

---

RICHARD A. JOHNSON • GOURI K. BHATTACHARYYA

---



# STATISTICS

PRINCIPLES AND METHODS

Third Edition

# Statistics

## Principles and Methods

---

THIRD EDITION

Richard A. Johnson  
Gouri K. Bhattacharyya  
*University of Wisconsin at Madison*



John Wiley and Sons, Inc.  
New York • Chichester • Brisbane • Toronto • Singapore

ACQUISITIONS EDITOR  
MARKETING MANAGER  
SENIOR PRODUCTION EDITORS  
DESIGNER  
MANUFACTURING MANAGER  
PHOTO RESEARCHER  
ILLUSTRATION COORDINATOR

Brad Wiley, II  
Debra Riegert  
Tracey Kuehn, Nancy Prinz  
Laura Nicholls  
Mark Cirillo  
Lisa Passmore  
Jamie Perea

This book was set in Berling by CRWaldman, and printed and bound by R.R. Donnelley/  
Harrisonburg.

The cover was printed by Lehigh Press.

Cover photo by N. Meyers/Bruce Coleman, Inc.

Recognizing the importance of preserving what has been written, it is a policy of John Wiley & Sons, Inc., to have books of enduring value published in the United States printed on acid-free paper, and we exert our best efforts to that end.

The paper on this book was manufactured by a mill whose forest management programs include sustained yield harvesting of its timberlands. Sustained yield harvesting principles ensure that the number of trees cut each year does not exceed the amount of new growth.

Copyright © 1987, 1992, 1996, by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Sections 107 and 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

***Library of Congress Cataloging-in-Publication Data:***

Johnson, Richard Arnold.

Statistics : principles and methods / Richard A. Johnson, Gouri K. Bhattacharyya.—3rd ed.

p. cm.

Includes index.

ISBN 0-471-04194-7 (cloth : alk. paper)

1. Statistics. I. Bhattacharyya, Gouri K., 1940– II. Title.

QA276.12.J63 1996

519.5—dc20

95-39204

CIP

Printed in the United States of America

10 9 8 7 6 5 4 3

# Preface

---

## THE NATURE OF THE BOOK

---

Statistics—the subject of data analysis and data-based reasoning—is playing an increasingly vital role in virtually all professions. Some familiarity with this subject is now an essential component of any college education. Yet, pressures to accommodate a growing list of academic requirements often necessitate that this exposure be brief. Keeping these conditions in mind, we have written this book to provide students with a first exposure to the powerful ideas of modern statistics. It presents the key statistical concepts and the most commonly applied methods of statistical analysis. Moreover, to keep it accessible to freshman and sophomores from a wide range of disciplines, we have avoided mathematical derivations. They usually pose a stumbling block to learning the essentials in a short period of time.

This book is intended for students who do not have a strong background in mathematics but seek to learn the basic ideas of statistics and their application in a variety of practical settings. The core material of this book is common to almost all first courses in statistics and is designed to be covered well within a one-semester course in introductory statistics for freshmen–seniors. It is supplemented with some additional special-topics chapters.

## ORIENTATION

---

The topics treated in this text are, by and large, the ones typically covered in an introductory statistics course. They span three major areas: (i) descriptive statistics, which deals with summarization and description of data; (ii) ideas of probability and an understanding of the manner in which sample-to-sample variation influences our conclusions; and (iii) a collection of statistical methods for analyzing the types of data that are of common occurrence. However, it is the treatment of these topics that makes the text distinctive. By means of good motivation, sound explanations, and an abundance of illustrations given in a real-world context, it emphasizes more than just a superficial understanding.

Each concept or technique is motivated by first setting out its goal and indicating its scope by an illustration of its application. The subsequent discussion is not only limited to showing how a method works but includes an explanation of the why. Even without recourse to mathematics, we are able to make the reader aware of possible pitfalls in the statistical analysis. Students can gain a proper appreciation of statistics only when they are provided with a careful explanation of the underlying logic. Without this understanding, a learning of elementary statistics is bound to be rote and transient.

When describing the various methods of statistical analysis, the reader is continually reminded that the validity of a statistical inference is contingent upon certain model assumptions. Misleading conclusions may result when these assumptions are violated. We feel that the teaching of statistics, even at an introductory level, should not be limited to the prescription of methods. Students should be encouraged to develop a critical attitude in applying the methods and to be cautious when interpreting the results. This attitude is especially important in the study of relationship among variables, which is perhaps the most widely used (and also abused) area of statistics. In addition to discussing inference procedures in this context, we have particularly stressed critical examination of the model assumptions and careful interpretation of the conclusions.

## SPECIAL FEATURES

---

1. Crucial elements are boxed to highlight important concepts and methods. These boxes provide an ongoing summary of the important items essential for learning statistics. At the end of each chapter, all of its **key ideas and formulas** are summarized.
2. A rich collection of examples and exercises is included. These are drawn from a large variety of **real-life settings**. In fact, many data sets stem from genuine experiments, surveys, or reports.
3. Exercises are provided at the end of **each major section**. These provide the reader with the opportunity to practice the ideas just learned. Occasionally, they supplement some points raised in the text. A larger collection of exercises appears at the **end of a chapter**. The starred problems are relatively difficult and suited to the more mathematically competent student.
4. Regression analysis is a primary statistical technique so we provide a more thorough coverage of the topic than is usual at this level. The basics of regression are introduced in Chapter 11, whereas Chapter 12 stretches the discussion to several issues of practical importance. These include methods of **model checking**, handling nonlinear relations, and multiple regression analysis. Complex formulas and calculations are judiciously replaced by computer output so the main ideas can be learned and appreciated with a minimum of stress.

5. **Computer Aided Statistical Analyses** use software packages that can remove much of the drudgery and hand calculation and plotting. They allow students to work with larger data sets where patterns are more pronounced and make complicated calculations. Besides discussion of some computer output in the text, computer exercises are included in all chapters where relevant.
6. **Convenient Electronic Data Bank** at the end of the book contains a substantial collection of data. These data sets, together with numerous others throughout the book, allow for considerable flexibility in the choice between concept-orientated and applications-orientated exercises. The Data Bank and the other larger data sets are available on floppy disk.
7. **Technical Appendix A** presents a few statistical facts of a mathematical nature. These are separated from the main text so that they can be left out if the instructor so desires.

## ABOUT THIS THIRD EDITION

---

The third edition of *STATISTICS—Principles and Methods* maintains the objectives and level of presentation of the earlier editions. The goals are the developing (i) of an understanding of the reasonings by which findings from sample data can be extended to general conclusions and (ii) a familiarity with some basic statistical methods. There are numerous data sets and computer outputs which give an appreciation of the role of the computer in modern data analysis.

Throughout, we have endeavored to give clear and concise explanations of the concepts and important statistical terminology and methods. Discussion of the statistical methods includes an explanation of their underlying assumptions and the dangers of ignoring them. Real-life settings are used to motivate the statistical ideas and well organized discussions proceed to cover statistical methods with heavy emphasis on examples. The third edition enhances these special features. More particularly, the major improvements are

**More Emphasis on Understanding Data.** We have expanded the first chapter by including an extended discussion of population and sample with more examples to tie to student experiences. A new section on “The Purposeful Collection of Data” helps focus on this first often overlooked step and helps improve the practicality of later examples and illustrations.

Chapter 2 contains a new section “Checking the Stability of the Observations Over Time” in which we introduce trend charts and then add control limits. These charts are introduced in the context of a real application.

“A Designed Experiment for Making a Comparison” section in Chapter 3 gives an extended discussion of a real application concerning proportions. The study introduces the idea of randomization and can provide the basis for student projects.

The new section “The Binomial Distribution in Context,” in Chapter 5, helps illustrate the use of a probability distribution in a practical setting.

**More Data Based Exercises.** Some of the new exercises are keyed to the new data based examples given an extended discussion. The Data Bank now includes a data set on alligators and one on sleeping and snoring which are connected to new exercises. As with all the data sets in the Data Bank, many additional techniques can be applied beyond those specifically detailed in any exercise. Overall, there are a large number of new or updated exercises.

**Improved Organization.** Some reorganization of topics has lead to a smoother flow in the presentation. Following the suggestion of many users, we have removed the early introduction to hypothesis testing from the old chapter 5. The introduction to hypothesis testing is now in Section 4 of Chapter 8 which has been rewritten.

The remainder of the old Chapter 5 has been combined with the previous chapter to form a new Chapter 4 where the ideas are better connected.

Finally, in this revision, we have changed some notation to help mark the distinction between a random variable and its value. The sample variance is denoted by  $S^2$  and student's  $t$  statistic by  $T$  where they are random variables.

## ORGANIZATION

This book is organized into fifteen chapters, an optional technical appendix (Appendix A), and a collection of tables (Appendix B). Although designed for a one-semester or a two-quarter course, it is enriched with ample additional material to allow the instructor some choices of topics. Beyond Chapter 1, which sets the theme of statistics, the subject matter could be classified as follows:

Topic	Chapter
Descriptive study of data	2, 3
Probability and distributions	4, 5, 6
Sampling variability	7
Core ideas and methods of statistical inference	8, 9, 10
Special topics of statistical inference	11, 12, 13, 14, 15

We regard Chapters 1 to 10 as constituting the core material of an introductory statistics course, with the exception of the starred sections in Chapter 6. Although this material is just about enough for a one-semester course, many instructors may wish to eliminate some sections in order to cover the basics of regression analysis in Chapter 11. This is most conveniently done by initially skipping Chapter 3 and then taking up only those portions that are linked to

Chapter 11. Also, instead of a thorough coverage of probability that is provided in Chapter 4, the later sections of that chapter may receive a lighter coverage.

## ACKNOWLEDGMENTS

---

We thank Minitab (State College, Pa.) and the SAS Institute (Cary, N.C.) for permission to include commands and output from their software packages. A special thanks to K. T. Wu and Kam Tsui for many helpful suggestions and comments. We also thank all those who have contributed the data sets which enrich the presentation and all those who have reviewed this and previous editions.

Richard A. Johnson  
Gouri K. Bhattacharyya



---

# *Photo Credits*

---

## ***Chapter Opening Photos***

*Chapters 1–5:* Gregory G. Dimijian, M.D./Photo Researchers. *Chapters 6–10:* Karl & Kay Ammann/Bruce Coleman, Inc. *Chapters 11–15:* N. Myers/Bruce Coleman, Inc. *Appendix:* Peter Davey/Bruce Coleman, Inc.

## ***Chapter 1***

*Page 2:* Peter Menzel/Stock, Boston. *Page 5:* Copyright 1995 GM Corp. *Page 6:* Mitch Wojnarowicz/The Image Works.

## ***Chapter 2***

*Page 20:* Courtesy Carl Watras, Wisconsin Dept. of Natural Resources and University of Wisconsin Center for Limnology. *Page 29:* Alan Carey/The Image Works.

## ***Chapter 3***

*Page 86:* Courtesy NASA.

## ***Chapter 4***

*Page 122:* Mark Antman/The Image Works.

## ***Chapter 5***

*Page 182:* Courtesy Viking Yacht. *Page 200:* Courtesy Sands Hotel and Casino. *Page 208:* Catherine Ursillo/Photo Researchers.

## ***Chapter 6***

*Page 273:* Courtesy U.S. Forest Service.

## ***Chapter 7***

*Page 282:* Larry Kolvoord/The Image Works.

## ***Chapter 8***

*Page 308:* Courtesy Reader's Digest/Photo: R. Michael Stucky/Photo File.

## ***Chapter 9***

*Page 364:* Adam Woolfitt/Woodfin Camp & Associates.

## ***Chapter 10***

*Page 398:* Richard Johnson. *Page 429:* Elizabeth Crews/Stock, Boston.

## ***Chapter 11***

*Page 460:* Nancy Sefton/Photo Researchers.

## ***Chapter 12***

*Page 506:* Jeffrey L. Rotman.

## ***Chapter 13***

*Page 534:* Richard Johnson.

## ***Chapter 14***

*Page 572:* Dion Ogust/The Image Works.

## ***Chapter 15***

*Page 608:* Renee Lynn/Photo Researchers.

---

# Contents

## **1 INTRODUCTION 1**

- 1 What Is Statistics? 3
- 2 Statistics in Our Everyday Life 3
- 3 Statistics in Aid of Scientific Inquiry 5
- 4 Two Basic Concepts—Population and Sample 8
- 5 The Purposeful Collection of Data 11
- 6 Objectives of Statistics 14
- 7 Review Exercises 15

## **2 ORGANIZATION AND DESCRIPTION OF DATA 19**

- 1 Introduction 21
- 2 Main Types of Data 21
- 3 Describing Data by Tables and Graphs 22
- 4 Measures of Center 41
- 5 Measures of Variation 51
- 6 Checking the Stability of the Observations over Time 63
- 7 More on Graphics 68
- 8 Review Exercises 71

## **3 DESCRIPTIVE STUDY OF BIVARIATE DATA 85**

- 1 Introduction 87
- 2 Summarization of Bivariate Categorical Data 87
- 3 A Designed Experiment for Making a Comparison 94
- 4 Scatter Diagram of Bivariate Measurement Data 96
- 5 The Correlation Coefficient—A Measure of Linear Relation 99
- 6 Prediction of One Variable from Another (Linear Regression) 108
- 7 Review Exercises 114

---

**4 PROBABILITY** **121**


---

- 1 Introduction 123
- 2 Probability of an Event 124
- 3 Methods of Assigning Probability 131
- 4 Event Relations and Two Laws of Probability 139
- 5 Conditional Probability and Independence 151
- 6 Random Sampling from a Finite Population 164
- 7 Review Exercises 173

---

**5 PROBABILITY DISTRIBUTIONS** **181**


---

- 1 Introduction 183
- 2 Random Variables 183
- 3 Probability Distribution of a Discrete Random Variable 187
- 4 Expectation (Mean) and Standard Deviation of a Probability Distribution 197
- 5 Successes and Failures—Bernoulli Trials 207
- 6 The Binomial Distribution 213
- 7 The Binomial Distribution in Context 225
- 8 Review Exercises 229

---

**6 THE NORMAL DISTRIBUTION** **239**


---

- 1 Probability Model for a Continuous Random Variable 241
- 2 The Normal Distribution—Its General Features 248
- 3 The Standard Normal Distribution 251
- 4 Probability Calculations with Normal Distributions 258
- 5 The Normal Approximation to the Binomial 263
- \*6 Checking the Plausibility of a Normal Model 269
- \*7 Transforming Observations to Attain Near Normality 272
- 8 Review Exercises 276

---

**7 VARIATION IN REPEATED SAMPLES—SAMPLING DISTRIBUTIONS** **281**


---

- 1 Introduction 283
- 2 The Sampling Distribution of a Statistic 284
- 3 Distribution of the Sample Mean and the Central Limit Theorem 291
- 4 Review Exercises 303

---

**8 DRAWING INFERENCES FROM LARGE SAMPLES 307**

---

- 1 Introduction 309
- 2 Point Estimation of a Population Mean 311
- 3 Confidence Interval for a Population Mean 317
- 4 Testing Hypotheses about a Population Mean 327
- 5 Inferences about a Population Proportion 342
- 6 Review Exercises 355

---

**9 SMALL-SAMPLE INFERENCES FOR NORMAL POPULATIONS 363**

---

- 1 Introduction 365
- 2 Student's  $t$  Distribution 365
- 3 Inferences about  $\mu$ —Small Sample Size 370
- 4 Relationship between Tests and Confidence Intervals 380
- 5 Inferences About the Standard Deviation  $\sigma$   
(The Chi-Square Distribution) 383
- 6 Robustness of Inference Procedures 389
- 7 Review Exercises 392

---

**10 COMPARING TWO TREATMENTS 397**

---

- 1 Introduction 399
- 2 Independent Random Samples from Two Populations 403
- 3 Randomization and Its Role in Inference 425
- 4 Matched Pair Comparisons 428
- 5 Choosing Between Independent Samples and a Matched Pair Sample 439
- 6 Comparing Two Population Proportions 440
- 7 Review Exercises 451

---

**11 REGRESSION ANALYSIS—I  
(Simple Linear Regression) 459**

---

- 1 Introduction 461
- 2 Regression with a Single Predictor 463
- 3 A Straight-Line Regression Model 466
- 4 The Method of Least Squares 468

- 5 The Sampling Variability of the Least Squares Estimators—  
Tools for Inference 476
- 6 Important Inference Problems 477
- 7 The Strength of a Linear Relation 491
- 8 Remarks About the Straight-Line Model Assumptions 495
- 9 Review Exercises 498

---

## **12 REGRESSION ANALYSIS—II**

### **Multiple Linear Regression and Other Topics**

---

**505**

- 1 Introduction 507
- 2 Nonlinear Relations and Linearizing Transformations 507
- 3 Multiple Linear Regression 512
- 4 Residual Plots to Check the Adequacy of a Statistical  
Model 523
- 5 Review Exercises 527

---

## **13 ANALYSIS OF CATEGORICAL DATA**

---

**533**

- 1 Introduction 535
- 2 Pearson's  $\chi^2$  Test for Goodness of Fit 538
- 3 Contingency Table with One Margin Fixed  
(Test of Homogeneity) 543
- 4 Contingency Table with Neither Margin Fixed  
(Test of Independence) 554
- 5 Review Exercises 563

---

## **14 ANALYSIS OF VARIANCE (ANOVA)**

---

**571**

- 1 Introduction 573
- 2 Comparison of Several Treatments—The Completely  
Randomized Design 573
- 3 Population Model and Inferences for a Completely  
Randomized Design 582
- 4 Simultaneous Confidence Intervals 585
- 5 Graphical Diagnostics and Displays to Supplement  
ANOVA 589
- 6 Randomized Block Experiments for Comparing  
 $k$  Treatments 592
- 7 Review Exercises 602

---

**15    NONPARAMETRIC INFERENCE** **607**


---

1	Introduction	609
2	The Wilcoxon Rank-Sum Test for Comparing Two Treatments	609
3	Matched Pair Comparisons	621
4	Measure of Correlation Based on Ranks	631
5	Concluding Remarks	635
6	Review Exercises	636

**APPENDIX A1    SUMMATION NOTATION** **639**
**APPENDIX A2    EXPECTATION AND STANDARD  
DEVIATION—PROPERTIES** **645**
**APPENDIX A3    THE EXPECTED VALUE AND STANDARD  
DEVIATION OF  $\bar{X}$**  **651**
**APPENDIX B    TABLES** **653**

Table 1	Random Digits	654
Table 2	Cumulative Binomial Probabilities	657
Table 3	Standard Normal Probabilities	664
Table 4	Percentage Points of $t$ Distributions	666
Table 5	Percentage Points of $\chi^2$ Distributions	667
Table 6	Percentage Points of $F(\nu_1, \nu_2)$ Distributions	668
Table 7	Selected Tail Probabilities for the Null Distribution of Wilcoxon's Rank-Sum Statistic	670
Table 8	Selected Tail Probabilities for the Null Distribution of Wilcoxon's Signed-Rank Statistic	675

**DATA BANK** **677**
**ANSWERS TO SELECTED ODD-NUMBERED EXERCISES** **692**
**INDEX** **718**



# Introduction

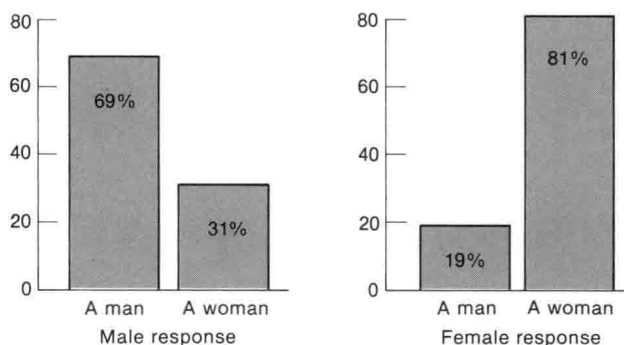
1. What Is Statistics?
2. Statistics in Our Everyday Life
3. Statistics in Aid of Scientific Inquiry
4. Two Basic Concepts—Population and Sample
5. The Purposeful Collection of Data
6. Objectives of Statistics
7. Review Exercises

# Gallup Opinion Index

**Trend in the popularity of the big three sports.** Football is the favorite spectator sport in the United States. What is your favorite spectator sport?

	1990	1981	1972	1960	1948
Football	35%	38%	36%	21%	17%
Baseball	16%	16%	21%	34%	39%
Basketball	15%	9%	8%	9%	10%
Other	33%	37%	35%	36%	34%

**Opposites Often Attract**  
Is your best friend a man or a woman?



Hometown fans attending today's game are but a sample of the population of all local baseball fans. A self-selected sample may not be entirely representative of the population on issues such as ticket price increases.



## 1. WHAT IS STATISTICS?

The word **statistics** originated from the Latin word “status” meaning “state.” For a long time, it was identified solely with the displays of data and charts pertaining to the economic, demographic, and political situations prevailing in a country. Even today, a major segment of the general public thinks of statistics as synonymous with forbidding arrays of numbers and myriads of graphs. This image is enhanced by numerous government reports that contain massive compilation of numbers and carry the word statistics in their titles: “Statistics of Farm Production,” “Statistics of Trade and Shipping,” “Labor Statistics,” to name a few. However, gigantic advances during the twentieth century have enabled statistics to grow and assume its present importance as a discipline of data-based reasoning. Passive display of numbers and charts is now a minor aspect of statistics, and few, if any, of today’s statisticians are engaged in the routine activities of tabulation and charting.

What, then, are the role and principal objectives of statistics as a scientific discipline? Stretching well beyond the confines of data display, statistics deals with collecting informative data, interpreting these data, and drawing conclusions about a phenomenon under study. The scope of this subject naturally extends to all processes of acquiring knowledge that involve fact finding through collection and examination of data. Opinion polls (surveys of households to study sociological, economic, or health-related issues), agricultural field experiments (with new seeds, pesticides, or farming equipment), clinical studies of vaccines, and cloud seeding for artificial rain production are just a few examples. The principles and methodology of statistics are useful in answering questions such as, What kind and how much data need to be collected? How should we organize and interpret the data? How can we analyze the data and draw conclusions? How do we assess the strength of the conclusions and gauge their uncertainty?

Statistics as a subject provides a body of principles and methodology for designing the process of data collection, summarizing and interpreting the data, and drawing conclusions or generalities.

## 2. STATISTICS IN OUR EVERYDAY LIFE

Fact finding through the collection and interpretation of data is not confined to professional researchers. In our attempts to understand issues of environmental protection, the state of unemployment, or the performance of competing football teams, numerical facts and figures need to be reviewed and interpreted. In our day-to-day life, learning takes place through an often implicit analysis of factual information.