Statistics

PRINCIPLES AND METHODS 5E

RICHARD A. JOHNSON GOURI K. BHATTACHARYYA

Thu Vien DHKTCN-TN
KNV.14001075

PUBLISHER	Laurie Rosatone
SENIOR ACQUISITIONS EDITOR	Angela Y. Battle
MARKETING MANAGER	Amy Sell
PRODUCTION EDITOR	Barbara Russiello
EDITORIAL ASSISTANT	Danielle Amico
SENIOR DESIGNER	Kevin Murphy
DESIGN DIRECTOR	Harry Nolan
SENIOR ILLUSTRATION EDITOR	Anna Melhorn
PHOTO EDITOR	Tara Sanford
MEDIA EDITOR	Stefanie Liebman

This book was set in Berling by Progressive Information Technologies and printed and bound by R. R. Donnelley. The cover was printed by Phoenix Color.

This book is printed on acid-free paper.

Copyright © 2006 John Wiley & Sons, Inc. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc. 222 Rosewood Drive, Danvers, MA 01923, (978)750-8400, fax (978)646-8600. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, (201)748-6011, fax (201)748-6008.
To order books or for customer service please, call 1-800-CALL WILEY (225-5945).

ISBN-13 978-0-471-65682-1
ISBN-10 0471-65682-8

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 freshmen 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. Statistics in Context sections, in four of the beginning chapters, each describe an important statistical application where a statistical approach to understanding variation is vital. These extended examples reveal, early on in the course, the value of understanding the subject of statistics.
5. *P*–values are emphasized in examples concerning tests of hypotheses. Graphs giving the relevant normal or t density curve, rejection region, and *P*–value are presented.

6. 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.
7. Computer Aided Statistical Analyses use software packages that can remove much of the drudgery of hand calculation and plotting. They allow students to work with larger data sets where patterns are more pronounced and to make complicated calculations. Besides discussion of some computer output in the text, computer exercises are included in all chapters where relevant.
8. 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 for download on the accompanying website located at www.wiley.com/college/johnson.
9. 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 THE FIFTH EDITION

The fifth 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 fifth edition enhances these special features. More particularly, the major improvements are:

Using Statistics Wisely Feature. Provides important guidelines for the appropriate use of statistics. Included at the end of each chapter.

Integrated Technology. A summary of the steps for using MINITAB, EXCEL, and the TI-84 calculator is included at the end of most chapters. This

concentrates the presentation of special purpose instructions so that, with few exceptions, only computer output is needed in the text.

More Data-Based Exercises. Some of the new exercises are keyed to new data based examples in the text. Others are based on the new grizzly bear data set added to the data bank.

New Exercises. Further new exercises providing practice on concepts, together with many new computational exercises, augment the already rich collection of exercises.

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, and distinguishes population and sample, 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.

SUPPLEMENTS

Instructor Solution Manual. (ISBN 0471-78869-4) This manual contains complete solutions to all exercises.

Test Bank. (Available on the accompanying website: www.wiley.com/college/johnson) Contains a large number of additional questions for each chapter.

Student Solutions Manual. (ISBN 0-471-71884-X) This manual contains complete solutions to all odd-numbered exercises.

Electronic Data Bank. (Available on the accompanying website: www.wiley.com/college/johnson) Contains data sets from varying sources that can be used to perform analyses with statistical software packages.

WileyPLUS. This powerful online tool provides a completely integrated suite of teaching and learning resources in one easy-to-use website. *WileyPLUS* offers an online assessment system with full gradebook capabilities and algorithmically generated skill building questions. To view a demo of *WileyPLUS*, contact your local Wiley Sales Representative or visit: www.wiley.com/college/wileyplus.

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 reviewed the previous editions. The following people gave their careful attention to this edition:

Hongshik Ahn, SUNY at Stony Brook
Casey T. Cremins, University of Maryland
Cecil Hallum, Sam Houston State University
Robert F. Herrmann, Elmhurst College
Mark McKibben, Goucher College
Ameina Summerlin, University of South Alabama
Gwen Terwilliger, University of Toledo
Quoc-Phong Vu, Ohio University
Yimin Xiao, Michigan State University

Special thanks go to Paul Lorczak who gave his input on the final pages for this fifth edition of *STATISTICS—Principles and Methods.*

Richard A. Johnson
Gouri K. Bhattacharyya

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 14
6 Statistics in Context 15
7 Objectives of Statistics 17
8 Using Statistics Wisely 18
9 Key Ideas 18
10 Review Exercises 19

2 ORGANIZATION AND DESCRIPTION OF DATA 21

1 Introduction 23
2 Main Types of Data 23
3 Describing Data by Tables and Graphs 24
3.1 Categorical Data 24
3.2 Discrete Data 28
3.3 Data on a Cotinuous Variable 29
4 Measures of Center 40
5 Measures of Variation 48
6 Checking the Stability of the Observations over Time 60
7 More on Graphics 64
8 Statistics in Context 66
9 Using Statistics Wisely 68
10 Key Ideas and Formulas 68
11 Technology 70
12 Review Exercises 73

3 DESCRIPTIVE STUDY OF BIVARIATE DATA 81

1 Introduction 83
2 Summarization of Bivariate Categorical Data 83
3 A Designed Experiment for Making a Comparison 88
4 Scatter Diagram of Bivariate Measurement Data 90
5 The Correlation Coefficient—A Measure of Linear Relation 93
6 Prediction of One Variable from Another (Linear Regression) 104
7 Using Statistics Wisely 109
8 Key Ideas and Formulas 109
9 Technology 110
10 Review Exercises 111

4 PROBABILITY 115

1 Introduction 117
2 Probability of an Event 118
3 Methods of Assigning Probability 124
  3.1 Equally Likely Elementary Outcomes—The Uniform Probability Model 124
  3.2 Probability As the Long-Run Relative Frequency 126
4 Event Relations and Two Laws of Probability 132
5 Conditional Probability and Independence 141
6 Random Sampling from a Finite Population 153
7 Using Statistics Wisely 160
8 Key Ideas and Formulas 160
9 Technology 162
10 Review Exercises 163

5 PROBABILITY DISTRIBUTIONS 169

1 Introduction 171
2 Random Variables 171
3 Probability Distribution of a Discrete Random Variable 174
4 Expectation (Mean) and Standard Deviation of a Probability Distribution 183
5 Successes and Failures—Bernoulli Trials 191
6 The Binomial Distribution 196
7 The Binomial Distribution in Context 206
8 Using Statistics Wisely 209
9 Key Ideas and Formulas 210
10 Technology 211
11 Review Exercises 213