



Mario F. Triola



**ESSENTIALS** of  
STATISTICS



Pearson

**6th  
EDITION**



# **ESSENTIALS OF STATISTICS**



6th  
EDITION



# ESSENTIALS OF STATISTICS

**MARIO F. TRIOLA**

Special Contributions by Laura Iossi,  
Broward College

 Pearson

*Director, Portfolio Management* Deirdre Lynch  
*Senior Portfolio Manager* Suzy Bainbridge  
*Content Producer* Peggy McMahon  
*Managing Producer* Scott Disanno  
*Manager, Courseware QA* Mary Durnwald  
*Manager, Content Development* Robert Carroll  
*Senior Producer* Vicki Dreyfus  
*Product Marketing Manager* Emily Ockay  
*Field Marketing Manager* Andrew Noble

*Field Marketing Assistant* Erin Rush  
*Senior Author Support/Technology Specialist* Joe Vetere  
*Manager, Rights and Permissions* Gina Cheselka  
*Text and Cover Design, Illustrations*  
*Production Coordination, Composition*  
Cenveo® Publisher Services  
*Cover Image:* © Laura A. Watt/Getty Images

Copyright © 2019, 2015, 2014 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/](http://www.pearsoned.com/permissions/).

Attributions of third-party content appear on page 613, which constitutes an extension of this copyright page.

PEARSON, ALWAYS LEARNING, and MyLab 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.

MICROSOFT AND/OR ITS RESPECTIVE SUPPLIERS MAKE NO REPRESENTATIONS ABOUT THE SUITABILITY OF THE INFORMATION CONTAINED IN THE DOCUMENTS AND RELATED GRAPHICS PUBLISHED AS PART OF THE SERVICES FOR ANY PURPOSE. ALL SUCH DOCUMENTS AND RELATED GRAPHICS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND. MICROSOFT AND/OR ITS RESPECTIVE SUPPLIERS HEREBY DISCLAIM ALL WARRANTIES AND CONDITIONS WITH REGARD TO THIS INFORMATION, INCLUDING ALL WARRANTIES AND CONDITIONS OF MERCHANTABILITY, WHETHER EXPRESS, IMPLIED OR STATUTORY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. IN NO EVENT SHALL MICROSOFT AND/OR ITS RESPECTIVE SUPPLIERS BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF INFORMATION AVAILABLE FROM THE SERVICES.

THE DOCUMENTS AND RELATED GRAPHICS CONTAINED HEREIN COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY ADDED TO THE INFORMATION HEREIN. MICROSOFT AND/OR ITS RESPECTIVE SUPPLIERS MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THE PRODUCT(S) AND/OR THE PROGRAM(S) DESCRIBED HEREIN AT ANY TIME. PARTIAL SCREEN SHOTS MAY BE VIEWED IN FULL WITHIN THE SOFTWARE VERSION SPECIFIED.

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

Names: Triola, Mario F., author. |

Title: Essentials of statistics / Mario F. Triola ; special contributions by Laura Iossi, Broward College.

Description: 6th edition. | New York, New York : Pearson, 2019. | Includes bibliographical references and index.

Identifiers: LCCN 2017027272 | ISBN 9780134685779 (student edition) | ISBN 0134685776 (student edition) | ISBN 9780134687094 (instructor's edition) | ISBN 0134687094 (instructor's edition)

Subjects: LCSH: Statistics. | Mathematical statistics.

Classification: LCC QA276.12 .T776 2017 | DDC 519.5--dc23

LC record available at <https://lcn.loc.gov/2017027272>



Student Edition  
ISBN-13: 978-0-13-468577-9  
ISBN-10: 0-13-468577-6

*To Ginny*  
*Marc, Dushana, and Marisa*  
*Scott, Anna, Siena, and Kaia*



# ABOUT THE AUTHOR



Mario F. Triola is a Professor Emeritus of Mathematics at Dutchess Community College, where he has taught statistics for over 30 years. Marty is the author of *Elementary Statistics*, 13th edition, *Elementary Statistics Using Excel*, 6th edition and *Elementary Statistics Using the TI-83/84 Plus Calculator*, 5th edition; he is a co-author of *Biostatistics for the Biological and Health Sciences*, 2nd edition, *Statistical Reasoning for Everyday Life*, 5th edition,

and *Business Statistics*. *Elementary Statistics* is currently available as an International Edition, and it has been translated into several foreign languages. Marty designed the original Statdisk statistical software, and he has written several manuals and workbooks for technology supporting statistics education. He has been a speaker at many conferences and colleges. Marty's consulting work includes the design of casino slot machines and fishing rods. He has worked with attorneys in determining probabilities in paternity lawsuits, analyzing data in medical malpractice lawsuits, identifying salary inequities based on gender, and analyzing disputed election results. He has also used statistical methods in analyzing medical school surveys, and in analyzing survey results for the New York City Transit Authority. Marty has testified as an expert witness in the New York State Supreme Court. The Text and Academic Authors Association has awarded Marty a "Texty" for Excellence for his work on *Elementary Statistics*.





# CONTENTS

<b>1</b>	<b>INTRODUCTION TO STATISTICS</b>	<b>1</b>
	1-1 Statistical and Critical Thinking 3	
	1-2 Types of Data 13	
	1-3 Collecting Sample Data 25	
<b>2</b>	<b>EXPLORING DATA WITH TABLES AND GRAPHS</b>	<b>40</b>
	2-1 Frequency Distributions for Organizing and Summarizing Data 42	
	2-2 Histograms 51	
	2-3 Graphs That Enlighten and Graphs That Deceive 57	
	2-4 Scatterplots, Correlation, and Regression 67	
<b>3</b>	<b>DESCRIBING, EXPLORING, AND COMPARING DATA</b>	<b>80</b>
	3-1 Measures of Center 82	
	3-2 Measures of Variation 97	
	3-3 Measures of Relative Standing and Boxplots 112	
<b>4</b>	<b>PROBABILITY</b>	<b>131</b>
	4-1 Basic Concepts of Probability 133	
	4-2 Addition Rule and Multiplication Rule 147	
	4-3 Complements, Conditional Probability, and Bayes' Theorem 159	
	4-4 Counting 169	
	4-5 Probabilities Through Simulations (download only) 177	
<b>5</b>	<b>DISCRETE PROBABILITY DISTRIBUTIONS</b>	<b>184</b>
	5-1 Probability Distributions 186	
	5-2 Binomial Probability Distributions 199	
	5-3 Poisson Probability Distributions 214	
<b>6</b>	<b>NORMAL PROBABILITY DISTRIBUTIONS</b>	<b>226</b>
	6-1 The Standard Normal Distribution 228	
	6-2 Real Applications of Normal Distributions 242	
	6-3 Sampling Distributions and Estimators 254	
	6-4 The Central Limit Theorem 265	
	6-5 Assessing Normality 275	
	6-6 Normal as Approximation to Binomial 284	
<b>7</b>	<b>ESTIMATING PARAMETERS AND DETERMINING SAMPLE SIZES</b>	<b>297</b>
	7-1 Estimating a Population Proportion 299	
	7-2 Estimating a Population Mean 316	
	7-3 Estimating a Population Standard Deviation or Variance 332	
	7-4 Bootstrapping: Using Technology for Estimates 342	
<b>8</b>	<b>HYPOTHESIS TESTING</b>	<b>356</b>
	8-1 Basics of Hypothesis Testing 358	
	8-2 Testing a Claim About a Proportion 373	
	8-3 Testing a Claim About a Mean 387	
	8-4 Testing a Claim About a Standard Deviation or Variance 399	
<b>9</b>	<b>INFERENCES FROM TWO SAMPLES</b>	<b>414</b>
	9-1 Two Proportions 416	
	9-2 Two Means: Independent Samples 428	
	9-3 Two Dependent Samples (Matched Pairs) 442	

<b>10</b>	<b>CORRELATION AND REGRESSION</b>	<b>459</b>
10-1	Correlation	461
10-2	Regression	480
10-3	Rank Correlation	494
<b>11</b>	<b>CHI-SQUARE AND ANALYSIS OF VARIANCE</b>	<b>509</b>
11-1	Goodness-of-Fit	511
11-2	Contingency Tables	522
11-3	One-Way Analysis of Variance	536
<b>APPENDIX A</b>	<b>TABLES</b>	<b>559</b>
<b>APPENDIX B</b>	<b>DATA SETS</b>	<b>567</b>
<b>APPENDIX C</b>	<b>WEBSITES AND BIBLIOGRAPHY OF BOOKS</b>	<b>579</b>
<b>APPENDIX D</b>	<b>ANSWERS TO ODD-NUMBERED SECTION EXERCISES</b>	<b>580</b>
	(and all Chapter Quick Quizzes, Chapter Review Exercises, and Cumulative Review Exercises)	
	<b>Credits</b>	<b>613</b>
	<b>Index of Applications</b>	<b>617</b>
	<b>Index</b>	<b>621</b>

# PREFACE

Statistics permeates nearly every aspect of our lives. From opinion polls, to clinical trials in medicine, self-driving cars, drones, and biometric security, statistics influences and shapes the world around us. *Essentials of Statistics* forges the relationship between statistics and our world through extensive use of a wide variety of real applications that bring life to theory and methods.

## Goals of This Sixth Edition

- Foster personal growth of students through critical thinking, use of technology, collaborative work, and development of communication skills.
- Incorporate the latest and best methods used by professional statisticians.
- Include features that address all of the recommendations included in the *Guidelines for Assessment and Instruction in Statistics Education (GAISE)* as recommended by the American Statistical Association.
- Provide an abundance of new and interesting data sets, examples, and exercises, such as those involving biometric security, cybersecurity, drones, and smartphone data speeds.
- Enhance teaching and learning with the most extensive and best set of supplements and digital resources.


## Audience/Prerequisites

*Essentials of Statistics* is written for students majoring in any subject. Algebra is used minimally. It is recommended that students have completed at least an elementary algebra course or that students learn the relevant algebra components through an integrated or co-requisite course available through MyLab™ Statistics. In many cases, underlying theory is included, but this book does not require the mathematical rigor more appropriate for mathematics majors.

## Hallmark Features

Great care has been taken to ensure that each chapter of *Essentials of Statistics* will help students understand the concepts presented. The following features are designed to help meet that objective of conceptual understanding.

### Real Data

Hundreds of hours have been devoted to finding data that are real, meaningful, and interesting to students. 97% of the examples are based on real data, and 93% of the exercises are based on real data. Some exercises refer to the 32 data sets listed in Appendix B, and 12 of those data sets are new to this edition. Exercises requiring use of the Appendix B data sets are located toward the end of each exercise set and are marked with a special data set icon .

Real data sets are included throughout the book to provide relevant and interesting real-world statistical applications including biometric security, self-driving cars, smartphone data speeds, and use of drones for delivery. Appendix B includes descriptions of the 32 data sets that can be downloaded from the companion website [www.pearsonhighered.com/triola](http://www.pearsonhighered.com/triola) or [www.TriolaStats.com](http://www.TriolaStats.com).

The companion website and TriolaStats.com include downloadable data sets in formats for technologies including Excel, Minitab, JMP, SPSS, and TI-83/84 Plus calculators. The data sets are also included in the free Statdisk software, which is also available on the website.

### Readability

Great care, enthusiasm, and passion have been devoted to creating a book that is readable, understandable, interesting, and relevant. Students pursuing any major are sure to find applications related to their future work.

### Website

This textbook is supported by [www.pearsonhighered.com/triola](http://www.pearsonhighered.com/triola) and the author's website [www.TriolaStats.com](http://www.TriolaStats.com) which are continually updated to provide the latest digital resources for the *Triola Statistics Series*, including:

- Statdisk: A free robust statistical software package designed for this book.
- Downloadable Appendix B data sets in a variety of technology formats.
- Downloadable textbook supplements including Section 4-5 *Probabilities Through Simulations*, *Glossary of Statistical Terms*, and *Formulas and Tables*.
- Online instructional videos created specifically for the 6th edition that provide step-by-step technology instructions.
- Triola Blog which highlights current applications of statistics, statistics in the news, and online resources.
- Contact link providing one-click access for instructors and students to contact the author, Marty Triola, with questions and comments.

### Chapter Features

#### *Chapter Opening Features*

- Chapters begin with a *Chapter Problem* that uses real data and motivates the chapter material.
- *Chapter Objectives* provide a summary of key learning goals for each section in the chapter.

**Exercises** Many exercises require the *interpretation* of results. Great care has been taken to ensure their usefulness, relevance, and accuracy. Exercises are arranged in order of increasing difficulty and exercises are also divided into two groups: (1) *Basic Skills and Concepts* and (2) *Beyond the Basics*. *Beyond the Basics* exercises address more difficult concepts or require a stronger mathematical background. In a few cases, these exercises introduce a new concept.

#### *End-of-Chapter Features*

- *Chapter Quick Quiz* provides 10 review questions that require brief answers.
- *Review Exercises* offer practice on the chapter concepts and procedures.
- *Cumulative Review Exercises* reinforce earlier material.
- *Technology Project* provides an activity that can be used with a variety of technologies.
- *From Data to Decision* is a capstone problem that requires critical thinking and writing.
- *Cooperative Group Activities* encourage active learning in groups.

### Other Features

**Margin Essays** There are 92 margin essays designed to highlight real-world topics and foster student interest. There are also many *Go Figure* items that briefly describe interesting numbers or statistics.

**Flowcharts** The text includes flowcharts that simplify and clarify more complex concepts and procedures. Animated versions of the text’s flowcharts are available within MyLab Statistics and MathXL®.

**Detachable Formula and Table Card** This insert, organized by chapter, gives students a quick reference for studying or for use when taking tests (if allowed by the instructor). It also includes the most commonly used tables. This is also available for download at [www.TriolaStats.com](http://www.TriolaStats.com).

### Technology Integration

As in the preceding edition, there are many displays of screens from technology throughout the book, and some exercises are based on displayed results from technology. Where appropriate, sections end with a new **Tech Center** subsection that includes new technology-specific videos and detailed instructions for Statdisk, Minitab®, Excel®, StatCrunch, or a TI-83/84 Plus® calculator. (Throughout this text, “TI-83/84 Plus” is used to identify a TI-83 Plus or TI-84 Plus calculator). The end-of-chapter features include a *Technology Project*.

The Statdisk statistical software package is designed specifically for this textbook and contains all Appendix B data sets. Statdisk is free to users of this book and it can be downloaded at [www.Statdisk.org](http://www.Statdisk.org).

## Changes in This Edition

### New Features

**Chapter Objectives** provide a summary of key learning goals for each section in the chapter.

**Your Turn:** Many examples include a new “your turn” feature that directs students to a relevant exercise so that they can immediately apply what they just learned from the example.

**Tech Center:** Improved technology instructions, supported by custom, author-created instructional videos and downloadable content available at [www.TriolaStats.com](http://www.TriolaStats.com).

**Technology Videos.** New, author-driven technology videos provide step-by-step details for key statistical procedures using Excel, TI-83/84 calculators and Statdisk.

**Larger Data Sets:** Some of the data sets in Appendix B are much larger than in previous editions. It is no longer practical to print all of the Appendix B data sets in this book, so the data sets are *described* in Appendix B, and they can be downloaded at [www.TriolaStats.com](http://www.TriolaStats.com).

**New Content:** New examples, exercises and chapter problems provide relevant and interesting real-world statistical applications including biometric security, self-driving cars, smartphone data speeds, and use of drones for delivery.

	Number	New to This Edition	Use Real Data
Exercises	1495	82% (1230)	93% (1392)
Examples	183	73% (133)	97% (177)
Chapter Problems	11	100% (11)	100% (11)

## Organization Changes

**New Chapter Objectives:** All chapters now begin with a list of key learning goals for that chapter. *Chapter Objectives* replaces the former *Review and Preview* numbered section. The first numbered section of each chapter now covers a major topic.

**New Subsection 1-3, Part 2: Big Data and Missing Data: Too Much and Not Enough**

**New Section 2-4: Scatterplots, Correlation, and Regression**

The previous edition included scatterplots in Chapter 2, but this new section includes scatterplots in Part 1, the linear correlation coefficient  $r$  in Part 2, and linear regression in Part 3. These additions are intended to greatly facilitate coverage for those professors who prefer some early coverage of correlation and regression concepts. Chapter 10 continues to include these topics discussed with much greater detail.

**New Subsection 4-3, Part 3: Bayes' Theorem**

**New Section 7-4: Bootstrapping: Using Technology for Estimates**

**Combined Sections:**

- **4-2: Addition Rule and Multiplication Rule**

Combines 5th edition Section 4-3 (*Addition Rule*) and Section 4-4 (*Multiplication Rule: Basics*).

- **5-2: Binomial Probability Distributions**

Combines 5th edition Section 5-3 (*Binomial Probability Distributions*) and Section 5-4 (*Parameters for Binomial Distributions*)

## Changed Terminology

**Significant:** References in the previous edition to “unusual” outcomes are now described in terms of “significantly low” or “significantly high,” so that the link to hypothesis testing is further reinforced.

**Multiplication Counting Rule:** References in Section 4-4 (*Counting*) to the “fundamental counting rule” now use “multiplication counting rule” so that the name of the rule better suggests how it is applied.

## Flexible Syllabus

This book’s organization reflects the preferences of most statistics instructors, but there are two common variations:

- **Early Coverage of Correlation and Regression:** Some instructors prefer to cover the basics of correlation and regression early in the course. Section 2-4 now includes basic concepts of scatterplots, correlation, and regression without the use of formulas and greater depth found in Sections 10-1 (*Correlation*) and 10-2 (*Regression*).
- **Minimum Probability:** Some instructors prefer extensive coverage of probability, while others prefer to include only basic concepts. Instructors preferring minimum coverage can include Section 4-1 while skipping the remaining sections of Chapter 4, as they are not essential for the chapters that follow. Many instructors prefer to cover the fundamentals of probability along with the basics of the addition rule and multiplication rule (Section 4-2).

**GAISE** This book reflects recommendations from the American Statistical Association and its *Guidelines for Assessment and Instruction in Statistics Education* (GAISE). Those guidelines suggest the following objectives and strategies.

1. **Emphasize statistical literacy and develop statistical thinking:** Each section exercise set begins with *Statistical Literacy and Critical Thinking* exercises. Many of the book's exercises are designed to encourage statistical thinking rather than the blind use of mechanical procedures.
2. **Use real data:** 97% of the examples and 93% of the exercises use real data.
3. **Stress conceptual understanding rather than mere knowledge of procedures:** Instead of seeking simple numerical answers, most exercises and examples involve conceptual understanding through questions that encourage practical interpretations of results. Also, each chapter includes a *From Data to Decision* project.
4. **Foster active learning in the classroom:** Each chapter ends with several *Cooperative Group Activities*.
5. **Use technology for developing conceptual understanding and analyzing data:** Computer software displays are included throughout the book. Special *Tech Center* subsections include instruction for using the software. Each chapter includes a *Technology Project*. When there are discrepancies between answers based on tables and answers based on technology, Appendix D provides *both* answers. The website [www.TriolaStats.com](http://www.TriolaStats.com) includes free, text-specific software (Statdisk), data sets formatted for several different technologies, and instructional videos for technologies.
6. **Use assessments to improve and evaluate student learning:** Assessment tools include an abundance of section exercises, *Chapter Quick Quizzes*, *Chapter Review Exercises*, *Cumulative Review Exercises*, *Technology Projects*, *From Data to Decision* projects, and *Cooperative Group Activities*.

## Acknowledgments

I would like to thank the thousands of statistics professors and students who have contributed to the success of this book. I thank the reviewers for their suggestions for this sixth edition: Eric Gorenstein, Bunker Hill Community College; Rhonda Hatcher, Texas Christian University; Ladorian Latin, Franklin University; Joseph Pick, Palm Beach State College; and Lisa Whitaker, Keiser University. Special thanks to Laura Iossi of Broward College for her comprehensive work in reviewing and contributing to this 6<sup>th</sup> edition.

Other recent reviewers have included Raid W. Amin, University of West Florida; Robert Black, United States Air Force Academy; James Bryan, Merced College; Donald Burd, Monroe College; Keith Carroll, Benedictine University; Monte Cheney, Central Oregon Community College; Christopher Donnelly, Macomb Community College; Billy Edwards, University of Tennessee—Chattanooga; Marcos Enriquez, Moorpark College; Angela Everett, Chattanooga State Technical Community College; Joe Franko, Mount San Antonio College; Rob Fusco, Broward College; Sanford Geraci, Broward College; Laura Heath, Palm Beach State College; Richard Herbst, Montgomery County Community College; Richard Hertz; Diane Hollister, Reading Area Community College; Michael Huber, Muhlenberg College; George Jahn, Palm Beach State College; Gary King, Ozarks Technical Community College; Kate Kozak, Coconino Community College; Dan Kumpf, Ventura College; Mickey Levendusky, Pima County Community College; Mitch Levy, Broward College; Tristan Londre, Blue River Community College; Alma Lopez, South Plains College; Kim McHale, Heartland Community College; Carla Monticelli, Camden County Community College; Ken Mulzet, Florida State College at Jacksonville; Julia Norton, California State University Hayward; Michael Oriolo,



Herkimer Community College; Jeanne Osborne, Middlesex Community College; Ali Saadat, University of California—Riverside; Radha Sankaran, Passaic County Community College; Steve Schwager, Cornell University; Pradipta Seal, Boston University; Kelly Smitch, Brevard College; Sandra Spain, Thomas Nelson Community College; Ellen G. Stutes, Louisiana State University, Eunice; Sharon Testone, Onondaga Community College; Chris Vertullo, Marist College; Dave Wallach, University of Findlay; Cheng Wang, Nova Southeastern University; Barbara Ward, Belmont University; Richard Weil, Brown College; Gail Wiltse, St. John River Community College; Claire Wladis, Borough of Manhattan Community College; Rick Woodmansee, Sacramento City College; Yong Zeng, University of Missouri at Kansas City; Jim Zimmer, Chattanooga State Technical Community College; Cathleen Zucco-Teveloff, Rowan University; Mark Z. Zuiker, Minnesota State University, Mankato.

This sixth edition of *Essentials of Statistics* is truly a team effort, and I consider myself fortunate to work with the dedication and commitment of the Pearson team. I thank Suzy Bainbridge, Deirdre Lynch, Peggy McMahon, Vicki Dreyfus, Christine O'Brien, Joe Vetere, and Rose Kernan of Cenveo Publisher Services.

I extend special thanks to Marc Triola, M.D., New York University School of Medicine, for his outstanding work on creating the new 13th edition of the Statdisk software. I thank Scott Triola for his very extensive help throughout the entire production process for this 6th edition.

I thank the following for their help in checking the accuracy of text and answers in this edition: James Lapp, Paul Lorcak, and Dirk Tempelaar.

M.F.T.  
*Madison, Connecticut*  
*September 2017*

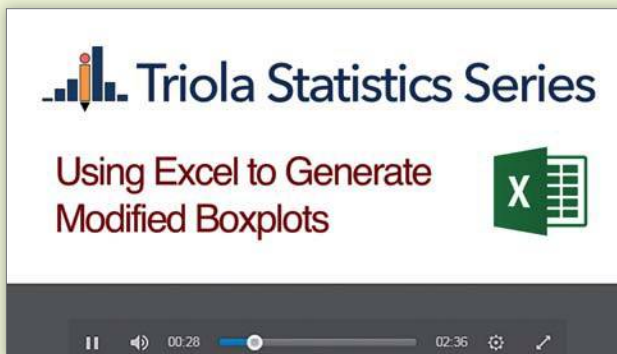
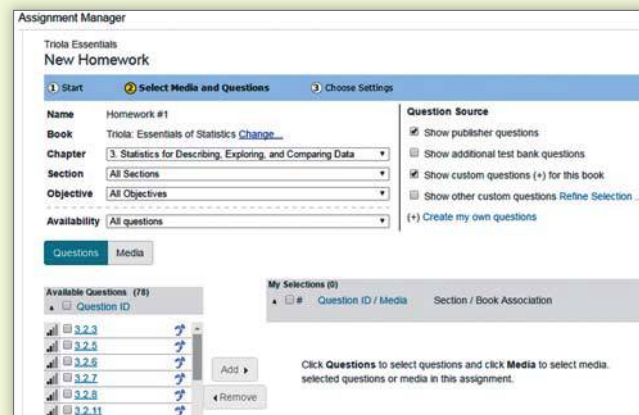
# Resources for Success

## MyLab Statistics Online Course for *Essentials of Statistics, 6e* by Mario F. Triola (access code required)

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

### Expanded objective-based MathXL coverage

MathXL® is newly mapped to improve student learning outcomes. Homework reinforces and supports students' understanding of key statistics topics.



### Enhanced video program to meet Introductory Statistics needs:

- **New! Tech-Specific Video Tutorials** - These short, topical videos address how to use various technologies to complete exercises.
- **Updated! Chapter Review Exercise Videos** - Watch the Chapter Review Exercises come to life with new review videos that help students understand key chapter concepts.
- **Updated! Section Lecture Videos** - Watch author, Marty Triola, work through examples and elaborate on key objectives of the chapter.

### Real-world data examples

Examples and exercises throughout the textbook and MyLab Statistics use current, real-world data to help students understand how statistics applies to everyday life.

Row	LEAD	AGE	SEX	YEAR1	YEAR2	IQV	IQP	IQF	var9
1	1	11	1	25	18	61	85	70	
2	1	9	1	31	28	82	90	85	
3	1	11	1	30	29	70	107	86	
4	1	6	1	29	30	72	85	76	
5	1	11	1	2	34	72	100	84	
6	1	6	1	29	25	95	97	96	
7	1	6	1	25	24	89	101	94	
8	1	15	2	24	15	57	64	56	
9	1	7	2	24	16	116	111	115	
10	1	7	1	31	24	95	100	97	
11	1	13	2	21	19	82	76	77	
12	1	10	2	29	27	116	136	128	
13	1	12	1	32	29	99	100	99	
14	1	12	1	36	32	74	90	80	
15	1	15	1	30	25	100	135	118	
16	1	10	1	29	23	72	104	86	
17	1	15	1	28	28	126	149	141	
18	1	9	2	28	19	80	99	88	
19	1	8	1	34	22	86	107	96	
20	1	11	1	21	22	94	99	96	
21	1	7	1	35	27	100	113	107	
22	1	11	2	39	38	72	104	86	

# Resources for Success

## Student Resources

**Student's Solutions Manual**, by James Lapp (Colorado Mesa University), provides detailed, worked-out solutions to all odd-numbered text exercises. (ISBN-13: 978-0-13-468707-0; ISBN-10: 0-13-468707-8)

**Student Workbook for the Triola Statistics Series**, by Laura Iossi (Broward College) offers additional examples, concept exercises, and vocabulary exercises for each chapter. (ISBN-13: 978-0-13-446423-7; ISBN 10: 0-13-446423-0)

*The following technology manuals contain detailed tutorial instructions and worked-out examples and exercises for many technologies. They correspond with the Triola Statistics Series and can be downloaded by students and instructors from [www.pearsonhighered.com/Triola](http://www.pearsonhighered.com/Triola) or from within MyLab Statistics.*

**Excel Student Laboratory Manual and Workbook (download only)**, by Laurel Chiappetta (University of Pittsburgh).

**Minitab® Student Laboratory Manual and Workbook (download only)**, by Mario F. Triola.

**Graphing Calculator Manual for the TI-83 Plus, TI-84 Plus, TI-84 Plus C and TI-84 Plus CE (download only)**, by Kathleen McLaughlin (University of Connecticut) and Dorothy Wakefield (University of Connecticut Health Center).

**Statdisk Student Laboratory Manual and Workbook (download only)**, by Mario F. Triola.

**SPSS Student Laboratory Manual and Workbook (download only)**, by James J. Ball (Indiana State University).

## Instructor Resources

**Annotated Instructor's Edition**, by Mario F. Triola, contains answers to exercises in the margin, plus recommended assignments, and teaching suggestions. (ISBN-13: 978-0-13-468709-4; ISBN-10: 0-13-468709-4)

**Instructor's Solutions Manual (download only)**, by James Lapp (Colorado Mesa University), contains solutions to all the exercises. These files are available to qualified instructors through Pearson's online catalog at [www.pearson.com/us/higher-education](http://www.pearson.com/us/higher-education) or within MyLab Statistics.

**Insider's Guide to Teaching with the Triola Statistics Series (download only)**, by Mario F. Triola, contains sample syllabi and tips for incorporating projects, as well as lesson overviews, extra examples, minimum outcome objectives, and recommended assignments for each chapter. This file is available to qualified instructors through Pearson's online catalog at [www.pearson.com/us/higher-education](http://www.pearson.com/us/higher-education) or within MyLab Statistics.

**TestGen® Computerized Test Bank** ([www.pearsoned.com/testgen](http://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's online catalog at [www.pearson.com/us/higher-education](http://www.pearson.com/us/higher-education). Test Forms (download only) are also available from the online catalog.

**PowerPoint Lecture Slides:** Free to qualified adopters, this classroom lecture presentation software is geared specifically to the sequence and philosophy of *Essentials of Statistics*. Key graphics from the book are included to help bring the statistical concepts alive in the classroom. These files are available to qualified instructors through Pearson's online catalog at [www.pearson.com/us/higher-education](http://www.pearson.com/us/higher-education) or within MyLab Statistics.

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

## Technology Resources

The following resources can be found at [www.pearsonhighered.com/triola](http://www.pearsonhighered.com/triola), the author-maintained Triola Statistics Series Web site (<http://www.triolastats.com>), and MyLab Statistics

- Appendix B data sets formatted for Minitab, SPSS, SAS, Excel, JMP, and as text files. Additionally, these data sets are available as data lists for the TI-83/84 Plus calculators; supplemental programs for the TI-83/84 Plus calculator are also available.
- Statdisk statistical software instructions for download. New features include the ability to directly use lists of data instead of requiring the use of their summary statistics.
- Extra data sets, *Probabilities Through Simulations*, *Bayes' Theorem*, an index of applications, and a symbols table.

**Video Resources** has been expanded and updated and now supplements most sections in the book, with many topics presented by the author. The videos aim to support both instructors and students through lecture, reinforcement of statistical basics through technology, and application of concepts:

- **Section Lecture Videos**
- **Chapter Review Exercise Videos** walk students through the exercises and help them understand key chapter concepts.
- **New! Technology Video Tutorials** These short, topical videos address how to use Excel, StatDisk, and the TI graphing calculator to complete exercises.
- **StatTalk Videos: 24 Conceptual Videos to Help You Actually Understand Statistics.** Fun-loving statistician Andrew Vickers takes to the streets of Brooklyn, NY, to demonstrate important statistical concepts through interesting stories and real-life events. These fun and engaging videos will help students actually understand statistical concepts. Available with an instructors user guide and assessment questions.

Videos also contain optional English and Spanish captioning. All videos are available through the MyLab Statistics online course.

### **MyLab Statistics Online Course (access code required)**

Used by nearly one million students a year, MyLab Statistics is the world's leading online program for teaching and learning statistics. Each course is developed to accompany Pearson's best-selling content, authored by thought

leaders across the statistics curriculum, and can be easily customized to fit any course format.

- **MyLab Statistics comprehensive online gradebook** automatically tracks students' results on tests, quizzes, homework, and in the study plan. Instructors can use the gradebook to provide positive feedback or intervene if students have trouble. Gradebook data can be easily exported to a variety of spreadsheet programs, such as Microsoft Excel. You can determine which points of data you want to export, and then analyze the results to determine success.

MyLab Statistics delivers assessment, tutorials, and multimedia resources that provide engaging and personalized experiences for each student, so learning can happen in any environment. In addition to the resources below, each course includes a full interactive online version of the accompanying textbook.

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

MyLab Statistics questions, into software such as StatCrunch, Minitab, Excel, and more. Students have access to a variety of support tools—Technology Tutorial Videos, Technology Study Cards, and Technology Manuals for select titles—to learn how to effectively use statistical software.

### **MathXL for Statistics Online Course (access code required)**

MathXL is the homework and assessment engine that runs MyLab Statistics. (MyLab Statistics is MathXL plus a learning management system.)

With MathXL for Statistics, instructors can:

- Create, edit, and assign online homework and tests using algorithmically generated exercises correlated at the objective level to the textbook.
- Create and assign their own online exercises and import TestGen tests for added flexibility.
- Maintain records of all student work, tracked in MathXL's online gradebook.

With MathXL for Statistics, students can:

- Take chapter tests in MathXL and receive personalized study plans and/or personalized homework assignments based on their test results.
- Use the study plan and/or the homework to link directly to tutorial exercises for the objectives they need to study.
- Access supplemental animations and video clips directly from selected exercises.
- Can easily copy our data sets, both from the ebook and the MyLab Statistics questions, into software like StatCrunch, Minitab, Excel, and more.

MathXL for Statistics is available to qualified adopters. For more information, visit our web site at [www.mathxl.com](http://www.mathxl.com), or contact your Pearson representative.

### **StatCrunch**

StatCrunch is powerful, web-based statistical software that allows users to perform complex analyses, share data

sets, and generate compelling reports. A vibrant online community offers tens of thousands of data sets for students to analyze.

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

Full access to StatCrunch is available with a MyLab Statistics kit, and StatCrunch is available by itself to qualified adopters. StatCrunch Mobile is also now available when you visit [www.StatCrunch.com](http://www.StatCrunch.com) from the browser on your smartphone or tablet. For more information, visit [www.StatCrunch.com](http://www.StatCrunch.com) or contact your Pearson representative.

**Minitab and Minitab Express™** make learning statistics easy and provide students with a skill set that's in demand in today's data-driven workforce. Bundling Minitab software with educational materials ensures students have access to the software they need in the classroom, around campus, and at home. And having the latest version of Minitab ensures that students can use the software for the duration of their course.

ISBN 13: 978-0-13-445640-9

ISBN 10: 0-13-445640-8 (Access Card only; not sold as stand alone.)

**JMP Student Edition** is an easy-to-use, streamlined version of JMP desktop statistical discovery software from SAS Institute, Inc., and is available for bundling with the text.

ISBN-13: 978-0-13-467979-2

ISBN-10: 0-13-467979-2



- 1-1 Statistical and Critical Thinking
- 1-2 Types of Data
- 1-3 Collecting Sample Data

# 1



## INTRODUCTION TO STATISTICS



### Survey Question: Do you prefer to read a *printed* book or an *electronic* book?

Surveys provide data that enable us to improve products or services. Surveys guide political candidates, shape business practices, influence social media, and affect many aspects of our lives. Surveys give us insight into the opinions and views of others. Let's consider one *USA Today* survey in which respondents were asked if they prefer to read a printed book or an electronic book. Among 281 respondents, 65% preferred a printed book and 35% preferred an electronic book. Figure 1-1 on the next page includes graphs that depict these results.

The survey results suggest that people overwhelmingly prefer reading printed books to reading ebooks. The graphs in Figure 1-1 visually depict the survey results, and they support a claim that people prefer printed books to ebooks by a wide margin. One of the most important objectives in this book is to encourage the use of critical thinking so that such results are not blindly accepted. We might question whether the survey results are valid. Who conducted the survey? How were respondents selected? Do the graphs in Figure 1-1

depict the results well, or are those graphs somehow misleading?

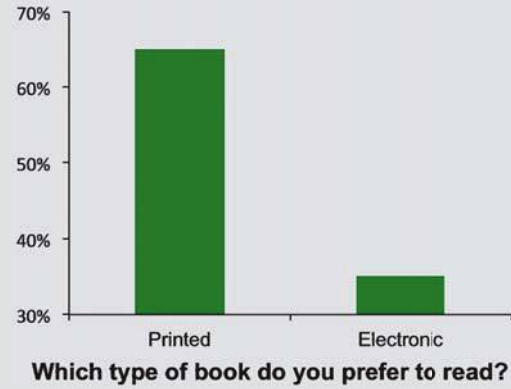
The survey results presented here have major flaws that are among the most commonly used, so they are especially important to recognize. Here are brief descriptions of each of the major flaws:

**Flaw 1: Misleading Graphs** The bar chart in Figure 1-1(a) is very deceptive. By using a vertical scale that does not start at zero, the difference between the two percentages is grossly exaggerated. Figure 1-1(a) makes it appear that about eight times as many people choose a printed book over an ebook, but with response rates of 65% and 35%, that ratio is very roughly 2:1, not 8:1.

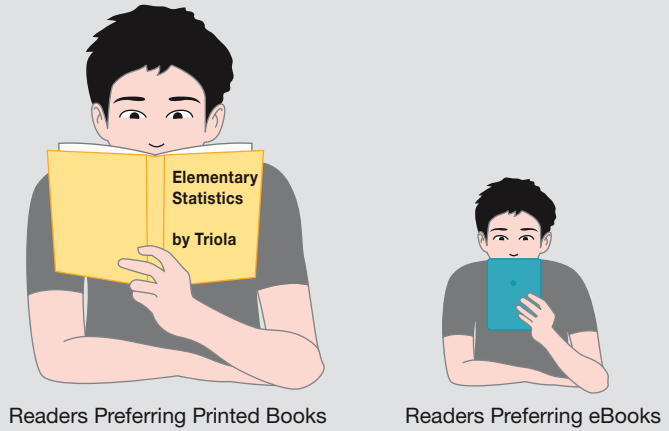
The illustration in Figure 1-1(b) is also deceptive. Again, the difference between the actual response rates of 65% for printed books and 35% for ebooks is a difference that is grossly distorted. The picture graph (or “pictograph”) in Figure 1-1(b) makes it appear that people prefer printed books to ebooks by a ratio of roughly 4:1 instead of being the correct ratio of 65:35, or roughly 2:1. (Objects with area or volume can distort perceptions because they can be drawn to be disproportionately larger or smaller than the data indicate.)

Deceptive graphs are discussed in more detail in Section 2-3, but we see here that the illustrations in Figure 1-1 grossly exaggerate the preference for printed books.

**Flaw 2: Bad Sampling Method** The aforementioned survey responses are from a *USA Today* survey of Internet users. The survey question was posted on a website and Internet users decided whether to respond. This is an example of a *voluntary response sample*—a sample in which respondents themselves decide whether to participate. With a voluntary response sample, it often happens that those with a strong interest in the topic are more likely to participate, so the results are very questionable. In this case, it is reasonable to suspect that Internet users might prefer ebooks at a rate higher than the rate in the general population. When using sample data to learn something about a population, it is *extremely* important to obtain sample data that are representative of the population from which the data are drawn. As we proceed



(a)



(b)

**FIGURE 1-1** Survey Results

through this chapter and discuss types of data and sampling methods, we should focus on these key concepts:

- **Sample data must be collected in an appropriate way, such as through a process of *random* selection.**
- **If sample data are not collected in an appropriate way, the data may be so completely useless that no amount of statistical torturing can salvage them.**

It would be easy to accept the preceding survey results and blindly proceed with calculations and statistical analyses, but we would miss the two critical flaws described above. We could then develop conclusions that are fundamentally wrong and misleading. Instead, we should develop skills in statistical thinking and critical thinking so that we can understand how the survey is so seriously flawed.

## CHAPTER OBJECTIVES

Here is the single most important concept presented in this chapter: When using methods of statistics with sample data to form conclusions about a population, it is absolutely essential to collect sample data in a way that is appropriate. Here are the chapter objectives:

### 1-1 Statistical and Critical Thinking

- Analyze sample data relative to context, source, and sampling method.
- Understand the difference between *statistical significance* and *practical significance*.
- Define and identify a *voluntary response sample* and know that statistical conclusions based on data from such a sample are generally not valid.

### 1-2 Types of Data

- Distinguish between a *parameter* and a *statistic*.
- Distinguish between *quantitative data* and *categorical* (or *qualitative* or *attribute*) *data*.
- Distinguish between *discrete* data and *continuous* data.
- Determine whether basic statistical calculations are appropriate for a particular data set.

### 1-3 Collecting Sample Data

- Define and identify a *simple random sample*.
- Understand the importance of sound sampling methods and the importance of good design of experiments.

## 1-1

## Statistical and Critical Thinking

**Key Concept** In this section we begin with a few very basic definitions, and then we consider an *overview* of the process involved in conducting a statistical study. This process consists of “prepare, analyze, and conclude.” “Preparation” involves consideration of the *context*, the *source* of data, and *sampling method*. In future chapters we construct suitable graphs, explore the data, and execute computations required for the statistical method being used. In future chapters we also form conclusions by determining whether results have statistical significance and practical significance.

Statistical thinking involves critical thinking and the ability to make sense of results. Statistical thinking demands so much more than the ability to execute complicated calculations. Through numerous examples, exercises, and discussions, this text will help you develop the statistical thinking skills that are so important in today’s world.



## Go Figure

78%: The percentage of *female* veterinarian students who are women, according to *The Herald* in Glasgow, Scotland.

We begin with some very basic definitions.

### DEFINITIONS

**Data** are collections of observations, such as measurements, genders, or survey responses. (A single data value is called a *datum*, a term rarely used. The term “data” is plural, so it is correct to say “data are . . .” not “data is . . .”)

**Statistics** is the science of planning studies and experiments; obtaining data; and organizing, summarizing, presenting, analyzing, and interpreting those data and then drawing conclusions based on them.

A **population** is the complete collection of *all* measurements or data that are being considered. Typically, a population is the complete collection of data that we would like to make inferences about.

A **census** is the collection of data from *every* member of the population.

A **sample** is a *subcollection* of members selected from a population.

Because populations are often very large, a common objective of the use of statistics is to obtain data from a sample and then use those data to form a conclusion about the population.

### EXAMPLE 1 Residential Carbon Monoxide Detectors

In the journal article “Residential Carbon Monoxide Detector Failure Rates in the United States” (by Ryan and Arnold, *American Journal of Public Health*, Vol. 101, No. 10), it was stated that there are 38 million carbon monoxide detectors installed in the United States. When 30 of them were randomly selected and tested, it was found that 12 of them failed to provide an alarm in hazardous carbon monoxide conditions. In this case, the population and sample are as follows:

**Population:** All 38 million carbon monoxide detectors in the United States

**Sample:** The 30 carbon monoxide detectors that were selected and tested

The objective is to use the sample data as a basis for drawing a conclusion about the population of all carbon monoxide detectors, and methods of statistics are helpful in drawing such conclusions.

**YOUR TURN** → Do part (a) of Exercise 2 “Reported Versus Measured.”

We now proceed to consider the process involved in a statistical study. See Figure 1-2 for a summary of this process and note that the focus is on critical thinking, not mathematical calculations. Thanks to wonderful developments in technology, we have powerful tools that effectively do the number crunching so that we can focus on understanding and interpreting results.

## Prepare

**Context** Figure 1-2 suggests that we begin our preparation by considering the *context* of the data, so let’s start with context by considering the data in Table 1-1. Table 1-1 includes the numbers of registered pleasure boats in Florida (tens of thousands) and the numbers of manatee fatalities from encounters with boats in Florida for each of several recent years. The format of Table 1-1 suggests the following goal: Determine whether there is a *relationship* between numbers of boats and numbers of manatee deaths from

**TABLE 1-1** Pleasure Boats and Manatee Fatalities from Boat Encounters

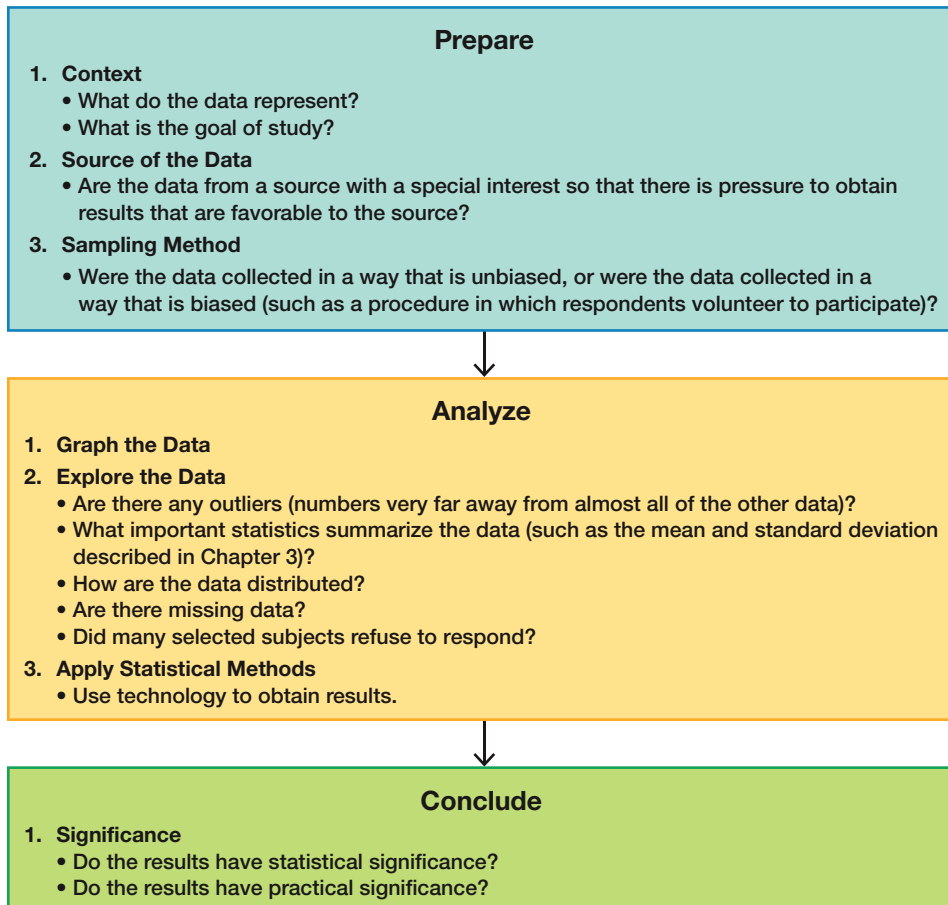
Pleasure Boats (tens of thousands)	99	99	97	95	90	90	87	90	90
Manatee Fatalities	92	73	90	97	83	88	81	73	68

boats. This goal suggests a reasonable hypothesis: As the numbers of boats increase, the numbers of manatee deaths increase.

**Source of the Data** The second step in our preparation is to consider the source (as indicated in Figure 1-2). The data in Table 1-1 are from the Florida Department of Highway Safety and Motor Vehicles and the Florida Marine Research Institute. The sources certainly appear to be reputable.

**Sampling Method** Figure 1-2 suggests that we conclude our preparation by considering the sampling method. The data in Table 1-1 were obtained from official government records known to be reliable. The sampling method appears to be sound.

Sampling methods and the use of randomization will be discussed in Section 1-3, but for now, we stress that a sound sampling method is absolutely essential for good results in a statistical study. It is generally a bad practice to use voluntary response (or self-selected) samples, even though their use is common.

**FIGURE 1-2** Statistical and Critical Thinking

## Survivorship Bias

In World War II, statistician Abraham Wald saved many lives with his work on the Applied

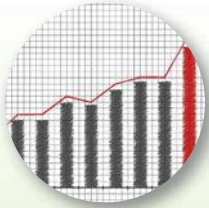


Mathematics Panel. Military leaders asked the panel how they could improve the chances of aircraft bombers returning after missions. They wanted to add some armor for protection, and they recorded locations on the bombers where damaging holes were found. They reasoned that armor should be placed in locations with the most holes, but Wald said that strategy would be a big mistake. He said that armor should be placed where returning bombers were *not* damaged. His reasoning was this: The bombers that made it back with damage were *survivors*, so the damage they suffered could be survived. Locations on the aircraft that were not damaged were the most vulnerable, and aircraft suffering damage in those vulnerable areas were the ones that did not make it back. The military leaders would have made a big mistake with survivorship bias by studying the planes that survived instead of thinking about the planes that did not survive.

## Go Figure

17%: The percentage of U.S. men between 20 and 40 years of age and taller than 7 feet who play basketball in the NBA.

## Origin of “Statistics”



The word *statistics* is derived from the Latin word *status* (meaning “state”).

Early uses of statistics involved compilations of data and graphs describing various aspects of a state or country. In 1662, John Graunt published statistical information about births and deaths. Graunt’s work was followed by studies of mortality and disease rates, population sizes, incomes, and unemployment rates. Households, governments, and businesses rely heavily on statistical data for guidance. For example, unemployment rates, inflation rates, consumer indexes, and birth and death rates are carefully compiled on a regular basis, and the resulting data are used by business leaders to make decisions affecting future hiring, production levels, and expansion into new markets.

## DEFINITION

A **voluntary response sample** (or **self-selected sample**) is one in which the respondents themselves decide whether to be included.

The following types of polls are common examples of voluntary response samples. By their very nature, all are seriously flawed because we should not make conclusions about a population on the basis of samples with a strong possibility of bias:

- Internet polls, in which people online can decide whether to respond
- Mail-in polls, in which people can decide whether to reply
- Telephone call-in polls, in which newspaper, radio, or television announcements ask that you voluntarily call a special number to register your opinion

The Chapter Problem involves a *USA Today* survey with a voluntary response sample. See also the following Example 2.

## EXAMPLE 2 Voluntary Response Sample

The ABC television show *Nightline* asked viewers to call with their opinion about whether the United Nations headquarters should remain in the United States. Viewers then decided themselves whether to call with their opinions, and 67% of 186,000 respondents said that the United Nations should be moved out of the United States. In a separate and independent survey, 500 respondents were randomly selected and surveyed, and 38% of this group wanted the United Nations to move out of the United States. The two polls produced dramatically different results. Even though the *Nightline* poll involved 186,000 volunteer respondents, the much smaller poll of 500 randomly selected respondents is more likely to provide better results because of the far superior sampling method.

**YOUR TURN** Do Exercise 1 “Online Medical Info.”

## Analyze

Figure 1-2 indicates that after completing our preparation by considering the context, source, and sampling method, we begin to *analyze* the data.

**Graph and Explore** An analysis should begin with appropriate graphs and explorations of the data. Graphs are discussed in Chapter 2, and important statistics are discussed in Chapter 3.

**Apply Statistical Methods** Later chapters describe important statistical methods, but application of these methods is often made easy with technology (calculators and/or statistical software packages). A good statistical analysis *does not* require strong computational skills. A good statistical analysis *does* require using common sense and paying careful attention to sound statistical methods.

## Conclude

Figure 1-2 shows that the final step in our statistical process involves conclusions, and we should develop an ability to distinguish between statistical significance and practical significance.

**Statistical Significance** *Statistical significance* is achieved in a study when we get a result that is very unlikely to occur by chance. A common criterion is that we have statistical significance if the likelihood of an event occurring by chance is 5% or less.

- Getting 98 girls in 100 random births *is* statistically significant because such an extreme outcome is not likely to result from random chance.
- Getting 52 girls in 100 births *is not* statistically significant because that event could easily occur with random chance.

**Practical Significance** It is possible that some treatment or finding is effective, but common sense might suggest that the treatment or finding does not make enough of a difference to justify its use or to be practical, as illustrated in Example 3.

### EXAMPLE 3 Statistical Significance Versus Practical Significance

ProCare Industries once supplied a product named Gender Choice that supposedly increased the chance of a couple having a baby with the gender that they desired. In the absence of any evidence of its effectiveness, the product was banned by the Food and Drug Administration (FDA) as a “gross deception of the consumer.” But suppose that the product was tested with 10,000 couples who wanted to have baby girls, and the results consist of 5200 baby girls born in the 10,000 births. This result is statistically significant because the likelihood of it happening due to chance is only 0.003%, so chance doesn’t seem like a feasible explanation. That 52% rate of girls is statistically significant, but it lacks practical significance because 52% is only slightly above 50%. Couples would not want to spend the time and money to increase the likelihood of a girl from 50% to 52%. (Note: In reality, the likelihood of a baby being a girl is about 48.8%, not 50%.)

**YOUR TURN** Do Exercise 15 “Gender Selection.”

## Analyzing Data: Potential Pitfalls

Here are a few more items that could cause problems when analyzing data.

**Misleading Conclusions** When forming a conclusion based on a statistical analysis, we should make statements that are clear even to those who have no understanding of statistics and its terminology. We should carefully avoid making statements not justified by the statistical analysis. For example, later in this book we introduce the concept of a correlation, or association between two variables, such as numbers of registered pleasure boats and numbers of manatee deaths from encounters with boats. A statistical analysis might justify the statement that there is a correlation between numbers of boats and numbers of manatee fatalities, but it would not justify a statement that an increase in the number of boats *causes* an increase in the number of manatee fatalities. Such a statement about causality can be justified by physical evidence, not by statistical analysis.

**Correlation does not imply causation.**

**Sample Data Reported Instead of Measured** When collecting data from people, it is better to take measurements yourself instead of asking subjects to *report* results. Ask people what they weigh and you are likely to get their *desired* weights, not their actual weights. People tend to round, usually down, sometimes *way* down. When asked, someone with a weight of 187 lb might respond that he or she weighs 160 lb. Accurate weights are collected by using a scale to *measure* weights, not by asking people what they weigh.

## Publication Bias

There is a “publication bias” in professional journals. It is the tendency to publish positive results (such as showing that some treatment is effective) much more often than negative results (such as showing that some treatment has no effect).



In the article “Registering Clinical Trials” (*Journal of the American Medical Association*, Vol. 290, No. 4), authors Kay Dickersin and Drummond Rennie state that “the result of not knowing who has performed what (clinical trial) is loss and distortion of the evidence, waste and duplication of trials, inability of funding agencies to plan, and a chaotic system from which only certain sponsors might benefit, and is invariably against the interest of those who offered to participate in trials and of patients in general.” They support a process in which *all* clinical trials are registered in one central system, so that future researchers have access to all previous studies, not just the studies that were published.

## Statistics Is Sexy



CareerCast.com is a job website, and its organizers analyzed professions using five criteria: environment, income, employment prospects, physical demands, and stress. Based on that study, here are the top 10 jobs: (1) mathematician, (2) actuary, (3) *statistician* (author's emphasis), (4) biologist, (5) software engineer, (6) computer system analyst, (7) historian, (8) sociologist, (9) industrial designer, (10) accountant. Lumberjacks are at the bottom of the list with very low pay, dangerous work, and poor employment prospects.

Reporter Steve Lohr wrote the article "For Today's Graduate, Just One Word: Statistics" in the *New York Times*. In that article he quoted the chief economist at Google as saying that "the sexy job in the next 10 years will be statisticians. And I'm not kidding."

**Loaded Questions** If survey questions are not worded carefully, the results of a study can be misleading. Survey questions can be "loaded," or intentionally worded to elicit a desired response. Here are the actual rates of "yes" responses for the two different wordings of a question:

97% yes: "Should the President have the line item veto to eliminate waste?"

57% yes: "Should the President have the line item veto, or not?"

**Order of Questions** Sometimes survey questions are unintentionally loaded by such factors as the order of the items being considered. See the following two questions from a poll conducted in Germany, along with the very different response rates:

"Would you say that traffic contributes more or less to air pollution than industry?"  
(45% blamed traffic; 27% blamed industry.)

"Would you say that industry contributes more or less to air pollution than traffic?"  
(24% blamed traffic; 57% blamed industry.)

In addition to the order of items within a question, as illustrated above, the order of separate questions could also affect responses.

**Nonresponse** A *nonresponse* occurs when someone either refuses to respond to a survey question or is unavailable. When people are asked survey questions, some firmly refuse to answer. The refusal rate has been growing in recent years, partly because many persistent telemarketers try to sell goods or services by beginning with a sales pitch that initially sounds as though it is part of an opinion poll. (This "selling under the guise" of a poll is called *sugging*.) In *Lies, Damn Lies, and Statistics*, author Michael Wheeler makes this very important observation:

**People who refuse to talk to pollsters are likely to be different from those who do not. Some may be fearful of strangers and others jealous of their privacy, but their refusal to talk demonstrates that their view of the world around them is markedly different from that of those people who will let poll-takers into their homes.**

**Percentages** Some studies cite misleading or unclear percentages. Note that 100% of some quantity is *all* of it, but if there are references made to percentages that exceed 100%, such references are often not justified. In an ad for The Club, a device used to discourage car thefts, it was stated that "The Club reduces your odds of car theft by 400%." If the Club eliminated *all* car thefts, it would reduce the odds of car theft by 100%, so the 400% figure is misleading and doesn't make sense.

The following list identifies some key principles to apply when dealing with percentages. These principles all use the basic concept that % or "percent" really means "divided by 100." The first principle that follows is used often in this book.

**Percentage of:** To find a percentage of an amount, replace the % symbol with division by 100, and then interpret "of" to be multiplication. This example shows that 6% of 1200 is 72:

$$6\% \text{ of } 1200 \text{ responses} = \frac{6}{100} \times 1200 = 72$$

**Decimal → Percentage:** To convert from a decimal to a percentage, multiply by 100%. This example shows that 0.25 is equivalent to 25%:

$$0.25 \rightarrow 0.25 \times 100\% = 25\%$$

**Fraction → Percentage:** To convert from a fraction to a percentage, divide the denominator into the numerator to get an equivalent decimal number; then multiply by 100%. This example shows that the fraction  $\frac{3}{4}$  is equivalent to 75%:

$$\frac{3}{4} = 0.75 \rightarrow 0.75 \times 100\% = 75\%$$

**Percentage → Decimal:** To convert from a percentage to a decimal number, replace the % symbol with division by 100. This example shows that 85% is equivalent to 0.85:

$$85\% = \frac{85}{100} = 0.85$$

## 1-1 Basic Skills and Concepts

### Statistical Literacy and Critical Thinking

**1. Online Medical Info** *USA Today* posted this question on its website: “How often do you seek medical information online?” Of 1072 Internet users who chose to respond, 38% of them responded with “frequently.” What term is used to describe this type of survey in which the people surveyed consist of those who decided to respond? What is wrong with this type of sampling method?

**2. Reported Versus Measured** In a survey of 1046 adults conducted by Bradley Corporation, subjects were asked how often they wash their hands when using a public restroom, and 70% of the respondents said “always.”

a. Identify the sample and the population.

b. Why would better results be obtained by observing the hand washing instead of asking about it?

**3. Statistical Significance Versus Practical Significance** When testing a new treatment, what is the difference between statistical significance and practical significance? Can a treatment have statistical significance, but not practical significance?

**4. Correlation** One study showed that for a recent period of 11 years, there was a strong correlation (or association) between the numbers of people who drowned in swimming pools and the amounts of power generated by nuclear power plants (based on data from the Centers for Disease Control and Prevention and the Department of Energy). Does this imply that increasing power from nuclear power plants is the cause of more deaths in swimming pools? Why or why not?

**Consider the Source.** In Exercises 5–8, determine whether the given source has the potential to create a bias in a statistical study.

**5. Physicians Committee for Responsible Medicine** The Physicians Committee for Responsible Medicine tends to oppose the use of meat and dairy products in our diets, and that organization has received hundreds of thousands of dollars in funding from the Foundation to Support Animal Protection.

**6. Arsenic in Rice** Amounts of arsenic in samples of rice grown in Texas were measured by the Food and Drug Administration (FDA).

**7. Brain Size** A data set in Appendix B includes brain volumes from 10 pairs of monozygotic (identical) twins. The data were collected by researchers at Harvard University, Massachusetts General Hospital, Dartmouth College, and the University of California at Davis.