Essentials of Statistics

FIFTH EDITION

Mario F. Triola

ALWAYS LEARNING

PEARSON

Symbol Table

$\overline{A}$	complement of event A
H_0	null hypothesis
H_1	alternative hypothesis
α	alpha; probability of a type I error or the area of the critical region
β	beta; probability of a type II error
r	sample linear correlation coefficient
ρ	rho; population linear correlation coefficient
r^2	coefficient of determination
r_s	Spearman's rank correlation coefficient
b_1	point estimate of the slope of the regression line
b_0	point estimate of the y-intercept of the regression line
$\hat{y}$	predicted value of y
d	difference between two matched values
$\overline{d}$	mean of the differences d found from matched sample data
s_d	standard deviation of the differences d found from matched sample data
s_e	standard error of estimate
$\mu_{\overline{x}}$	mean of the population of all possible sample means $\overline{x}$
$\sigma_{\overline{x}}$	standard deviation of the population of all possible sample means $\overline{x}$
E	margin of error of the estimate of a population parameter, or expected value
Q_1, Q_2, Q_3	quartiles
$D_1, D_2, \ldots, D_9$	deciles
$P_1, P_2, \ldots, P_{99}$	percentiles
x	data value
f	frequency with which a value occurs
Σ	capital sigma; summation
Σx	sum of the values
Σx^2	sum of the squares of the values
$(\Sigma x)^2$	square of the sum of all values
Σxy	sum of the products of each x value multiplied by the corresponding y value
n	number of values in a sample
$n!$	n factorial
N	number of values in a finite population; also used as the size of all samples combined
k	number of samples or populations or categories
$\overline{x}$	mean of the values in a sample
μ	mu; mean of all values in a population
s	standard deviation of a set of sample values
σ	lowercase sigma; standard deviation of all values in a population
s^2	variance of a set of sample values
σ^2	variance of all values in a population
z	standard score
$z_{\alpha/2}$	critical value of z
t	t distribution
$t_{\alpha/2}$	critical value of t
df	number of degrees of freedom
F	F distribution
χ^2	chi-square distribution
χ^2_R	right-tailed critical value of chi-square
χ^2_L	left-tailed critical value of chi-square
p	probability of an event or the population proportion
q	probability or proportion equal to $1 - p$
$\hat{p}$	sample proportion
$\hat{q}$	sample proportion equal to $1 - \hat{p}$
$\overline{p}$	proportion obtained by pooling two samples
$\overline{q}$	proportion or probability equal to $1 - \overline{p}$
$P(A)$	probability of event A
$P(A \mid B)$	probability of event A, assuming event B has occurred
${}_nP_r$	number of permutations of n items selected r at a time
${}_nC_r$	number of combinations of n items selected r at a time

Essentials of Statistics

Essentials of Statistics

5TH EDITION GLOBAL EDITION

MARIO F. TRIOLA

PEARSON

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

Editor in Chief: Deirdre Lynch
Executive Editor: Christopher Cummings
Senior Content Editors: Rachel Reeve and Chere Bemelmans
Assistant Editor: Sonia Ashraf
Senior Managing Editor: Karen Wernholm
Production Project Managers: Tracy Patruno and Mary Sanger
Associate Director of Design: Andrea Nix
Art Director and Cover Designer: Beth Paquin
Digital Assets Manager: Marianne Groth
Media Producer: Vicki Dreyfus
Software Developers: Mary Durnwald and Bob Carroll
Senior Marketing Manager: Erin Lane
Marketing Assistant: Kathleen DeChavez
Senior Author Support/Technology Specialist: Joe Vetere
Image Manager: Rachel Youdelman
Procurement Specialist: Debbie Rossi
Production Coordination, Composition, Illustrations: Cenveo® Publisher Services
Text Design: Leslie Haimes
Cover Image: tadamichi/Shutterstock;
Head of Learning Asset Acquisition, Global Edition:Laura Dent
Associate Acquisitions Editor, Global Edition: Murchana Borthakur
Senior Manufacturing Controller, Global Edition: Trudy Kimber
Project Editor, Global Edition: Aaditya Bugga

Credits appear on pages 673–674, which constitute a continuation of the copyright page.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and Pearson was aware of a trademark claim, the designations have been printed in initial caps or all caps.

Pearson Education Limited
Edinburgh Gate
Harlow
Essex CM20 2JE
England

and Associated Companies throughout the world

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

© Pearson Education Limited, 2015

The right of Mario F. Triola to be identified as the author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Authorized adaptation from the United States edition, entitled Essentials of Statistics, 5th edition, ISBN 978-0-321-92459-9, by Mario F. Triola, published by Pearson Education © 2015.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without either the prior written permission of the publisher or a license permitting restricted copying in the United Kingdom issued by the Copyright Licensing Agency Ltd, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

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

ISBN 10: 1-292-05876-5
ISBN 13: 978-1-292-05876-4

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

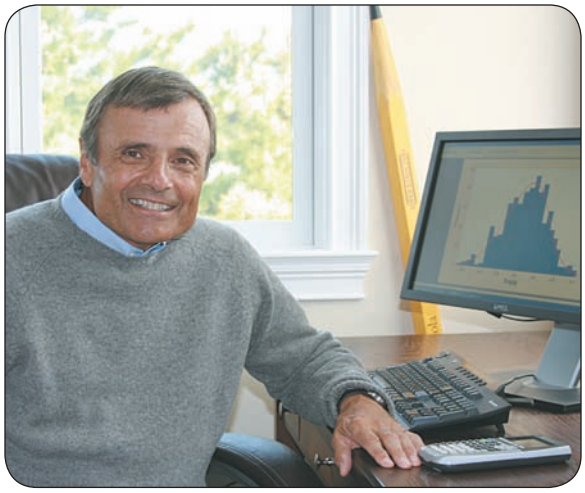
10 9 8 7 6 5 4 3 2 1

Typeset in 11 AGaramondPro-Regular by Cenveo® Publisher Services

Printed and bound by Courier Kendallville in the United States of America.

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*, 12th edition, *Elementary Statistics Using Excel*, 5th edition, *Elementary Statistics Using the TI-83/84 Plus Calculator*, 4th edition, and he is a co-author of *Biostatistics for the Biological and Health Sciences, Statistical Reasoning for Everyday Life*, 4th edition, *Business Statistics,* and *Introduction to Technical Mathematics*, 5th edition. *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, and 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 analyzing survey results for the New York City Transit Authority. Marty has testified as an expert witness in 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

1 Introduction to Statistics **20**
- **1-1** Review and Preview 22
- **1-2** Statistical and Critical Thinking 23
- **1-3** Types of Data 33
- **1-4** Collecting Sample Data 41

2 Summarizing and Graphing Data **60**
- **2-1** Review and Preview 62
- **2-2** Frequency Distributions 62
- **2-3** Histograms 72
- **2-4** Graphs That Enlighten and Graphs That Deceive 78

3 Statistics for Describing, Exploring, and Comparing Data **96**
- **3-1** Review and Preview 98
- **3-2** Measures of Center 98
- **3-3** Measures of Variation 114
- **3-4** Measures of Relative Standing and Boxplots 130

4 Probability **150**
- **4-1** Review and Preview 152
- **4-2** Basic Concepts of Probability 153
- **4-3** Addition Rule 167
- **4-4** Multiplication Rule: Basics 174
- **4-5** Multiplication Rule: Complements and Conditional Probability 186
- **4-6** Counting 193
- **4-7** Probabilities through Simulations (on companion Web site)
- **4-8** Bayes' Theorem (on companion Web site)

5 Discrete Probability Distributions **212**
- **5-1** Review and Preview 214
- **5-2** Probability Distributions 214
- **5-3** Binomial Probability Distributions 228
- **5-4** Parameters for Binomial Distributions 241

6 Normal Probability Distributions **254**
- **6-1** Review and Preview 256
- **6-2** The Standard Normal Distribution 257
- **6-3** Applications of Normal Distributions 270
- **6-4** Sampling Distributions and Estimators 284
- **6-5** The Central Limit Theorem 296
- **6-6** Assessing Normality 309
- **6-7** Normal as Approximation to Binomial 317

7 Estimates and Sample Sizes **334**
- **7-1** Review and Preview 336
- **7-2** Estimating a Population Proportion 336
- **7-3** Estimating a Population Mean 355
- **7-4** Estimating a Population Standard Deviation or Variance 373

8 Hypothesis Testing **392**
8-1 Review and Preview 394
8-2 Basics of Hypothesis Testing 394
8-3 Testing a Claim about a Proportion 411
8-4 Testing a Claim about a Mean 424
8-5 Testing a Claim about a Standard Deviation or Variance 435

9 Inferences from Two Samples **452**
9-1 Review and Preview 454
9-2 Two Proportions 454
9-3 Two Means: Independent Samples 465
9-4 Two Dependent Samples (Matched Pairs) 479

10 Correlation and Regression **498**
10-1 Review and Preview 500
10-2 Correlation 500
10-3 Regression 521
10-4 Rank Correlation 536

11 Chi-Square and Analysis of Variance **552**
11-1 Review and Preview 554
11-2 Goodness-of-Fit 554
11-3 Contingency Tables 565
11-4 Analysis of Variance 578

APPENDIX A Tables **601**
APPENDIX B Data Sets **609**
APPENDIX C Bibliography **638**
APPENDIX D Answers to Odd-Numbered Section Exercises **639**
(and all Quick Quizzes, all Review Exercises, and all Cumulative Review Exercises)

Credits 673
Index of Applications 675
Index 679

Preface

This Fifth Edition was written with several goals:

- To provide an abundance of new and interesting data sets, examples, and exercises.
- To foster personal growth of students through critical thinking, use of technology, collaborative work, and development of communication skills.
- To incorporate the latest and best methods used by professional statisticians.
- To include information personally helpful to students, such as the best job search methods and the importance of avoiding mistakes on résumés.
- To provide the largest and best set of supplements to enhance teaching and learning.

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:** 92% of the examples and 89% of the exercises use real data.
3. **Stress conceptual understanding rather than mere knowledge of procedures:** Instead of seeking simple numerical answers, exercises and examples involve conceptual understanding through questions that encourage practical interpretations of results. Also, each chapter includes a *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 *Using Technology* 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 companion Web site for the book includes instructions for downloading free text-specific software (STATDISK) and data sets formatted for several different technologies, which are also listed in Appendix B.
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, "Data to Decision" projects, and Cooperative Group Activity projects.

Audience/Prerequisites

Essentials of Statistics is written for students majoring in any subject. Algebra is used minimally, but students should have completed at least a high school or college elementary algebra course. In many cases, underlying theory is included, but this book does not require the mathematical rigor more suitable for mathematics majors.

Changes in This Edition

As in previous editions, this fifth edition includes a substantial revision of examples, exercises, and Chapter Problems, as shown in the following table:

	Number	New to This Edition	Use Real Data
Exercises	1585	86% (1362)	89% (1411)
Examples	196	85% (166)	92% (181)
Chapter Problems	11	100% (11)	100% (11)

Organization

Combined Sections

- The 4th edition Section 1-2 ("Statistical Thinking") and Section 1-4 ("Critical Thinking") have been combined into one section in this 5th edition:

  Section 1-2 Statistical and Critical Thinking

- The 4th edition Section 2-4 ("Statistical Graphics") and Section 2-5 ("Critical Thinking: Bad Graphs") have been combined into one section in this 5th edition:

  Section 2-4 Graphs That Enlighten and Graphs That Deceive

- The 4th edition Section 7-3 ("Estimating a Population Mean: σ Known") and Section 7-4 ("Estimating a Population Mean: σ Not Known") have been combined into one section in this 5th edition:

  Section 7-3 Estimating a Population Mean

  This change is motivated by two factors: (1) Technology makes use of the t distribution relatively simple, and (2) professional statisticians almost never use the normal distribution when constructing confidence interval estimates of population means.

- The 4th edition Section 8-4 ("Testing a Claim about a Mean: σ Known") and Section 8-5 ("Testing a Claim about a Mean: σ Not Known") have been combined into one section in this 5th edition:

  Section 8-4 Testing a Claim about a Mean

  This change is motivated by two factors: (1) Technology makes use of the t distribution relatively simple, and (2) professional statisticians almost never use the normal distribution when testing claims about a population mean.

Switched Sections Sections 6-6 and 6-7 from the previous edition have been switched so that Section 6-6 is now "Assessing Normality" and Section 6-7 is now "Normal as Approximation to Binomial." This change is motivated by the widespread availability of technology that facilitates methods for assessing normality, while the same technology has diminished the importance of using a normal approximation for a binomial distribution.

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 by dividing them 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.

Real data

Hundreds of hours have been devoted to finding data that are real, meaningful, and interesting to students. All of the Chapter Problems are based on real data, 92% of the examples are based on real data, and 89% of the exercises are based on real data. Some exercises refer to the 23 large data sets listed in Appendix B, and 10 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, where they are clearly identified.

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. *Sections 10-2 ("Correlation") and 10-3 ("Regression") can be covered early*. Simply limit coverage to Part 1 ("Basic Concepts") in each of those two sections.
- **Minimum probability:** Some instructors prefer extensive coverage of probability, while others prefer to include only basic concepts. Instructors preferring minimum coverage can include Section 4-2 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, and those topics can be covered with Sections 4-1 through 4-4.

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:

Chapter-opening features:

- A list of chapter sections previews the chapter for the student.
- A chapter-opening problem, using real data, motivates the chapter material. Examples and exercises that further explore the ideas presented in the opening problem are marked with an icon.
- The first section is a brief review of relevant earlier concepts and previews the chapter's objectives.

End-of-chapter features:

A chapter **Review** summarizes the key concepts and topics of the chapter.

A **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.

A **Technology Project** provides an activity for STATDISK, MINITAB®, Excel®, a TI-83/84 Plus calculator, or StatCrunch®.

From Data to Decision is a capstone problem that requires critical thinking and writing.

Cooperative Group Activities encourage active learning in groups.

Other features:

Real Data Sets Appendix B contains printed versions of 23 large data sets referenced throughout the book, including 10 that are new. These data sets are also available on the companion Web site (http://www.pearsonglobaleditions.com/triola) and MyStatLab®.

Margin Essays Of 107 margin essays, 20% are new and several others have been updated. New topics include *Statistics for Online Dating, DNA Evidence Misused, Bar Code,* and *How Many People Do You Know?*

Flowcharts The text includes 12 flowcharts that simplify and clarify more complex concepts and procedures. Animated versions of the text's flowcharts are available within MyStatLab.

Top 20 Topics The most important topics in any introductory statistics course are identified in the text with an icon. Students using MyStatLab have access to additional resources for learning these topics with definitions, animations, and video lessons.

Quick-Reference Tables A-2 and A-3 (the normal and t distributions) are reproduced at the end of this book.

Formula and Table Supplement This supplement, organized by chapter, gives students a quick reference for studying. It also includes the most commonly used tables.

Companion Web site The companion Web site contains the data sets from Appendix B available as txt files, MINITAB worksheets, SPSS files, SAS files, JMP files, Excel workbooks, and a TI-83/84 Plus application. The Web site also includes sections on *Probabilities Through Simulations* and *Bayes' Theorem*, an index of applications, a symbols table, programs for the TI-83/84 Plus graphing calculator, and instructions for obtaining STATDISK Statistical Software (Version 12).

Technology

New: This edition now includes instructions and displays from the StatCrunch technology, and XLSTAT is now used in Excel screenshots.

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 *Using Technology* subsection that includes instruction 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, TI-84 Plus, or TI-Nspire calculator with the TI-84 Plus keypad installed.) The end-of-chapter features include a *Technology Project.*

The STATDISK (Version 12) statistical software package is designed specifically for this textbook. STATDISK is free to users of this book and instructions for downloading it are included on the companion Web site.

Supplements

For the Student

Student Workbook for the Triola Statistics Series, by Anne Landry (Florida Community College at Jacksonville) offers additional examples, concept exercises, and vocabulary exercises for each chapter.

(ISBN-13: 978-0-321-89196-9; ISBN-10: 0-321-89196-1)

The following technology manuals include instructions, examples from the main text, and interpretations to complement those given in the text.

Excel Student Laboratory Manual and Workbook for the Triola Statistics Series, by Beverly Dretzke (University of Minnesota).

(ISBN-13: 978-0-321-83799-8; ISBN-10: 0-321-83799-1)

MINITAB Student Laboratory Manual and Workbook for the Triola Statistics Series, by Mario F. Triola.

(ISBN-13: 978-0-321-83379-2; ISBN-10: 0-321-83379-1)

Graphing Calculator Manual for the TI-83 Plus, TI-84 Plus, TI-89 and TI-Nspire, by Kathleen McLaughlin (University of Connecticut) and Dorothy Wakefield (University of Connecticut Health Center).

(ISBN-13: 978-0-321-83803-2; ISBN-10: 0-321-83803-3)

STATDISK Student Laboratory Manual and Workbook for the Triola Statistics Series (Download Only), by Mario F. Triola. These files are available to instructors and students through the Triola Statistics Series Web site, www.pearsonglobaleditions.com/triola, and MyStatLab.

SPSS Student Laboratory Manual and Workbook for the Triola Statistics Series (Download Only), by James J. Ball (Indiana State University). These files are available to instructors and students through the Triola Statistics Series Web site, www.pearsonglobaleditions.com/triola, and MyStatLab.

StatCrunch Manual (Download Only) for the Triola Statistics Series, by Diane Hollister (Reading Area Community College). These files are available to instructors and students through the Triola Statistics Series Web site, www.pearsonglobaleditions.com/triola, and MyStatLab.

For the Instructor

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 at www.pearsonglobaleditions.com/Triola or within MyStatLab.

Insider's Guide to Teaching with the Triola Statistics Series, 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.

(ISBN-13: 978-0-321-83372-3; ISBN-10: 0-321-83372-4)

Testing System: TestGen® (www.pearsoned.com/testgen) enables instructors to build, edit, and 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 testbank are available for download from www.pearsonglobaleditions.com/Triola.

PowerPoint® Lecture Slides: Free to qualified adopters, this classroom lecture presentation software is geared specifically to the sequence and philosophy of *Essentials of S tatistics.* 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 www.pearsonglobaleditions.com/Triola or within MyStatLab.

Active Learning Questions: Prepared in PowerPoint®, these multiple-choice questions are available for each chapter of the book, allowing instructors to quickly assess mastery of material in class. The Active Learning Questions are available to download from within MyStat Lab® and from www.pearsonglobaleditions.com/Triola.

Technology Resources

On the Triola Statistics Series Web site (http://www.pearsonglobaleditions.com/Triola), and **MyStatLab:**

- Appendix B data sets formatted for Minitab, SPSS, SAS, Excel, JMP, and as text files. Additionally, these data sets are available as an APP for the TI-83/84 Plus calculators, and supplemental programs for the TI-83/84 Plus calculator are also available.
- **STATDISK** statistical software instructions for download.
- Extra data sets, *Probabilities through Simulations*, *Bayes' Theorem*, and a symbols table.

Video Resources have been expanded and now supplement most sections in the book, with many topics presented by the author. The videos feature technologies found in the book and the worked-out Chapter Review exercises. This is an excellent resource for students who have missed class or wish to review a topic. It is also an excellent resource for instructors involved with distance learning, individual study, or self-paced learning programs. These videos also contain optional English and Spanish captioning. All videos are available through the MyStatLab online course.

MyStatLab™ Online Course (access code required) MyStatLab is a course management system that delivers **proven results** in helping individual students succeed. It provides **engaging experiences** that personalize, stimulate, and measure learning for each student. Tools are embedded to make it easy to integrate statistical software into the course. And, it comes from a **trusted partner** with educational expertise and an eye on the future.

- MyStatLab's 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.

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

- **Tutorial Exercises with Multimedia Learning Aids:** The homework and practice exercises in MyStatLab align with the exercises in the textbook, and most regenerate algorithmically to give students unlimited opportunity for practice and mastery. Exercises offer immediate helpful feedback, including guided solutions, sample problems, animations, and videos.
- **Adaptive Study Plan:** Pearson now offers an optional focus on adaptive learning in the study plan to allow students to work on just what they need to learn when it makes the most sense to learn it. The adaptive study plan maximizes students' potential for understanding and success.
- **Additional Question Libraries:** In addition to algorithmically regenerated questions that are aligned with your textbook, MyStatLab courses come with two additional question libraries. **450 Getting Ready for Statistics** questions cover the developmental math topics students need for the course. These can be assigned as a prerequisite to other assignments, if desired. The **1000 Conceptual Question Library** requires students to apply their statistical understanding.
- **StatCrunch®:** MyStatLab includes Web-based statistical software, StatCrunch, within the online assessment platform so that students can analyze data sets from exercises and the text. In addition, MyStatLab includes access to **www.StatCrunch.com,** a Web site where users can access thousands of shared data sets, conduct online surveys, perform complex analyses using the powerful statistical software, and generate compelling reports.
- **Integration of Statistical Software:** We make it easy to copy our data sets, both from the ebook and the MyStatLab questions, into software such as StatCrunch, MINITAB, Excel, and more. Students have access to a variety of support tools—Technology Instruction Videos, Technology Study Cards, and Manuals for select titles—to learn how to use statistical software.
- **StatTalk Videos:** Fun-loving statistician Andrew Vickers takes to the streets of Brooklyn, NY, to demonstrate important statistical concepts through interesting stories and real-life events. This series of 24 videos includes available assessment questions and an instructor's guide.

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

StatCrunch®
StatCrunch is powerful Web-based statistical software that allows users to perform complex analyses, share data sets, and generate compelling reports of their data. The vibrant online community offers 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 MyStatLab kit, and StatCrunch is available by itself to qualified adopters. For more information, visit our Web site at www.StatCrunch.com, or contact your Pearson representative.

The Student Edition of MINITAB is a condensed version of the professional release of MINITAB statistical software. It offers the full range of statistical methods and graphical capabilities, along with worksheets that can include up to 10,000 data points. Individual copies of the software can be bundled with the text (ISBN-13 978-0-13-143661-9; ISBN-10: 0-13-143661-9).

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: 978-0-321-89164-8; ISBN-10: 0-321-89164-3).

Acknowledgments

I would like to thank the thousands of statistics professors and students who have contributed to the success of this book. I would like to extend special thanks to Mitch Levy, Broward College; Kate Kozak, Coconino Community College; Steve Schwager, Cornell University; Rick Woodmansee, Sacramento City College; Rob Fusco, Broward College; Joe Pick, Palm Beach State College; Richard Weil, Brown College; Donald Burd, Monroe College; James Bryan, Merced College; Richard Herbst, Montgomery County Community College; Diane Hollister, Reading Area Community College; George Jahn, Palm Beach State College; Dan Kumpf, Ventura College; Kim McHale, Heartland Community College; Ken Mulzet, Florida State College at Jacksonville; Sandra Spain, Thomas Nelson Community College; Ellen G. Stutes, Louisiana State University, Eunice; Barbara Ward, Belmont University; Richard Hertz; Chris Vertullo, Marist College; Kelly Smitch, Brevard College; Robert Black, United States Air Force Academy; Michael Huber.

This fifth 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 Arts and Sciences team. I thank Chris Cummings, Deirdre Lynch, Rachel Reeve, Chere Bemelmans, John Orr (of Cenveo Publisher Services), Tracy Patruno, Mary Sanger, Sonia Ashraf, Christina Lepre, Joe Vetere, and Beth Paquin. I extend special thanks to Marc Triola, M.D., New York University, for his outstanding work on the STATDISK software, and Scott Triola for his great help in creating this new edition.

I thank the following for their help in checking the accuracy of text and answers in this fifth edition: James Lapp, David Lund, and Kimberley Polly.

M.F. T.
Madison, Connecticut
September 2013

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

Contributor:
Neeraj Misra, *Indian Institute of Technology, Kanpur*

Reviewers:
Dilip C. Nath, *Gauhati University*
Ruma Talukdar, *Cotton College State University*

Essentials of Statistics

1 Introduction to Statistics

1-1 Review and Preview

1-2 Statistical and Critical Thinking

1-3 Types of Data

1-4 Collecting Sample Data

chapter problem

Survey: Have you ever been hit with a computer virus?

The world in which we live is now saturated with surveys. Surveys are essential tools used in marketing. Surveys determine what television shows we watch. Surveys guide political candidates. Surveys shape business practices and many other aspects of our lives. Surveys provide us with understanding about the thinking of the rest of the world. Let's consider one particular survey dealing with a topic of great concern to all of us who have embraced the use of computer technology. The survey question and responses are given below, and Figure 1-1 graphically depicts the survey results. (Figure 1-1 was generated using Minitab statistical software.)

"Have you ever been hit by a computer virus?"

- Yes: 106,685
- No: 63,378

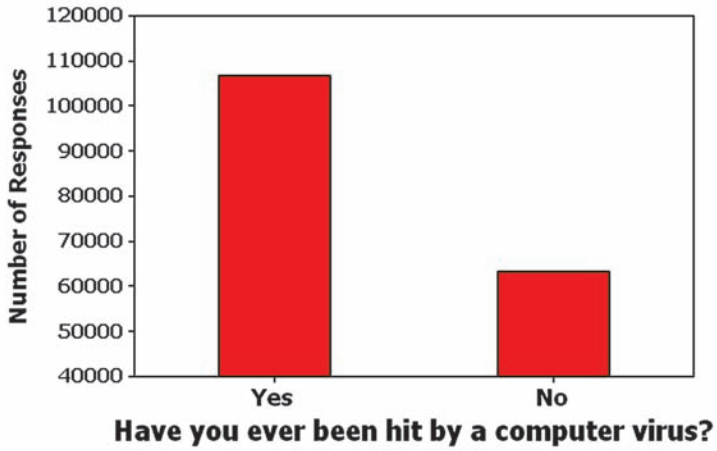


Figure 1-1 **Survey Results**

The results of the survey appear to be quite dramatic. The total number of respondents is 170,063 adults, and that is a very large number of respondents. Many polls have only about one thousand or two thousand respondents. Also, by looking at the bars in Figure 1-1, we see that roughly three times as many respondents have been hit by computer viruses as have not been hit. One important objective of this text 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? Does the graph in Figure 1-1 depict the results in a way that is not misleading?

The survey results presented here have two major flaws. Because these two flaws are among the most common, it is especially important to recognize them. Following are brief descriptions of each of the two major flaws.

Flaw 1: Misleading Graph Figure 1-1 is deceptive. Using a vertical scale that does not start at zero exaggerates the difference between the two numbers of responses. Thus Figure 1-1 makes it appear that the "yes" responses are about three times the number of "no" responses, but examination of the actual response counts shows that the "yes" responses are really about 1.7 times the "no" responses. Deceptive graphs are discussed in more detail in Section 2-4.

Flaw 2: Bad Sampling Method The survey responses are from a recent America OnLine survey of Internet users. The survey question was posted on the America OnLine Web site and Internet users decided whether to respond. This is an example of a *voluntary response sample*—a sample in which respondents decide themselves 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. The large number of respondents does not overcome this flaw of having a voluntary response sample. When we want to use 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 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 if we did so, we would miss the critical two flaws described above. We might then develop conclusions that are fundamentally wrong and misleading. Instead, we should develop skills in statistical thinking and critical thinking so we can understand why the survey is so seriously flawed and why we should not rely on it to yield any valid information.

1-1 Review and Preview

The first section of each of Chapter 1 through Chapter 11 begins with a brief review of what preceded the chapter, and a preview of what the chapter includes. This first chapter isn't preceded by much of anything except the Preface, and we won't review that (most people don't even read it). However, we can review and formally define some statistical terms that are commonly used. The Chapter Problem discussed an America OnLine poll that collected sample data. Polls collect data from a small part of a larger group so that we can learn something about the larger group. This is a common and important goal of statistics: Learn about a large group by examining sample data from some of its members. In this context, the terms *sample* and *population* have special meanings. Formal definitions for these and other basic terms are given here.

DEFINITIONS

Data are collections of observations, such as measurements, genders, or survey responses. (A single data value is called a *datum*, a term that does not see very much use.)

Statistics is the science of planning studies and experiments; obtaining data; and then 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.

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. See Example 1.

Example 1 **Gallup Poll: Identity Theft**

In a poll conducted by the Gallup corporation, 1013 adults in the United States were randomly selected and surveyed about identity theft. Results showed that 66% of the respondents worried about identity theft frequently or occasionally.

Gallup pollsters decided who would be asked to participate in the survey and they used a sound method of randomly selecting adults. The respondents are not a voluntary response sample, and the results are likely to be better than those obtained from the America OnLine survey discussed earlier.

In this case, the population consists of all 241,472,385 adults in the United States, and it is not practical to survey each of them. The sample consists of the 1013 adults who were surveyed. The objective is to use the sample data as a basis for drawing a conclusion about the population of all adults, and methods of statistics are helpful in drawing such conclusions.

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.

1-2 Statistical and Critical Thinking

Key Concept This section provides an overview of the process involved in conducting a statistical study. This process consists of "prepare, analyze, and conclude." We begin with a preparation that involves consideration of the context, consideration of the source of data, and consideration of the sampling method. Next, we construct suitable graphs, explore the data, and execute computations required for the statistical method being used. Finally, we form conclusions by determining whether results have statistical significance and practical significance. See Figure 1-2 for a summary of this process.

Figure 1-2 includes key elements in a statistical study. Note that the procedure outlined in Figure 1-2 does not focus on mathematical calculations. Thanks to wonderful developments in technology, we now have tools that effectively do the number crunching so that we can focus on understanding and interpreting results.

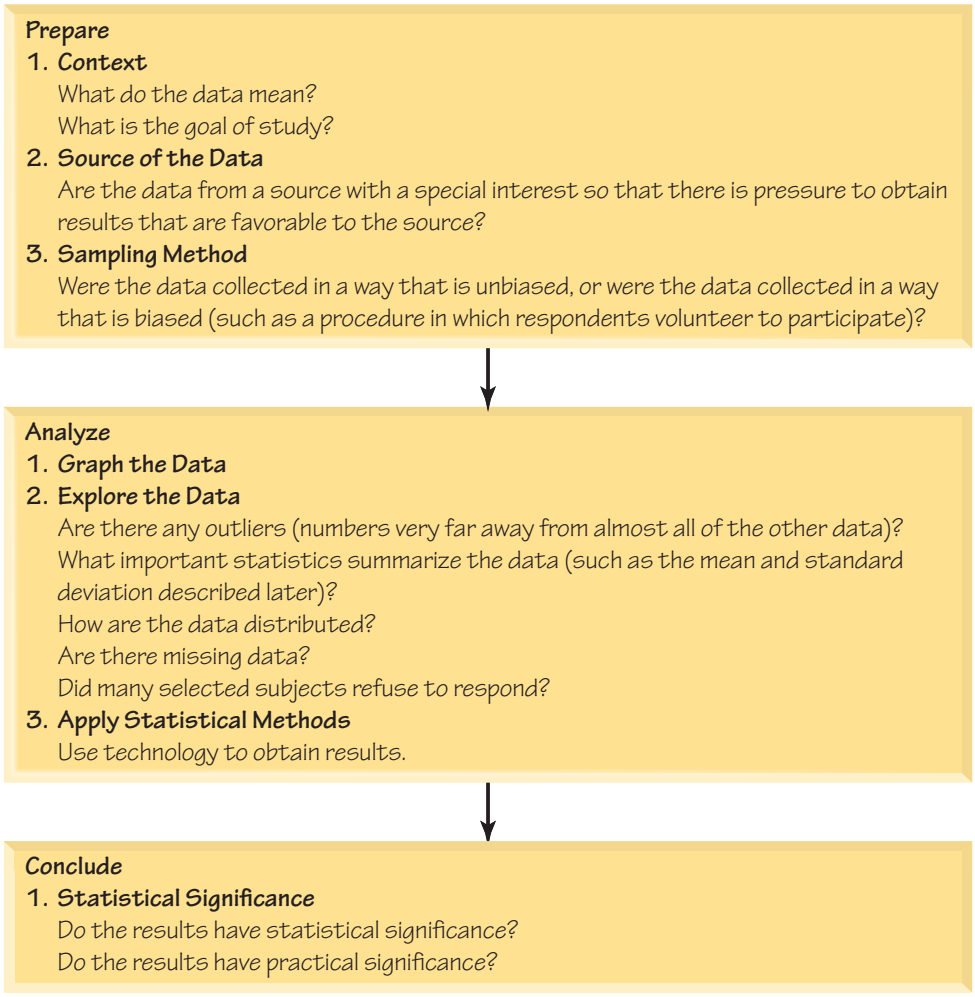


Figure 1-2 **Statistical Thinking**

Prepare

Context Let's consider the data in Table 1-1. (The data are from Data Set 6 in Appendix B.) The data in Table 1-1 consist of measured IQ scores and measured brain volumes from 10 different subjects. The data are matched in the sense that each individual "IQ/brain volume" pair of values is from the same subject. The first subject had a measured IQ score of 96 and a brain volume of 1005 cm^3. The format of Table 1-1 suggests the following goal: Determine whether there is a relationship between IQ score and brain volume. This goal suggests a possible hypothesis: People with larger brains tend to have higher IQ scores.

Table 1-1 IQ Scores and Brain Volumes (cm^3)

IQ	96	87	101	103	127	96	88	85	97	124
Brain Volume (cm^3)	1005	1035	1281	1051	1034	1079	1104	1439	1029	1160

Source of the Data The data in Table 1-1 were provided by M. J. Tramo, W. C. Loftus, T. A. Stukel, J. B. Weaver, and M. S. Gazziniga, who discuss the data in the article "Brain Size, Head Size, and IQ in Monozygotic Twins," *Neurology*, Vol. 50. The researchers are from reputable medical schools and hospitals, and they would not gain by putting spin on the results. In contrast, Kiwi Brands, a maker of shoe polish, commissioned a study that resulted in this statement, which was printed in some newspapers: "According to a nationwide survey of 250 hiring professionals, scuffed shoes was the most common reason for a male job seeker's failure to make a good first impression." We should be very wary of such a survey in which the sponsor can somehow profit from the results. When physicians who conduct clinical experiments on the efficacy of drugs receive funding from drug companies they have an incentive to obtain favorable results. Some professional journals, such as the *Journal of the American Medical Association*, now require that physicians report such funding in journal articles. We should be skeptical of studies from sources that may be biased.

Sampling Method The data in Table 1-1 were obtained from subjects who were recruited by researchers, and the subjects were paid for their participation. All subjects were between 24 years and 43 years of age, they all had at least a high school education, and the medical histories of subjects were reviewed in an effort to ensure that no subjects had neurologic or psychiatric disease. In this case, the sampling method appears to be sound.

Sampling methods and the use of randomization will be discussed in Section 1-4, but for now, we simply emphasize 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.

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 such a biased sample:

- Internet polls, in which people online can decide whether to respond
- Mail-in polls, in which subjects 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

With such voluntary response samples, we can draw valid conclusions only about the specific group of people who chose to participate; nevertheless, such samples are often incorrectly used to assert or imply conclusions about a larger population. From a statistical viewpoint, such a sample is fundamentally flawed and should not be used for making general statements about a larger population. The Chapter Problem involves an America OnLine poll with a voluntary response sample. See also Examples 1 and 2, which follow.

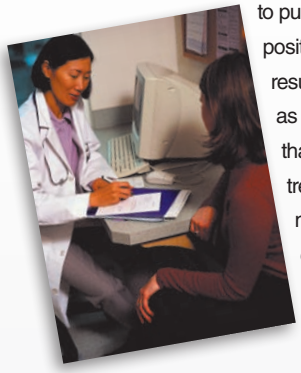
Value of a Statistical Life

The *value of a statistical life* (VSL) is a measure routinely calculated and used for making decisions in fields such as medicine, insurance, environmental health, and transportation safety. As of this writing, the value of a statistical life is $6.9 million.

Many people oppose the concept of putting a value on a human life, but the word *statistical* in the "value of a statistical life" is used to ensure that we don't equate it with the true worth of a human life. Some people legitimately argue that every life is priceless, but others argue that there are conditions in which it is impossible or impractical to save every life, so a value must be somehow assigned to a human life in order that sound and rational decisions can be made. Not far from the author's home, a parkway was modified at a cost of about $3 million to improve safety at a location where car occupants had previously died in traffic crashes. In the cost-benefit analysis that led to this improvement in safety, the value of a statistical life was surely considered.

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.



Example 1 Voluntary Response Sample

Literary Digest magazine conducted a poll for the 1936 presidential election by sending out 10 million ballots. The magazine received 2.3 million responses. The poll results suggested incorrectly that Alf Landon would win the presidency. In a much smaller poll of 50,000 people, George Gallup correctly predicted that Franklin D. Roosevelt would win. The lesson here is that it is not necessarily the *size* of the sample that makes it effective, but the *sampling method*. The *Literary Digest* ballots were sent to magazine subscribers as well as to registered car owners and those who used telephones. On the heels of the Great Depression, this group included disproportionately more wealthy people, who were Republicans. But the real flaw in the *Literary Digest* poll is that it resulted in a voluntary response sample. In contrast, Gallup used an approach in which he obtained a representative sample based on demographic factors. (Gallup modified his methods when he made a wrong prediction in the famous 1948 Dewey/Truman election. Gallup stopped polling too soon, and he failed to detect a late surge in support for Truman.) The *Literary Digest* poll is a classic illustration of the flaws inherent in basing conclusions on a voluntary response sample.

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 poll, 500 respondents were randomly selected and 72% of them wanted the United Nations to *stay* in 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 superior sampling method.

Analyze

Graph and Explore After carefully considering context, source of the data, and sampling method, we can proceed with an analysis that 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 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

Statistical Significance *Statistical significance* is achieved in a study when we get a result that is very unlikely to occur by chance.

- Getting 98 girls in 100 random births is statistically significant because such an extreme event is not likely to be the result of 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**

In a test of the Atkins weight loss program, 40 subjects using that program had a mean weight loss of 2.1 kg (or 4.6 pounds) after one year (based on data from "Comparison of the Atkins, Ornish, Weight Watchers, and Zone Diets for Weight Loss and Heart Disease Risk Reduction," by Dansinger et al., *Journal of the American Medical Association*, Vol. 293, No. 1). Using formal methods of statistical analysis, we can conclude that the mean weight loss of 2.1 kg is statistically significant. That is, based on statistical criteria, the diet appears to be effective. However, using common sense, it does not seem very worthwhile to pursue a weight loss program resulting in such relatively insignificant results. Someone starting a weight loss program would probably want to lose considerably more than 2.1 kg. Although the mean weight loss of 2.1 kg is statistically significant, it does not have practical significance. The statistical analysis suggests that the weight loss program is effective, but *practical* considerations suggest that the program is basically ineffective.

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, Section 10-2 introduces the concept of a *correlation*, or association between two variables, such as smoking and pulse rate. A statistical analysis might justify the statement that there is a correlation between the number of cigarettes smoked and pulse rate, but it would not justify a statement that the number of cigarettes smoked *causes* a person's pulse rate to change. Such a statement about causality can be justified by physical evidence, not by statistical analysis.

Correlation does not imply causation.

Reported Results 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. Accurate weights are collected by using a scale to measure weights, not by asking people to report their weights.

Small Samples Conclusions should not be based on samples that are far too small. The Children's Defense Fund published *Children Out of School in America*, in which it was reported that among secondary school students suspended in one region, 67% were suspended at least three times. But that figure is based on a sample of only *three* students! Media reports failed to mention that this sample size was so small.

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

Detecting Phony Data

A class is given the homework assignment of recording the results when a coin is tossed 500 times. One dishonest student decides to save time by just making up the results instead of actually flipping a coin. Because people generally cannot make up results that are really random, we can often identify such phony data. With 500 tosses of an actual coin, it is extremely likely that at some point, you will get a run of six heads or six tails, but people almost never include such a run when they make up results.

Another way to detect fabricated data is to establish that the results violate Benford's law: For many collections of data, the leading digits are not uniformly distributed. Instead, the leading digits of 1, 2, . . . , 9 occur with rates of 30%, 18%, 12%, 10%, 8%, 7%, 6%, 5%, and 5%, respectively. (See "The Difficulty of Faking Data," by Theodore Hill, *Chance*, Vol. 12, No. 3.)

Statistics Is Sexy

CareerCast.com is a job Web site, and its organizers analyzed professions using five criteria: environment, income, employment prospects, physical demands, and stress. Based on that study, here are the top ten 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."

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.)

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 like it is part of an opinion poll. (This "selling under the guise" of a poll is now 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.

Missing Data Results can sometimes be dramatically affected by missing data. Sometimes sample data values are missing because of random factors (such as subjects dropping out of a study for reasons unrelated to the study), but some data are missing because of special factors, such as the tendency of people with low incomes to be less likely to report their incomes. It is well known that the U.S. Census suffers from missing people, and the missing people are often from the homeless or low income groups.

Precise Numbers Example 1 in Section 1-1 included a statement that there are 241,472,385 adults in the United States. Because that figure is very precise, many people incorrectly assume that it is also *accurate*. In this case, that number is an estimate, and it would be better to state that the number of adults in the United States is about 240 million.

Percentages Some studies cite misleading or unclear percentages. Keep in mind 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 referring to lost baggage, Continental Airlines ran ads claiming that this was "an area where we've already improved 100% in the last six months." In an editorial criticizing this statistic, the *New York Times* correctly interpreted the 100% improvement to mean that no baggage is now being lost—an accomplishment that was not achieved by Continental Airlines.

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

Percentage of: To find a *percentage of* an amount, drop the % symbol and divide the percentage value by 100, then multiply. This example shows that 6% of 1200 is 72:

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

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 and affix the % symbol. This example shows that the fraction 3/4 is equivalent to 75%:

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

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

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

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

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

There are many examples of the misuse of statistics. Books such as Darrell Huff's classic *How to Lie with Statistics,* Robert Reichard's *The Figure Finaglers,* and Cynthia Crossen's *Tainted Truth* describe some of those other cases. Understanding these practices will be extremely helpful in evaluating the statistical data encountered in everyday situations.

What Is Statistical Thinking? Statisticians universally agree that statistical thinking is good, but there are different views of what actually constitutes statistical thinking. If you ask the 18,000 members of the American Statistical Association to define statistical thinking, you will probably get 18,001 different definitions. In this section we have described statistical thinking in terms of the ability to see the big picture; to consider such relevant factors as context, source of data, and sampling method; and to form conclusions and identify practical implications. Statistical thinking involves critical thinking and the ability to make sense of results. Statistical thinking might involve determining whether results are statistically significant and practically significant. 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.

1-2 Basic Skills and Concepts

Statistical Literacy and Critical Thinking

1. Statistical Significance versus Practical Significance What is the difference between statistical significance and practical significance? Can a statistical study have statistical significance, but not practical significance?

2. Source of Data In conducting a statistical study, why is it important to consider the source of the data?

3. Voluntary Response Sample What is a voluntary response sample, and why is such a sample generally not suitable for a statistical study?

4. Correlation and Causation What is meant by the statement that "correlation does not imply causation"?

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. Body Measurements Data Set 1 in Appendix B includes body measurements obtained by the U.S. Department of Health and Human Services, National Center for Health Statistics.

7. Word Counts Data Set 17 in Appendix B includes word counts obtained by members of the Departments of Psychology at the University of Arizona, Washington University, and the University of Texas at Austin.

8. Chocolate An article in *Journal of Nutrition* (Vol. 130, No. 8) noted that chocolate is rich in flavonoids. The article notes that "regular consumption of foods rich in flavonoids may reduce the risk of coronary heart disease." The study received funding from Mars, Inc., the candy company, and the Chocolate Manufacturers Association.

Sampling Method. ***In Exercises 9–12, determine whether the sampling method appears to be sound or is flawed.***

9. Wi-Fi Security A survey of 721 subjects involved the providing of personal information when using Wi-Fi hotspots. The survey subjects were Internet users who responded to a question that was posted on the electronic edition of *USA Today.*

10. Text Messaging In a survey of 109 subjects, each was asked to indicate how many text messages they send and receive each day. The sample consisted of those who chose to respond to the request posted on the StatCrunch Web site.

11. Applying for a Job In a survey of 514 human resource professionals, each was asked about the importance of the appearance of a job applicant. The survey subjects were randomly selected by Harris Interactive pollsters.

12. Evolution In a survey of beliefs about evolution, Gallup pollsters randomly selected and telephoned 1018 adults in the United States.

Statistical Significance and Practical Significance. ***In Exercises 13–16, determine whether the results appear to have statistical significance, and also determine whether the results have practical significance.***

13. Waite Diet In a study of the Marisa Waite diet, four subjects lost an average of 45 pounds. It is found that there is about a 30% chance of getting such results with a diet that has no effect.

14. Gender Selection In a study of the Gender Aide method of gender selection, 1000 users of the method gave birth to 540 boys and 460 girls. There is about a 1% chance that such extreme results would occur if the method had no effect.

15. Guessing Technique When making random guesses for difficult multiple-choice test questions with possible answers of a, b, c, d, and e, we expect to get about 20% of the answers correct. The Ashton Prep Program claims to have developed a better method of guessing. In a test of that program, guesses were made for 100 answers, and 23 were found to be correct. There is a 23% chance of getting such results if the program has no effect.