

GLOBAL
EDITION



Statistics

The Art and Science of Learning from Data

FOURTH EDITION

Agresti
Franklin
Klingenberg



 Pearson

Statistics

The Art and Science of Learning from Data

Fourth Edition

Global Edition

Alan Agresti

University of Florida

Christine Franklin

University of Georgia

Bernhard Klingenberg

Williams College

With Contributions by

Michael Posner

Villanova University



Pearson

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • Sao Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

Editorial Director: Chris Hoag
Editor in Chief: Deirdre Lynch
Acquisitions Editor: Suzanna Bainbridge
Editorial Assistant: Justin Billing
Acquisitions Editor, Global Editions: Sourabh Maheshwari
Program Manager: Danielle Simbajon
Project Manager: Rachel S. Reeve
Assistant Project Editor, Global Editions: Vikash Tiwari
Senior Manufacturing Controller, Global Editions: Kay Holman
Program Management Team Lead: Karen Wernholm
Project Management Team Lead: Christina Lepre
Media Producer: Jean Choe
Media Production Manager, Global Editions: Vikram Kumar

TestGen Content Manager: Marty Wright
MathXL Content Manager: Robert Carroll
Product Marketing Manager: Tiffany Bitzel
Field Marketing Manager: Andrew Noble
Marketing Assistant: Jennifer Myers
Senior Author Support/Technology Specialist: Joe Vetere
Rights and Permissions Project Manager: Gina M. Cheselka
Procurement Specialist: Carol Melville
Associate Director of Design: Andrea Nix
Program Design Lead: Beth Paquin
Production Coordination, Text Design, Composition, and Illustrations: Integra Software Services Pvt Ltd.
Cover Design: Lumina Datamatics
Cover Image: Vipada Kanajod/Shutterstock.com

Acknowledgements of third-party content appear on page C-1, which constitutes an extension of this copyright page.

PEARSON, ALWAYS LEARNING, and MYSTATLAB are exclusive trademarks owned by Pearson Education, Inc., or its affiliates in the U.S. and/or other countries.

Pearson Education Limited
Edinburgh Gate
Harlow
Essex CM20 2JE
England

and Associated Companies throughout the world

Visit us on the World Wide Web at:
www.pearsonglobaleditions.com

© Pearson Education Limited 2018

The rights of Alan Agresti, Christine Franklin, and Bernhard Klingenberg to be identified as the authors of this work have been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

Authorized adaptation from the United States edition, entitled Statistics: The Art and Science of Learning from Data, Fourth Edition, ISBN 978-0-321-99783-8, by Alan Agresti, Christine Franklin, and Bernhard Klingenberg, published by Pearson Education © 2017.

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 or otherwise, without either the prior written permission of the publisher or a license permitting restricted copying in the United Kingdom issued by the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

All trademarks used herein are the property of their respective owners. The use of any trademark in this text does not vest in the author or publisher any trademark ownership rights in such trademarks, nor does the use of such trademarks imply any affiliation with or endorsement of this book by such owners.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

10 9 8 7 6 5 4 3 2 1

ISBN 10: 1-292-16477-8
ISBN 13: 978-1-292-16477-9

Typeset by Integra Software Services Pvt Ltd.
Printed and bound in Malaysia

Dedication

To my wife Jacki for her extraordinary support, including making numerous suggestions and putting up with the evenings and weekends I was working on this book.

ALAN AGRESTI

To Corey and Cody, who have shown me the joys of motherhood, and to my husband, Dale, for being a dear friend and a dedicated father to our boys. You have always been my biggest supporters.

CHRIS FRANKLIN

To my wife Sophia and our children Franziska, Florentina, Maximilian, and Mattheus, who are a bunch of fun to be with, and to Jean-Luc Picard for inspiring me.

BERNHARD KLINGENBERG

Contents

Preface 9

Part One Gathering and Exploring Data

Chapter 1 Statistics: The Art and Science of Learning from Data 28

- 1.1 Using Data to Answer Statistical Questions 29
- 1.2 Sample Versus Population 34
- 1.3 Using Calculators and Computers 43

Chapter Summary 49
Chapter Problems 50

Chapter 2 Exploring Data with Graphs and Numerical Summaries 52

- 2.1 Different Types of Data 53
- 2.2 Graphical Summaries of Data 58
- 2.3 Measuring the Center of Quantitative Data 76
- 2.4 Measuring the Variability of Quantitative Data 84
- 2.5 Using Measures of Position to Describe Variability 92
- 2.6 Recognizing and Avoiding Misuses of Graphical Summaries 102

Chapter Summary 108
Chapter Problems 109

Chapter 3 Association: Contingency, Correlation, and Regression 117

- 3.1 The Association Between Two Categorical Variables 119
- 3.2 The Association Between Two Quantitative Variables 127
- 3.3 Predicting the Outcome of a Variable 139
- 3.4 Cautions in Analyzing Associations 154

Chapter Summary 170
Chapter Problems 171

Chapter 4 Gathering Data 179

- 4.1 Experimental and Observational Studies 180
- 4.2 Good and Poor Ways to Sample 188
- 4.3 Good and Poor Ways to Experiment 198
- 4.4 Other Ways to Conduct Experimental and Nonexperimental Studies 204

Chapter Summary 216
Chapter Problems 216

Part Review 1 Online

Part Two Probability, Probability Distributions, and Sampling Distributions

Chapter 5 Probability in Our Daily Lives 226

- 5.1 How Probability Quantifies Randomness 227
- 5.2 Finding Probabilities 234
- 5.3 Conditional Probability 247
- 5.4 Applying the Probability Rules 258

Chapter Summary 273
Chapter Problems 273

Chapter 6 Probability Distributions 280

- 6.1 Summarizing Possible Outcomes and Their Probabilities 281
- 6.2 Probabilities for Bell-Shaped Distributions 293

6.3 Probabilities When Each Observation Has Two Possible Outcomes 305

Chapter Summary 315

Chapter Problems 316

Chapter 7 Sampling Distributions 324

7.1 How Sample Proportions Vary Around the Population Proportion 325

7.2 How Sample Means Vary Around the Population Mean 337

Chapter Summary 352

Chapter Problems 352

Part Review 2 Online

Part Three Inferential Statistics

Chapter 8 Statistical Inference: Confidence Intervals 360

8.1 Point and Interval Estimates of Population Parameters 361

8.2 Constructing a Confidence Interval to Estimate a Population Proportion 367

8.3 Constructing a Confidence Interval to Estimate a Population Mean 380

8.4 Choosing the Sample Size for a Study 391

8.5 Using Computers to Make New Estimation Methods Possible 400

Chapter Summary 404

Chapter Problems 404

Chapter 9 Statistical Inference: Significance Tests About Hypotheses 412

9.1 Steps for Performing a Significance Test 413

9.2 Significance Tests About Proportions 418

9.3 Significance Tests About Means 434

9.4 Decisions and Types of Errors in Significance Tests 444

9.5 Limitations of Significance Tests 449

9.6 The Likelihood of a Type II Error and the Power of a Test 456

Chapter Summary 463

Chapter Problems 464

Chapter 10 Comparing Two Groups 470

10.1 Categorical Response: Comparing Two Proportions 472

10.2 Quantitative Response: Comparing Two Means 486

10.3 Other Ways of Comparing Means, Including a Permutation Test 498

10.4 Analyzing Dependent Samples 513

10.5 Adjusting for the Effects of Other Variables 524

Chapter Summary 530

Chapter Problems 531

Part Review 3 Online

Part Four Analyzing Association and Extended Statistical Methods

Chapter 11 Analyzing the Association Between Categorical Variables 542

11.1 Independence and Dependence (Association) 543

11.2 Testing Categorical Variables for Independence 548

11.3 Determining the Strength of the Association 563

11.4 Using Residuals to Reveal the Pattern of Association 572

11.5 Fisher's Exact and Permutation Tests 576

Chapter Summary 585

Chapter Problems 585

Chapter 12 Analyzing the Association
Between Quantitative Variables:
Regression Analysis 592

- 12.1** Modeling How Two Variables Are Related 593
- 12.2** Inference About Model Parameters and the Association 603
- 12.3** Describing the Strength of Association 610
- 12.4** How the Data Vary Around the Regression Line 620
- 12.5** Exponential Regression: A Model for Nonlinearity 631

Chapter Summary 637
Chapter Problems 638

Chapter 13 Multiple Regression 644

- 13.1** Using Several Variables to Predict a Response 645
- 13.2** Extending the Correlation and R^2 for Multiple Regression 651
- 13.3** Inferences Using Multiple Regression 657
- 13.4** Checking a Regression Model Using Residual Plots 668
- 13.5** Regression and Categorical Predictors 674
- 13.6** Modeling a Categorical Response 680

Appendix A-1

Answers A-7

Index I-1

Index of Applications I-9

Credits C-1

A Guide to Learning From the Art in This Text D-1

Dataset Files D-2

A Guide to Choosing a Statistical Method D-3

Summary of Key Notations and Formulas D-4

Chapter Summary 689
Chapter Problems 690

Chapter 14 Comparing Groups:
Analysis of Variance
Methods 695

- 14.1** One-Way ANOVA: Comparing Several Means 696
- 14.2** Estimating Differences in Groups for a Single Factor 706
- 14.3** Two-Way ANOVA 716

Chapter Summary 730
Chapter Problems 730

Chapter 15 Nonparametric Statistics 736

- 15.1** Compare Two Groups by Ranking 737
- 15.2** Nonparametric Methods for Several Groups and for Matched Pairs 748

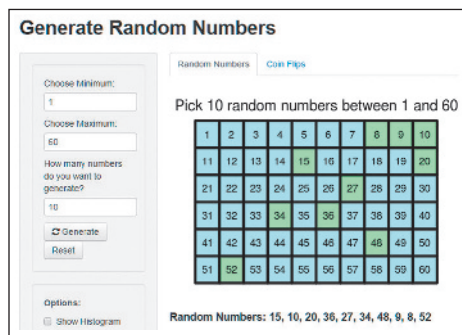
Chapter Summary 759
Chapter Problems 759

Part Review 4 Online

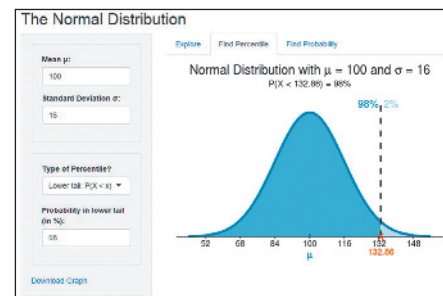
An Introduction to the Web Apps

The book's website, www.pearsonglobaleditions.com/agresti, links to several new and interactive web-based applets (or web apps) that run in a browser. These apps are designed to help students understand a wide range of statistical concepts and carry out statistical inference. Many of these apps are featured (often including screenshots) in Activities throughout the book. The apps allow saving output (such as graphs or tables) for potential inclusion in homework or projects.

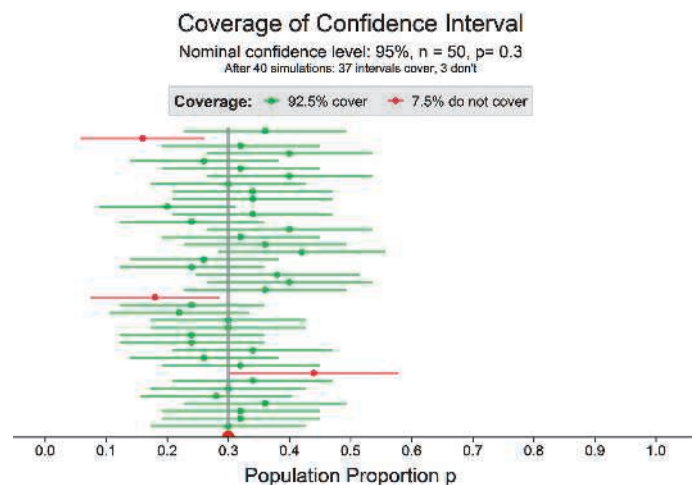
- The **Random Numbers** app generates uniform random numbers (with or without replacement) from a user-defined range of integer values and simulates flipping a (potentially biased) coin.



- The **Mean vs. Median** app allows users to add or delete points from a dot plot as the users explore the effect of outliers or skew on these two statistics.
- The **Explore Categorical Data** and **Explore Quantitative Data** apps provide basic statistics and plots for user-supplied data.
- The **Explore Linear Regression** app allows users to add or delete points from a scatterplot and observe how the regression line changes for different patterns or is affected by outliers. The **Fit Linear Regression** app allows users to supply their own data, fit a linear regression model and explore residuals.
- The **Guess the Correlation** app lets users guess the correlation for a given scatterplot (and find the correlation between guesses and the true values).
- The **Binomial, Normal, t -, Chi-square, and F Distribution** apps visually explore the meaning of parameters for these distributions. Users can also find probabilities and percentiles and check them visually on the graph.



- The various **Sampling Distribution** apps generate sampling distributions of the sample proportion or the sample mean. These apps let users generate samples of various sizes from a wide range of distributions such as skewed, uniform, bell-shaped, bimodal, or custom-built. The apps display the population distribution, the data distribution of a randomly generated sample, and the sampling distribution of the sample mean or proportion. With the (repeated) click on a button, one can see how the sampling distribution builds up one simulated random sample at a time and, for large sample sizes, assumes a bell shape. Users can move sliders for sample size and various population parameters to see the effect on the sampling distribution. Chapter 7 shows many screenshots of these apps.
- The **Inference for a Proportion** and the **Inference for a Mean** app carry out statistical inference. They provide graphs, confidence intervals and results from z - or t -tests for data supplied in summary or original form.
- The **Explore Coverage** app uses simulation to demonstrate the concept of the confidence coefficient, both



for confidence intervals for the proportion and for the mean. Different sliders for true population parameter, sample size or confidence coefficient show their effect on coverage and width of confidence intervals.

- The **Errors and Power** app explores Type I and Type II errors and the concept of power visually and interactively. Users can move sliders to connect these concepts to sample size, significance level, and true parameter value for one-sample tests about proportions or means.
- The **Inference Comparing Proportions** and the **Inference Comparing Means** apps construct appropriate graphs for a visual comparison and carry out two-sample inference. Confidence intervals and results of hypotheses tests for two independent (or two dependent) samples are displayed. Data can be supplied in summary or original form.
- The **Bootstrap** app finds a bootstrap confidence interval for a mean, median, or standard deviation.
- The **Permutation Test (cont. data)** app compares quantitative responses between two groups using a permutation approach. By repeatedly clicking a button, the sampling distribution using permutations is generated step-by-step, which is useful when first introducing the topic. Both the original and the (randomly) permuted datasets are shown.
- The **Permutation Test for Independence** app tests for independence in contingency tables using the permutation sampling distribution of the Chi-squared statistic X^2 . It displays the original contingency table and bar chart along with the table and chart for the permuted dataset, as well as the sampling distribution of X^2 .
- The **Fisher Exact Test** app can be used for exact inference in 2×2 contingency tables.
- The **ANOVA (One-Way)** app allows comparison of several means, including post-hoc pairwise multiple comparisons.

Permutation Test Comparing Two Groups

Enter observations for each group, separated by spaces:

Group 1:

Group 2:

Permutation Test: Select Test Statistic

Difference of sample means

Select number of permutations to generate:

1
 10
 1,000
 10,000

Generate Random Permutation(s)

Reset

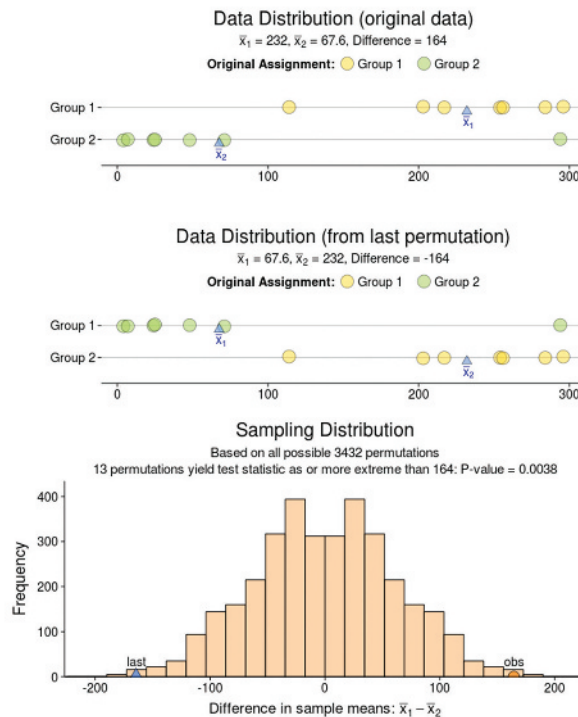
Options:

Find all possible permutations

Select alternative hypothesis:

not equal
 greater
 less

[Download dotplot of original data](#)
[Download dotplot of permuted data](#)
[Download histogram of sampling distribution](#)



Preface

We have each taught introductory statistics for many years, and we have witnessed the welcome evolution from the traditional formula-driven mathematical statistics course to a concept-driven approach. This concept-driven approach places more emphasis on why statistics is important in the real world and places less emphasis on mathematical probability. One of our goals in writing this book was to help make the conceptual approach more interesting and more readily accessible to students. At the end of the course, we want students to look back at their statistics course and realize that they learned practical concepts that will serve them well for the rest of their lives.

We also want students to come to appreciate that in practice, assumptions are not perfectly satisfied, models are not exactly correct, distributions are not exactly normally distributed, and different factors should be considered in conducting a statistical analysis. The title of our book reflects the experience of data analysts, who soon realize that statistics is an art as well as a science.

What's New in This Edition

Our goal in writing the fourth edition of our textbook was to improve the student and instructor user experience. We have:

- Clarified terminology and streamlined writing throughout the text to improve ease of reading and facilitate comprehension.
- Used real data and real examples to illustrate almost all concepts discussed. Throughout the book, within three to five consecutive pages, an example is presented that depicts a real-world scenario to illustrate the statistical concept discussed.
- Introduced new web-based applets (referred to as *web apps* or *apps*) illustrating and helping students interact with key statistical concepts and techniques. These apps invite students to explore consequences of changing parameters and to carry out statistical inference. Among other relevant concepts and techniques, students are introduced to:
 - Sampling distributions
 - Central limit theorem
 - Bootstrapping for interval estimation (Chapter 8)
 - Randomization or permutation tests for significance testing (Chapter 10 for difference in two means and Chapter 11 for two categorical variables).
- Inserted brief overviews to set the stage for each chapter, introducing students to chapter concepts and helping them see how previous chapters' concepts, tools, and techniques are related.
- Included computer output from the most recent versions of MINITAB and the TI calculator.
- Expanded Chapter 1, providing key terminology to establish a foundation to understand the big picture of the statistical investigative process—the importance of asking good statistical questions, designing an appropriate study, performing descriptive and inferential analysis, and making a conclusion.

- Reflected the latest trends in statistical education, including:
 - Measures of association for categorical variables in Chapter 3
 - Permutation testing in Chapters 10 and 11
 - Updated coverage of McNemar’s test in Chapter 10 (previously Chapter 11)
- Moved important coverage of risk difference and relative risk to Chapter 3 (instead of first introducing these measures in Chapter 11). We believe that understanding these two statistics is a necessary part of statistical literacy for the everyday citizen as they are pervasive in mass media and the medical literature.
- Updated or replaced over 25 percent of the exercises and examples. In addition, we have updated all General Social Services (GSS) data with the most current data available.

Our Approach

In 2005, the American Statistical Association (ASA) endorsed guidelines and recommendations for the introductory statistics course as described in the report, “Guidelines for Assessment and Instruction in Statistics Education (GAISE) for the College Introductory Course” (www.amstat.org/education/gaise). The report states that the overarching goal of all introductory statistics courses is to produce statistically educated students, which means that students should develop statistical literacy and the ability to think statistically. The report gives six key recommendations for the college introductory course:

- Emphasize statistical literacy and develop statistical thinking.
- Use real data.
- Stress conceptual understanding rather than mere knowledge of procedures.
- Foster active learning in the classroom.
- Use technology for developing concepts and analyzing data.
- Use assessment to evaluate and improve student learning.

We wholeheartedly endorse these recommendations, and our textbook takes every opportunity to support these guidelines.

Ask and Answer Interesting Questions

In presenting concepts and methods, we encourage students to think about the data and the appropriate analyses by posing questions. Our approach, learning by framing questions, is carried out in various ways, including (1) presenting a structured approach to examples that separates the question and the analysis from the scenario presented, (2) providing homework problems that encourage students to think and write, and (3) asking questions in the figure captions that are answered in the Chapter Review.

Present Concepts Clearly

Students have told us that this book is more “readable” and interesting than other introductory statistics texts because of the wide variety of intriguing real data examples and exercises. We have simplified our prose wherever possible, without sacrificing any of the accuracy that instructors expect in a textbook.

A serious source of confusion for students is the multitude of inference methods that derive from the many combinations of confidence intervals and tests, means and proportions, large sample and small sample, variance known

and unknown, two-sided and one-sided inference, independent and dependent samples, and so on. We emphasize the most important cases for practical application of inference: large sample, variance unknown, two-sided inference, and independent samples. The many other cases are also covered (except for known variances), but more briefly, with the exercises focusing mainly on the way inference is commonly conducted in practice. We present the traditional probability distribution–based inference but now also include inference using simulation through bootstrapping and permutation tests.

Connect Statistics to the Real World

We believe it's important for students to be comfortable with analyzing a balance of both quantitative and categorical data so students can work with the data they most often see in the world around them. Every day in the media, we see and hear percentages and rates used to summarize results of opinion polls, outcomes of medical studies, and economic reports. As a result, we have increased the attention paid to the analysis of proportions. For example, we use contingency tables early in the text to illustrate the concept of association between two categorical variables and to show the potential influence of a lurking variable.

Organization of the Book

The statistical investigative process has the following components: (1) asking a statistical question; (2) designing an appropriate study to collect data; (3) analyzing the data; and (4) interpreting the data and making conclusions to answer the statistical questions. With this in mind, the book is organized into four parts.

Part 1 focuses on gathering and exploring data. This equates to components 1, 2, and 3, when the data is analyzed descriptively (both for one variable and the association between two variables).

Part 2 covers probability, probability distributions, and the sampling distribution. This equates to component 3, when the student learns the underlying probability necessary to make the step from analyzing the data descriptively to analyzing the data inferentially (for example, understanding sampling distributions to develop the concept of a margin of error and a P-value).

Part 3 covers inferential statistics. This equates to components 3 and 4 of the statistical investigative process. The students learn how to form confidence intervals and conduct significance tests and then make appropriate conclusions answering the statistical question of interest.

Part 4 covers analyzing associations (inferentially) and looks at extended statistical methods.

The chapters are written in such a way that instructors can teach out of order. For example, after Chapter 1, an instructor could easily teach Chapter 4, Chapter 2, and Chapter 3. Alternatively, an instructor may teach Chapters 5, 6, and 7 after Chapters 1 and 4.

Features of the Fourth Edition

Promoting Student Learning

To motivate students to think about the material, ask appropriate questions, and develop good problem-solving skills, we have created special features that distinguish this text.

Student Support

To draw students to important material we highlight key definitions, guidelines, procedures, “In Practice” remarks, and other summaries in boxes throughout the text. In addition, we have four types of margin notes:

- **In Words:** This feature explains, in plain language, the definitions and symbolic notation found in the body of the text (which, for technical accuracy, must be more formal).
- **Caution:** These margin boxes alert students to areas to which they need to pay special attention, particularly where they are prone to make mistakes or incorrect assumptions.
- **Recall:** As the student progresses through the book, concepts are presented that depend on information learned in previous chapters. The Recall margin boxes direct the reader back to a previous presentation in the text to review and reinforce concepts and methods already covered.
- **Did You Know:** These margin boxes provide information that helps with the contextual understanding of the statistical question under consideration.

Graphical Approach

Because many students are visual learners, we have taken extra care to make the **text figures** informative. We’ve annotated many of the figures with labels that clearly identify the noteworthy aspects of the illustration. Further, most figure captions include a question (answered in the Chapter Review) designed to challenge the student to interpret and think about the information being communicated by the graphic. The graphics also feature a pedagogical use of color to help students recognize patterns and distinguish between statistics and parameters. The use of color is explained on page D-1 for easy reference.

Hands-On Activities and Simulations

Each chapter contains diverse and dynamic activities that allow students to become familiar with a number of statistical methodologies and tools. The instructor can elect to carry out the activities in class, outside of class, or a combination of both. The activity often involves simulation, commonly using a web app available through the book’s website or MyStatLab. Similar activities can also be found within MyStatLab. These hands-on activities and simulations encourage students to learn by doing.

Connection to History: On the Shoulders of . . .

We believe that knowledge pertaining to the evolution and history of the statistics discipline is relevant to understanding the methods we use for designing studies and analyzing data. Throughout the text, several chapters feature a spotlight on people who have made major contributions to the statistics discipline. These spotlights are titled **On the Shoulders of . . .**

Real-World Connections

Chapter-Opening Example

Each chapter begins with a **high-interest example** that raises key questions and establishes themes that are woven throughout the chapter. Illustrated with engaging photographs, this example is designed to grab students’ attention and draw them into the chapter. The issues discussed in the chapter’s opening example are referred to and revisited in examples within the chapter. All chapter-opening examples use real data from a variety of applications.

Statistics: In Practice

We realize that there is a difference between proper academic statistics and what is actually done in practice. Data analysis in practice is an art as well as a science. Although statistical theory has foundations based on precise assumptions and conditions, in practice the real world is not so simple. **In Practice** boxes and text references alert students to the way statisticians actually analyze data in practice. These comments are based on our extensive consulting experience and research and by observing what well-trained statisticians do in practice.

Exercises and Examples

Innovative Example Format

Recognizing that the worked examples are the major vehicle for engaging and teaching students, we have developed a unique structure to help students learn to model the question-posing and investigative thought process required to examine issues intelligently using statistics. The five components are as follows:

- **Picture the Scenario** presents background information so students can visualize the situation. This step places the data to be investigated in context and often provides a link to previous examples.
- **Questions to Explore** reference the information from the scenario and pose questions to help students focus on what is to be learned from the example and what types of questions are useful to ask about the data.
- **Think It Through** is the heart of each example. Here, the questions posed are investigated and answered using appropriate statistical methods. Each solution is clearly matched to the question so students can easily find the response to each Question to Explore.
- **Insight** clarifies the central ideas investigated in the example and places them in a broader context that often states the conclusions in less technical terms. Many of the Insights also provide connections between seemingly disparate topics in the text by referring to concepts learned previously and/or foreshadowing techniques and ideas to come.
- **Try Exercise:** Each example concludes by directing students to an end-of-section exercise that allows immediate practice of the concept or technique within the example.

TRY

Concept tags are included with each example so that students can easily identify the concept demonstrated in the example.


Relevant and Engaging Exercises


The text contains a strong emphasis on real data in both the examples and exercises. We have updated the exercise sets in the fourth edition to ensure that students have ample opportunity to practice techniques and apply the concepts. Nearly all of the chapters contain more than 100 exercises, and more than 25 percent of the exercises are new to this edition or have been updated with current data. These exercises are realistic and ask students to provide interpretations of the data or scenario rather than merely to find a numerical solution. We show how statistics addresses a wide array of applications, including opinion polls, market research, the environment, and health and human behavior. Because we believe that most students benefit more from focusing on the underlying concepts and interpretations of data analyses than from the actual calculations, the exercises often show summary statistics and print-outs and ask what can be learned from them.

We have exercises in three places:

- **At the end of each section.** These exercises provide immediate reinforcement and are drawn from concepts within the section.

- **At the end of each chapter.** This more comprehensive set of exercises draws from all concepts across all sections within the chapter.
- **Part Reviews.** These exercises draw connections among a part's chapters and summarize the overarching themes and concepts. Part exercises reinforce primary learning objectives. These are all available in MyStatLab.

Each exercise has a descriptive label. Exercises for which technology is recommended (such as using software or an app to carry out the analysis) are indicated with the  icon. Larger data sets used in examples and exercises are referenced in the text, listed on page D-2, and made available on the book's website. The exercises are divided into the following three categories:

- **Practicing the Basics** are the section exercises and the first group of end-of-chapter exercises; they reinforce basic application of the methods.
- **Concepts and Investigations** exercises require the student to explore real data sets and carry out investigations for mini-projects. They may ask students to explore concepts and related theory or be extensions of the chapter's methods. This section contains some multiple-choice and true-false exercises to help students check their understanding of the basic concepts and prepare for tests. A few more difficult, optional exercises (highlighted with the ) are included to present some additional concepts and methods. Concepts and Investigations exercises are found in the end-of-chapter exercises.
- **Student Activities** are designed for group work based on investigations each of the students performs on a team. Student Activities are found in the end-of-chapter exercises, and additional activities may be found within chapters as well.

Technology Integration


Up-to-Date Use of Technology

The availability of technology enables instruction that is less calculation-based and more concept-oriented. Output from software applications and calculators is displayed throughout the textbook, and discussion focuses on interpretation of the output rather than on the keystrokes needed to create the output. Although most of our output is from MINITAB[®] and the TI calculators, we also show screen captures from IBM[®] SPSS[®] and Microsoft Excel[®] as appropriate.

Web Apps

Web apps referred to in the text are found on the book's website (www.pearsonglobaleditions.com/agresti) and in MyStatLab. These apps have great value because they demonstrate concepts to students visually. For example, creating a sampling distribution is accomplished more readily with a dynamic and interactive web app than with a static text figure. (Description and list of the apps may be found on page 7.)

Data Sets

We use a wealth of real data sets throughout the textbook. These data sets are available on the www.pearsonglobaleditions.com/agresti website. The same data set is often used in several chapters, helping reinforce the four components of the statistical investigative process and allowing the students to see the big picture of statistical reasoning. Exercises requiring students to download the data set from the book's website are noted with this icon: 

Learning Catalytics

Learning Catalytics is a web-based engagement and assessment tool. As a “bring-your-own-device” direct response system, Learning Catalytics offers a diverse library of dynamic question types that allow students to interact with and think critically about statistical concepts. As a real-time resource, instructors can take advantage of critical teaching moments both in the classroom or through assignable and gradeable homework.

UPDATED! Example-Level Videos

Select examples from the text have guided videos. These updated videos provide excellent support for students who require additional assistance or want reinforcement on topics and concepts learned in class.

MyStatLab™ Online Course (access code required)

MyStatLab is a course management system that delivers **proven results** in helping individual students succeed.

- MyStatLab can be successfully implemented in any environment—lab-based, hybrid, fully online, traditional—and demonstrates the quantifiable difference that integrated usage has on student retention, subsequent success, and overall achievement.
- MyStatLab’s comprehensive online gradebook automatically tracks students’ results on tests, quizzes, homework, and in the study plan. Instructors can use the gradebook to intervene if students have trouble or to provide positive feedback. Data can be easily exported to a variety of spreadsheet programs, such as Microsoft Excel.

MyStatLab provides **engaging experiences** that personalize, stimulate, and measure learning for each student.

- **Tutorial Exercises with Multimedia Learning Aids:** The homework and practice exercises in MyStatLab align with the exercises in the textbook, and they regenerate algorithmically to give students unlimited opportunity for practice and mastery. Exercises offer immediate helpful feedback, guided solutions, sample problems, animations, videos, and eText clips for extra help at point-of-use.
- **Getting Ready for Statistics:** A library of questions now appears within each MyStatLab course to offer the developmental math topics students need for the course. These can be assigned as a prerequisite to other assignments if desired.
- **Conceptual Question Library:** In addition to algorithmically regenerated questions that are aligned with your textbook, a library of 1,000 Conceptual Questions is available in the assessment managers that require students to apply their statistical understanding.
- **StatCrunch:** MyStatLab includes a web-based statistical software, StatCrunch, within the online assessment platform so that students can easily analyze data sets from exercises and the text. In addition, MyStatLab includes access to **www.StatCrunch.com**, a website where users can access more than 20,000 shared data sets, conduct online surveys, perform complex analyses using the powerful statistical software, and generate compelling reports.
- **Integration of Statistical Software:** Knowing that students often use external statistical software, we make it easy to copy our data sets from the

MyStatLab questions into software like StatCrunch, MINTAB, Excel, and more. Students have access to a variety of support—Technology Tutorials Videos and Technology Study Cards—to learn how to use statistical software effectively.

And, MyStatLab comes from a **trusted partner** with educational expertise and an eye on the future.

Knowing that you are using a Pearson product means knowing that you are using quality content. That means that our eTexts are accurate, that our assessment tools work, and that our questions are error-free. And whether you are just getting started with MyStatLab, or have a question along the way, we're here to help you learn about our technologies and how to incorporate them into your course.

To learn more about how MyStatLab combines proven learning applications with powerful assessment, visit **www.mystatlab.com** or contact your Pearson representative.

StatCrunch®

StatCrunch® is powerful web-based statistical software that allows users to perform complex analyses, share data sets, and generate compelling reports of their data. The vibrant online community offers more than 20,000 data sets for instructors to use and students to analyze.

- **Collect.** Users can upload their own data to StatCrunch or search a large library of publicly shared data sets, spanning almost any topic of interest. Also, an on-line survey tool allows users to collect data quickly through web-based surveys.
- **Crunch.** A full range of numerical and graphical methods allows users to analyze and gain insights from any data set. Interactive graphics help users understand statistical concepts and are available for export to enrich reports with visual representations of data.
- **Communicate.** Reporting options help users create a wide variety of visually appealing representations of their data.

Full access to StatCrunch is available with a MyStatLab kit, and StatCrunch is available by itself to qualified adopters. For more information, visit our website at www.statcrunch.com or contact your Pearson representative.

An Invitation Rather Than a Conclusion

We hope that students using this textbook will gain a lasting appreciation for the vital role the art and science of statistics plays in analyzing data and helping us make decisions in our lives. Our major goals for this textbook are that students learn how to:

- Recognize that we are surrounded by data and the importance of becoming statistically literate to interpret these data and make informed decisions based on data.
- Become critical readers of studies summarized in mass media and of research papers that quote statistical results.
- Produce data that can provide answers to properly posed questions.
- Appreciate how probability helps us understand randomness in our lives and grasp the crucial concept of a sampling distribution and how it relates to inference methods.
- Choose appropriate descriptive and inferential methods for examining and analyzing data and drawing conclusions.

- Communicate the conclusions of statistical analyses clearly and effectively.
- Understand the limitations of most research, either because it was based on an observational study rather than a randomized experiment or survey or because a certain lurking variable was not measured that could have explained the observed associations.

We are excited about sharing the insights that we have learned from our experience as teachers and from our students through this text. Many students still enter statistics classes on the first day with dread because of its reputation as a dry, sometimes difficult, course. It is our goal to inspire a classroom environment that is filled with creativity, openness, realistic applications, and learning that students find inviting and rewarding. We hope that this textbook will help the instructor and the students experience a rewarding introductory course in statistics.

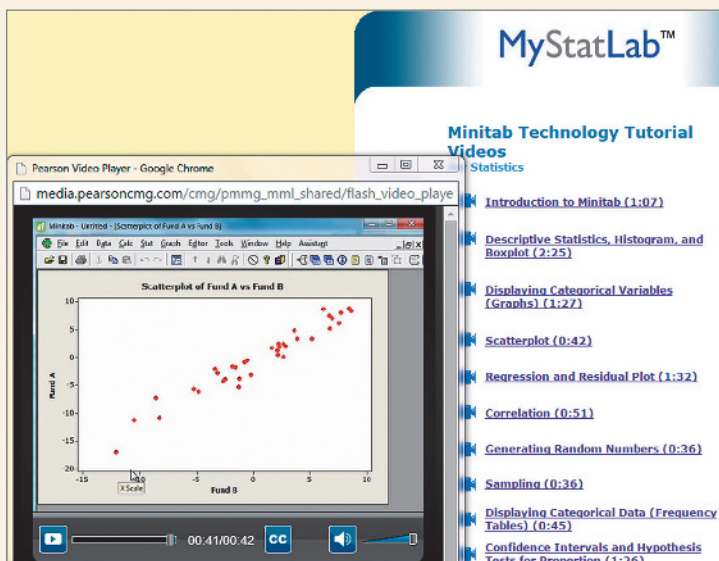
Resources for Success

MyStatLab™ Online Course for *Statistics: The Art and Science of Learning from Data* by Agresti, Franklin, and Klingenberg (access code required)

MyStatLab is available to accompany Pearson’s market leading text offerings. To give students a consistent tone, voice, and teaching method, each text’s flavor and approach is tightly integrated throughout the accompanying MyStatLab course, making learning the material as seamless as possible.

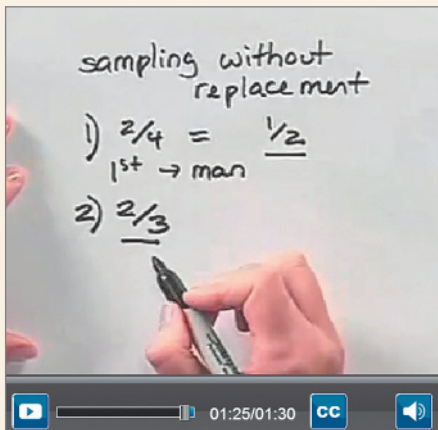
Technology Tutorials and Study Cards

Technology tutorials provide brief video walkthroughs and step-by-step instructional study cards on common statistical procedures for MINITAB®, Excel®, and the TI family of graphing calculators.



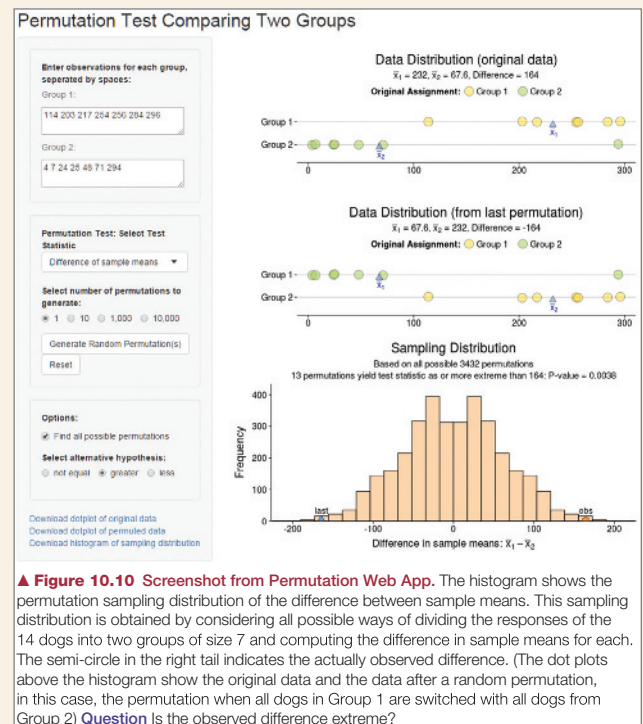
Example-Level Resources

Students looking for additional support can use the example-based videos to help solve problems, provide reinforcement on topics and concepts learned in class, and support their learning.



New! Apps: Examples, Exercises, and Simulations

Author-created web apps allow students to interact with key statistical concepts and techniques, including statistical distributions, inference for one and two samples, permutation tests, bootstrapping, and sampling distributions. Students can explore consequences of changing parameters and carry out simulations to explore coverage, or simply obtain descriptive statistics or a proper statistical graph. All these in a highly user-friendly and well designed app, where results can be downloaded for inclusion in homework or projects.



Resources for Success

Instructor Resources

Additional resources can be downloaded from www.pearsonglobaleditions.com/agresti

Text-specific website, www.pearsonglobaleditions.com/agresti New to this edition, students and instructors will have a full library of resources, including apps developed for in-text activities and data sets (.csv, TI-83/84 Plus C, and .txt).

Updated! Instructor to Instructor Videos provide an opportunity for adjuncts, part-timers, TAs, or other instructors who are new to teaching from this text or have limited class prep time to learn about the book's approach and coverage from the authors. The videos, available through MyStatLab, focus on those topics that have proven to be most challenging to students. The authors offer suggestions, pointers, and ideas about how to present these topics and concepts effectively based on their many years of teaching introductory statistics. They also share insights on how to help students use the textbook in the most effective way to realize success in the course.

Instructor's Solutions Manual, by James Lapp, contains fully worked solutions to every textbook exercise. Available for download from Pearson's online catalog at www.pearsonglobaleditions.com/agresti and through MyStatLab.

Answers to the Student Laboratory Workbook are available for download from www.pearsonglobaleditions.com/agresti and through MyStatLab.

PowerPoint Lecture Slides are fully editable and printable slides that follow the textbook. These slides can be used during lectures or posted to a website in an online course. The PowerPoint Lecture Slides are available from www.pearsonglobaleditions.com/agresti and through MyStatLab.

TestGen[®] (www.pearsoned.com/testgen) enables instructors to build, edit, print, and administer tests using a computerized bank of questions developed to cover all the objectives of the text. TestGen is algorithmically based, allowing instructors to create multiple but equivalent

versions of the same question or test with the click of a button. Instructors can also modify test bank questions or add new questions. The test bank is available for download from www.pearsonglobaleditions.com/agresti and through MyStatLab.

The Online Test Bank is a test bank derived from TestGen[®]. It includes multiple choice and short answer questions for each section of the text, along with the answer keys. Available for download from www.pearsonglobaleditions.com/agresti and through MyStatLab.

Student Resources

Additional resources to help student success.

Text-specific website, www.pearsonglobaleditions.com/agresti New to this edition, students and instructors will have a full library of resources, including apps developed for in-text activities and data sets (.csv, TI-83/84 Plus C, and .txt).

Updated! Example-level videos explain how to work examples from the text. The videos provide excellent support for students who require additional assistance or want reinforcement on topics and concepts learned in class (available in MyStatLab).

Student Laboratory Workbook, by Megan Mocko (University of Florida) and Maria Ripol (University of Florida), is a study tool for the first ten chapters of the text. This workbook provides section-by-section review and practice and additional activities that cover fundamental statistical topics (ISBN-10: 0-13-386089-2; ISBN-13: 978-0-13-386089-4).

Study Cards for Statistics Software This series of study cards, available for Excel[®], MINITAB[®], JMP[®], SPSS[®], R[®], StatCrunch[®], and the TI family of graphing calculators provides students with easy, step-by-step guides to the most common statistics software. Available in MyStatLab.

Acknowledgments

We are indebted to the following individuals, who provided valuable feedback for the fourth edition:

Evelyn Bailey, *Oxford College of Emory University*

Erica Bernstein, *University of Hawaii at Hilo*

Phyllis Curtiss, *Grand Valley State*

Katherine C. Earles, *Wichita State University*

Rob Eby, *Blinn College—Bryan Campus*

Matthew Jones, *Austin Perry Statue University*

Ann Kalinoskii, *San Jose University*

Michael Roty, *Mercer University*

Ping-Shou Zhong, *Michigan State University*

We are also indebted to the many reviewers, class testers, and students who gave us invaluable feedback and advice on how to improve the quality of the book.

ARIZONA Russel Carlson, University of Arizona; Peter Flanagan-Hyde, Phoenix Country Day School ■ **CALIFORNIA** James Curl, Modesto Junior College; Christine Drake, University of California at Davis; Mahtash Esfandiari, UCLA; Brian Karl Finch, San Diego State University; Dawn Holmes, University of California Santa Barbara; Rob Gould, UCLA; Rebecca Head, Bakersfield College; Susan Herring, Sonoma State University; Colleen Kelly, San Diego State University; Marke Mavis, Butte Community College; Elaine McDonald, Sonoma State University; Corey Manchester, San Diego State University; Amy McElroy, San Diego State University; Helen Noble, San Diego State University; Calvin Schmall, Solano Community College ■ **COLORADO** David Most, Colorado State University ■ **CONNECTICUT** Paul Bugl, University of Hartford; Anne Doyle, University of Connecticut; Pete Johnson, Eastern Connecticut State University; Dan Miller, Central Connecticut State University; Kathleen McLaughlin, University of Connecticut; Nalini Ravishanker, University of Connecticut; John Vangar, Fairfield University; Stephen Sawin, Fairfield University ■ **DISTRICT OF COLUMBIA** Hans Engler, Georgetown University; Mary W. Gray, American University; Monica Jackson, American University ■ **FLORIDA** Nazanin Azarnia, Santa Fe Community College; Brett Holbrook; James Lang, Valencia Community College; Karen Kinard, Tallahassee Community College; Megan Mocko, University of Florida; Maria Ripol, University of Florida; James Smart, Tallahassee Community College; Latricia Williams, St. Petersburg Junior College, Clearwater; Doug Zahn, Florida State University ■ **GEORGIA** Carrie Chmielarski, University of Georgia; Ouida Dillon, Oconee County High School; Kim Gilbert, University of Georgia; Katherine Hawks, Meadowcreek High School; Todd Hendricks, Georgia Perimeter College; Charles LeMarsh, Lakeside High School; Steve Messig, Oconee County High School; Broderick Oluyede, Georgia Southern University; Chandler Pike, University of Georgia; Kim Robinson, Clayton State University; Jill Smith, University of Georgia; John Seppala, Valdosta State University; Joseph Walker, Georgia State University ■ **IOWA** John Cryer, University of Iowa; Kathy Rogotzke, North Iowa Community College; R. P. Russo, University of Iowa; William Duckworth, Iowa State University ■ **ILLINOIS** Linda Brant Collins, University of Chicago; Dagmar Budikova, Illinois State University; Ellen Fireman, University of Illinois; Jinadasa Gamage, Illinois State; University; Richard Maher, Loyola University Chicago; Cathy Poliak, Northern Illinois University; Daniel Rowe, Heartland Community College ■ **KANSAS** James Higgins, Kansas State University; Michael Mosier, Washburn University ■ **KENTUCKY** Lisa Kay, Eastern Kentucky University

■ **MASSACHUSETTS** Richard Cleary, Bentley University; Katherine Halvorsen, Smith College; Xiaoli Meng, Harvard University; Daniel Weiner, Boston University ■ **MICHIGAN** Kirk Anderson, Grand Valley State University; Phyllis Curtiss, Grand Valley State University; Roy Erickson, Michigan State University; Jann-Huei Jinn, Grand Valley State University; Sango Otieno, Grand Valley State University; Alla Sikorskii, Michigan State University; Mark Stevenson, Oakland Community College; Todd Swanson, Hope College; Nathan Tintle, Hope College ■ **MINNESOTA** Bob Dobrow, Carleton College; German J. Pliego, University of St. Thomas; Peihua Qui, University of Minnesota; Engin A. Sungur, University of Minnesota–Morris ■ **MISSOURI** Lynda Hollingsworth, Northwest Missouri State University; Robert Paige, Missouri University of Science and Technology; Larry Ries, University of Missouri–Columbia; Suzanne Tourville, Columbia College ■ **MONTANA** Jeff Banfield, Montana State University ■ **NEW JERSEY** Harold Sackowitz, Rutgers, The State University of New Jersey; Linda Tappan, Montclair State University ■ **NEW MEXICO** David Daniel, New Mexico State University ■ **NEW YORK** Brooke Fridley, Mohawk Valley Community College; Martin Lindquist, Columbia University; Debby Lurie, St. John’s University; David Mathiason, Rochester Institute of Technology; Steve Stehman, SUNY ESF; Tian Zheng, Columbia University ■ **NEVADA**: Alison Davis, University of Nevada–Reno ■ **NORTH CAROLINA** Pamela Arroyo, North Carolina State University; E. Jacquelin Dietz, North Carolina State University; Alan Gelfand, Duke University; Gary Kader, Appalachian State University; Scott Richter, UNC Greensboro; Roger Woodard, North Carolina State University ■ **NEBRASKA** Linda Young, University of Nebraska ■ **OHIO** Jim Albert, Bowling Green State University; John Holcomb, Cleveland State University; Jackie Miller, The Ohio State University; Stephan Pelikan, University of Cincinnati; Teri Rysz, University of Cincinnati; Deborah Rumsey, The Ohio State University; Kevin Robinson, University of Akron; Dottie Walton, Cuyahoga Community College - Eastern Campus ■ **OREGON** Michael Marciniak, Portland Community College; Henry Mesa, Portland Community College, Rock Creek; Qi-Man Shao, University of Oregon; Daming Xu, University of Oregon ■ **PENNSYLVANIA** Winston Crawley, Shippensburg University; Douglas Frank, Indiana University of Pennsylvania; Steven Gendler, Clarion University; Bonnie A. Green, East Stroudsburg University; Paul Lupinacci, Villanova University; Deborah Lurie, Saint Joseph’s University; Linda Myers, Harrisburg Area Community College; Tom Short, Villanova University; Kay Somers, Moravian College; Sister Marcella Louise Wallowicz, Holy Family University ■ **SOUTH CAROLINA** Beverly Diamond, College of Charleston; Martin Jones, College of Charleston; Murray Siegel, The South Carolina Governor’s School for Science and Mathematics; ■ **SOUTH DAKOTA** Richard Gayle, Black Hills State University; Daluss Siewert, Black Hills State University; Stanley Smith, Black Hills State University ■ **TENNESSEE** Bonnie Daves, Christian Academy of Knoxville; T. Henry Jablonski, Jr., East Tennessee State University; Robert Price, East Tennessee State University; Ginger Rowell, Middle Tennessee State University; Edith Seier, East Tennessee State University ■ **TEXAS** Larry Ammann, University of Texas, Dallas; Tom Bratcher, Baylor University; Jianguo Liu, University of North Texas; Mary Parker, Austin Community College; Robert Paige, Texas Tech University; Walter M. Potter, Southwestern University; Therese Shelton, Southwestern University; James Surlles, Texas Tech University; Diane Resnick, University of Houston-Downtown ■ **UTAH** Patti Collings, Brigham Young University; Carolyn Cuff, Westminster College; Lajos Horvath, University of Utah; P. Lynne Nielsen, Brigham Young University ■ **VIRGINIA** David Bauer, Virginia Commonwealth University; Ching-Yuan Chiang, James Madison University; Jonathan Duggins, Virginia Tech; Steven Garren, James Madison University; Hasan Hamdan, James Madison University; Debra Hydorn, Mary Washington College; Nusrat Jahan, James Madison University; D’Arcy Mays, Virginia Commonwealth University; Stephanie Pickle, Virginia Polytechnic Institute and State University

■ **WASHINGTON** Rich Alldredge, Washington State University; Brian T. Gill, Seattle Pacific University; June Morita, University of Washington ■ **WISCONSIN** Brooke Fridley, University of Wisconsin–LaCrosse; Loretta Robb Thielman, University of Wisconsin–Stout. ■ **WYOMING** Burke Grandjean, University of Wyoming ■ **CANADA** Mike Kowalski, University of Alberta; David Loewen, University of Manitoba

We appreciate all of the thoughtful contributions to the fourth edition by Michael Posner. We also thank the following individuals, who made invaluable contributions to the third edition:

Ellen Breazel, *Clemson University*

Linda Dawson, *Washington State University, Tacoma*

Bernadette Lanciaux, *Rochester Institute of Technology*

Scott Nickleach, *Sonoma State University*

The detailed assessment of the text fell to our accuracy checkers, Ann Cannon and Joan Sanuik.

Thank you to James Lapp, who took on the task of revising the solutions manuals to reflect the many changes to the fourth edition. We also want to thank Jackie Miller (Ohio State University) for her contributions to the Instructor's Notes, and our student technology manual and workbook authors, Megan Mocko (University of Florida) and Maria Ripol (University of Florida).

We would like to thank the Pearson team who has given countless hours in developing this text; without their guidance and assistance, the text would not have come to completion. We thank Suzanna Bainbridge, Rachel Reeve, Danielle Simbajon, Justin Billing, Jean Choe, Tiffany Bitzel, Andrew Noble, Jennifer Myers, and especially Joe Vetere. We also thank Kristin Jobe, Senior Project Manager at Integra-Chicago, for keeping this book on track throughout production.

Alan Agresti would like to thank those who have helped us in some way, often by suggesting data sets or examples. These include Anna Gottard, Wolfgang Jank, Bernhard Klingenberg, René Lee-Pack, Jacalyn Levine, Megan Lewis, Megan Meece, Dan Nettleton, Yongyi Min, and Euijung Ryu. Many thanks also to Tom Piazza for his help with the General Social Survey. Finally, Alan Agresti would like to thank his wife Jacki Levine for her extraordinary support throughout the writing of this book. Besides putting up with the evenings and weekends he was working on this book, she offered numerous helpful suggestions for examples and for improving the writing.

Chris Franklin gives a special thank you to her husband and sons, Dale, Corey, and Cody Green. They have patiently sacrificed spending many hours with their spouse and mom as she has worked on this book through four editions. A special thank you also to her parents Grady and Helen Franklin and her two brothers, Grady and Mark, who have always been there for their daughter and sister. Chris also appreciates the encouragement and support of her colleagues and her many students who used the book, offering practical suggestions for improvement. Finally, Chris thanks her coauthors Alan and Bernhard for the amazing journey of writing a textbook together.

Bernhard Klingenberg wants to thank his statistics teachers in Graz, Austria; Sheffield, UK; and Gainesville, Florida, who showed him all the fascinating facets of statistics throughout his education. Thanks also to the Department of Mathematics & Statistics at Williams College for being such a wonderful place to work. Finally, thanks to Chris Franklin and Alan Agresti for a wonderful and inspiring collaboration.

ALAN AGRESTI, *Gainesville, Florida*

CHRISTINE FRANKLIN, *Athens, Georgia*

BERNHARD KLINGENBERG, *Williamstown, Massachusetts*

Acknowledgments for the Global Edition

Pearson would like to thank and acknowledge the following people for their work on the Global Edition.

Contributors

Mohammad Kacim, *Holy Spirit University of Kaslik*

Abhishek K. Umrawal, *Delhi University*

Reviewers

Ruben Garcia Berasategui, *Jakarta International College*

Amit Kumar Misra, *Babasaheb Bhimrao Ambedkar University*

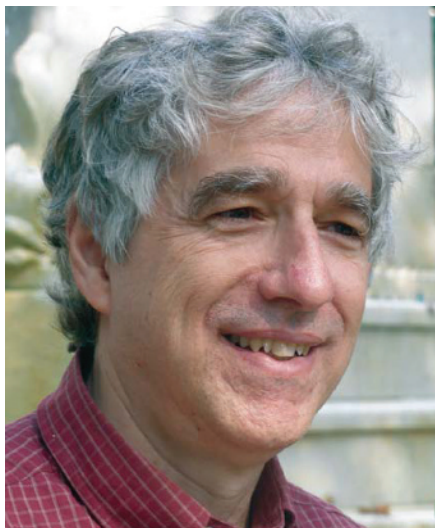
Ranjita Pandey, *Delhi University*

Louise M. Ryan, *University of Technology Sydney*

C. V. Vinay, *JSS Academy of Technical Education*

This page intentionally left blank

About the Authors



Alan Agresti is Distinguished Professor Emeritus in the Department of Statistics at the University of Florida. He taught statistics there for 38 years, including the development of three courses in statistical methods for social science students and three courses in categorical data analysis. He is author of more than 100 refereed articles and six texts, including *Statistical Methods for the Social Sciences* (with Barbara Finlay, Prentice Hall, 4th edition, 2009) and *Categorical Data Analysis* (Wiley, 3rd edition, 2013). He is a Fellow of the American Statistical Association and recipient of an Honorary Doctor of Science from De Montfort University in the UK. He has held visiting positions at Harvard University, Boston University, the London School of Economics, and Imperial College and has taught courses or short courses for universities and companies in about 30 countries worldwide. He has also received teaching awards from the University of Florida and an excellence in writing award from John Wiley & Sons.



Christine Franklin is a Senior Lecturer and Lothar Tresp Honoratus Honors Professor in the Department of Statistics at the University of Georgia. She has been teaching statistics for more than 35 years at the college level. Chris has been actively involved at the national and state level with promoting statistical education at Pre-K–16 since the 1980s. She is a past Chief Reader for AP Statistic. Chris served as the lead writer for the ASA-endorsed Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre- K–12 Curriculum Framework and chaired the ASA Statistical Education of Teachers (SET) Report. Chris has been honored by her selection as a Fellow of the American Statistical Association, recipient of the 2006 Mu Sigma Rho National Statistical Education Award, the 2013 USCOTS Lifetime Achievement Award, the 2014 ASA Founders Award, a 2014–2015 U.S. Fulbright Scholar, and numerous teaching and advising awards at the University of Georgia. Most important for Chris is her family, who love to hike and attend baseball games together.



Bernhard Klingenberg is Associate Professor of Statistics in the Department of Mathematics & Statistics at Williams College, where he has been teaching introductory and advanced statistics classes for the past 11 years. In 2013, Bernhard was instrumental in creating an undergraduate major in statistics at Williams, one of the first for a liberal arts college. A native of Austria, Bernhard frequently returns there to hold visiting positions at universities and gives short courses on categorical data analysis in Europe and the United States. He has published several peer-reviewed articles in statistical journals and consults regularly with academia and industry. Bernhard enjoys photography (some of his pictures appear in this book), scuba diving, and time with his wife and four children.

This page intentionally left blank

Gathering and Exploring Data



Chapter 1

Statistics: The Art and Science of Learning from Data

Chapter 2

Exploring Data with Graphs and Numerical Summaries

Chapter 3

Association: Contingency, Correlation, and Regression

Chapter 4

Gathering Data

- 1.1 Using Data to Answer Statistical Questions
- 1.2 Sample Versus Population
- 1.3 Using Calculators and Computers



Statistics: The Art and Science of Learning from Data

Example 1

How Statistics Helps Us Learn About the World

Picture the Scenario

In this book, you will explore a wide variety of everyday scenarios. For example, you will evaluate media reports about opinion surveys, medical research studies, the state of the economy, and environmental issues. You'll face financial decisions such as choosing between an investment with a sure return and one that could make you more money but could possibly cost you your entire investment. You'll learn how to analyze the available information to answer necessary questions in such scenarios. One purpose of this book is to show you why an understanding of statistics is essential for making good decisions in an uncertain world.

Questions to Explore

This book will show you how to collect appropriate information and how to apply statistical methods so you can better evaluate that information

and answer the questions posed. Here are some examples of questions we'll investigate in coming chapters:

- How can you evaluate evidence about global warming?
- Are cell phones dangerous to your health?
- What's the chance your tax return will be audited?
- How likely are you to win the lottery?
- Is there bias against women in appointing managers?
- What 'hot streaks' should you expect in basketball?
- How can you analyze whether a diet really works?
- How can you predict the selling price of a house?

Thinking Ahead

Each chapter uses questions like these to introduce a topic and then introduces tools for making sense of the available information. We'll see that **statistics** is the art and science of designing studies and analyzing the information that those studies produce.

In the business world, managers use statistics to analyze results of marketing studies about new products, to help predict sales, and to measure employee performance. In finance, statistics is used to study stock returns and investment opportunities. Medical studies use statistics to evaluate whether new ways to treat disease are better than existing ways. In fact, most professional occupations today rely heavily on statistical methods. In a competitive job market, understanding statistics provides an important advantage.

But it's important to understand statistics even if you will never use it in your job. Understanding statistics can help you make better choices. Why? Because every day you are bombarded with statistical information from news reports, advertisements, political campaigns, and surveys. How do you know what to heed and what to ignore? An understanding of the statistical reasoning—and in some cases statistical misconceptions—underlying these pronouncements will help. For instance, this book will enable you to evaluate claims about medical research studies more effectively so that you know when you should be skeptical. For example, does taking an aspirin daily truly lessen the chance of having a heart attack?

We realize that you are probably not reading this book in the hope of becoming a statistician. (That's too bad, because there's a severe shortage of statisticians—more jobs than trained people. And with the ever-increasing ways in which statistics is being applied, it's an exciting time to be a statistician.) You may even suffer from math phobia. Please be assured that to learn the main concepts of statistics, logical thinking and perseverance are more important than high-powered math skills. Don't be frustrated if learning comes slowly and you need to read about a topic a few times before it starts to make sense. Just as you would not expect to sit through a single foreign language class session and be able to speak that language fluently, the same is true with the language of statistics. It takes time and practice. But we promise that your hard work will be rewarded. Once you have completed even part of this text, you will understand much better how to make sense of statistical information and, hence, the world around you.

1.1 Using Data to Answer Statistical Questions

Does a low-carbohydrate diet result in significant weight loss? Are people more likely to stop at a Starbucks if they've seen a recent Starbucks TV commercial? Information gathering is at the heart of investigating answers to such questions. The information we gather with experiments and surveys is collectively called **data**.

For instance, consider an experiment designed to evaluate the effectiveness of a low-carbohydrate diet. The data might consist of the following measurements for the people participating in the study: weight at the beginning of the study, weight at the end of the study, number of calories of food eaten per day, carbohydrate intake per day, body-mass index (BMI) at the start of the study, and gender. A marketing survey about the effectiveness of a TV ad for Starbucks could collect data on the percentage of people who went to a Starbucks since the ad aired and analyze how it compares for those who saw the ad and those who did not see it.

Defining Statistics

You already have a sense of what the word *statistics* means. You hear statistics quoted about sports events (number of points scored by each player on a basketball team), statistics about the economy (median income, unemployment rate), and statistics about opinions, beliefs, and behaviors (percentage of students who