ‘Best Presenter Award’ and ‘Young Researcher Award’ in International Conferences in 2020 and 2021, respectively.

Presently, she is working as Assistant Professor (Economics) in SRM University, Delhi-NCR, Sonapat since August 2019. She also shared her knowledge and insights as a Visiting Professor at Centre for Advanced Studies in Policy Research Foundation, Pune, from February 2021 to January 2022.



Sultan Chand & Sons

Publishers of Standard Educational Textbooks

23 Daryaganj, New Delhi-110002

Phones (S): 011-23281876, 23266105, 41625022

(O): 011-23247051, 40234454

Email : sultanchand74@yahoo.com

info@sultanchandandsons.com



Scan to Visit Us

TC 1307



ISBN 978-93-91820-80-0

9 789391 820800