

A string of red pennants flying against a clear blue sky. The pennants are arranged in a descending line from the top right towards the bottom left. The largest pennant is at the top right, and the string continues down to the bottom left with many smaller pennants. The background is a solid, clear blue sky.

Mario F. Triola

ELEMENTARY
STATISTICS

5th edition

using Excel[®]



Elementary Statistics Using Excel

5TH 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

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

Editor in Chief: Deirdre Lynch
Acquisitions Editor: Christopher Cummings
Senior Content Editor: Rachel Reeve
Editorial Assistant: Sonia Ashraf
Senior Managing Editor: Karen Wernholm
Senior Production Project Manager: Tracy Patruno
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
Marketing Manager: Erin Lane
Marketing Assistant: Kathleen DeChavez
Senior Author Support/Technology Specialist: Joe Vetere
Image Manager: Rachel Youdelman
Procurement Manager: Evelyn Beaton
Procurement Specialist: Debbie Rossi
Senior Media Procurement Specialist: Ginny Michaud
Production Coordination, Composition, Illustrations:
 Cenveo® Publisher Services
Text Design: Leslie Haimes
Cover Images: (kites) Manuel Fernandes/Shutterstock;
 (pencils) Diane Miller/iStockphoto

Credits appear on pages 857–858, 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.

Cataloging-in-Publication Data is on file with the Library of Congress.

Copyright © 2014, 2010, 2007 Pearson Education, Inc. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America. For information on obtaining permission for use of material in this work, please submit a written request to Pearson Education, Inc., Rights and Contracts Department, 501 Boylston Street, Suite 900, Boston, MA 02116, fax your request to 617-671-3447, or e-mail at <http://www.pearsoned.com/legal/permissions.htm>.

1 2 3 4 5 6 7 8 9 10—CRK—16 15 14 13 12

PEARSON

www.pearsonhighered.com

ISBN-10: 0-321-85166-8

ISBN-13: 978-0-321-85166-6

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, *Essentials of Statistics*, 4th edition, *Elementary Statistics Using the TI-83/84 Plus Calculator*, 3rd 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	2
	1-1 Review and Preview 4	
	1-2 Statistical and Critical Thinking 5	
	1-3 Types of Data 15	
	1-4 Collecting Sample Data 23	
	1-5 Introduction to Excel 35	
2	Summarizing and Graphing Data	50
	2-1 Review and Preview 52	
	2-2 Frequency Distributions 52	
	2-3 Histograms 64	
	2-4 Graphs That Enlighten and Graphs That Deceive 71	
3	Statistics for Describing, Exploring, and Comparing Data	94
	3-1 Review and Preview 96	
	3-2 Measures of Center 96	
	3-3 Measures of Variation 113	
	3-4 Measures of Relative Standing and Boxplots 130	
4	Probability	152
	4-1 Review and Preview 154	
	4-2 Basic Concepts of Probability 155	
	4-3 Addition Rule 170	
	4-4 Multiplication Rule: Basics 177	
	4-5 Multiplication Rule: Complements and Conditional Probability 189	
	4-6 Counting 196	
	4-7 Probabilities Through Simulations (on CD-ROM)	
	4-8 Bayes' Theorem (on CD-ROM)	
5	Discrete Probability Distributions	216
	5-1 Review and Preview 218	
	5-2 Probability Distributions 218	
	5-3 Binomial Probability Distributions 233	
	5-4 Parameters for Binomial Distributions 246	
	5-5 Poisson Probability Distributions 251	
6	Normal Probability Distributions	266
	6-1 Review and Preview 268	
	6-2 The Standard Normal Distribution 269	
	6-3 Applications of Normal Distributions 282	
	6-4 Sampling Distributions and Estimators 296	
	6-5 The Central Limit Theorem 308	
	6-6 Assessing Normality 323	
	6-7 Normal as Approximation to Binomial 331	
7	Estimates and Sample Sizes	348
	7-1 Review and Preview 350	
	7-2 Estimating a Population Proportion 350	
	7-3 Estimating a Population Mean 370	
	7-4 Estimating a Population Standard Deviation or Variance 389	
8	Hypothesis Testing	408
	8-1 Review and Preview 410	

	8-2	Basics of Hypothesis Testing	410	
	8-3	Testing a Claim About a Proportion	427	
	8-4	Testing a Claim About a Mean	440	
	8-5	Testing a Claim About a Standard Deviation or Variance	453	
9		Inferences from Two Samples		470
	9-1	Review and Preview	472	
	9-2	Two Proportions	472	
	9-3	Two Means: Independent Samples	484	
	9-4	Two Dependent Samples (Matched Pairs)	499	
	9-5	Two Variances or Standard Deviations	510	
10		Correlation and Regression		528
	10-1	Review and Preview	530	
	10-2	Correlation	530	
	10-3	Regression	552	
	10-4	Prediction Intervals and Variation	568	
	10-5	Multiple Regression	575	
	10-6	Nonlinear Regression	587	
11		Goodness-of-Fit and Contingency Tables		602
	11-1	Review and Preview	604	
	11-2	Goodness-of-Fit	604	
	11-3	Contingency Tables	616	
12		Analysis of Variance		636
	12-1	Review and Preview	638	
	12-2	One-Way ANOVA	639	
	12-3	Two-Way ANOVA	653	
13		Nonparametric Statistics		676
	13-1	Review and Preview	678	
	13-2	Sign Test	680	
	13-3	Wilcoxon Signed-Ranks Test for Matched Pairs	692	
	13-4	Wilcoxon Rank-Sum Test for Two Independent Samples	699	
	13-5	Kruskal-Wallis Test	705	
	13-6	Rank Correlation	712	
	13-7	Runs Test for Randomness	721	
14		Statistical Process Control		736
	14-1	Review and Preview	738	
	14-2	Control Charts for Variation and Mean	738	
	14-3	Control Charts for Attributes	749	
15		Ethics, Projects, Procedures, and Perspectives		760
	15-1	Ethics in Statistics	760	
	15-2	Projects	766	
	15-3	Procedures	767	
	15-4	Perspectives	767	
APPENDIX A		Tables		769
APPENDIX B		Data Sets		783
APPENDIX C		Bibliography		812
APPENDIX D		Answers to Odd-Numbered Section Exercises		813
		(and all Quick Quizzes, all Review Exercises, and all Cumulative Review Exercises)		
	Credits	857		
	Index	859		

Preface

The 21st century has brought with it an explosion of data that pervades every aspect of our lives, and as a result, the relevance and application of statistics have grown exponentially. *Elementary Statistics Using Excel* illustrates the relationship between statistics and our new data-driven world.

Why Excel?

Microsoft Excel continues to be the dominant spreadsheet program used in business and industry. Motivated by a desire to better serve their students by better preparing them for their professional careers, many professors include Excel as the technology used throughout the statistics course. By using Excel, students acquire important professional skills along with concepts of statistics. This marriage of statistics concepts and spreadsheet applications is giving birth to a generation of students who can enter professional careers armed with knowledge and skills that were once desired but are now required.

The marriage of statistics and Excel is not without its pitfalls. Excel's statistical capabilities are sometimes inadequate, so supplemental "add-ins" to Excel must be used. This book is designed to guide professors and students in choosing and using the best features of Excel and the appropriate add-in.

Which Version of Excel?

Excel instructions are given for Excel 2013, Excel 2010, Excel 2007, and Excel for Mac 2011. These versions of Excel are similar, so that instructions for one version typically apply to the other versions. Differences between versions will be highlighted in this book only when necessary. Older versions of Excel, including Excel 2003, are also similar in design and function, so users of these versions should be able to follow most instructions in this book.



Using Excel

Elementary Statistics Using Excel, Fifth Edition, is designed to be an introduction to basic statistics. Instead of being a manual of computer instructions, this book places strong emphasis on understanding concepts of statistics, with Excel included throughout as the key supplement. Topics are presented with illustrative examples, identification of required assumptions, and underlying theory. Excel instructions are provided along with typical displays of results. In some cases, such as examples involving *formulas* and *graphs*, detailed instructions are presented so that Excel can be used effectively in all applications, instead of those relating only to statistics.

Features for Excel

This book has key features that are critical to an effective use of Excel as a statistical tool, including the following.

- **Introduction to Excel** Section 1-5 introduces important and basic features of using Excel.

- **Excel Instructions** Detailed Excel instructions are included throughout the book.
- **Excel Displays** More than 300 displays of Excel results are included throughout the book (compared to 125 displays in the previous edition).
-  • **Excel Exercises** Exercises that are suitable for use with Excel are identified with a special icon.
- **Excel Answers** In some cases, exercise answers based on table values may be slightly different from answers based on the use of Excel. For those exercises, Appendix D generally gives both answers.
- **Excel Quick References** A concise list of Excel Quick References includes brief commands for executing most of the key statistical methods used in this book. The Excel Quick References are located near the end of the book.
-  • **Excel Projects** The wide variety of end-of-chapter features includes special Excel Projects.
- **XLSTAT Add-In** Excel lacks some important features, such as the ability to generate confidence intervals or to conduct hypothesis tests involving proportions. XLSTAT is a software add-in that enhances and expands the capability of Excel. XLSTAT can be installed to become part of the Excel program so that it is automatically available when Excel is launched. Instructions for installing XLSTAT are given in Section 1-5.
- **Data Analysis Add-In** Section 1-5 includes instructions for installing Excel's Data Analysis add-in. Instructions for using that add-in are included in appropriate sections throughout the book.
- **STATDISK** For those who choose to further supplement Excel, the separate program STATDISK has been designed specifically for the content of this book. Because STATDISK is easy to use and includes procedures for almost every major method in this book, it is extremely valuable as a tool for verifying results from Excel or obtaining results that cannot be obtained with Excel. Data can be easily transported between Excel and STATDISK by using Copy and Paste features. STATDISK is free to users of this book. Instructions for downloading STATDISK are on the CD-ROM that is included with new copies of this book.
- **CD** The data sets in Appendix B are included on this book's CD-ROM as Excel worksheets. Some exercises use those data sets, so students can open Excel files instead of manually entering long lists of data.

For those who choose to supplement Excel with other software, this book's CD-ROM includes the Appendix B data sets in a variety of formats. In addition to Excel, the data sets are available in formats for Minitab, SPSS, SAS, and the TI-83 Plus calculator or TI-84 Plus calculator, and they are also available as text files.

Goals

In addition to the inclusion of Excel, 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:** 91% 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 *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:** Excel displays are included throughout the book. Instructions for using Excel are also included. Each chapter includes an *Excel Project*. When there are discrepancies between answers based on tables and answers based on Excel, Appendix D provides *both* answers. The CD-ROM included with the book includes instructions for downloading free text-specific software (STATDISK) and data sets formatted for Excel, 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, Excel Projects, "From Data to Decision" projects, and Cooperative Group Activity projects.

Audience/Prerequisites

Elementary Statistics Using Excel 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

This fifth edition includes more extensive integration of Excel and substantially more Excel screen displays. 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	1851	85% (1582)	89% (1640)
Examples	227	84% (191)	91% (206)
Chapter Problems	14	100% (14)	100% (14)

Organization

New Section Section 15-1: "Ethics in Statistics" discusses ethics in data collection, analysis, and reporting, as well as samples of ethical guidelines that have been established by various organizations.

Combined Sections

- The fourth edition Section 1-2 (“Statistical Thinking”) and Section 1-4 (“Critical Thinking”) have been combined into one section in this fifth edition:

Section 1-2 Statistical and Critical Thinking

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

Section 2-4 Graphs That Enlighten and Graphs That Deceive

- The fourth 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 fifth 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 fourth 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 fifth 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.

- Section 11-4 (“McNemar’s Test for Matched Pairs”) in the fourth edition has been moved to Part 2 of Section 11-3 in this fifth edition.

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.

Moved to CD Section 4-6 (“Probabilities Through Simulations”) in the fourth edition has been moved from the text to the CD that is included with this book. Section 4-7 (“Counting”) from the fourth edition now becomes Section 4-6 in this fifth edition.

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.

Real data

All of the Chapter Problems are based on real data, 91% 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 *Elementary Statistics Using Excel* 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.
- An **Excel Project** provides an activity for using Excel.
- **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.pearsonhighered.com/triola>), the CD-ROM bound in the back of new copies of the book, and MyStatLab.

Margin Essays Of 117 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 17 flowcharts that simplify and clarify more complex concepts and procedures. Animated versions of the text's flowcharts are available within MyStatLab and MathXL.



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 Endpapers Tables A-2 and A-3 (the normal and t distributions) are reproduced on the rear inside cover pages.

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.

CD-ROM The CD-ROM was prepared by Mario F. Triola and is bound into the back of every new copy of the book. It contains the data sets from Appendix B available as text files, Minitab worksheets, SPSS files, SAS files, JMP files, Excel workbooks, and a TI-83/84 Plus application. The CD 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 uses XLSTAT as the Excel add-in, which is available for bundling with this product.

As in the preceding edition, there are many displays of screens from Excel throughout the book, and some exercises are based on displayed results from Excel. The end-of-chapter features include an *Excel Project*.

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.

Recent reviewers have included Raid W. Amin, University of West Florida; 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; Sanford Geraci, Broward College; Laura Heath, Palm Beach State College; Laura Hillerbrand, Broward College; Gary King, Ozarks Technical Community College; Mickey Levendusky, Pima County Community College; Tristan Londre, Blue River Community College; Alma Lopez, South Plains College; Carla Monticelli, Camden County Community College; 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; Pradipta Seal, Boston University;

Sharon Testone, Onondaga Community College; Dave Wallach, University of Findlay; Cheng Wang, Nova Southeastern University; Gail Wiltse, St. John River Community College; Claire Wladis, Borough of Manhattan Community 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 fifth edition of *Elementary Statistics Using Excel* is truly a team effort, and I consider myself fortunate to work with the dedication and commitment of the Pearson Arts & Sciences team. I thank Chris Cummings, Deirdre Lynch, Rachel Reeve, Sherry Berg (of Cenveo Publisher Services), Elizabeth Bernardi, Peggy McMahon, Tracy Patrino, Sonia Ashraf, Christina Lepre, Joe Vetere, and Beth Paquin. I thank Marc Triola, M.D., New York University, for his work on the STATDISK software, and I extend special thanks to Scott Triola, who was so instrumental in the creation of this text.

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
January 2013

Supplements

For the Student

Student's Solutions Manual, by James Lapp, provides detailed, worked-out solutions to all odd-numbered text exercises.

(ISBN-13: 978-0-321-85167-3; ISBN-10: 0-321-85167-6)

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)

XLSTAT™ for Pearson. Used by leading businesses and universities, XLSTAT is an Excel® add-in that offers a wide variety of functions to enhance the analytical capabilities of Microsoft® Excel® making it the ideal tool for your everyday data analysis and statistics requirements. XLSTAT™ is compatible with all Excel versions (except Mac 2008). Available for bundling with the text.

(ISBN-13: 978-0-321-75932-0; ISBN-10: 0-321-75932-X)

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, 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, 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 (Download Only), by Mario F. Triola. These files are available to instructors and students through the Triola Statistics Series Web site, www.pearsonhighered.com/triola, and MyStatLab.

SPSS Student Laboratory Manual and Workbook (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.pearsonhighered.com/triola, and MyStatLab.

StatCrunch™ Manual (Download Only), by Diane Hollister (Reading Area Community College). These files

are available to instructors and students through the Triola Statistics Series Web site, www.pearsonhighered.com/triola, and MyStatLab.

For the Instructor

Annotated Instructor's Edition, by Mario F. Triola, contains answers to exercises in the margin, plus recommended assignments, and teaching suggestions.

(ISBN-13: 978-0-321-85172-7; ISBN-10: 0-321-85172-2)

Instructor's Solutions Manual (Download Only), by James Lapp, contains solutions to all the exercises. These files are available to qualified instructors through Pearson Education's online catalog at www.pearsonhighered.com/irc 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, print, and administer tests using a computerized bank of questions developed to cover all the objectives of the text. TestGen is algorithmically based, allowing instructors to create multiple but equivalent versions of the same question or test with the click of a button. Instructors can also modify test bank questions or add new questions. The software and test bank are available for download from Pearson Education's online catalog at www.pearsonhighered.com. **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 *Elementary Statistics Using Excel*. 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 Education's online catalog at www.pearsonhighered.com/irc or within MyStatLab.

Active Learning Questions: Prepared in PowerPoint®, these questions are intended for use with classroom response systems. Several multiple-choice questions are available for each section of the book, allowing instructors to quickly assess mastery of material in class. These files are available to qualified instructors through Pearson Education's online catalog at www.pearsonhighered.com/irc or within MyStatLab.

Technology Resources

On the CD-ROM, Triola Statistics Series Web site (<http://www.pearsonhighered.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*, an index of applications, 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.

- MyStatLab can be successfully implemented in any environment—lab-based, hybrid, fully online, traditional—and demonstrates the quantifiable difference that integrated usage has on student retention, subsequent success, and overall achievement.
- MyStatLab's comprehensive online gradebook automatically tracks students' results on tests, quizzes, homework, and in the study plan. Instructors can use the gradebook to 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.

MyStatLab provides **engaging experiences** that personalize, stimulate, and measure learning for each student. 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 MyStatLab align with the exercises in the textbook, and they regenerate algorithmically to give students unlimited opportunity for practice and mastery. Exercises offer immediate helpful feedback, guided solutions, sample problems, animations, videos, and eText clips for extra help at point-of-use.
- **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 instructor's user guide and assessment questions.
- **Getting Ready for Statistics:** A library of questions now appears within each MyStatLab course to offer the developmental math topics students need for the course. These can be assigned as a prerequisite to other assignments, if desired.
- **Conceptual Question Library:** In addition to algorithmically regenerated questions that are aligned with your textbook, there is a library of 1000 Conceptual Questions available in the assessment manager that require students to apply their statistical understanding.
- **StatCrunch™:** MyStatLab 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, MyStatLab includes access to www.StatCrunch.com, a Web site where users can access more than 15,000 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, 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 Tutorial Videos, Technology Study Cards, and Technology Manuals for select titles—to learn how to effectively use statistical software.
- **Expert Tutoring:** Although many students describe the whole of MyStatLab as “like having your own personal tutor,” students also have access to live tutoring from Pearson. Qualified statistics instructors provide tutoring sessions for students via MyStatLab.

And, MyStatLab comes from a **trusted partner** with educational expertise and an eye on the future. Knowing that you are using a Pearson product means knowing that you are using quality content. That means that our eTexts are accurate and our assessment tools work. Whether you are just getting started with MyStatLab, or have a question along the way, we're here to help you learn about our technologies and how to incorporate them into your course.

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

MyStatLab™ Ready to Go Course (access code required)

These new Ready to Go courses provide students with all the same great MyStatLab features but make it easier for instructors to get started. Each course includes preassigned homework and quizzes to make creating a course even simpler. Ask your Pearson representative about the details for this particular course or to see a copy of this course.

MathXL® for Statistics Online Course (access code required)

MathXL® is the homework and assessment engine that runs MyStatLab. (MyStatLab 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.

Knowing that students often use external statistical software, we make it easy to copy our data sets, both from the ebook and the MyStatLab 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, 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 more than 15,000 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 allows users to analyze and gain insights from any data set. Interactive graphics help users understand statistical concepts and are available for export to enrich reports with visual representations of data.
- **Communicate.** Reporting options help users create a wide variety of visually appealing representations of their data.

Full access to StatCrunch is available with a MyStatLab kit, and StatCrunch is available by itself to qualified adopters. StatCrunch Mobile is now available to access from your mobile device. 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: 978-0-321-11313-9; ISBN-10: 0-321-11313-6) (CD only).

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).

This page intentionally left blank



Elementary Statistics Using Excel



1

Introduction to Statistics



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

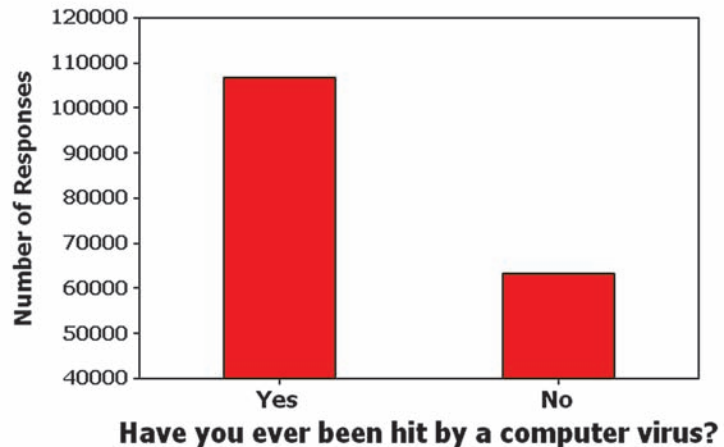


Figure 1-1 Survey Results

- 1-1 Review and Preview
- 1-2 Statistical and Critical Thinking
- 1-3 Types of Data
- 1-4 Collecting Sample Data
- 1-5 Introduction to Excel

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 14 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.

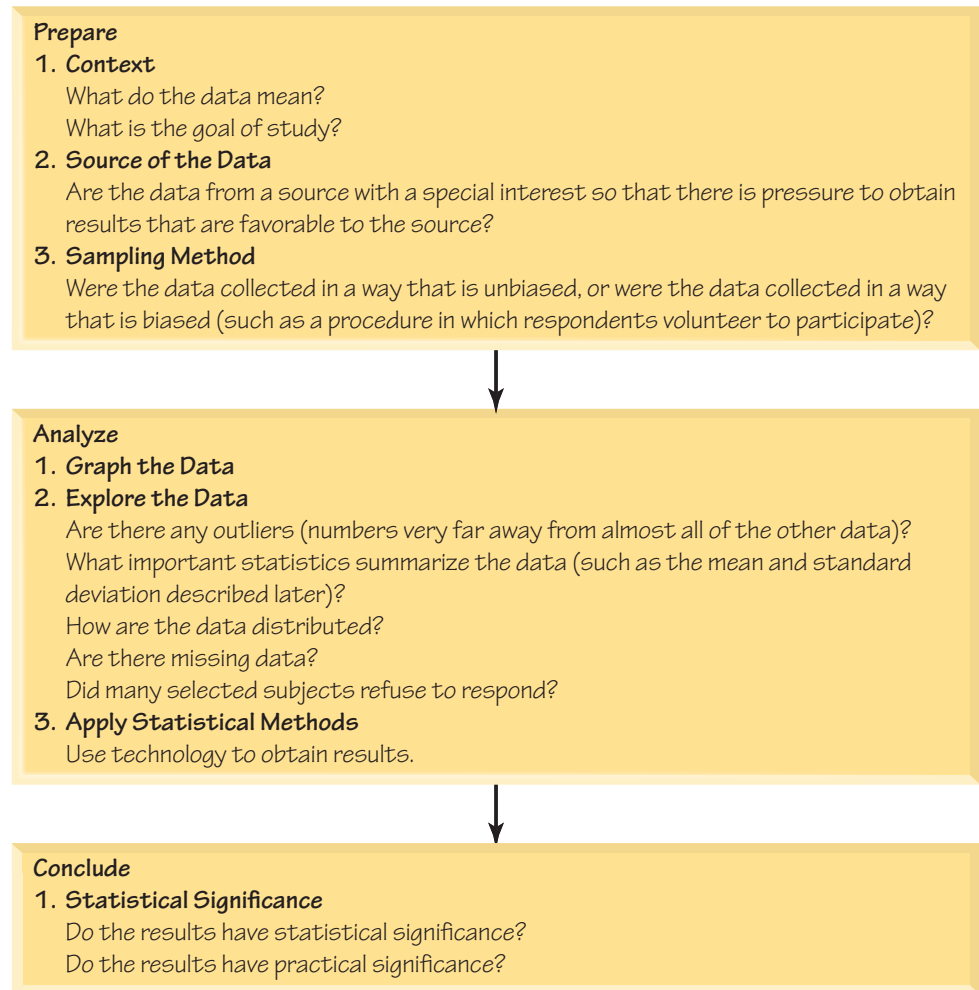


Figure 1-2 Statistical Thinking

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 cm³. 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 (cm³)

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

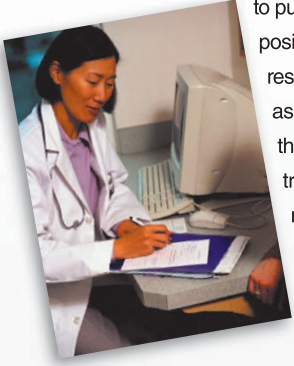
The *value of a statistical life* (VSL) is a measure routinely calculated and used for making decisions in fields such as medicine, insurance, environmental health, and transportation safety. As of this writing, the value of a statistical life is \$6.9 million.

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.)

