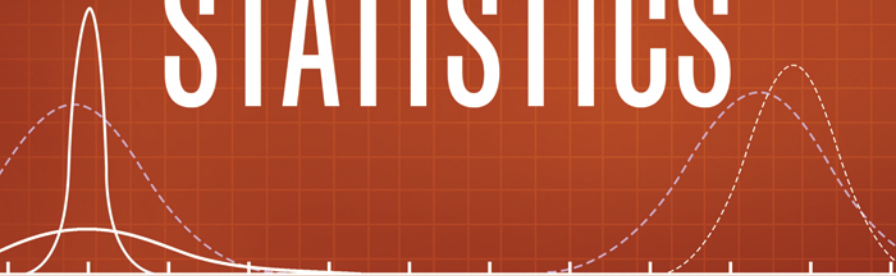


JAMES McCLAVE | TERRY SINCICH

A FIRST COURSE IN

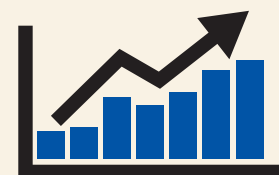
STATISTICS



TWELFTH EDITION

This page intentionally left blank

Available in MyStatLab™ for Your Introductory Statistics Courses

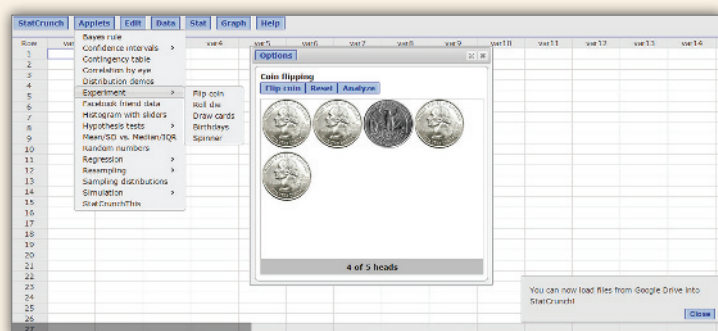
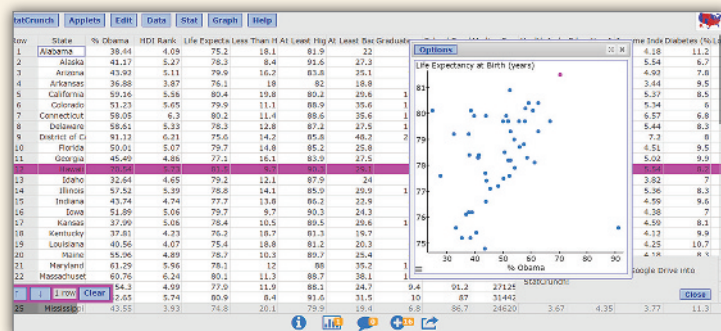


MyStatLab is the market-leading online resource for learning and teaching statistics.

Leverage the Power of StatCrunch

MyStatLab leverages the power of StatCrunch—powerful, web-based statistics software. Integrated into MyStatLab, students can easily analyze data from their exercises and etext.

In addition, access to the full online community allows users to take advantage of a wide variety of resources and applications at www.statcrunch.com.

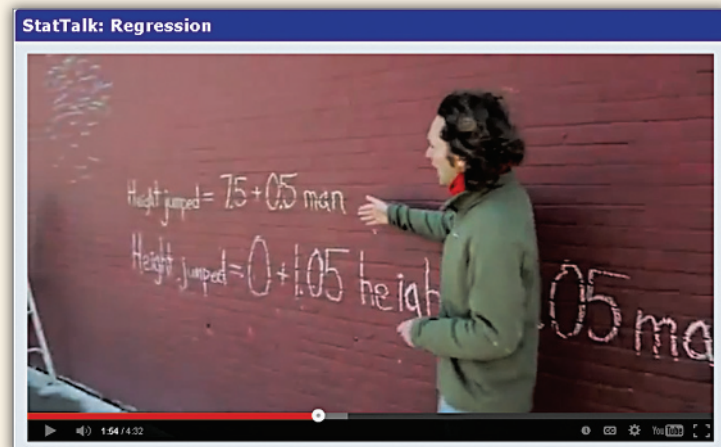


Bring Statistics to Life

Virtually flip coins, roll dice, draw cards, and interact with animations on your mobile device with the extensive menu of experiments and applets in StatCrunch. Offering a number of ways to practice resampling procedures, such as permutation tests and bootstrap confidence intervals, StatCrunch is a complete and modern solution.

Real-World Statistics

MyStatLab video resources help foster conceptual understanding. StatTalk Videos, hosted by fun-loving statistician Andrew Vickers, demonstrate important statistical concepts through interesting stories and real-life events. This series of 24 videos includes assignable questions built in MyStatLab and an instructor's guide.



APPLET CORRELATION

Applet	Concept Illustrated	Description	Applet Activity
Sample from a population	Assesses how well a sample represents the population and the role that sample size plays in the process.	Produces random sample from population from specified sample size and population distribution shape. Reports mean, median, and standard deviation; applet creates plot of sample.	4.4 , 192; 4.6 , 207
Sampling distributions	Compares means and standard deviations of distributions; assesses effect of sample size; illustrates unbiasedness.	Simulates repeatedly choosing samples of a fixed size n from a population with specified sample size, number of samples, and shape of population distribution. Applet reports means, medians, and standard deviations; creates plots for both.	4.7 , 236; 4.8 , 236
Random numbers	Uses a random number generator to determine the experimental units to be included in a sample.	Generates random numbers from a range of integers specified by the user.	1.1 , 19; 1.2 , 20; 3.6 , 159; 4.1 , 178
Long-run probability demonstrations illustrate the concept that theoretical probabilities are long-run experimental probabilities.			
Simulating probability of rolling a 6	Investigates relationship between theoretical and experimental probabilities of rolling 6 as number of die rolls increases.	Reports and creates frequency histogram for each outcome of each simulated roll of a fair die. Students specify number of rolls; applet calculates and plots proportion of 6s.	3.1 , 127; 3.3 , 138; 3.4 , 139; 3.5 , 153
Simulating probability of rolling a 3 or 4	Investigates relationship between theoretical and experimental probabilities of rolling 3 or 4 as number of die rolls increases.	Reports outcome of each simulated roll of a fair die; creates frequency histogram for outcomes. Students specify number of rolls; applet calculates and plots proportion of 3s and 4s.	3.3 , 138; 3.4 , 139
Simulating the probability of heads: fair coin	Investigates relationship between theoretical and experimental probabilities of getting heads as number of fair coin flips increases.	Reports outcome of each fair coin flip and creates a bar graph for outcomes. Students specify number of flips; applet calculates and plots proportion of heads.	3.2 , 127; 4.2 , 179
Simulating probability of heads: unfair coin ($P(H) = .2$)	Investigates relationship between theoretical and experimental probabilities of getting heads as number of unfair coin flips increases.	Reports outcome of each flip for a coin where heads is less likely to occur than tails and creates a bar graph for outcomes. Students specify number of flips; applet calculates and plots the proportion of heads.	4.3 , 192
Simulating probability of heads: unfair coin ($P(H) = .8$)	Investigates relationship between theoretical and experimental probabilities of getting heads as number of unfair coin flips increases.	Reports outcome of each flip for a coin where heads is more likely to occur than tails and creates a bar graph for outcomes. Students specify number of flips; applet calculates and plots the proportion of heads.	4.3 , 192
Simulating the stock market	Theoretical probabilities are long run experimental probabilities.	Simulates stock market fluctuation. Students specify number of days; applet reports whether stock market goes up or down daily and creates a bar graph for outcomes. Calculates and plots proportion of simulated days stock market goes up.	4.5 , 192
Mean versus median	Investigates how skewedness and outliers affect measures of central tendency.	Students visualize relationship between mean and median by adding and deleting data points; applet automatically updates mean and median.	2.1 , 61; 2.2 , 61; 2.3 , 61

Applet	Concept Illustrated	Description	Applet Activity
Standard deviation	Investigates how distribution shape and spread affect standard deviation.	Students visualize relationship between mean and standard deviation by adding and deleting data points; applet updates mean and standard deviation.	2.4 , 68; 2.5 , 69; 2.6 , 69; 2.7 , 91
Confidence intervals for a proportion	Not all confidence intervals contain the population proportion. Investigates the meaning of 95% and 99% confidence.	Simulates selecting 100 random samples from the population and finds the 95% and 99% confidence intervals for each. Students specify population proportion and sample size; applet plots confidence intervals and reports number and proportion containing true proportion.	5.5 , 279; 5.6 , 280
Confidence intervals for a mean (the impact of confidence level)	Not all confidence intervals contain the population mean. Investigates the meaning of 95% and 99% confidence.	Simulates selecting 100 random samples from population; finds 95% and 99% confidence intervals for each. Students specify sample size, distribution shape, and population mean and standard deviation; applet plots confidence intervals and reports number and proportion containing true mean.	5.1 , 261; 5.2 , 261
Confidence intervals for a mean (not knowing standard deviation)	Confidence intervals obtained using the sample standard deviation are different from those obtained using the population standard deviation. Investigates effect of not knowing the population standard deviation.	Simulates selecting 100 random samples from the population and finds the 95% z-interval and 95% t-interval for each. Students specify sample size, distribution shape, and population mean and standard deviation; applet plots confidence intervals and reports number and proportion containing true mean.	5.3 , 271; 5.4 , 271
Hypothesis tests for a proportion	Not all tests of hypotheses lead correctly to either rejecting or failing to reject the null hypothesis. Investigates the relationship between the level of confidence and the probabilities of making Type I and Type II errors.	Simulates selecting 100 random samples from population; calculates and plots z-statistic and P-value for each. Students specify population proportion, sample size, and null and alternative hypotheses; applet reports number and proportion of times null hypothesis is rejected at 0.05 and 0.01 levels.	6.5 , 343; 6.6 , 344
Hypothesis tests for a mean	Not all tests of hypotheses lead correctly to either rejecting or failing to reject the null hypothesis. Investigates the relationship between the level of confidence and the probabilities of making Type I and Type II errors.	Simulates selecting 100 random samples from population; calculates and plots t statistic and P-value for each. Students specify population distribution shape, mean, and standard deviation; sample size, and null and alternative hypotheses; applet reports number and proportion of times null hypothesis is rejected at both 0.05 and 0.01 levels.	6.1 , 317; 6.2 , 327; 6.3 , 327; 6.4 , 327
Correlation by eye	Correlation coefficient measures strength of linear relationship between two variables. Teaches user how to assess strength of a linear relationship from a scattergram.	Computes correlation coefficient r for a set of bivariate data plotted on a scattergram. Students add or delete points and guess value of r ; applet compares guess to calculated value.	9.2 , 539
Regression by eye	The least squares regression line has a smaller SSE than any other line that might approximate a set of bivariate data. Teaches students how to approximate the location of a regression line on a scattergram.	Computes least squares regression line for a set of bivariate data plotted on a scattergram. Students add or delete points and guess location of regression line by manipulating a line provided on the scattergram; applet plots least squares line and displays the equations and the SSEs for both lines.	9.1 , 512

This page intentionally left blank

A FIRST COURSE IN

STATISTICS

This page intentionally left blank

A FIRST COURSE IN

STATISTICS

TWELFTH EDITION

James T. McClave

Info Tech, Inc.

University of Florida

Terry Sincich

University of South Florida

PEARSON

Boston Columbus Indianapolis New York San Francisco
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Editorial Director: Chris Hoag
Editor in Chief: Deirdre Lynch
Acquisitions Editor: Patrick Barbera
Editorial Assistant: Justin Billing
Program Manager: Tatiana Anacki
Project Manager: Christine O'Brien
Program Management Team Lead:
Karen Wernholm
Project Management Team Lead: Peter Silvia
Media Producer: Jean Choe
TestGen Content Manager: John Flanagan
MathXL Content Manager: Bob 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 Cheselka
Procurement Specialist: Carol Melville
Associate Director of Design:
Andrea Nix
Program Design Lead: Barbara Atkinson
Text Design: Integra
Composition: Integra
Illustrations: Integra
Cover Design: Studio Montage
Cover Images: Future Technology-
Illustrator—Hong Li/Getty Images;
Business people discussing the charts
and graphs showing the results of their
successful teamwork—Pressmaster/
Shutterstock; Woman using
iPhone and iPad Mini—Ivanko80/Fotolia

Copyright © 2017, 2013, 2011, 2008 by Pearson Education, Inc. All Rights Reserved. Printed in the United States of America. This publication is protected by copyright, and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise. For information regarding permissions, request forms and the appropriate contacts within the Pearson Education Global Rights & Permissions department, please visit www.pearsoned.com/permissions/.

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

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

Unless otherwise indicated herein, any third-party trademarks that may appear in this work are the property of their respective owners and any references to third-party trademarks, logos or other trade dress are for demonstrative or descriptive purposes only. Such references are not intended to imply any sponsorship, endorsement, authorization, or promotion of Pearson's products by the owners of such marks, or any relationship between the owner and Pearson Education, Inc. or its affiliates, authors, licensees or distributors.

Library of Congress Cataloging-in-Publication Data

McClave, James T.

A first course in statistics / James T. McClave, Info Tech, Inc., University of Florida, Terry Sincich, University of South Florida. — Twelfth edition.

pages cm

ISBN 978-0-13-408062-8 ((pbk.))

1. Statistics. I. Sincich, Terry. II. Title.

QA276.M378 2017

519.5—dc23

2015027150

1 2 3 4 5 6 7 8 9 10—CRK—18 17 16 15

PEARSON

www.pearsonhighered.com

ISBN 10: 0-13-408062-9

ISBN 13: 978-0-13-408062-8

Contents

Preface xi

Applications Index xix

Chapter 1

Statistics, Data, and Statistical Thinking 1

- 1.1 The Science of Statistics 2
- 1.2 Types of Statistical Applications 3
- 1.3 Fundamental Elements of Statistics 5
- 1.4 Types of Data 9
- 1.5 Collecting Data: Sampling and Related Issues 11
- 1.6 The Role of Statistics in Critical Thinking and Ethics 16
- Statistics in Action:** Social Media Network Usage—Are You Linked In? 2
- Using Technology:** MINITAB: Accessing and Listing Data 25

Chapter 2

Methods for Describing Sets of Data 29

- 2.1 Describing Qualitative Data 31
- 2.2 Graphical Methods for Describing Quantitative Data 42
- 2.3 Numerical Measures of Central Tendency 54
- 2.4 Numerical Measures of Variability 65
- 2.5 Using the Mean and Standard Deviation to Describe Data 71
- 2.6 Numerical Measures of Relative Standing 79
- 2.7 Methods for Detecting Outliers: Box Plots and z -Scores 83
- 2.8 Graphing Bivariate Relationships (Optional) 93
- 2.9 Distorting the Truth with Descriptive Statistics 98
- Statistics in Action:** Body Image Dissatisfaction: Real or Imagined? 30
- Using Technology:** MINITAB: Describing Data 112
- TI-83/TI-84 Plus Graphing Calculator: Describing Data 113

Chapter 3

Probability 115

- 3.1 Events, Sample Spaces, and Probability 117
- 3.2 Unions and Intersections 130
- 3.3 Complementary Events 133
- 3.4 The Additive Rule and Mutually Exclusive Events 135
- 3.5 Conditional Probability 142
- 3.6 The Multiplicative Rule and Independent Events 145

Statistics in Action: Lotto Buster! Can You Improve Your Chance of Winning? 116

Using Technology: TI-83/TI-84 Plus Graphing Calculator: Combinations and Permutations 165

Chapter 4 Random Variables and Probability Distributions 166

- 4.1 Two Types of Random Variables 168
- 4.2 Probability Distributions for Discrete Random Variables 171
- 4.3 The Binomial Random Variable 183
- 4.4 Probability Distributions for Continuous Random Variables 194
- 4.5 The Normal Distribution 196
- 4.6 Descriptive Methods for Assessing Normality 209
- 4.7 Approximating a Binomial Distribution with a Normal Distribution (Optional) 218
- 4.8 Sampling Distributions 223
- 4.9 The Sampling Distribution of \bar{x} and the Central Limit Theorem 230

Statistics in Action: Super Weapons Development—Is the Hit Ratio Optimized? 167

Using Technology: MINITAB: Binomial Probabilities, Normal Probability, and Simulated Sampling Distribution 247

Chapter 5 Inferences Based on a Single Sample 252

- 5.1 Identifying and Estimating the Target Parameter 253
- 5.2 Confidence Interval for a Population Mean: Normal (z) Statistic 255
- 5.3 Confidence Interval for a Population Mean: Student's t -Statistic 265
- 5.4 Large-Sample Confidence Interval for a Population Proportion 275
- 5.5 Determining the Sample Size 282
- 5.6 Confidence Interval for a Population Variance (Optional) 289

Statistics in Action: Medicare Fraud Investigations 253

Using Technology: MINITAB: Confidence Intervals 302

TI-83/TI-84 Plus Graphing Calculator: Confidence Intervals 304

Chapter 6 Inferences Based on a Single Sample 306

- 6.1 The Elements of a Test of Hypothesis 307
- 6.2 Formulating Hypotheses and Setting Up the Rejection Region 313
- 6.3 Observed Significance Levels: p -Values 318
- 6.4 Test of Hypothesis about a Population Mean: Normal (z) Statistic 323
- 6.5 Test of Hypothesis about a Population Mean: Student's t -Statistic 331
- 6.6 Large-Sample Test of Hypothesis about a Population Proportion 338
- 6.7 Test of Hypothesis about a Population Variance (Optional) 346
- 6.8 A Nonparametric Test about a Population Median (Optional) 352

Statistics in Action: Diary of a KLEENEX® User How Many Tissues in a Box? 307

Using Technology: MINITAB: Tests of Hypotheses 364

TI-83/TI-84 Plus Graphing Calculator: Tests of Hypotheses 366

Chapter 7

Comparing Population Means 367

7.1 Identifying the Target Parameter 368

7.2 Comparing Two Population Means: Independent Sampling 369

7.3 Comparing Two Population Means: Paired Difference Experiments 387

7.4 Determining the Sample Size 399

7.5 A Nonparametric Test for Comparing Two Populations: Independent Samples 403

7.6 A Nonparametric Test for Comparing Two Populations: Paired Difference Experiment (Optional) 412

7.7 Comparing Three or More Population Means: Analysis of Variance (Optional) 421

Statistics in Action: ZixIt Corp. v. Visa USA Inc.—A Libel Case 368

Using Technology: MINITAB: Comparing Means 443

TI-83/TI-84 Plus Graphing Calculator: Comparing Means 446

Chapter 8

Comparing Population Proportions 449

8.1 Comparing Two Population Proportions: Independent Sampling 451

8.2 Determining the Sample Size 458

8.3 Testing Category Probabilities: Multinomial Experiment 461

8.4 Testing Categorical Probabilities: Two-Way (Contingency) Table 470

Statistics in Action: The Case of the Ghoulish Transplant Tissue 450

Using Technology: MINITAB: Categorized Data Analyses 496

TI-83/TI-84 Plus Graphing Calculator: Categorical Data Analyses 497

Chapter 9

Simple Linear Regression 499

9.1 Probabilistic Models 501

9.2 Fitting the Model: The Least Squares Approach 505

9.3 Model Assumptions 518

9.4 Assessing the Utility of the Model: Making Inferences about the Slope β_1 523

9.5 The Coefficients of Correlation and Determination 532

9.6 Using the Model for Estimation and Prediction 542

9.7 A Complete Example 550

9.8 A Nonparametric Test for Correlation (Optional) 554

Statistics in Action: Can “Dowsers” Really Detect Water? 500

Using Technology: MINITAB: Simple Linear Regression 573

TI-83/TI-84 Plus Graphing Calculator: Simple Linear Regression 575

Appendices

Appendix A Summation Notation 577

Appendix B Tables 579

Table I Binomial Probabilities 580

Table II Normal Curve Areas 584

Table III Critical Values of t 585

Table IV Critical Values of χ^2 586

Table V Critical Values of T_L and T_U for the Wilcoxon Rank Sum Test 588

Table VI Critical Values of T_0 in the Wilcoxon Signed Rank Test 589

Table VII Percentage Points of the F -Distribution, $\alpha = .10$ 590

Table VIII Percentage Points of the F -Distribution, $\alpha = .05$ 592

Table IX Percentage Points of the F -Distribution, $\alpha = .025$ 594

Table X Percentage Points of the F -Distribution, $\alpha = .01$ 596

Table XI Critical Values of Spearman's Rank Correlation Coefficient 598

Appendix C Calculation Formulas for Analysis of Variance
(Independent Sampling) 599

Short Answers to Selected Odd-Numbered Exercises 600

Index 607

Photo Credits 612

Preface

A First Course in Statistics, 12th edition, is an introductory text designed for one-semester courses that emphasizes inference and sound decision-making through extensive coverage of data collection and analysis. As in earlier editions, the 12th edition text stresses the development of statistical thinking, the assessment of credibility, and value of the inferences made from data, both by those who consume and those who produce them. It assumes a mathematical background of basic algebra.

The text incorporates the following features, developed from the American Statistical Association's (ASA) Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project:

- Emphasize statistical literacy and develop statistical thinking
- Use real data in applications
- Use technology for developing conceptual understanding and analyzing data
- Foster active learning in the classroom
- Stress conceptual understanding rather than mere knowledge of procedures
- Emphasize intuitive concepts of probability

New in the 12th Edition



- **Over 1,000 exercises, with revisions and updates to 30%.** Many new and updated exercises, based on contemporary studies and real data, have been added. Most of these exercises foster and promote critical thinking skills.
- **Updated technology.** All printouts from statistical software (SAS, SPSS, MINITAB, and the TI-83/TI-84 Plus Graphing Calculator) and corresponding instructions for use have been revised to reflect the latest versions of the software.
- **New Statistics in Action Cases.** Almost half of the 9 Statistics in Action cases are new or updated, each based on real data from a recent study.
- **Continued emphasis on Ethics.** Where appropriate, boxes have been added emphasizing the importance of ethical behavior when collecting, analyzing, and interpreting data with statistics.


Content-Specific Changes to This Edition

- **Chapter 1 (Statistics, Data, and Statistical Thinking).** Material on all basic sampling concepts (e.g., random sampling and sample survey designs) has been streamlined and moved to Section 1.5 (Collecting Data: Sampling and Related Issues).
- **Chapter 2 (Methods for Describing Sets of Data).** The section on summation notation has been moved to the appendix (Appendix A). Also, recent examples of misleading graphics have been added to Section 2.10 (Distorting the Truth with Descriptive Statistics).
- **Chapter 4 (Random Variables and Probability Distributions).** Use of technology for computing probabilities of random variables with known probability distributions (e.g., binomial and normal distributions) has been incorporated into the relevant sections of this chapter. This reduces the use of tables of probabilities for these distributions.
- **Chapter 6 (Tests of Hypothesis).** The section on p -values in hypothesis testing (Section 6.3) has been moved up to emphasize the importance of their use in real-life studies. Throughout the remainder of the text, conclusions from a test of hypothesis are based on p -values.

Hallmark Strengths

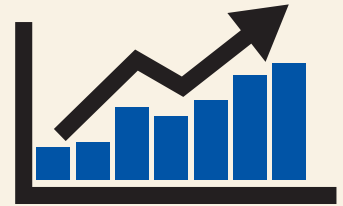
We have maintained or strengthened the pedagogical features of *A First Course in Statistics* that make it unique among introductory statistics texts. These features, which assist the student in achieving an overview of statistics and an understanding of its relevance in both the business world and everyday life, are as follows:

- **Use of Examples as a Teaching Device**—Almost all new ideas are introduced and illustrated by data-based applications and examples. We believe that students better understand definitions, generalizations, and theoretical concepts *after* seeing an application. All examples have three components: (1) “Problem,” (2) “Solution,” and (3) “Look Back” (or “Look Ahead”). This step-by-step process provides students with a defined structure by which to approach problems and enhances their problem-solving skills. The “Look Back” feature often gives helpful hints to solving the problem and/or provides a further reflection or insight into the concept or procedure that is covered.
- **Now Work**—A “Now Work” exercise suggestion follows each example. The Now Work exercise (marked with the icon  in the exercise sets) is similar in style and concept to the text example. This provides the student with an opportunity to immediately test and confirm their understanding.
- **Statistics in Action**—Each chapter begins with a case study based on an actual contemporary, controversial or high-profile issue. Relevant research questions and data from the study are presented and the proper analysis demonstrated in short “Statistics in Action Revisited” sections throughout the chapter. These motivate students to critically evaluate the findings and think through the statistical issues involved.
- **Applet Exercises**—The text is accompanied by applets (short computer programs) available at www.pearsonhighered.com/mathstatsresources and within MyStatLab. These point-and-click applets allow students to easily run simulations that visually demonstrate some of the more difficult statistical concepts (e.g., sampling distributions and confidence intervals.) Each chapter contains several optional applet exercises in the exercise sets. They are denoted with the following icon: .
- **Real Data-Based Exercises**—The text includes more than 1,000 exercises based on applications in a variety of disciplines and research areas. All the applied exercises employ the use of current real data extracted from a current publications (e.g., newspapers, magazines, current journals, and the Internet). Some students have difficulty learning the mechanics of statistical techniques when all problems are couched in terms of realistic applications. For this reason, all exercise sections are divided into four parts:
 - Learning the Mechanics.** Designed as straightforward applications of new concepts, these exercises allow students to test their ability to comprehend a mathematical concept or a definition.
 - Applying the Concepts—Basic.** Based on applications taken from a wide variety of journals, newspapers, and other sources, these short exercises help students begin developing the skills necessary to diagnose and analyze real-world problems.
 - Applying the Concepts—Intermediate.** Based on more detailed real-world applications, these exercises require students to apply their knowledge of the technique presented in the section.
 - Applying the Concepts—Advanced.** These more difficult real-data exercises require students to use their critical thinking skills.
- **Critical Thinking Challenges**—Placed at the end of the “Supplementary Exercises” section only, students are asked to apply their critical thinking skills to solve one or two challenging real-life problems. These exercises expose students to real-world problems with solutions that are derived from careful, logical thought and selection of the appropriate statistical analysis tool.

- **Exploring Data with Statistical Computer Software and the Graphing Calculator**—Each statistical analysis method presented is demonstrated using output from three leading Windows-based statistical software packages: SAS, SPSS, and MINITAB. Students are exposed early and often to computer printouts they will encounter in today’s hi-tech world.
- **“Using Technology” Tutorials**—MINITAB software tutorials appear at the end of each chapter and include point-and-click instructions (with screen shots). These tutorials are easily located and show students how to best use and maximize MINITAB statistical software. In addition, output and keystroke instructions for the TI-84 Graphing Calculator are presented.
- **Profiles of Statisticians in History (Biography)**—Brief descriptions of famous statisticians and their achievements are presented in side boxes. With these profiles, students will develop an appreciation of the statistician’s efforts and the discipline of statistics as a whole.
- **Data and Applets**—The Web site www.pearsonhighered.com/mathstatsresources has files for all the data sets marked with the dataset icon  in the text. These include data sets for text examples, exercises, Statistics in Action and Real-World cases. All data files are saved in three different formats: SAS, MINITAB, and SPSS. This site also contains the applets that are used to illustrate statistical concepts.

Get the most out of MyStatLab™

MyStatLab is the world's leading online resource for teaching and learning statistics. MyStatLab helps students and instructors improve results, and provides engaging experiences and personalized learning for each student so learning can happen in any environment. Plus, it offers flexible and time-saving course management features to allow instructors to easily manage their classes while remaining in complete control, regardless of course format.



Personalized Support for Students

- MyStatLab comes with many learning resources—eText, animations, videos, and more—all designed to support your students as they progress through their course.
- The Adaptive Study Plan acts as a personal tutor, updating in real time based on student performance to provide personalized recommendations on what to work on next. With the new Companion Study Plan assignments, instructors can now assign the Study Plan as a prerequisite to a test or quiz, helping to guide students through concepts they need to master.
- Personalized Homework allows instructors to create homework assignments tailored to each student's specific needs, focused on just the topics they have not yet mastered.

Used by nearly 4 million students each year, the MyStatLab and MyMathLab family of products delivers consistent, measurable gains in student learning outcomes, retention, and subsequent course success.

Resources for Success

Student Resources

Student's Solutions Manual, by Nancy Boudreau (Emeritus Associate Professor, Bowling Green State University), includes complete worked-out solutions to all odd-numbered text exercises (ISBN-13: 978-0-13-408101-4, ISBN-10: 0-13-408101-3).

Excel[®] Manual (download only), by Mark Dummeldinger (University of South Florida). Available for download from www.pearsonhighered.com/mathstatsresources.

Study Cards for Statistics Software. This series of study cards, available for Excel[®], MINITAB, JMP[®], SPSS, R, StatCrunch[®], and TI-83/84 Plus Graphing Calculators, provides students with easy step-by-step guides to the most common statistics software. Visit myPearsonstore.com for more information.

Instructor Resources

Annotated Instructor's Edition contains answers to text exercises. Annotated marginal notes include Teaching Tips, suggested exercises to reinforce the statistical concepts discussed in the text, and short answers to exercises and examples (ISBN-13: 978-0-13-408081-9; ISBN-10: 0-13-408081-5).

Instructor's Solutions Manual (download only), by Nancy Boudreau (Emeritus Associate Professor, Bowling Green State University), includes complete worked-out solutions to all even-numbered text exercises. Careful attention has been paid to ensure that all methods of solution and notation are consistent with those used in the core text.

PowerPoint[®] Lecture Slides include figures and tables from the textbook. Available for download from Pearson's online catalog at www.pearsonhighered.com/irc and in 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 software and test bank are available for download from Pearson Education's online catalog at www.pearsonhighered.com/irc and in MyStatLab.

Online Test Bank, a test bank derived from TestGen[®], is available for download from Pearson's online catalog at www.pearsonhighered.com/irc and in MyStatLab.

Technology Resources

A companion website (www.pearsonhighered.com/mathstatsresources) holds a number of support materials, including:

- **Data sets** formatted as .csv, .txt, .sas7bdat (SAS), .sav (SPSS), .mtp (minitab), .xls (Excel), and TI files
- **Applets** (short computer programs) that allow students to run simulations that visually demonstrate statistical concepts

Acknowledgments

This book reflects the efforts of a great many people over a number of years. First, we would like to thank the following professors, whose reviews and comments on this and prior editions have contributed to the 12th edition:

Ali Arab, *Georgetown University*
 Jen Case, *Jacksonville State University*
 Maggie McBride, *Montana State University—Billings*
 Surajit Ray, *Boston University*
 JR Schott, *University of Central Florida*
 Susan Schott, *University of Central Florida*
 Lewis Shoemaker, *Millersville University*
 Engin Sungur, *University of Minnesota—Morris*
 Sherwin Toribio, *University of Wisconsin—La Crosse*
 Michael Zwilling, *Mt. Union College*

Reviewers of Previous Editions

Bill Adamson, *South Dakota State*; Ibrahim Ahmad, *Northern Illinois University*; Roddy Akbari, *Guilford Technical Community College*; David Atkinson, *Olivet Nazarene University*; Mary Sue Beersman, *Northeast Missouri State University*; William H. Beyer, *University of Akron*; Marvin Bishop, *Manhattan College*; Patricia M. Buchanan, *Pennsylvania State University*; Dean S. Burbank, *Gulf Coast Community College*; Ann Cascarelle, *St. Petersburg College*; Kathryn Chaloner, *University of Minnesota*; Hanfeng Chen, *Bowling Green State University*; Gerardo Chin-Leo, *The Evergreen State College*; Linda Brant Collins, *Iowa State University*; Brant Deppa, *Winona State University*; John Dirkse, *California State University—Bakersfield*; N. B. Ebrahimi, *Northern Illinois University*; John Egenolf, *University of Alaska—Anchorage*; Dale Everson, *University of Idaho*; Christine Franklin, *University of Georgia*; Khadiga Gamgoum, *Northern Virginia CC*; Rudy Gideon, *University of Montana*; Victoria Marie Gribshaw, *Seton Hill College*; Larry Griffey, *Florida Community College*; David Groggel, *Miami University at Oxford*; Sneh Gulati, *Florida International University*; John E. Groves, *California Polytechnic State University—San Luis Obispo*; Dale K. Hathaway, *Olivet Nazarene University*; Shu-ping Hodgson, *Central Michigan University*; Jean L. Holton, *Virginia Commonwealth University*; Soon Hong, *Grand Valley State University*; Ina Parks S. Howell, *Florida International University*; Gary Itzkowitz, *Rowan College of New Jersey*; John H. Kellermeier, *State University College at Plattsburgh*; Golan Kibria, *Florida International University*; Timothy J. Killeen, *University of Connecticut*; William G. Koellner, *Montclair State University*; James R. Lackritz, *San Diego State University*; Diane Lambert, *AT&T/Bell Laboratories*; Edwin G. Landauer, *Clackamas Community College*; James Lang, *Valencia Junior College*; Glenn Larson, *University of Regina*; John J. Lefante, Jr., *University of South Alabama*; Pi-Erh Lin, *Florida State University*; R. Bruce Lind, *University of Puget Sound*; Rhonda Magel, *North Dakota State University*; Linda C. Malone, *University of Central Florida*; Allen E. Martin, *California State University—Los Angeles*; Rick Martinez, *Foothill College*; Brenda Masters, *Oklahoma State University*; Leslie Matekaitis, *Cal Genetics*; E. Donice McCune, *Stephen F. Austin State University*; Mark M. Meerschaert, *University of Nevada—Reno*; Greg Miller, *Steven F. Austin State University*; Satya Narayan Mishra, *University of South Alabama*; Kazemi Mohammed, *UNC—Charlotte*; Christopher Morrell, *Loyola College in Maryland*; Mir Mortazavi, *Eastern New Mexico University*; A. Mukherjea, *University of South Florida*; Steve Nimmo, *Morningside College (Iowa)*; Susan Nolan, *Seton Hall University*; Thomas O’Gorman, *Northern Illinois University*; Bernard Ostle, *University of Central Florida*; William B. Owen, *Central Washington University*; Won J. Park, *Wright State University*; John J. Peterson, *Smith Kline & French*

Laboratories; Ronald Pierce, Eastern Kentucky University; Betty Rehfuss, North Dakota State University—Bottineau; Andrew Rosalsky, University of Florida; C. Bradley Russell, Clemson University; Rita Schillaber, University of Alberta; James R. Schott, University of Central Florida; Susan C. Schott, University of Central Florida; George Schultz, St. Petersburg Junior College; Carl James Schwarz, University of Manitoba; Mike Seyfried, Shippensburg University; Arvind K. Shah, University of South Alabama; Lewis Shoemaker, Millersville University; Sean Simpson, Westchester CC; Charles W. Sinclair, Portland State University; Robert K. Smidt, California Polytechnic State University—San Luis Obispo; Vasanth B. Solomon, Drake University; W. Robert Stephenson, Iowa State University; Thaddeus Tarpey, Wright State University; Kathy Taylor, Clackamas Community College; Barbara Treadwell, Western Michigan University; Dan Voss, Wright State University; Augustin Vukov, University of Toronto; Dennis D. Wackerly, University of Florida; Barbara Wainwright, Salisbury University; Matthew Wood, University of Missouri—Columbia.

Other Contributors

Special thanks are due to our ancillary authors, Nancy Boudreau and Mark Dummeldinger, both of whom have worked with us for many years. Accuracy checkers Dave Bregenzer and Engin Sungur helped ensure a highly accurate, clean text. Finally, the Pearson Education staff of Deirdre Lynch, Patrick Barbera, Christine O'Brien, Chere Bemelmans, Justin Billing, Tiffany Bitzel, Jennifer Myers, Barbara Atkinson and Jean Choe as well as Integra-Chicago's Alverne Ball helped greatly with all phases of the text development, production, and marketing effort.

This page intentionally left blank

Applications Index

Agricultural/gardening/farming applications:

chickens with fecal contamination, 243
colored string preferred by chickens, 264, 361
crop damage by wild boars, 128, 153
crop yield comparisons, 399–400
dehorning of dairy calves, 344
fungi in beech forest trees, 160
killing insects with low oxygen, 346, 491–492
maize seeds, 162
pig castration, 441
plants that grow on Swiss cliffs, 97, 541
rat damage to sugarcane, 460
subarctic plants, 494
USDA chicken inspection, 128
zinc phosphide in pest control, 111

Archaeological applications:

ancient pottery, 106, 159, 297, 489
bone fossils, 329–330
defensibility of a landscape, 345–346, 493
radon exposure in Egyptian tombs, 272, 294, 336–337, 357
shaft graves in ancient Greece, 50, 69, 273–274, 287, 350

Astronomy/space science applications:

astronomy students and the Big Bang theory, 346
lunar soil, 362
measuring the moon's orbit, 504, 513, 521, 547–548, 560
redshifts of quasi-stellar objects, 514
Satellite Database, 21
satellites in orbit, 40
space shuttle disaster, 245
speed of light from galaxies, 109, 110–111
tracking missiles with satellite imagery, 241

Automotive/motor vehicle applications.

See also Aviation applications; Travel applications

air bag danger to children, 300–301
air-pollution standards for engines, 332–334
ammonia in car exhaust, 108–109
car battery guarantee, 74–75
car crash testing, 107, 159–160, 170, 179, 242, 438
critical-part failures in NASCAR vehicles, 237
gas mileage, 202, 210–212
improving driving performance while fatigued, 433
motorcycle detection while driving, 345
motorcyclists and helmets, 17
railway track allocation, 40, 129

red light cameras and car crashes, 398–399, 418–419
safety of hybrid cars, 488–489
satellite radio in cars, 17–18
selecting new-car options, 162
speeding and fatal car crashes, 154
speeding and young drivers, 328
traffic sign maintenance, 457, 469
unleaded fuel costs, 237
variable speed limit control for freeways, 181

Aviation applications:

aircraft bird strikes, 281, 288
classifying air threats with heuristics, 483
“cry wolf” effect in air traffic controlling, 482
flight response of geese to helicopter traffic, 492–493
shared leadership in airplane crews, 381–382
unoccupied seats per flight, 259

Behavioral applications. See also

Gender applications; Psychological applications; Sociological applications

accountants and Machiavellian traits, 360
attempted suicide methods, 140
bullying, 456
cell phone handoff behavior, 141
dating and disclosure, 23, 329
divorced couples, 123–124
employee behavior problems, 141
eye and head movement relationship, 570
fish feeding, 96, 569
interactions in children's museum, 41, 280, 469, 484
Jersey City drug market, 23
last name effect, 180, 382, 402, 539
laughter among deaf signers, 396, 402
married women, 241
money spent on gifts (buying love), 23
parents' behavior at gym meet, 243
personality and aggressive behavior, 263–264
planning-habits survey, 456–457
rudeness in the workplace, 385–386, 411
service without a smile, 386
shock treatment to learners (Milgram experiment), 146
shopping vehicle and judgment, 78, 207, 384
spanking, parents who condone, 241, 362
teacher perceptions of child behavior, 360
time required to complete a task, 330
walking in circles when lost, 338
working on summer vacation, 192, 222

Beverage applications:

alcohol, threats, and electric shocks, 209
alcohol consumption by college students, 264, 490, 491
alcoholic fermentation in wine, 399
bacteria in bottled water, 288
bursting strength of bottles, 410
coffee, caffeine content of, 288, 356
coffee, organic, 345
coffee, overpriced Starbucks, 280
drinking water quality, 21
lead in drinking water, 82
“Pepsi challenge” marketing campaign, 360
Pepsi vs. Coca-Cola, 7–8
restoring self-control when intoxicated, 433
soft-drink bottles, 245
temperature and ethanol production, 433
undergraduate problem drinking, 264

Biology/life science applications. See also Dental applications; Forestry applications; Marine/marine life applications

African rhinos, 128
aircraft bird strikes, 281, 288
antigens for parasitic roundworm in birds, 274, 294
armyworm pheromones, 457
bacteria in bottled water, 288
bacteria-infected spider mites, reproduction of, 274
beetles and slime molds, 467
blond hair types in the Southwest Pacific, 91
body length of armadillos, 107
chemical insect attractant, 160
chemical signals of mice, 141, 193, 233
chickens with fecal contamination, 243
colored string preferred by chickens, 264
comparing measuring instruments, 400–401
crab spiders hiding on flowers, 51–52, 336, 357
crop damage by wild boars, 128, 153
dehorning of dairy calves, 344
DNA-reading tool for quick identification of species, 317
ecotoxicological survival, 222–223
environmental vulnerability of amphibians, 180
extinct birds, 21, 42, 78, 82, 155, 297
fallow deer bucks' probability of fighting, 140–141, 155
fish feeding, 96
fish feeding behavior, 569
flight response of geese to helicopter traffic, 492–493

Biology/life science applications.*(continued)*

giraffe vision, 272, 287, 530–531, 540–541
 great white shark lengths, 338
 habitats of endangered species, 216
 identifying organisms using
 computers, 345
 inbreeding of tropical wasps, 299, 361
 killing insects with low oxygen, 346,
 491–492
 Mongolian desert ants, 63, 97, 170–171,
 440, 514, 522, 548, 560
 mortality of predatory birds, 570
 parrot fish weights, 361
 pig castration, 441
 radioactive lichen, 108, 298, 362
 rainfall and desert ants, 272, 560
 rat damage to sugarcane, 460
 rat-in-maze experiment, 72–73
 rhino population, 39
 roaches and Raid fumigation, 264
 salmonella in food, 300, 457
 shrimp quality, 356
 supercooling temperature of frogs, 244
 swim maze study of rat pups, 441
 USDA chicken inspection, 128
 water hyacinth control, 179–180
 zoo animal training, 300

Business applications:

accountant salary survey, 300
 accountants and Machiavellian traits, 360
 blood diamonds, 153, 222
 brokerage analyst forecasts, 139
 brown-bag lunches at work, 299
 child labor in diamond mines, 541
 consumer sentiment on state of
 economy, 277–278
 corporate sustainability, 22, 50, 61–62,
 77, 92, 236, 262, 293, 328
 employee behavior problems, 141
 employee performance ratings, 208
 executive coaching and meeting
 effectiveness, 209
 executives who cheat at golf, 143
 expected value of insurance, 174–175
 facial structure of CEOs, 263, 294, 329
 gender and salaries, 88–89, 392–393
 goodness-of-fit test with monthly
 salaries, 494–495
 job satisfaction of women in
 construction, 483
 lawyer salaries, 100
 museum management, 41–42, 102,
 129, 467
 nannies who worked for celebrities, 280
 nice guys finish last, 512–513, 521, 547, 562
 overpriced Starbucks coffee, 280
 “Pepsi challenge” marketing
 campaign, 360
 rudeness in the workplace, 385–386, 411
 salary linked to height, 540
 self-managed work teams and family
 life, 442
 shopping on Black Friday, 263, 288
 shopping vehicle and judgment, 78,
 207, 384
 trading skills of institutional
 investors, 349–350

work-life balance, 553–554
 Zillow.com estimates of home values, 22

Chemicals/chemistry applications.

*See also Medical/medical research/
 alternative medicine applications*

chemical insect attractant, 160
 chemical signals of mice, 141, 193, 233
 drug content assessment, 215–216,
 351, 384–385
 firefighters’ use of gas detection
 devices, 154
 mineral flotation in water, 64, 216, 387
 oxygen bubbles in molten salt, 274
 roaches and Raid fumigation, 264
 Teflon-coated cookware hazards, 238
 toxic chemical incidents, 160–161
 zinc phosphide in pest control, 111

**Computer applications. See Electronics/
 computer applications****Construction/home improvement/home
 purchases and sales applications:**

bending strength of wooden roof, 298
 land purchase decision, 79
 levelness of concrete slabs, 244
 load on frame structures, 209
 predicting sale prices of homes, 566
 spall damage in bricks, 572
 strand bond performance of pre-
 stressed concrete, 351

**Crime applications. See also Legal/
 legislative applications**

burglary risk in cul-de-sacs, 287
 casino employment and crime, 534–535
 computer, 21
 domestic abuse victims, 193–194
 effectiveness ratings by crime
 prevention experts, 415–416
 gender attitudes toward corruption
 and tax evasion, 409–410
 Jersey City drug market, 23
 masculinity and crime, 386, 492
 Medicare fraud investigations, 253,
 270–271, 279, 286, 301
 motivation of drug dealers, 77, 82, 171,
 237, 262, 293–294, 351
 post office violence, 159
 victims of violent crime, 278–279

Dental applications:

acidity of mouthwash, 397–398
 anesthetics, dentists’ use of, 77, 91
 cheek teeth of extinct primates, 38, 50,
 62, 70, 128–129, 294, 336, 356
 dental bonding agent, 361
 dental visit anxiety, 207, 336
 laughing gas usage, 241
 teeth defects and stress in prehistoric
 Japan, 458

Earth science applications. See also

**Agricultural/gardening/farming
 applications; Environmental
 applications; Forestry applications**
 albedo of ice melt ponds, 262
 alkalinity of river water, 360

daylight duration in western
 Pennsylvania, 273, 288
 deep mixing of soil, 207
 dissolved organic compound in lakes,
 337–338
 dowsers for water detection, 500,
 510–511, 527, 538, 546
 earthquake aftershocks, 59–60
 earthquake ground motion, 20
 estimating well scale deposits, 397
 glacial drifts, 107
 glacier elevations, 215
 ice melt ponds, 40, 281, 468
 identifying urban land cover, 360
 permeability of sandstone during
 weathering, 63–64, 70, 78, 92–93, 217
 quantum tunneling, 571
 rockfall rebound length, 61, 69–70, 92,
 293, 350, 357
 shear strength of rock fractures, 215
 soil scouring and overturned trees, 432

**Education/school applications. See also
 Library/book applications**

blue vs. red exam, 82
 bullying behavior, 456
 calories in school lunches, 317
 children’s attitude toward reading,
 242–243
 college application, 20
 college entrance exam scores, 204
 college protests of labor exploitation,
 566–568
 compensatory advantage in education,
 154–155
 delinquent children, 101
 ESL reading ability, 569
 ESL students and plagiarism, 129
 family involvement in homework, 412
 FCAT math test, 242
 FCAT scores and poverty, 515–516,
 522, 530
 food availability at middle schools,
 420, 563
 gambling in high schools, 441–442
 grades in statistics courses, 111
 humane education and classroom
 pets, 38–39
 insomnia and education status, 22
 instructing English-as-a-first-language
 learners, 330–331
 interactions in children’s museum, 41,
 280, 469, 484
 IQ and *The Bell Curve*, 245
 Japanese reading levels, 106–107
 math scores, 82
 passing grade scores, 194
 ranking Ph.D. programs in economics,
 83, 217
 RateMyProfessors.com, 539
 reading comprehension, 419
 SAT scores, 30, 52–53, 80, 92, 95, 108,
 242, 422
 sensitivity of teachers towards racial
 intolerance, 398, 419
 standardized test “average,” 111
 STEM experiences for girls, 20, 39, 128
 student gambling on sports, 243
 student GPAs, 20–21, 83

teacher perceptions of child behavior, 360
 teaching method comparisons, 369–379
 teaching software effectiveness, 382
 teenagers' use of emoticons in writing, 281, 344
 text messaging in class, 409
 untutored second language acquisition, 93
 using game simulation to teach a course, 129–130
 visually impaired students, 243

Elderly/older-person applications:

Alzheimer's detection, 468, 483
 Alzheimer's treatment, 299–300
 dementia and leisure activities, 398
 personal networks of older adults, 297
 wheelchair users, 162

Electronics/computer applications:

cell phone charges, 200–201
 cell phone defects, 285–286
 cell phone handoff behavior, 141
 cell phone use, 246
 cell phones, Short Message Service (SMS) for, 438–439
 charge length of iPod batteries, 354–355
 college tennis recruiting with Web site, 439
 computer crimes, 21
 cyberchondria, 159
 downloading apps to cell phone, 179
 encoding variability in software, 142
 encryption systems with erroneous ciphertexts, 157
 flicker in an electrical power system, 208
 forecasting movie revenues with Twitter, 505, 550
 identifying organisms using computers, 345
 Internet addiction, 15
 intrusion detection systems, 156, 318
 Microsoft program security issues, 39
 mobile device typing strategies, 468, 483
 monitoring quality of power equipment, 162
 network forensic analysis, 244–245
 paper friction in photocopier, 195
 paying for music downloads, 38, 280, 344
 phishing attacks to email accounts, 53, 236–237, 295
 requests to a Web server, 237
 robot-sensor system configuration, 182
 robots trained to behave like ants, 432
 satellite radio in cars, 17–18
 scanning errors at Wal-Mart, 139, 297–298, 359
 series and parallel systems, 162–163
 social robots walking and rolling, 38, 76–77, 127, 139, 153, 179, 273, 281, 287, 467
 software file updates, 215
 solder joint inspections, 362–363
 teaching software effectiveness, 382
 testing electronic circuits, 441

text messaging in class, 409
 transmission delays in wireless technology, 242
 versatility with resistor-capacitor circuits, 484
 visual attention of video game players, 238, 384, 402–403
 voltage sags and swells, 82, 92, 208, 236
 vulnerability of relying party Web sites, 458
 wear-out failure time display panels, 244
 Web survey response rates, 456

Entertainment applications. See also

Gambling applications

ages of Broadway ticketbuyers, 7
 cable-TV home shoppers, 460
 children's recall of TV ads, 383, 410
 coin toss, 118–119, 122, 127, 134–137, 164, 171–173
 craps game outcomes, 172–173
 data in the news, 24
 die toss, 121–122, 127, 131, 148–149
 effectiveness of TV program on marijuana use, 464–465
 forecasting movie revenues with Twitter, 505, 550
 game show “Monty Hall Dilemma” choices, 485
 Howard Stern on Sirius radio, 17–18
 “Let's Make a Deal,” 164
 life expectancy of Oscar winners, 439
 media and attitudes toward tanning, 431–432
 movie selection, 125
 music performance anxiety, 50, 61, 69, 272, 335–336, 355
 “name game,” 434, 517, 531, 541, 550, 561–562
 newspaper reviews of movies, 125
 Odd Man Out game, 164
 parlay card betting, 182
 paying for music downloads, 38, 280, 344
 recall of TV commercials, 432
 religious symbolism in TV commercials, 458
 scary movies, 299
 Scrabble game analysis, 469
 size of TV households, 179
 sports news on local TV broadcasts, 566
 TV buyers, 460
 TV subscription streaming, 344
 20/20 survey exposés, 23–24
 using game simulation to teach a course, 129–130
 visual attention of video game players, 238, 384, 402–403
 “winner's curse” in auction bidding, 489

Environmental applications. See also

Earth science applications; Forestry applications

air-pollution standards for engines, 332–334
 aluminum cans contaminated by fire, 287
 ammonia in car exhaust, 108–109
 beach erosional hot spots, 160, 181
 contaminated fish, 289–292

contaminated river, 10–11
 dissolved organic compound in lakes, 337–338
 drinking water quality, 21
 environmental vulnerability of amphibians, 180
 fecal pollution, 245–246
 fire damage, 550–553
 groundwater contamination in wells, 42, 108, 356, 562
 hotel water conservation, 121
 ice melt ponds, 40, 281, 468
 lead in drinking water, 82
 natural-gas pipeline accidents, 157
 oil spill and seabirds, 102, 109–110, 438, 488
 PCB in plant discharge, 361
 power plant environmental impact, 439
 removing nitrogen from toxic wastewater, 548–549
 sea turtles and beach nourishment, 419–420
 soil scouring and overturned trees, 432
 water pollution testing, 298
 whales entangled in fishing gear, 431

Exercise applications. See Sports/exercise/fitness applications

Farming applications. See Agricultural/gardening/farming applications

Fitness applications. See Sports/exercise/fitness applications

Food applications. See also Agricultural/gardening/farming applications; Beverage applications; Health/health care applications

calories in school lunches, 317
 colors of M&Ms candies, 128
 food availability at middle schools, 420, 563
 honey as cough remedy, 51, 62, 70, 92, 294–295, 410–411, 433–434
 Hot Tamale caper, 363
 oil content of fried sweet potato chips, 294, 351
 oven cooking, 298–299
 package design and taste, 482
 packaging of children's health food, 329, 395
 red snapper served in restaurants, 155, 281
 red vs. yellow gummy bears and their flavors, 344
 salmonella, 300, 457
 shrimp quality, 356
 steak as favorite barbecue food, 456
 sweetness of orange juice, 516, 522, 530, 548, 549
 taste test rating protocols, 383
 taste-testing scales, 539, 562
 tomato as taste modifier, 207, 237

Forestry applications. See also

Environmental applications

forest fragmentation, 97, 163, 530
 fungi in beech forest trees, 160
 tractor skidding distance, 274, 337, 357

Gambling applications. See also**Entertainment applications**

casino gaming, 207
 chance of winning at blackjack, 164
 chance of winning at craps, 164, 226–227
 craps game outcomes, 172–173
 Galileo's passe-dix game, 142
 gambling in high schools, 494
 game show "Monty Hall Dilemma" choices, 485
 jai alai Quinella betting, 129
 "Let's Make a Deal," 164
 odds of winning a horse race, 164
 odds of winning Lotto, 116, 126, 137, 151–152, 181
 parlay card betting, 182
 roulette, odds of winning at, 161, 181–182
 student gambling on sports, 243

Gardening applications. See Agricultural/gardening/farming applications**Gender applications:**

distribution of boys in families, 194
 gender attitudes toward corruption and tax evasion, 409–410
 gender in two-child families, 180, 468
 height, 209, 516–517
 job satisfaction of women in construction, 483
 masculinity and crime, 386, 492
 masculinizing human faces, 360
 salaries and gender, 88–89, 392–393
 sex composition patterns of children in families, 163
 voting on women's issues, 528

Genetics applications:

birth order and IQ, 329
 dominant vs. recessive traits, 130
 gene expression profiling, 139
 IQ and *The Bell Curve*, 245
 light-to-dark transition of genes, 440–441
 maize seeds, 162
 Punnett square for earlobes, 182
 random mutation of cells, 156

Health/health care applications.

See also Beverage applications;

Dental applications; Environmental applications; Food applications;

Genetics applications; Medical/medical research/alternative medicine applications; Safety applications

air bag danger to children, 300–301
 birth weights of cocaine babies, 351
 blood pressure, 262, 267–268
 body fat in men, 223
 CDC health survey, 297
 cigar smoking and cancer, 161
 cigarette advertisements, 314
 cigarette smoking, 143–145, 557–558
 cruise ship sanitation inspection, 51, 77, 82, 92, 217
 cyberchondria, 159
 dementia and leisure activities, 398

drinking water quality, 21
 hand washing vs. hand rubbing, 78, 238
 health risks to beachgoers, 128, 154
 heart rate variability (HRV) of police officers, 261
 hygiene of handshakes, high fives, and fist bumps, 385, 402
 inflammation in children, 409
 insomnia and education status, 22
 latex allergy in health care workers, 262, 300, 350–351
 lung cancer CT scanning, 22
 media and attitudes toward tanning, 431–432
 Medicare fraud investigations, 253, 270–271, 279, 286, 301
 MS and exercise, 442
 muscle, fat, and bone issues while aging, 155–156
 neurological impairment of POWs, 420
 packaging of children's health food, 329, 395
 passing physical fitness examination, 184–188
 physical activity of obese young adults, 237, 540
 sleep and mental performance, 458
 sleep deprivation, 359
 summer weight-loss camp, 395
 Teflon-coated cookware hazards, 238
 undergraduate problem drinking, 264
 waking sleepers early, 274–275
 walking to improve health, 317
 weight loss diets, 369–373
 when sick at home, 281

Home improvement. See Construction/home improvement/home purchases and sales applications**Home maintenance applications:**

burglary risk in cul-de-sacs, 287
 dye discharged in paint, 245
 portable grill displays selection, 129, 181, 362
 roaches and Raid fumigation, 264
 tissues, number in box, 307, 316, 327, 342–343

Home purchases and sales applications.

See Construction/home improvement/home purchases and sales applications

Legal/legislative applications. Crime applications

cloning credit or debit cards, 141–142
 credit card lawsuit, 368, 379–380, 408
 curbing street gang gun violence, 41, 281, 468
 drivers stopped by police, 82
 eyewitnesses and mug shots, 481
 federal civil trial appeals, 161, 361–362
 fingerprint expertise, 193, 223
 gender attitudes toward corruption and tax evasion, 409–410
 heart rate variability (HRV) of police officers, 261
 jury trial outcomes, 318

lead bullets as forensic evidence, 130
 legal advertising, 554
 lie detector test, 162, 362
 patent infringement case, 439–440
 polygraph test error rates, 362
 racial profiling by the LAPD, 489
 recall notice sender and lawsuits, 477–479
 scallop harvesting and the law, 301

Library/book applications:

importance of libraries, 37
 library book checkouts, 93
 library cards, 160
 reading Japanese books, 106–107
 reading tongue twisters, 439

Life science applications. See Biology/life science applications; Marine/marine life applications**Manufacturing applications. See**

also Automotive/motor vehicle applications; Construction/home improvement/home purchases and sales applications

active nuclear power plants, 64–65, 70
 aluminum smelter pot life span, 569–570
 bursting strength of bottles, 410
 child labor in diamond mines, 541
 consumer complaints, 146, 149
 contaminated gun cartridges, 180
 cooling method for gas turbines, 330, 351
 corrosion prevention of buried steel structures, 20
 cutting tool life span tests, 523, 550
 defect rate comparison between machines, 458–459
 defective batteries, 340–341
 estimating repair and replacement costs of water pipes, 515, 528
 flexography printing plates, evaluation of, 432
 freckling of superalloy ingots, 109
 increasing hardness of polyester composites, 337
 lot acceptance sampling, 220–221
 metal lathe quality control, 314
 preventing production of defective items, 288
 quality control monitoring, 314
 reliability of a manufacturing network, 181
 settlement of shallow foundations, 396–397, 418
 soft-drink bottles, 245
 softness ratings of paper, 412–413
 solar energy cells, 180, 397, 402, 419, 553
 spall damage in bricks, 572
 temperature and ethanol production, 433
 testing manufacturer's claim, 234–235
 thickness of steel sheets, 227–228
 twinned drill holes, 395–396
 weapons development, 167, 205–206, 212–213
 when to replace a maintenance system, 243

wind turbine blade stress, 566
yield strength of steel connecting bars, 337

Marine/marine life applications, 96

contaminated fish, 289–292
lobster fishing, 529, 540, 561
lobster trap placement, 273, 287, 294, 335, 357, 383–384
mercury levels in wading birds, 318
oil spill and seabirds, 102, 109–110, 438, 488
rare underwater sounds, 128
scallop harvesting and the law, 301
sea turtles and beach nourishment, 419–420
sea turtles' shell lengths, 69, 207, 237, 264, 273, 294
shrimp quality, 356
underwater acoustic communication, 194, 345
underwater sound-locating abilities of alligators, 344
whales entangled in fishing gear, 431
whistling dolphins, 109

Medical/medical research/alternative medicine applications. See also Dental applications; Genetics applications; Health/health care applications

abortion provider survey, 140
accuracy of pregnancy tests, 164
Alzheimer's detection, 468, 483
Alzheimer's treatment, 299–300
ambulance response time, 156, 208
angioplasty's benefits challenged, 457, 460
animal-assisted therapy for heart patients, 78–79, 434, 439
asthma drug, 299–300
blood typing method, 96, 514, 521–522
brain specimen research, 52, 91, 299
bulimia, 383, 402
Caesarian births, 193, 222
cancer and smoking, 143–145
cardiac stress testing, 153
characterizing bone with fractal geometry, 531
comparing measuring instruments, 400–401
contact lenses for myopia, 64
dance/movement therapy, 571
dementia and leisure activities, 398
depression treatment, 437–438
drug content assessment, 215–216, 351, 384–385
drug designed to reduce blood loss, 33–35
drug response time, 315–316, 324–325, 520, 525, 537, 543–544
drug testing, 130, 492
dust mite allergies, 242
eating disorders, 52, 243
effectiveness of TV program on marijuana use, 464–465
emergency room waiting time, 223
errors in medical tests, 361
ethnicity and pain perception, 386–387
eye refraction, 64

eye shadow, mascara, and nickel allergies, 282, 288
FDA mandatory new-drug testing, 359
fitness of cardiac patients, 244
gestation time for pregnant women, 243–244
guided bone regeneration, 410
healing potential of handling museum objects, 396, 418
heart patients, healing with music, imagery, touch, and prayer, 481–482
heart rate during laughter, 329
herbal medicines and therapy, 21, 359
HIV vaccine efficacy, 484–485
honey as cough remedy, 51, 62, 70, 92, 294–295, 433–434
hospital administration of malaria patients, 456
hospital admissions, 135–136
hospital stay, length of, 94–95, 255–257, 326
interocular eye pressure, 362
jaw dysfunction, 466–467
LASIK surgery complications, 222
latex allergy in health care workers, 262, 300, 350–351
lung cancer CT scanning, 22
male fetal deaths following 9/11/2001, 346
MS and exercise, 442
multiple sclerosis drug, 492
olfactory reference syndrome (ORS), 282, 288
pain empathy and brain activity, 531, 563
pain tolerance, 541
perceptions of emergency medical residents, 411–412
placebo effect and pain, 396
post-op nausea, 130
psoriasis treatment with “Doctorfish of Kangal,” 92, 417–418
reaction time to drugs, 405–407, 503, 507–510
skin cancer treatment, 176–177
skin cream effectiveness, 363
sleep apnea and sleep stage transitioning, 139–140, 154
splinting in mountain-climbing accidents, 281–282
stability of compounds in drugs, 49–50, 82, 328
sterile couples in Jordan, 159
teamwork between doctors and nurses, 411
tendon pain treatment, 420–421
transplants, 164, 450, 477–479
virtual reality hypnosis for pain, 318
yoga for cancer patients, 431

Miscellaneous applications:

Benford's Law of Numbers, 110
box plots and standard normal distribution, 209
customers in line at Subway shop, 170
evaporation from swimming pools, 262–263
fill weight variance, 346–349
impact of dropping ping-pong balls, 553
jitter in water power system, 300

marine selection, 125
matching socks, 130
National Bridge Inventory, 21
national firearms survey, 153, 280–281
psychic ability, 156, 194
quantitative models of music, 522
questionnaire mailings, 245
random numbers, 19–20
randomly sampling households, 13
regression through the origin, 571–572
selecting a random sample of students, 160
sound waves from a basketball, 52, 96–97, 170, 516, 549
spreading rate of spilled liquid, 97–98, 517–518, 532, 550
symmetric vs. skewed data sets, 63
testing normality, 495
TNT detection, 156
Winchester bullet velocity, 78

Motor vehicle applications. See Automotive/motor vehicle applications

Nuclear applications. See under Manufacturing applications

Political applications:

beauty and electoral success, 529
blood diamonds, 153, 222
border protection avatar, 318
consumer sentiment on state of economy, 277–278
countries allowing free press, 243
electoral college votes, 208
exit polls, 165
Iraq War casualties, 102
political poll size, 460
political representation of religious groups, 469
politics and religion, 489–490
public opinion polls, 275
rigged election, 495
trust in president, 275
U.S. Treasury deficit prior to Civil War, 21
verifying voter petitions, 363
voting for mayor, 190–191
voting in primary elections, 193
voting on women's issues, 528

Psychological applications. See also

Behavioral applications; Gender applications; Religion applications; Sociological applications
alcohol, threats, and electric shocks, 209
appraisals and negative emotions, 154
attention time given to twins, 298
birth order and IQ, 329
body image dissatisfaction, 30, 35–37, 48, 75–76, 90
bulimia, 383, 402
characteristics of antiwar demonstrators, 77–78, 215, 238
children's perceptions of their neighborhood, 480
children's recall of TV ads, 383, 410
choosing a mother, 23

Psychological applications.*(continued)*

cognitive impairment of schizophrenics, 382
 cognitive skills for successful arguing, 385
 dental visit anxiety, 207, 336
 detecting rapid visual targets and attentional blink, 522
 divorced couples, 150–151
 eating disorders, 52, 243
 effectiveness of TV program on marijuana use, 464–465
 emotional empathy in young adults, 329, 356
 free recall memory strategy, 337, 357
 guilt in decision making, 22, 140, 153–154, 494
 influencing performance in a serial addition task, 456, 460, 482
 interactions in children's museum, 41, 280, 469, 484
 Internet addiction, 15
 IQ and mental deficiency, 493
 irrelevant speech effects, 49, 77, 102–103, 171, 214–215, 261–262, 330, 350
 lie detector test, 162, 362
 listening time of infants, 317
 married women, 241
 money spent on gifts (buying love), 23
 motivation and right-oriented bias, 41
 motivation of drug dealers, 77, 82, 171, 237, 262, 293–294, 351
 music performance anxiety, 50, 61, 69, 272, 335–336, 355
 olfactory reference syndrome (ORS), 282, 288
 personality and aggressive behavior, 263–264
 pitch memory of amusiacs, 273, 288, 337
 post-traumatic stress of POWs, 360–361
 rat-in-maze experiment, 72–73
 recall of TV commercials, 432
 restoring self-control when intoxicated, 433
 rotating objects, view of, 540
 shock treatment to learners (Milgram experiment), 146
 shopping vehicle and judgment, 78, 207, 384
 sleep deprivation, 359
 social interaction of mental patients, 330
 spanking, parents who condone, 241, 362
 stimulus reaction, 87–88
 superstition survey, 192–193

susceptibility to hypnosis, 15–16, 244, 493–494
 task deviations, 46–47
 time required to complete a task, 330
 “tip-of-the-tongue” phenomenon, 457
 undergraduate problem drinking, 264
 virtual reality hypnosis for pain, 318
 visual search and memory, 398
 water-level task, 46–47, 86

Religion applications:

belief in an afterlife, 242
 belief in Bible, 41
 marital status and religion, 475–476
 political representation of religious groups, 469
 politics and religion, 489–490
 religious symbolism in TV commercials, 458

Safety applications. See also Health/health care applications

hybrid cars, 488–489
 underground tunnels, 208

School applications. See Education/school applications**Sociological applications. See also Behavioral applications; Gender applications; Psychological applications**

acquiring a pet, 193, 222, 242
 family planning, 132–133
 fieldwork methods, 108, 163, 490
 genealogy research, 39–40
 Generation Y's entitlement mentality, 528–529
 Hite Report, 111
 ideal height of mate, 516–517, 523, 530, 549–550
 identical twins reared apart, 440
 marital name change, 193, 223
 mother's race and maternal age, 132–133
 salary linked to height, 540
 single-parent families, 361
 social network usage, 2, 9, 16, 18, 139
 stereotyping deceptive and authentic news stories, 481
 welfare workers, 146–148

Space science applications. See**Astronomy/space science applications****Sports/exercise/fitness applications:**

altitude effects on climbers, 438
 baseball batting averages, 216

baseball batting averages vs. wins, 566–568
 basketball shooting free throws, 162
 bowler's hot hand, 421
 drafting football quarterbacks, 20
 drug testing of athletes, 492
 elevation and baseball hitting performance, 96, 531–532
 executives who cheat at golf, 143
 exercise workout dropouts, 299
 favorite sport, 317
 football fourth down tactics, 539–540
 football speed training, 264, 298
 game performance of water polo players, 504–505, 513–514, 521, 549, 560–561
 golf ball specifications, 194
 golf ball tests, 299
 golfers' driving performance, 62–63, 98, 216–217, 410, 515, 530, 548
 inflation pressure of footballs, 284
 long-jump takeoff error, 572
 marathon winning times, 570–571
 massage, effect on boxers, 22, 530, 541, 561
 odds of winning a horse race, 164
 parents' behavior at a gym meet, 243
 physical activity of obese young adults, 237, 540
 Play Golf America program, 317
 point spreads of football games, 351
 professional athlete salaries, 111
 scouting a football free agent, 403
 soccer goal target, 208
 sports news on local TV broadcasts, 566
 sprint speed training, 20
 student gambling on sports, 243
 walking to improve health, 317

Travel applications. See also**Automotive/motor vehicle applications; Aviation applications**

cruise ship sanitation inspection, 51, 77, 82, 92, 217
 hotels, ratings of five-star, 438
 purchasing souvenirs, 490–491
 travel manager salaries, 242
 unleaded fuel costs, 237

Weather applications:

chance of rainfall, 129
 rainfall and desert ants, 272, 560
 rainfall estimation, 569
 Texas droughts, 173–174



1 Statistics, Data, and Statistical Thinking

CONTENTS

- 1.1 The Science of Statistics
- 1.2 Types of Statistical Applications
- 1.3 Fundamental Elements of Statistics
- 1.4 Types of Data
- 1.5 Collecting Data: Sampling and Related Issues
- 1.6 The Role of Statistics in Critical Thinking and Ethics

Where We're Going

- Introduce the field of statistics (1.1)
- Demonstrate how statistics applies to real-world problems (1.2)
- Introduce the language of statistics and the key elements to any statistical problem (1.3)
- Differentiate between population and sample data (1.3)
- Differentiate between descriptive and inferential statistics (1.3)
- Identify the different types of data and data collection methods (1.4–1.5)
- Discover how critical thinking through statistics can help improve our quantitative literacy (1.6)

Statistics IN Action Social Media Network Usage— Are You Linked In?

The Pew Research Center, a nonpartisan organization funded by a Philadelphia-based charity, has conducted more than 100 surveys on Internet usage in the United States as part of the Pew Internet & American Life Project (PIALP). In a recent report titled *Social Media Update, 2013*, the PIALP examined adults' (ages 18 and up) attitudes and behavior toward online social media networks. Regarded merely as a fun, online activity for high school and college students just a few years ago, social media now exert tremendous influence over the way people around the world—of all ages—get and share information. The five social media sites investigated in this report include Facebook, Twitter, Instagram, Pinterest, and LinkedIn. The Pew Research Center contacted 1,445 Internet users via landline telephone or cell phone for the survey.

Several of the many survey questions asked are provided here as well as the survey results:

- **Social Networking:**

When asked if they ever use an online social networking site, adults responded:

Yes	73%
No	27%

- **Facebook Usage:**

When Facebook users were asked how often they visit the social media site, they responded:

Several times a day	40%
About once a day	24%
3–5 days a week	10%
1–2 days a week	13%
Every few weeks	6%
Less often	7%

- **Twitter Usage:**

When asked if they ever use Twitter, adults responded:

Yes	18%
No	82%



- **Overall Social Media Usage:**

When asked about how many of the five social networking sites they use, adults responded:

0	22%
1	36%
2	23%
3	12%
4	5%
5	2%

(Average = 1.48 sites)

In the following “Statistics in Action Revisited” sections, we discuss several key statistical concepts covered in this chapter that are relevant to the Pew Internet & American Life Project survey.

Statistics IN Action Revisited

- Identifying the Population, Sample, and Inference (p. 9)
- Identifying the Data Collection Method and Data Type (p. 16)
- Critically Assessing the Ethics of a Statistical Study (p. 18)

1.1 The Science of Statistics

What does statistics mean to you? Does it bring to mind batting averages, Gallup polls, unemployment figures, or numerical distortions of facts (lying with statistics!)? Or is it simply a college requirement you have to complete? We hope to persuade you that statistics is a meaningful, useful science whose broad scope of applications to business, government, and the physical and social sciences is almost limitless. We also want to show that statistics can lie only when they are misapplied. Finally, we wish to demonstrate the key role statistics plays in critical thinking—whether in the classroom, on the job, or in everyday life. Our objective is to leave you with the impression that the time you spend studying this subject will repay you in many ways.

The *Random House College Dictionary* defines **statistics** as “the science that deals with the collection, classification, analysis, and interpretation of information or data.”

Thus, a statistician isn't just someone who calculates batting averages at baseball games or tabulates the results of a Gallup poll. Professional statisticians are trained in *statistical science*. That is, they are trained in collecting information in the form of **data**, evaluating the information, and drawing conclusions from it. Furthermore, statisticians determine what information is relevant in a given problem and whether the conclusions drawn from a study are to be trusted.

Statistics is the science of data. This involves collecting, classifying, summarizing, organizing, analyzing, presenting, and interpreting numerical and categorical information.

In the next section, you'll see several real-life examples of statistical applications that involve making decisions and drawing conclusions.

1.2 Types of Statistical Applications

“Statistics” means “numerical descriptions” to most people. Monthly housing starts, the failure rate of liver transplants, and the proportion of African-Americans who feel brutalized by local police all represent statistical descriptions of large sets of data collected on some phenomenon. (Later, in Section 1.4, we learn that not all data is numerical in nature.) Often the data are selected from some larger set of data whose characteristics we wish to estimate. We call this selection process *sampling*. For example, you might collect the ages of a sample of customers who shop for a particular product online to estimate the average age of *all* customers who shop online for the product. Then you could use your estimate to target the Web site's advertisements to the appropriate age group. Notice that statistics involves two different processes: (1) describing sets of data and (2) drawing conclusions (making estimates, decisions, predictions, etc.) about the sets of data on the basis of sampling. So, the applications of statistics can be divided into two broad areas: **descriptive statistics** and **inferential statistics**.

Descriptive statistics utilizes numerical and graphical methods to look for patterns in a data set, to summarize the information revealed in a data set, and to present that information in a convenient form.

Inferential statistics utilizes sample data to make estimates, decisions, predictions, or other generalizations about a larger set of data.

BIOGRAPHY FLORENCE NIGHTINGALE (1820–1910)

The Passionate Statistician

In Victorian England, the “Lady of the Lamp” had a mission to improve the squalid field hospital conditions of the British army during the Crimean War. Today, most historians consider Florence Nightingale to be the founder of the nursing profession. To convince members of the British Parliament of the need for supplying nursing and medical care to soldiers in the field, Nightingale compiled massive amounts of data from army files. Through a remarkable series of graphs (which included the first pie chart), she demonstrated that most of the deaths in the war either were due to illnesses contracted outside the battlefield or occurred long after battle action from wounds that went untreated. Florence Nightingale's compassion and self-sacrificing nature, coupled with her ability to collect, arrange, and present large amounts of data, led some to call her the Passionate Statistician. ■

Although we'll discuss both descriptive and inferential Statistics in the chapters that follow, the primary theme of the text is **inference**.

Let's begin by examining some studies that illustrate applications of statistics.

Study 1.1 “Best-Selling Girl Scout Cookies” (Source: www.girlscouts.org)

Since 1917, the Girl Scouts of America have been selling boxes of cookies. Currently, there are 12 varieties for sale: Thin Mints, Samoas, Lemonades, Tagalongs, Do-si-dos, Trefoils,

Savannah Smiles, Thanks-A-Lot, Dulce de Leche, Cranberry Citrus Crisps, Chocolate Chip, and Thank U Berry Much. Each of the approximately 150 million boxes of Girl Scout cookies sold each year is classified by variety. The results are summarized in Figure 1.1. From the graph, you can clearly see that the best-selling variety is Thin Mints (25%), followed by Samoas (19%) and Tagalongs (13%). Since the figure describes the various categories of boxes of Girl Scout cookies sold, the graphic is an example of descriptive statistics.

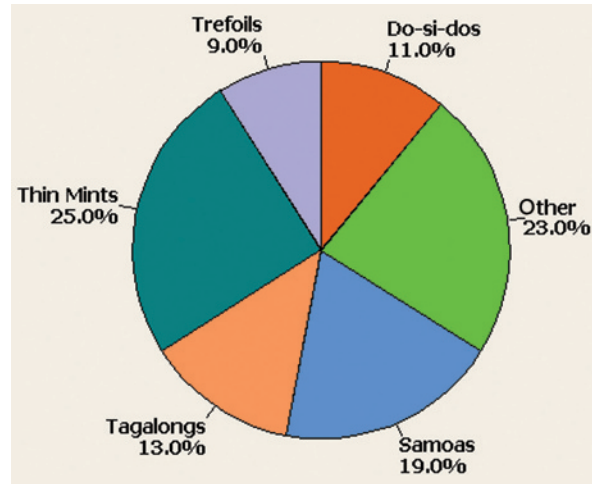


Figure 1.1
MINITAB graph of best-selling Girl Scout cookies (Based on www.girlscouts.org, 2011–12 sales.)

Study 1.2 “Are Action Video Game Players Better than Non-gamers at Complex, Divided Attention Tasks?” (Source: *Human Factors*, Vol. 56, No. 31, May 2014)

Researchers at the Universities of Illinois (Urbana-Champaign) and Central Florida conducted a study to determine whether video game players are better than non-video game players at crossing the street when presented with distractions. Each in a sample of 60 college students was classified as a video game player or a non-gamer. Participants entered a street crossing simulator and were asked to cross a busy street at an unsigned intersection. The simulator was designed to have cars traveling at various high rates of speed in both directions. During the crossing, the students also performed a memory task as a distraction. The researchers found no differences in either the street crossing performance or memory task score of video game players and non-gamers. “These results,” say the researchers, “suggest that action video game players [and non-gamers] are equally susceptible to the costs of dividing attention in a complex task.” Thus, inferential statistics was applied to arrive at this conclusion.

Study 1.3 “Does Rudeness Really Matter in the Workplace?” (Source: *Academy of Management Journal*, Oct. 2007)

Previous studies have established that rudeness in the workplace can lead to retaliatory and counterproductive behavior. However, there has been little research on how rude behaviors influence a victim’s task performance. Consider a study where college students enrolled in a management course were randomly assigned to one of two experimental conditions: rudeness condition (45 students) and control group (53 students). Each student was asked to write down as many uses for a brick as possible in five minutes; this value (total number of uses) was used as a performance measure for each student. For those students in the rudeness condition, the facilitator displayed rudeness by berating the students in general for being irresponsible and unprofessional (due to a late-arriving confederate). No comments were made about the late-arriving confederate for students in the control group. As you might expect, the researchers discovered that the performance levels for students in the rudeness condition were generally lower than the performance levels for students in the control group; thus, they concluded that rudeness in the workplace negatively affects job performance. As in Study 1.2, this study is an example of the use of inferential statistics. The researchers used data collected on 98 college students in a simulated work environment to make an inference about the performance levels of all workers exposed to rudeness on the job.

These studies provide three real-life examples of the uses of statistics. Notice that each involves an analysis of data, either for the purpose of describing the data set (Study 1.1) or for making inferences about a data set (Studies 1.2 and 1.3).

1.3 Fundamental Elements of Statistics

Statistical methods are particularly useful for studying, analyzing, and learning about **populations of experimental units**.

An **experimental** (or **observational**) **unit** is an object (e.g., person, thing, transaction, or event) about which we collect data.

A **population** is a set of all units (usually people, objects, transactions, or events) that we are interested in studying.

For example, populations may include (1) *all* employed workers in the United States, (2) *all* registered voters in California, (3) *everyone* who is afflicted with AIDS, (4) *all* the cars produced last year by a particular assembly line, (5) the *entire* stock of spare parts available at Southwest Airlines' maintenance facility, (6) *all* sales made at the drive-in window of a McDonald's restaurant during a given year, or (7) the set of *all* accidents occurring on a particular stretch of interstate highway during a holiday period. Notice that the first three population examples (1–3) are sets (groups) of people, the next two (4–5) are sets of objects, the next (6) is a set of transactions, and the last (7) is a set of events. Notice also that *each set includes all the units in the population*.

In studying a population, we focus on one or more characteristics or properties of the units in the population. We call such characteristics **variables**. For example, we may be interested in the variables age, gender, and number of years of education of the people currently unemployed in the United States.

A **variable** is a characteristic or property of an individual experimental (or observational) unit in the population.

The name *variable* is derived from the fact that any particular characteristic may vary among the units in a population.

In studying a particular variable, it is helpful to be able to obtain a numerical representation for it. Often, however, numerical representations are not readily available, so measurement plays an important supporting role in statistical studies. **Measurement** is the process we use to assign numbers to variables of individual population units. We might, for instance, measure the performance of the president by asking a registered voter to rate it on a scale from 1 to 10. Or we might measure the age of the U.S. workforce simply by asking each worker, "How old are you?" In other cases, measurement involves the use of instruments such as stopwatches, scales, and calipers.

If the population you wish to study is small, it is possible to measure a variable for every unit in the population. For example, if you are measuring the GPA for all incoming first-year students at your university, it is at least feasible to obtain every GPA. When we measure a variable for every unit of a population, it is called a **census** of the population. Typically, however, the populations of interest in most applications are much larger, involving perhaps many thousands, or even an infinite number, of units. Examples of large populations are those following the definition of population above, as well as all graduates of your university or college, all potential buyers of a new iPhone, and all pieces of first-class mail handled by the U.S. Post Office. For such populations, conducting a census would be prohibitively time consuming or costly. A reasonable alternative would be to select and study a *subset* (or portion) of the units in the population.

A **sample** is a subset of the units of a population.

For example, instead of polling all 145 million registered voters in the United States during a presidential election year, a pollster might select and question a sample of just 1,500 voters. (See Figure 1.2.) If he is interested in the variable “presidential preference,” he would record (measure) the preference of each vote sampled.

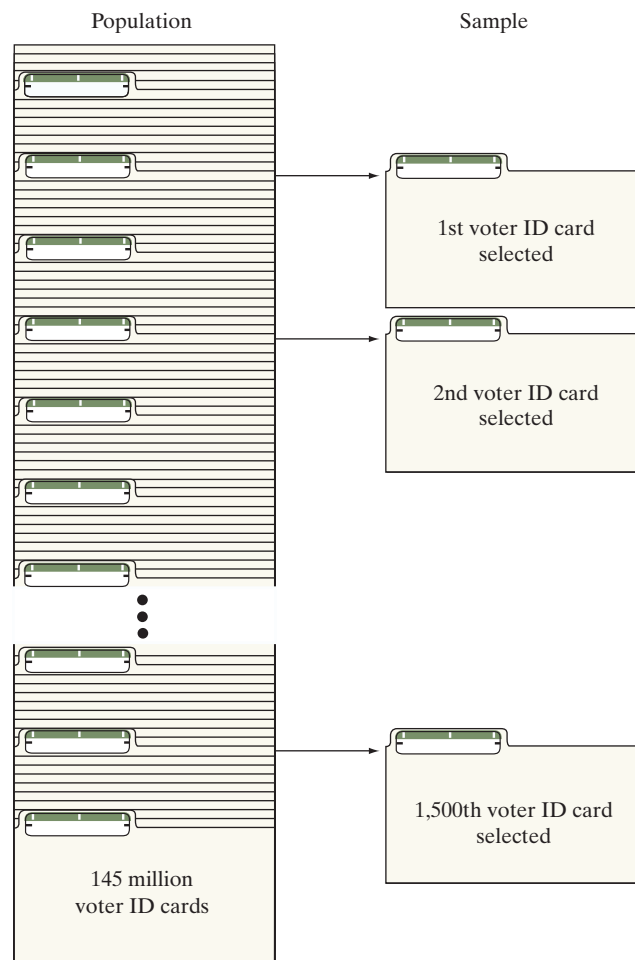


Figure 1.2
A sample of voter registration cards for all registered voters

After the variables of interest for every unit in the sample (or population) are measured, the data are analyzed, either by descriptive or inferential statistical methods. The pollster, for example, may be interested only in *describing* the voting patterns of the sample of 1,500 voters. More likely, however, he will want to use the information in the sample to make inferences about the population of all 145 million voters.

A **statistical inference** is an estimate, prediction, or some other generalization about a population based on information contained in a sample.

That is, *we use the information contained in the smaller sample to learn about the larger population.** Thus, from the sample of 1,500 voters, the pollster may estimate the percentage of all the voters who would vote for each presidential candidate if the election were held on the day the poll was conducted, or he might use the results to predict the outcome on election day.

*The terms *population* and *sample* are often used to refer to the sets of measurements themselves as well as to the units on which the measurements are made. When a single variable of interest is being measured, this usage causes little confusion. But when the terminology is ambiguous, we'll refer to the measurements as *population data sets* and *sample data sets*, respectively.

Example 1.1**Key Elements of a Statistical Problem—Ages of Broadway Ticketbuyers**

Problem According to *Variety* (Jan. 10, 2014), the average age of Broadway ticketbuyers is 42.5 years. Suppose a Broadway theatre executive hypothesizes that the average age of ticketbuyers to her theatre’s plays is less than 42.5 years. To test her hypothesis, she samples 200 ticketbuyers to her theatre’s plays and determines the age of each.

- Describe the population.
- Describe the variable of interest.
- Describe the sample.
- Describe the inference.

Solution

- The population is the set of all units of interest to the theatre executive, which is the set of all ticketbuyers to her theatre’s plays.
- The age (in years) of each ticketbuyer is the variable of interest.
- The sample must be a subset of the population. In this case, it is the 200 ticketbuyers selected by the executive.
- The inference of interest involves the *generalization* of the information contained in the sample of 200 ticketbuyers to the population of all her theatre’s ticketbuyers. In particular, the executive wants to *estimate* the average age of the ticketbuyers to her theatre’s plays in order to determine whether it is less than 42.5 years. She might accomplish this by calculating the average age of the sample and using that average to estimate the average age of the population.

Look Back A key to diagnosing a statistical problem is to identify the data set collected (in this example, the ages of the 200 ticketbuyers) as a population or a sample.

■ **Now Work Exercise 1.13**

Example 1.2**Key Elements of a Statistical Problem—Pepsi vs. Coca-Cola**

Problem “Cola wars” is the popular term for the intense competition between Coca-Cola and Pepsi displayed in their marketing campaigns, which have featured movie and television stars, rock videos, athletic endorsements, and claims of consumer preference based on taste tests. Suppose, as part of a Pepsi marketing campaign, 1,000 cola consumers are given a blind taste test (i.e., a taste test in which the two brand names are disguised). Each consumer is asked to state a preference for brand A or brand B.

- Describe the population.
- Describe the variable of interest.
- Describe the sample.
- Describe the inference.

Solution

- Since we are interested in the responses of cola consumers in a taste test, a cola consumer is the experimental unit. Thus, the population of interest is the collection or set of all cola consumers.
- The characteristic that Pepsi wants to measure is the consumer’s cola preference, as revealed under the conditions of a blind taste test, so *cola preference* is the variable of interest.
- The sample is the 1,000 cola consumers selected from the population of all cola consumers.
- The inference of interest is the *generalization* of the cola preferences of the 1,000 sampled consumers to the population of all cola consumers. In particular, the preferences of the consumers in the sample can be used to *estimate* the percentages of cola consumers who prefer each brand.

Look Back In determining whether the study is inferential or descriptive, we assess whether Pepsi is interested in the responses of only the 1,000 sampled customers (descriptive statistics) or in the responses of the entire population of consumers (inferential statistics).

■ **Now Work Exercise 1.16b**

The preceding definitions and examples identify four of the five elements of an inferential statistical problem: a population, one or more variables of interest, a sample, and an inference. But making the inference is only part of the story; we also need to know its **reliability**—that is, how good the inference is. The only way we can be certain that an inference about a population is correct is to include the entire population in our sample. However, because of *resource constraints* (i.e., insufficient time or money), we usually can't work with whole populations, so we base our inferences on just a portion of the population (a sample). Thus, we introduce an element of *uncertainty* into our inferences. Consequently, whenever possible, it is important to determine and report the reliability of each inference made. Reliability, then, is the fifth element of inferential statistical problems.

The **measure of reliability** that accompanies an inference separates the science of statistics from the art of fortune-telling. A palm reader, like a statistician, may examine a sample (your hand) and make inferences about the population (your life). However, unlike statistical inferences, the palm reader's inferences include no measure of reliability.

Suppose, like the theatre executive in Example 1.1, we are interested in the *error of estimation* (i.e., the difference between the average age of a population of ticketbuyers and the average age of a sample of ticketbuyers). Using statistical methods, we can determine a *bound on the estimation error*. This bound is simply a number that our estimation error (the difference between the average age of the sample and the average age of the population) is not likely to exceed. We'll see in later chapters that this bound is a measure of the uncertainty of our inference. The reliability of statistical inferences is discussed throughout this text. For now, we simply want you to realize that an inference is incomplete without a measure of its reliability.

A **measure of reliability** is a statement (usually quantitative) about the degree of uncertainty associated with a statistical inference.

Let's conclude this section with a summary of the elements of descriptive and of inferential statistical problems and an example to illustrate a measure of reliability.

Four Elements of Descriptive Statistical Problems

1. The population or sample of interest
2. One or more variables (characteristics of the population or sample units) that are to be investigated
3. Tables, graphs, or numerical summary tools
4. Identification of patterns in the data

Five Elements of Inferential Statistical Problems

1. The population of interest
2. One or more variables (characteristics of the population units) that are to be investigated
3. The sample of population units
4. The inference about the population based on information contained in the sample
5. A measure of the reliability of the inference

Example 1.3**Reliability of an Inference—Pepsi vs. Coca-Cola**

Problem Refer to Example 1.2, in which the preferences of 1,000 cola consumers were indicated in a taste test. Describe how the reliability of an inference concerning the preferences of all cola consumers in the Pepsi bottler’s marketing region could be measured.

Solution When the preferences of 1,000 consumers are used to estimate those of all consumers in a region, the estimate will not exactly mirror the preferences of the population. For example, if the taste test shows that 56% of the 1,000 cola consumers preferred Pepsi, it does not follow (nor is it likely) that exactly 56% of all cola drinkers in the region prefer Pepsi. Nevertheless, we can use sound statistical reasoning (which we’ll explore later in the text) to ensure that the sampling procedure will generate estimates that are almost certainly within a specified limit of the true percentage of all cola consumers who prefer Pepsi. For example, such reasoning might assure us that the estimate of the preference for Pepsi is almost certainly within 5% of the preference of the population. The implication is that the actual preference for Pepsi is between 51% [i.e., $(56 - 5)\%$] and 61% [i.e., $(56 + 5)\%$] — that is, $(56 \pm 5)\%$. This interval represents a measure of the reliability of the inference.

Look Ahead The interval 56 ± 5 is called a *confidence interval*, since we are confident that the true percentage of cola consumers who prefer Pepsi in a taste test falls into the range (51, 61). In Chapter 7, we learn how to assess the degree of confidence (e.g., a 90% or 95% level of confidence) in the interval.

Statistics IN Action Revisited**Identifying the Population, Sample, and Inference**

Consider the 2013 Pew Internet & American Life Project survey on social networking. In particular, consider the survey results on the use of social networking sites like Facebook. The experimental unit for the study is an adult (the person answering the question), and the variable measured is the response (“yes” or “no”) to the question.

The Pew Research Center reported that 1,445 adult Internet users participated in the study. Obviously, that number is not all of the adult Internet users in the United States. Consequently, the 1,445 responses represent a sample selected from the much larger population of all adult Internet users.

Earlier surveys found that 55% of adults used an online social networking site in 2006 and 65% in 2008. These

are descriptive statistics that provide information on the popularity of social networking in past years. Since 73% of the surveyed adults in 2013 used an online social networking site, the Pew Research Center inferred that usage of social networking sites continues its upward trend, with more and more adults getting online each year. That is, the researchers used the descriptive statistics from the sample to make an inference about the current population of U.S. adults’ use of social networking.

**1.4 Types of Data**

You have learned that statistics is the science of data and that data are obtained by measuring the values of one or more variables on the units in the sample (or population). All data (and hence the variables we measure) can be classified as one of two general types: **quantitative data** and **qualitative data**.

Quantitative data are data that are measured on a naturally occurring numerical scale.* The following are examples of quantitative data:

1. The temperature (in degrees Celsius) at which each piece in a sample of 20 pieces of heat-resistant plastic begins to melt

*Quantitative data can be subclassified as either *interval data* or *ratio data*. For ratio data, the origin (i.e., the value 0) is a meaningful number. But the origin has no meaning with interval data. Consequently, we can add and subtract interval data, but we can’t multiply and divide them. Of the four quantitative data sets listed as examples, (1) and (3) are interval data while (2) and (4) are ratio data.