7
EDITION

Statistics for People Who (Think They) Hate Statistics

Neil J. Salkind
Bruce B. Frey

Praise for *Statistics for People Who (Think They) Hate Statistics*

"The Seventh Edition of Neil J. Salkind's text, *Statistics for People Who (Think They) Hate Statistics*, builds on an already solid pedagogical foundation for teaching statistics. Grounded in the history and development of statistics as a tool to study populations, it facilitates leading students from the idea that statistics is a tool only for those rare mathematicians and programmers, into an understanding that statistics can help us see a clearer picture of the world around us."

—Russell Brandon, Mississippi State University

"Two of my favorite statistical authors being together in one book may be a dream come true. Both Salkind's and Frey's texts have been a survival manual both for me and for my students. This is a masterpiece of statistical reference data that meshes the best parts of both authors and fills in the gap following the passing of Salkind. I can't wait to get this new text into the hands of my students, and I know that this new book is going to become a foundational pillar in all of my classes."

—Jesse Buchholz, Northwest Nazarene University

"In its seventh edition, Salkind and Frey's book provides a scaffolding journey for *anyone* that has a desire to learn the principles of statistics. This book begins with a foundation of defining the rudimentary principles of measurement, explains the principles of descriptive statistics, and disentangles the challenging principles of hypothesis testing and inferential statistics. It uses a very easy reading format, and it contains clear instructions to running all statistical procedures in SPSS. The book has an amazing retrieval system of resources (data sets, problem exercises, and more) that revitalizes teaching and learning. Therefore, I recommend this book without reservation."

—Tyrone Bynoe, University of Michigan–Flint

"When I mention statistics in the research methods course, some students get very anxious and remind me that they are not 'math people.' *Statistics for People Who (Think They) Hate Statistics* is a helpful supplemental text for a research methods course. It provides a different perspective regarding how statistics are used and helps students retrieve and build on their statistics knowledge. The text uses humor and interesting examples and helps illustrate why certain research issues are important."

—Adele Crudden, Mississippi State University

"Salkind and Frey have written an informative and comprehensive text for the introductory statistics course that is also funny and disarming. My graduate students—many of whom exhibit an initial wariness toward math courses and long-dormant math skills—have found it to be an unexpected pleasure and an accessible read."

—Diana Dansereau, Boston University

"Much like the concern for adding mayo to the brownie recipe in this book's appendix, Salkind and Frey's *Statistics for People Who (Think They) Hate Statistics*, Seventh Edition, takes the worrisome topic of statistics and turns it into an enjoyable enterprise. Through a dialogic and humorous approach, Salkind and Frey provide step-by-step instructions for calculating various statistics by hand as well as within SPSS. Without sacrificing rigor, Salkind and Frey help to dissolve the fear surrounding statistics through demonstrating statistic's every day uses and by employing both historical and real-world examples."

—Amanda Graham, University of Cincinnati

"Of all the statistics textbooks that I have reviewed, *Statistics for People Who (Think They) Hate Statistics* is by far the best. The step-by-step approach is easy to follow and it presents the information in a humorous but also serious way that encourages students with math phobias to take a second look. I have been adopting this text for my introductory statistics class since the third edition and never looked back. This text takes the cake (or should I say brownies!)."

—Brittany Landrum, University of Dallas

"This version continues to be the best textbook for students not comfortable with statistics. While not sacrificing depth, the text makes difficult topics approachable. This text meets my students where they are and allows them to gain the needed knowledge and appreciation of statistics."

—Christopher Ortega, SUNY Cortland

"This book has a successful conversion mission. Indeed, it succeeds in converting statistics-shy students into statistics-savvy ardent learners. The reappearance of this mind-catching treasure is a major plus in the effective teaching and easy learning of an introductory course in descriptive and inferential studies."

—Abdolhossein Abdollahy Zarandi, University of Texas at El Paso

Statistics for People Who (Think They) Hate Statistics

Seventh Edition

In honor and memory of my friend and mentor, Neil J. Salkind

Sara Miller McCune founded SAGE Publishing in 1965 to support the dissemination of usable knowledge and educate a global community. SAGE publishes more than 1000 journals and over 600 new books each year, spanning a wide range of subject areas. Our growing selection of library products includes archives, data, case studies and video. SAGE remains majority owned by our founder and after her lifetime will become owned by a charitable trust that secures the company's continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne

Statistics for People Who (Think They) Hate Statistics

Seventh Edition

Neil J. Salkind
Bruce B. Frey
University of Kansas

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
18 Cross Street #10-10/11/12
China Square Central
Singapore 048423

Acquisitions Editor: Helen Salmon
Content Development Editor: Chelsea Neve
Editorial Assistant: Megan O'Heffernan
Production Editor: Jane Martinez
Copy Editor: Gillian Dickens
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Wendy Jo Dymond
Indexer: Jeanne R. Busemeyer
Cover Designer: Candice Harman
Marketing Manager: Shari Countryman

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Names: Salkind, Neil J., author. | Frey, Bruce B., author.

Title: Statistics for people who (think they) hate statistics / Neil J. Salkind, Bruce B. Frey.

Description: Seventh edition. | Los Angeles : Sage, [2020] | Includes index.

Identifiers: LCCN 2019020133 | ISBN 9781544381855 (paperback)

Subjects: LCSH: Statistics.

Classification: LCC HA29 .S2365 2020 | DDC 519.5—dc23
LC record available at https://lccn.loc.gov/2019020133

This book is printed on acid-free paper.

19 20 21 22 23 10 9 8 7 6 5 4 3 2 1

BRIEF CONTENTS

DETAILED CONTENTS

A NOTE TO THE STUDENT

Why We Wrote This Book

With another new edition (now the seventh), we welcome you to what we hope will be, in all ways, a good learning experience. I, Bruce, who joins this project as Neil's coauthor, am touched and honored that my old friend, colleague, and mentor chose me to carry on his work on this book and his other popular SAGE publications. Neil passed away in 2017. I became a writer because he was. I care about my teaching because he did. And I'm dedicated to simplifying and explaining statistics and research methods because he was. Thank you, Neil. For everything. Here we go!

What many students of introductory statistics (be they new to the subject or just reviewing the material) have in common (at least at the beginning of their studies) is a relatively high level of anxiety, the origin of which is, more often than not, what they've heard from their fellow students. Often, a small part of what they have heard is true—learning statistics takes an investment of time and effort (and there's the occasional monster for a teacher).

But most of what they've heard (and where most of the anxiety comes from)—that statistics is unbearably difficult and confusing—is just not true. Thousands of fear-struck students have succeeded where they thought they would fail. They did it by taking one thing at a time, pacing themselves, seeing illustrations of basic principles as they are applied to real-life settings, and even having some fun along the way. That's what Neil tried to do in writing the first six editions of *Statistics for People Who (Think They) Hate Statistics,* and I, Bruce, have done my best to continue that tradition with this revision.

After a great deal of trial and error, some successful and many unsuccessful attempts, and a ton of feedback from students and teachers at all levels of education, we attempt with this book to teach statistics in a way that we (and many of our students) think is unintimidating and informative. We have tried our absolute best to incorporate all that experience into this book.

What you will learn from this book is the information you need to understand what the field and study of basic statistics is all about. You'll learn about the fundamental ideas and the most commonly used techniques to organize and make sense of data. There's very little theory (but some), and there are few mathematical proofs or discussions of the rationale for certain mathematical routines.

Why isn't this theory stuff and more in *Statistics for People Who (Think They) Hate Statistics*? Simple. Right now, you don't need it. It's not that we don't think it's important. Rather, at this point and time in your studies, we want to offer you material at a level we think you can understand and learn with some reasonable amount of effort while at the same time not be scared off from taking additional courses in the future. We (including your professor) want you to succeed.

So, if you are looking for a detailed unraveling of the derivation of the analysis of variance *F* ratio, go find another good book from SAGE (Bruce will be glad to refer you to one). But if you want to learn why and how statistics can work for you, you're in the right place. This book will help you understand the material you read in journal articles, explain what the results of many statistical analyses mean, and teach you how to perform basic statistical tasks.

And, if you want to talk about any aspect of teaching or learning statistics, feel free to contact me. You can do this through Bruce's email address at school (bfrey@ku.edu). Good luck, and let him know how he can improve this book to even better meet the needs of the beginning statistics student. And, if you want the data files that will help you succeed, go to the SAGE website at **edge.sagepub.com/salkindfrey7e**.

AND A (LITTLE) NOTE TO THE INSTRUCTOR

We would like to share two things.

First, we applaud your efforts at teaching basic statistics. Although this topic may be easier for some students, most find the material very challenging. Your patience and hard work are appreciated by all, and if there is anything we can do to help, please send Bruce a note.

Second, *Statistics for People Who (Think They) Hate Statistics* is not meant to be a dumbed-down book similar to others you may have seen. Nor is the title meant to convey anything other than the fact that many students new to the subject are very anxious about what's to come. This is not an academic or textbook version of a "book for dummies" or anything of the kind. We have made every effort to address students with the respect they deserve, not to patronize them, and to ensure that the material is approachable. How well we did in these regards is up to you, but we want to convey our very clear intent that this book contains the information needed in an introductory course, and even though our approach involves some humor, nothing about our intent is anything other than serious. Thank you.

ACKNOWLEDGMENTS

Everybody at SAGE deserves a great deal of thanks for providing us with the support, guidance, and professionalism that takes a mere idea (way back before the first edition) and makes it into a book like the one you are now reading—and then makes it successful. From Johnny Garcia, who heads up the distribution center, to Vonessa Vondera, who handles financials—this book would not be possible without their hard work.

However, some people have to be thanked individually for their special care and hard work. Helen Salmon, Senior Acquisitions Editor—Research Methods and Statistics, has shepherded this edition, being always available to discuss new ideas and seeing to it that everything got done on time and done well. She is the editor whom every author wants. C. Deborah Laughton, Lisa Cuevas Shaw, and Vicki Knight—all previous editors—helped this book along the way, and to them, we are forever grateful. Others who deserve a special note are Megan O'Heffernan, editorial assistant, and Jane Martinez, production editor supreme. Special, special thanks go to Paula Fleming for her sharp eye and sound copyediting, which make this material read as well as it does.

Bruce worked with some great folks in making the many videos that are included with this edition—Chelsea Neve, content development editor; Mitch Yapko, producer; and Juan Reyna, videographer, all made for a couple days of high adventure, hard work, and great fun. I particularly am grateful for their interest (mock, feigned, or genuine) in the obscure genre of popular music that means so much to me—bubblegum pop from 1966 to 1970. Additionally, two of my doctoral students have been particularly helpful in revising this edition, the very smart Bo Hu and the very smart Alicia Stoltenberg. Thanks so much!

AND NOW, ABOUT THE SEVENTH EDITION . . .

What you read earlier about this book reflects our thoughts about why Neil wrote this book way back when in the first place. But it tells you little about this revised seventh edition.

Any book is always a work in progress, and this latest edition of *Statistics for People Who (Think They) Hate Statistics* is no exception. Over the past 20 years or so, many people told Neil how helpful this book is, and others have told him how they would like it to change and why. In revising this book, we are trying to meet the needs of all audiences. Some things remain the same, and some have indeed changed.

There are always new things worth consideration and different ways to present old themes and ideas. Every chapter was revised in one way or another, as we continue to find and fix errors and the occasional typo, as well as respond to the many thoughtful comments from reviewers, students, and instructors. Here's a list more specific improvements you'll find that's new in the seventh edition of *Statistics for People Who (Think They) Hate Statistics*.

- A new pedagogical feature spotlights people from history who liked statistics—and not just the same old boring list of dead white men.
- There's now even more in-depth discussions of effect sizes as stand-alone features in most chapters.
- A greater focus is on the importance of level of measurement in choosing a statistical procedure.
- Additional discussion is provided about contemporary approaches to research and defining variables (e.g., treating gender as a continuous variable, not as binary or categorical).
- Many of the real-world examples have been updated to keep up with the present.
- Some chapters that cover broad areas are now appendices that can be referenced as needed.

This seventh edition features IBM® SPSS® Statistics 25, the latest version that SPSS offers. For the most part, you can use a version of SPSS that is as early as Version 11 to do most of the work, and these earlier versions can read the data files created with the later versions. If the reader needs help with SPSS, we offer a mini course in Appendix A, which is also available online.

- Maybe the most interesting (and perhaps coolest) thing about this edition is that it continues to be available as an **interactive eBook**. The **Interactive eBook** can be packaged with the text for just $5 or purchased separately. The Interactive eBook offers hyperlinks to original videos, including **Core Concepts in Stats animation videos, Lightboard Lecture videos featuring short lessons by author Bruce B. Frey, and SPSS tutorials for select end-of-chapter "Time to Practice" problems**. Users will also have immediate access to study tools such as highlighting, bookmarking, note-taking/sharing, and more. We all hope you find it as exciting as we do and hope it helps you in understanding the material.

Any typos and such that appear in this edition of the book are entirely Bruce's fault, and we apologize to the professors and students who are inconvenienced by their appearance. And Bruce would so appreciate any letters, calls, and emails pointing out these errors. We have all made every effort in this edition to correct previous errors and hope we did a reasonably good job. Let us hear from you with suggestions, criticisms, nice notes, and so on. Good luck.

Neil J. Salkind
Bruce B. Frey
bfrey@ku.edu
University of Kansas

DIGITAL RESOURCES

SAGE coursepacks

Instructor Resources:

SAGE coursepacks and SAGE edge online resources are included FREE with this text. For a brief demo, contact your sales representative today.

edge.sagepub.com/salkindfrey7e

SAGE coursepacks for instructors makes it easy to import our quality content into your school's learning management system (LMS)*. Intuitive and simple to use, it allows you to

SPSS is a registered trademark of International Business Machines Corporation.

***For use in:** Blackboard, Canvas, Brightspace by Desire2Learn (D2L), and Moodle

Say NO to . . .

- required access codes
- learning a new system

Say YES to . . .

- using only the content you want and need
- high-quality assessment and multimedia exercises

Don't use an LMS platform? No problem, you can still access many of the online resources for your text via SAGE edge: **edge.sagepub.com/salkindfrey7e**

With SAGE coursepacks, you get:

- quality textbook content delivered **directly into your LMS**;
- an **intuitive, simple format** that makes it easy to integrate the material into your course with minimal effort;
- **assessment tools** that foster review, practice, and critical thinking, including:
  - diagnostic chapter **pre-tests and post-tests** that identify opportunities for improvement, track student progress, and ensure mastery of key learning objectives
  - **test banks** built on Bloom's Taxonomy that provide a diverse range of test items with ExamView test generation
  - **activity and quiz options** that allow you to choose only the assignments and tests you want
  - **instructions** on how to use and integrate the comprehensive assessments and resources provided
- **assignable SAGE Premium Video** (available via the interactive eBook version, linked through SAGE coursepacks) that is tied to learning objectives, and produced exclusively for this text to bring concepts to life, featuring:
  - Videos that include **animations of key concepts and tutorial videos for end-of-chapter SPSS problems**
  - **Lightboard Lecture Videos** from Bruce B. Frey (one short lecture video per chapter) that demonstrate each statistical procedure, captured using innovative LED glass technology
  - **Corresponding multimedia assessment options** that automatically feed to your gradebook

  - Comprehensive, downloadable, easy-to-use ***Media Guide in the Coursepack for every video resource***, listing the chapter to which the video content is tied, matching learning objective(s), a helpful description of the video content, and assessment questions

- **chapter-specific discussion questions** to help launch engaging classroom interaction while reinforcing important content;
- **class activities** designed for individual or group settings that encourage active learning and allow students to delve more into the subject matter;
- editable, chapter-specific **PowerPoint® slides** that offer flexibility when creating multimedia lectures so you don't have to start from scratch;
- **sample course syllabi** with suggested models for structuring your course that give you options to customize your course to your exact needs;
- **lecture notes** that summarize key concepts on a chapter-by-chapter basis to help you with preparation for lectures and class discussions;
- **integrated links to the interactive eBook** that make it easy for students to maximize their study time with this "anywhere, anytime" mobile-friendly version of the text. It also offers access to more digital tools and resources, including SAGE Premium Video; and
- **all tables and figures** from the textbook.

SAGE edge™

Student Resources:

edge.sagepub.com/salkindfrey7e

SAGE edge for students enhances learning, it's easy to use, and offers:

- an **open-access site** that makes it easy for students to maximize their study time, anywhere, anytime;
- **datasets** that are available for download and use with exercises in the book;
- **eFlashcards** that strengthen understanding of key terms and concepts;
- **eQuizzes** that allow students to practice and assess how much they've learned and where they need to focus their attention;
- **links to videos and podcasts** for further exploration of statistical topics.

ABOUT THE AUTHORS

Neil J. Salkind received his PhD in human development from the University of Maryland, and after teaching for 35 years at the University of Kansas, he was Professor Emeritus in the Department of Psychology and Research in Education, where he collaborated with colleagues and worked with students. His early interests were in the area of children's cognitive development, and after research in the areas of cognitive style and (what was then known as) hyperactivity, he was a postdoctoral fellow at the University of North Carolina's Bush Center for Child and Family Policy. His work then changed direction to focus on child and family policy, specifically the impact of alternative forms of public support on various child and family outcomes. He delivered more than 150 professional papers and presentations, wrote more than 100 trade and textbooks, and is the author of *Statistics for People Who (Think They) Hate Statistics, Theories of Human Development,* and *Exploring Research*. He has edited several encyclopedias, including the *Encyclopedia of Human Development,* the *Encyclopedia of Measurement and Statistics,* and the *Encyclopedia of Research Design.* He was editor of *Child Development Abstracts and Bibliography* for 13 years. He lived in Lawrence, Kansas, where he liked to read, swim with the River City Sharks, work as the proprietor and sole employee of big boy press, bake brownies (see Appendix I for the recipe), and poke around old Volvos and old houses. He died in 2017 at the age of 70.

Bruce B. Frey, PhD, is an award-winning researcher, teacher, and professor of educational psychology at the University of Kansas. He is the editor of *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* and author of *There's a Stat for That!, Modern Classroom Assessment*, and *100 Questions (and Answers) About Tests and Measurement*. He is the associate editor of SAGE's *Encyclopedia of Research Design*. He also wrote *Statistics Hacks* for O'Reilly Media. His primary research interests include classroom assessment, instrument development, and program evaluation. In his spare time, Bruce leads a secret life as Professor Bubblegum, host of *Echo Valley*, a podcast that celebrates bubblegum pop music of the late 1960s. The show is wildly popular with the young people.

PART I

YIPPEE! I'M IN STATISTICS

"Well, if it isn't a significant subset of our study group!"

Not much to shout about, you might say? Let me take a minute and show you how some real-life scientists use this widely used set of tools we call *statistics*:

- Michelle Lampl is a physician, the Samuel Candler Dobbs Professor of Anthropology and the co-director of the Emory-Georgia Tech Predictive Health Initiative. She was having coffee with a friend, who commented on how quickly her young infant was growing. In fact, the new mother spoke as if her son was "growing like a weed." Being a curious scientist (as all scientists should be), Dr. Lampl thought she might actually examine how rapid this child's growth, and that of others, was during infancy. She proceeded to measure a group of children's growth on a daily basis and found, much to her surprise, that some infants grew as much as 1 inch overnight! Some growth spurt.

  Want to know more? Why not read the original work? You can find more about this in Lampl, M., Veldhuis, J. D., & Johnson, M. L. (1992). Saltation and stasis: A model of human growth. *Science, 258,* 801–803.

- Why do some people eat meat and others don't? The traditional explanation has been that meat eaters like the taste of meat. Christopher A. Monteiro at Cornell University wondered if that was so or whether there was more to it than that. Might philosophical beliefs about animals play a role, as well? Dr. Monteiro created a measure to assess the level of two different beliefs: whether it is okay to eat animals and whether it is okay to kill animals if you are going to eat them. He collected data and used statistics to find relationships among the questions on his measure to identify that there seem to be these different attitudes that work with things like "meat tastes good" to explain the practice of meat eating. Interestingly, those who scored the highest on the "okay to kill animals" scale also tended to score the highest on measures of racism and sexism.

  Want to know more? You can read the whole study! Look this up:

  Monteiro, C. A., Pfeiler, T. M., Patterson, M. D., & Milburn, M. A. (2017). The Carnism Inventory: Measuring the ideology of eating animals. *Appetite, 113,* 51–62.

- Sue Kemper is the Roberts Distinguished Professor of Psychology at the University of Kansas and has worked on the most interesting of projects. She and several other researchers studied a group of nuns and examined how their early experiences, activities, personality characteristics, and other information related to their health during their late adult years. Most notably, this diverse group of scientists (including psychologists, linguists, neurologists, and others) wanted to know how well all this information predicts the occurrence of Alzheimer's disease. Kemper and her colleagues found that the complexity of the nuns' writing during their early 20s was related to the nuns' risk for Alzheimer's 50, 60, and 70 years later.

  Want to know more? Why not read the original work? You can find more about this in Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein,

D. R., & Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings from the nun study. *Journal of the American Medical Association, 275,* 528–532.

- Do you suffer from migraine headaches? Millions do. Doctors have long been concerned about the risks associated with migraines because there are often serious diseases associated with them. Dr. Tobias Kurth, a German researcher working with Harvard, analyzed data from thousands of women who have migraine headaches and had information about their health across decades. After equalizing the women on a bunch of variables, he discovered that women with migraines were 50% more likely to have heart disease than those who did not have migraine headaches. And they are 62% more likely to have a stroke! The suggestion by Dr. Kurth and his coauthors is that women who suffer from migraines get more frequent checkups that include an evaluation for these other risks.

  Want to know more? Find out for yourself how the study was conducted and how statistics were used to get these estimates. Here are the details: Kurth, T., Winter, A. C., Eliassen, A. H., Dushkes, R., Mukamal, K. J., Rimm, E. B., . . . Rexrode, K. M. (2016). Migraine and risk of cardiovascular disease in women: Prospective cohort study. *BMJ, 353,* i2610.

All of these researchers had a specific question they found interesting and used their intuition, curiosity, and excellent training to answer it. As part of their investigations, they used this set of tools we call *statistics* to make sense out of all the information they collected. Without these tools, all this information would have been just a collection of unrelated outcomes. The outcomes would be nothing that Lampl could have used to reach a conclusion about children's growth or Kemper could have used to better understand aging and cognition (and perhaps Alzheimer's disease).

Statistics—the science of organizing and analyzing information to make it more easily understood—made these tasks doable. The reason that any of the results from such studies are useful is that we can use statistics to make sense out of them. And that's exactly the goal of this book—to provide you with an understanding of these basic tools and how researchers use them and, of course, how to use them yourself.

In this first part of *Statistics for People Who (Think They) Hate Statistics,* you will be introduced to what the study of statistics is about and why it's well worth your efforts to master the basics—the important terminology and ideas that are central to the field. This part gives you a solid preparation for the rest of the book.

STATISTICS OR SADISTICS? IT'S UP TO YOU

Difficulty Scale ☺ ☺ ☺ ☺ ☺
(really easy)

WHAT YOU WILL LEARN IN THIS CHAPTER

- What statistics is all about
- Why you should take statistics
- How to succeed in this course

WHY STATISTICS?

You've heard it all before, right?

"Statistics is difficult."

"I'm not a math person."

"I don't know how to use statistics software."

"What do I need this stuff for?"

"What do I do next?"

And the famous cry of the introductory statistics student: "I don't get it!"

Well, relax. Students who study introductory statistics find themselves, at one time or another, thinking at least one of the above and quite possibly sharing the thought with another student, their spouse, a colleague, or a friend.

And all kidding aside, some statistics courses can easily be described as *sadistics*. That's because the books are repetitiously boring, the examples don't seem to apply to real life, and too much math is thrown at you too quickly.

That's not the case for you. The fact that you or your instructor has selected *Statistics for People Who (Think They) Hate Statistics* shows that you're ready to take the right approach—one that is unintimidating, informative, and applied (and even a little fun) and that tries to teach you what you need to know about using statistics as the valuable tool that it is.

If you're using this book in a class, it also means that your instructor is clearly on your side. He or she knows that statistics can be intimidating but has taken steps to see that it is not intimidating for you. As a matter of fact, we'll bet there's a good chance (as hard as it may be to believe) that you'll be enjoying this class in just a few short weeks.

And Why SPSS?

Throughout this book, you'll be shown how to use SPSS, a statistical analysis tool, for the analysis of data. No worries; you'll also be shown how to do the same analysis by hand to assure you of an understanding of both.

Why SPSS? Simple. It's one of the most popular, most powerful analytic tools available today, and it can be an exceedingly important and valuable tool for learning how to use basic and some advanced statistics. In fact, many stats courses taught at the introductory level use SPSS as their primary computational tool, and you can look to Appendix A for a refresher on some basic SPSS tasks. Also, the way that technology is advancing, few opportunities to use statistics in research, administration, and everyday work will not require some knowledge of how and when to use tools such as SPSS. That's why we're including it in this book! We will show you how to use it to make your statistics learning experience a better one.

A 5-MINUTE HISTORY OF STATISTICS

Before you read any further, it would be useful to have some historical perspective about this topic called statistics. After all, almost every undergraduate in the social, behavioral, and biological sciences and every graduate student in education, nursing, psychology, social welfare and social services, anthropology, and . . . (you get the picture) are required to take this course. Wouldn't it be nice to have some idea from whence the topic it covers came? Of course it would.

Way, way back, as soon as humans realized that counting was a good idea (as in "How many of these do I need to trade for one of those?" and "Uh oh, there are more of them than there are of us!" and "Yes, little one, you may keep one pet saber-toothed tiger, but not more than one"), collecting information became a useful skill.

If counting counted, then one could define the seasons by how often the sun rose and set, how much food was needed to last the winter, and what amount of resources belonged to whom.

That was just the beginning. Once numbers became part of language, it seemed like the next step was to attach these numbers to outcomes. That started in earnest during the 17th century, when the first set of data pertaining to populations was collected. This was when the use of *descriptive statistics* began, which we will talk about later. From that point on, scientists (mostly mathematicians at first but, later, physical and biological scientists) needed to develop specific tools to answer specific questions. For example, Francis Galton (a half-cousin of Charles Darwin, by the way), who lived from 1822 to 1911, was very interested in the nature of human intelligence. He also speculated that hair loss was due to the intense energy that went into thinking. No, really! (It's probably not.) To explore one of his primary questions regarding the similarity of intelligence among family members, he used a specific statistical tool called the correlation coefficient (first developed by mathematicians), and then he popularized its use in the behavioral and social sciences. You'll learn all about this tool in Chapter 5. In fact, most of the basic statistical procedures that you will learn about were first developed and used in the fields of agriculture, astronomy, and even politics. Their application to human behavior came much later.

PEOPLE WHO LOVED STATISTICS

Inferential statistics, the use of sample observations or data that we can see to make guesses about the likely characteristics of populations that we cannot see, probably started with Blaise Pascal (1623–1662), a French mathematician and religious philosopher. He developed the mathematical formulas that can predict important things like the probability of dice rolls and the likelihood of flipping a coin three times and having it come up heads each time. He even proved that if the coin almost always comes up heads, someone is cheating. The application of these statisti-

cal inventions was of immediate practical use to gamblers, and this might be the first time in history that statistics was seen as having practical applications. (Because one could make money by understanding them.) You may notice that Pascal did not live very long. He suffered from various illnesses during the later years of his life, and his cause of death isn't even known for sure, although after his death, he was found to have had stomach cancer and some brain damage. As a deeply religious man, Pascal believed that suffering was necessary for a good life. Thus, he probably would have enjoyed being a stats professor.

The past 100 years have seen great strides in the invention of new ways to use old ideas. The simplest test for examining the differences between the averages of two groups was first advanced during the early 20th century. Techniques that build on this idea were offered decades later and have been greatly refined. And the introduction of personal computers and such programs as Excel has opened up the use of sophisticated techniques to anyone who wants to explore these fascinating topics.

The introduction of these powerful personal computers has been both good and bad. It's good because most statistical analyses no longer require access to a huge and expensive mainframe computer. Instead, a simple personal computer costing less than $250 or a cloud account can do 95% of what 95% of the people need. On the other hand, less than adequately educated students (such as your fellow students who chose not to take this course!) will take any old data they have and think that by running them through some sophisticated analysis, they will have reliable, trustworthy, and meaningful outcomes—not true. What your professor would say is "Garbage in, garbage out"; if you don't start with data you can trust, what you'll have after your data are analyzed are results you cannot trust.

Today, statisticians in all different areas, from criminal justice to geophysics to psychology to determining whether the "hot" hand really exists in the NBA, find themselves using basically the same techniques to answer different questions. There are, of course, important differences in how data are collected, but for the most part, the analyses (the plural of *analysis*) that are done following the collection of data (the plural of *datum,* which means one piece of information) tend to be very similar, even if called something different. The moral here? This class will provide you with the tools to understand how statistics are used in almost any discipline. Pretty neat, and all for just the cost of a few credit hours.

Okay. Five minutes is up, and you know as much as you need to know for now about the history of statistics. You'll get some more history here and there as we learn about different procedures. Let's move on to what statistics is (and isn't).

STATISTICS: WHAT IT IS (AND ISN'T)

Statistics for People Who (Think They) Hate Statistics is a book about basic statistics and how to apply them to a variety of different situations, including the analysis and understanding of information, especially when that information is expressed as numbers and quantities.

In the most general sense, **statistics** describes a set of tools and techniques that are used for describing, organizing, and interpreting information or data. Those data might be the scores on a test taken by students participating in a special math curriculum, the speed with which problems are solved, the number of side effects when patients use one type of drug rather than another, the number of errors in each inning of a World Series game, or the average price of a dinner in an upscale restaurant in Santa Fe, New Mexico (not cheap).

In all these examples, and the million more we could think of, data are collected, organized, summarized, and then interpreted. In this book, you'll learn about collecting, organizing, and summarizing data as part of descriptive statistics. And then you'll learn about interpreting data when you learn about the usefulness of inferential statistics.

What Are Descriptive Statistics?

Descriptive statistics are used to organize and describe the characteristics of a collection of data. The collection is sometimes called a **data set** or just **data**. Scientists would say that descriptive statistics describe a *sample*—a collection of data that you have in front of you.

For example, the following list shows you the names of 22 college students, their major areas of study, and their ages. If you needed to describe what the most popular college major is, you could use a descriptive statistic that summarizes their most frequent choice (called the mode). In this case, the most common major is psychology. And if you wanted to know the average age, you could easily compute another descriptive statistic that identifies this variable (that one's called the mean). Both of these simple descriptive statistics are used to describe data. They do a fine job of allowing us to represent the characteristics of a large collection of data such

as the 22 cases in our example. Any time there are more than just a few people or things you want to describe, descriptive statistics make that easier. Much easier! That's why descriptive statistics are so popular in the social and natural sciences.

Name	Major	Age	Name	Major	Age
Deja	Education	19	Aliyah	English	21
Sara	Psychology	18	Mateo	Psychology	22
Asma	Education	19	Hadley	Psychology	23
Trevon	Psychology	21	Alejandro	Education	21
Jordan	Education	20	Chip	Education	19
Pam	Education	24	Homer	Psychology	18
Xavier	Psychology	21	Li	English	22
Liz	Psychology	19	Darius	Psychology	24
Nicole	Chemistry	19	Leonard	Psychology	21
Zhang	Nursing	20	Jeffrey	Chemistry	18
Kent	History	18	Emily	Spanish	19

So watch how simple this is. To find the most frequently selected major, just find the one that occurs most often. And to find the average age, just add up all the age values and divide by 22. You're right—the most often occurring major is psychology (9 times), and the average age is 20.3 (actually 20.27). Look, Ma! No hands—you're a statistician.

What Are Inferential Statistics?

Inferential statistics are often (but not always) the next step after you have collected and summarized data. Inferential statistics are used to make inferences based on a smaller group of data (such as our group of 22 students) about a possibly larger one (such as all the undergraduate students in the College of Arts and Sciences).

A smaller group of data is often called a **sample**, which is a portion, or a subset, of a **population**. For example, all the fifth graders in Newark (Neil's fair city of origin), New Jersey, would be a population (the population is all the occurrences with certain characteristics, in this case, being in fifth grade and attending school in Newark), whereas a selection of 150 of these students would be a sample. If we think this sample represents the population well, we can make guesses about the population.

Let's look at another example. Your marketing agency asks you (a newly hired researcher) to determine which of several names is most appealing for a new brand of potato chip.

Will it be Chipsters? FunChips? Crunchies? As a statistics pro (we know we're moving a bit ahead of ourselves, but keep the faith), you need to find a small group of potato chip eaters who are representative of all potato chip fans and ask these people to tell you which one of the three names they like the most. Then, if you do things right, you can easily extrapolate the findings to the huge group of potato chip eaters.

Or let's say you're interested in the best treatment for a particular type of disease. Perhaps you'll try a new drug as one alternative, a placebo (a substance that is known not to have any effect) as another alternative, and nothing as the third alternative to see what happens. Well, you find out that more patients get better when no action is taken and nature (and we assume that's the only factor or set of factors that differentiate the groups) just takes its course! The drug does not have any effect. Then, with that information, you can extrapolate to the larger group of patients who suffer from the disease, given the results of your experiment.

Inferring from a sample to a population makes a lot of sense, especially when you are sure the sample represents the population. That's why, as you'll see later, scientists spend a lot of effort getting a representative sample.

In Other Words . . .

Statistics is a tool that helps us understand the world around us. It does so by organizing information we've collected and then letting us make certain statements about how characteristics of those data are applicable to new settings. Descriptive and inferential statistics work hand in hand, and which statistic you use and when depends on the question you want answered and how you happened to measure your variables.

And today, a knowledge of statistics is more important than ever because it provides us with the tools to make decisions that are based on empirical (observed) evidence and not our own biases or beliefs. Want to know whether early intervention programs work? Then test whether they work and provide that evidence to the court that will make a ruling on the viability of a new school bond issue that could pay for those programs.

WHAT AM I DOING IN A STATISTICS CLASS?

You might find yourself using this book for many reasons. You might be enrolled in an introductory statistics class. Or you might be reviewing for your comprehensive exams. Or you might even be reading this on summer vacation (horrors!) in preparation for a more advanced class.

In any case, you are a statistics student, whether you have to take a final exam at the end of a formal course or you're just in it of your own accord. But there are

plenty of good reasons to be studying this material—some fun, some serious, and some both.

Here's the list of some of the things that our students hear at the beginning of our introductory statistics courses:

1. Statistics 101 or Statistics 1 or whatever it's called at your school looks great listed on your transcript. Kidding aside, this may be a required course for you to complete your major. But even if it is not, having these skills is definitely a big plus when it comes time to apply for a job or for further schooling. And with more advanced courses, your résumé will be even more impressive.
2. If this is not a required course, taking basic statistics sets you apart from those who do not. It shows that you are willing to undertake a course that is above average with regard to difficulty and commitment. And, as the political and economic (and sports!) worlds become more "accountable," more emphasis is being placed on analytic skills. Who knows, this course may be your ticket to a job!
3. Basic statistics is an intellectual challenge of a kind that you might not be used to. There's a good deal of thinking that's required, a bit of math, and some integration of ideas and application. The bottom line is that all this activity adds up to what can be an invigorating intellectual experience because you learn about a whole new area or discipline.
4. There's no question that having some background in statistics makes you a better student in the social or behavioral sciences, because you will have a better understanding not only of what you read in journals but also of what your professors and colleagues may be discussing and doing in and out of class. You will be amazed the first time you say to yourself, "Wow, I actually understand what they're talking about." And it will happen over and over again, because you will have the basic tools necessary to understand exactly how scientists reach the conclusions they do.
5. If you plan to pursue a graduate degree in education, anthropology, economics, nursing, sociology, or any one of many other social, behavioral, and biological pursuits, this course will give you the foundation you need to move further.
6. There are many different ways of thinking about, and approaching, different types of problems. The set of tools you learn about in this book (and this course) will help you look at interesting problems from a new perspective. And, while the possibilities may not be apparent now, this new way of thinking can be brought to new situations.
7. Finally, you can brag that you completed a course that everyone thinks is the equivalent of building and running a nuclear reactor.

TEN WAYS TO USE THIS BOOK (AND LEARN STATISTICS AT THE SAME TIME!)

Yep. Just what the world needs—another statistics book. But this one is different. It is directed at the student, is not condescending, is informative, and is as basic as possible in its presentation. It makes no presumptions about what you should know before you start and proceeds in slow, small steps, which lets you pace yourself.

However, there has always been a general aura surrounding the study of statistics that it's a difficult subject to master. And we don't say otherwise, because parts of it are challenging. On the other hand, millions and millions of students have mastered this topic, and you can, too. Here are 10 hints to close this introductory chapter before we move on to our first topic:

1. **You're not dumb.** That's true. If you were, you would not have gotten this far in school. So, treat statistics as you would any other new course. Attend the lectures, study the material, do the exercises in the book and from class, and you'll do fine. Rocket scientists know statistics, but you don't have to be a rocket scientist to succeed in statistics.
2. **How do you know statistics is hard?** Is statistics difficult? Yes and no. If you listen to friends who have taken the course and didn't do well, they'll surely volunteer to tell you how hard it was and how much of a disaster it made of their entire semester, if not their lives. And let's not forget—we always tend to hear from complainers. So, we'd suggest that you start this course with the attitude that you'll wait and see how it is and judge the experience for yourself. Better yet, talk to several people who have had the class and get a good idea of what they think. Don't base your expectations on just one spoilsport's experience. Get a bigger sample!
3. **Don't skip lessons—work through the chapters in sequence.** *Statistics for People Who (Think They) Hate Statistics* is written so that each chapter provides a foundation for the next one in the book. When you are all done with the course, you will (I hope) continue to use this book as a reference. So if you need a particular value from a table, you might consult Appendix B. Or if you need to remember how to compute the standard deviation, you might turn to Chapter 3. But for now, read each chapter in the sequence that it appears. It's okay to skip around and see what's offered down the road. Just don't study later chapters before you master earlier ones.
4. **Form a study group.** This is a big hint and one of the most basic ways to ensure some success in this course. Early in the semester, arrange to study with friends or classmates. If you don't have any friends who are in the same class as you, then make some new ones or offer to study with someone who looks as happy to be there as you are. Studying with others allows you to help them if you know the material better or to benefit from

those who know some material better than you. Set a specific time each week to get together for an hour and go over the exercises at the end of the chapter or ask questions of one another. Take as much time as you need. Studying with others is an invaluable way to help you understand and master the material in this course.

5. **Ask your teacher questions, and then ask a friend.** If you do not understand what you are being taught in class, ask your professor to clarify it. Have no doubt—if you don't understand the material, then you can be sure that others do not as well. More often than not, instructors welcome questions. And especially because you've read the material before class, your questions should be well informed and help everyone in class to better understand the material.

6. **Do the exercises at the end of a chapter.** The exercises are based on the material and the examples in the chapter they follow. They are there to help you apply the concepts that were taught in the chapter and build your confidence at the same time. If you can answer these end-of-chapter exercises, then you are well on your way to mastering the content of the chapter. Correct answers to each exercise are provided in Appendix D.

7. **Practice, practice, practice.** Yes, it's a very old joke:

   Q. How do you get to Carnegie Hall?

   A. Practice, practice, practice.

   Well, it's no different with basic statistics. You have to use what you learn and use it frequently to master the different ideas and techniques. This means doing the exercises at the end of Chapters 1 through 17 and Chapter 19, as well as taking advantage of any other opportunities you have to understand what you have learned.

8. **Look for applications to make it more real.** In your other classes, you probably have occasion to read journal articles, talk about the results of research, and generally discuss the importance of the scientific method in your own area of study. These are all opportunities to see how your study of statistics can help you better understand the topics under class discussion as well as the area of beginning statistics. The more you apply these new ideas, the fuller your understanding will be.

9. **Browse.** Read over the assigned chapter first; then go back and read it with more intention. Take a nice leisurely tour of *Statistics for People Who (Think They) Hate Statistics* to see what's contained in the various chapters. Don't rush yourself. It's always good to know what topics lie ahead as well as to familiarize yourself with the content that will be covered in your current statistics class.

10. **Have fun.** This might seem like a strange thing to say, but it all boils down to you mastering this topic rather than letting the course and its demands master you. Set up a study schedule and follow it, ask questions in class, and consider

this intellectual exercise to be one of growth. Mastering new material is always exciting and satisfying—it's part of the human spirit. You can experience the same satisfaction here—just keep your eye on the ball and make the necessary commitment to stay current with the assignments and work hard.

ABOUT THE BOOK'S FEATURES

Throughout the book, there are short biographies of People Who Loved Statistics. All the statistical tricks and procedures we will discover in this book were invented by real people, and it's good to realize that they were just like you and me! (Well, a little like you and me.)

Throughout this book, you'll find a small-steps icon like the one you see here. This indicates that a set of steps is coming up that will direct you through a particular process. Sometimes you will use SPSS to do these steps. These steps have been tested and approved by whatever federal agency approves these things.

Real-World Stats will appear at the end of every chapter as appropriate and, it is hoped, will provide you with a demonstration of how a particular method, test, idea, or some aspect of statistics is used in the everyday workplace of scientists, physicians, policy makers, government folks, and others. In this first such exploration, we look at a very short paper where the author recalls and shares the argument that the National Academy of Sciences (first chartered in 1863, by the way!) "shall, whenever called upon by any department of the Government, investigate, examine, experiment, and report upon any subject of science or art." This charter, some 50 years later in 1916, led to the formation of the National Research Council, another federal body that helped provide information that policy makers need to make informed decisions. And often this "information" takes the form of quantitative data—also referred to as statistics—that assist people in evaluating alternative approaches to problems that have a wide-ranging impact on the public. So, this article, as does your book and the class you are taking, points out how important it is to think clearly and use data to support your arguments.

Want to know more? Go online or go to the library, and find . . .

Cicerone, R. (2010). The importance of federal statistics for advancing science and identifying policy options. *The Annals of the American Academy of Political and Social Science, 631,* 25–27.

Appendix A contains an introduction to SPSS. Working through this appendix is all you really need to do to be ready to use SPSS. If you have an earlier version of SPSS (or the Mac version), you will still find this material to be very helpful. In fact, the latest Windows and Mac versions of SPSS are almost identical in appearance and functionality.

Appendix B contains important tables you will learn about and need throughout the book.

And, in working through the exercises in this book, you will use the data sets in Appendix C. In the exercises, you'll find references to data sets with names like "Chapter 2 Data Set 1," and each of these sets is shown in Appendix C. You can either enter the data manually or download them from the publisher's site at **edge.sagepub.com/salkindfrey7e**.

Appendix D contains answers to end-of-chapter questions.

Appendix E contains a primer on math for those who could use a refresher.

Appendix F describes some statistical software other than SPSS that you might find useful.

Appendix G contains statistics websites that are fun to play around in.

Appendix H contains the most helpful hints for gathering your own data.

And Appendix I offers the long-sought-after brownie recipe (yes, you finally found it).

KEY TO DIFFICULTY ICONS

To help you along a bit, we placed a difficulty index at the beginning of each chapter. This adds some fun to the start of each chapter, but it's also a useful tip to let you know what's coming and how difficult chapters are in relation to one another. Because the index uses smiley faces, the more smiles, the merrier!

- ☺ (very hard)
- ☺☺ (hard)
- ☺☺☺ (not too hard but not easy either)
- ☺☺☺☺ (easy)
- ☺☺☺☺☺ (very easy)

GLOSSARY

Bolded terms in the text are included in the glossary at the back of the book.

Summary

That couldn't have been that bad, right? We want to encourage you to continue reading and not worry about what's difficult or time-consuming or too complex for you to understand and apply. Just take one chapter at a time, as you did this one.

Time to Practice

Because there's no substitute for the real thing, Chapters 1 through 17 and Chapter 19 will end with a set of exercises that will help you review the material that was covered in the chapter. As noted earlier, the answers to these exercises can be found near the end of the book in Appendix D.

For example, here is the first set of exercises (but don't look for any answers for these because these are kind of "on your own" answers—each question and answer are highly tied to your own experiences and interest).

1. Interview someone who uses statistics in his or her everyday work. It might be your advisor, an instructor, a researcher who lives on your block, a health care analyst, a marketer for a company, a city planner, or . . . Ask the person what his or her first statistics course was like. Find out what the person liked and didn't like. See if this individual has any suggestions to help you succeed. And most important, ask the person about how he or she uses these new-to-you tools at work.
2. We hope that you are part of a study group or, if that is not possible, that you have a phone, email, instant messaging, or webcam study buddy (or even more than one). And, of course, plenty of texting and Facebook friends. Talk to your group or a fellow student in your class about similar likes, dislikes, fears, and so on about the statistics course. What do you have in common? Not in common? Discuss with your fellow student strategies to overcome your fears.
3. Search through your local newspaper (or any other publication) and find the results of a survey or interview about any topic. Summarize the results and do the best job you can describing how the researchers who were involved, or the authors of the survey, came to the conclusions they did. Their methods and reasoning may or may not be apparent. Once you have some idea of what they did, try to speculate as to what other ways the same information might be collected, organized, and summarized.
4. Go to the library (either in person or online) and find a copy of a journal article in your own discipline. Then, go through the article and highlight the section (usually the "Results" section) where statistical procedures were used to organize and analyze the data. You don't know much about the specifics of this yet, but how many different statistical procedures (such as *t* test, mean, and calculation of the standard deviation) can you identify? Can you take the next step and tell your instructor how the results relate to the research question or the primary topic of the research study?

5. Find five websites that contain data on any topic and write a brief description of what type of information is offered and how it is organized. For example, if you go to the mother of all data sites, the U.S. Census (http://www.census.gov), you'll find links to hundreds of databases, tables, and other informative tools. Try to find data and information that fit in your own discipline.
6. And the big extra-credit assignment is to find someone who actually uses SPSS for daily data analysis needs. Ask if there is anything specific about SPSS that makes it stand out as a tool for their type of data analysis. You may very well find these good folks in everything from political science to nursing, so search widely!
7. Finally, as your last in this first set of exercises, come up with five of the most interesting questions you can about your own area of study or interest. Do your best to come up with questions for which you would want real, existing information or data to answer. Be a scientist!

Student Study Site

Get the tools you need to sharpen your study skills! Visit **edge.sagepub.com/salkindfrey7e** to access practice quizzes, eFlashcards, original and curated videos, data sets, and more!

PART II

ΣIGMA FREUD AND DESCRIPTIVE STATISTICS

"Yes, it's all becoming clear."

One of the things that Sigmund Freud, the founder of psychoanalysis, did quite well was to observe and describe the nature of his patients' conditions. An astute observer, he used his skills to develop the first systematic and comprehensive theory of personality. The validity of his ideas is questioned today by some, but he was a good scientist.

Back in the early 20th century, courses in statistics (like the one you are taking) were not offered as part of undergraduate or graduate curricula. The field was relatively new, and the nature of scientific explorations did not demand the precision that this set of tools brings to the scientific arena.

But things have changed. Now, in almost any endeavor, numbers count (as Francis Galton, the inventor of correlation and a first cousin to Charles Darwin, said as well). This section of *Statistics for People Who (Think They) Hate Statistics* is devoted to understanding how we can use statistics to describe an outcome and better understand it, once the information about the outcome is organized.

Chapter 2 discusses measures of central tendency, how computing one of several different types of averages gives you the one best data point that represents a set of scores and when to use what. Chapter 3 completes the coverage of tools we need to fully describe a set of data points in its discussion of how scores differ from each other, including the standard deviation and variance. When you get to Chapter 4, you will be ready to learn how distributions, or sets of scores, differ from one another and what this difference means. Chapter 5 deals with the nature of relationships between variables, namely, correlations. And finally, Chapter 6 introduces you to the importance of reliability, the randomness in scores, and validity, whether scores represent what they are supposed to, when we are describing some of the qualities of effective measurement tools.

When you finish Part II, you'll be in excellent shape to start understanding how we use an understanding of chance to make inferences in the social, behavioral, and other sciences.

COMPUTING AND UNDERSTANDING AVERAGES

Means to an End

Difficulty Scale ☺ ☺ ☺ ☺
(moderately easy)

WHAT YOU WILL LEARN IN THIS CHAPTER

- Understanding measures of central tendency
- Computing the mean for a sample of scores
- Computing the median for a sample of scores
- Computing the mode for a sample of scores
- Understanding and applying scales or levels of measurement
- Selecting a measure of central tendency

You've been very patient, and now it's finally time to get started working with some real, live data. That's exactly what you'll do in this chapter. Once data are collected, a usual first step is to begin to understand all those bits of information using single numbers to describe the data. The easiest way to do this is through computing an average, of which there are several different types.

An **average** is the one value that best represents an entire group of scores. It doesn't matter whether the group of scores represents the number correct on a spelling test for 30 fifth graders or the typical batting percentage for all the baseball players on the New York Yankees or how voters feel about a congressional candidate. In all of these examples, a big group of data can be summarized using an average. You can usually think of an average as the "middle" space or as a fulcrum on a seesaw. It's the point in a range of values that seems to most fairly represent all the values.

Averages, also called **measures of central tendency**, come in three flavors: the mean, the median, and the mode. Each provides you with a different type of information about a distribution of scores and is simple to compute and interpret.

COMPUTING THE MEAN

The **mean** is the most common type of average that is computed. It is so popular that scientists sometimes sloppily treat the word *average* as if it means *mean* when it only sometimes means *mean*. The mean is simply the sum of all the values in a group, divided by the number of values in that group. So, if you had the spelling scores for 30 fifth graders, you would simply add up all the scores to get a total and then divide by the number of students, which is 30.

We are about to show a formula or equation for the first time in this book. Don't panic. Equations are just statements or sentences that use symbols instead of words. We will always tell you what words the symbols stand for. The formula for computing the mean is shown in Formula 2.1:

$$\bar{X} = \frac{\sum X}{n}, \tag{2.1}$$

where

- the letter X with a line above it (also sometimes called "X bar") is the mean value of the group of scores;
- the $\sum$, or the Greek letter sigma, is the summation sign, which tells you to "sum up" or add together whatever follows it;
- the X is each individual score in a group of scores; and
- the n is the size of the sample from which you are computing the mean, the number of scores.

To compute the mean, follow these steps:

List the entire set of values in one or more columns. These are all the *X*s.

Compute the sum or total of all the values.

Divide the total or sum by the number of values.

For example, if you needed to compute the average number of shoppers at three different locations, you would compute a mean for that value.

Location	Number of Shoppers Last Year
Lanham Park Store	2,150
Williamsburg Store	1,534
Downtown Store	3,564

The mean or average number of shoppers in each store is 2,416. Formula 2.2 shows how this average was computed using the formula you saw in Formula 2.1:

$$\bar{X} = \frac{\sum X}{n} = \frac{2{,}150 + 1{,}534 + 3{,}564}{3} = \frac{7{,}248}{3} = 2{,}416. \qquad (2.2)$$

Or, if you needed to compute the average number of students in each grade in a school building, you would follow the same procedure.

Grade	Number of Students
Kindergarten	18
1	21
2	24
3	23
4	22
5	24
6	25

The mean or average number of students in each class is 22.43. Formula 2.3 shows how this average was computed using the formula you saw in Formula 2.1:

$$\bar{X} = \frac{\sum X}{n} = \frac{18 + 21 + 24 + 23 + 22 + 24 + 25}{7} = \frac{157}{7} = 22.43. \qquad (2.3)$$

See, we told you it was easy. No big deal. By the way, when you calculated that mean just now, you may have gotten a number with lots more digits in it: 22.42857143 or something like that. Statisticians are usually okay with you shortening numbers to just a couple digits past the decimal. So, we felt fine reporting the mean as 22.43 (rounding up for that last digit).

- The mean is sometimes represented by the letter *M* and is also called the typical, average, or most central score. If you are reading another statistics book or a research report and you see something like $M = 45.87$, it probably means that the mean is equal to 45.87. Technically, that capital letter *M* is used when you are talking about the mean of the larger population represented by the sample in front of you. Those sorts of distinctions aren't important right now but might be interesting later on.

- In the formula, a small *n* represents the sample size for which the mean is being computed. A large *N* (← like this) would represent the population size. In some books and in some journal articles, no distinction is made between the two. Notice, as with the capital *M* when talking about the mean of a population, statistical types often capitalize a letter symbol to refer to a population and keep the letter as lowercase when talking about samples.
- The mean is like the fulcrum on a seesaw. It's the centermost point where all the values on one side of the mean are equal in weight to all the values on the other side of the mean.
- Finally, for better or worse, the mean is very sensitive to extreme scores. An extreme score can pull the mean in one or the other direction and make it less representative of the set of scores and less useful as a measure of central tendency. This, of course, all depends on the values for which the mean is being computed. And, if you have extreme scores and the mean won't work as well as you want, we have a solution! More about that later.

The mean is also referred to as the **arithmetic mean**, and there are other types of means that you may read about, such as the harmonic mean. Those are used in special circumstances and need not concern you here. And if you want to be technical about it, the arithmetic mean (which is the one that we have discussed up to now) is also defined as the point about which the sum of the deviations is equal to zero (whew!). Each score in a sample is some distance from the mean. If you add up all those distances, it will equal zero. Always. Every time. That's why we like the mean. For instance, if you have scores like 3, 4, and 5 (whose mean is 4), the sum of the deviations about the mean (−1, 0, and +1) is 0.

Remember that the word *average* means only the one measure that best represents a set of scores and that there are many different types of averages. Which type of average you use depends on the question that you are asking and the type of data that you are trying to summarize. This is a levels of measurement issue that we will cover later in this chapter when we talk about when to use which statistic.

In basic statistics, an important distinction is made between those values associated with samples (a part of a population) and those associated with populations. To do this, statisticians use the following conventions. For a sample statistic (such as the mean of a sample), Roman letters are used. For a population parameter (such as the mean of a population), Greek letters are used. So, for example, the mean for the spelling score for a sample of 100 fifth graders is represented as $\bar{X} = 5$, whereas the mean for the spelling score for the entire population of fifth graders is represented, using the Greek letter *mu*, as μ_5.

COMPUTING THE MEDIAN

The median is also an average, but of a very different kind. The **median** is defined as the midpoint in a set of scores. It's the point at which one half, or 50%, of the scores fall above and one half, or 50%, fall below. It's got some special qualities that we will talk about later in this section, but for now, let's concentrate on how it is computed. There's not really a formula for computing the median but instead a set of steps.

To compute the median, follow these steps:

List the values in order, from either highest to lowest or lowest to highest.

Find the middle-most score. That's the median.

For example, here are the annual incomes from five different households:

$135,456

$45,500

$62,456

$54,365

$37,668

Here is the list ordered from highest to lowest:

$135,456

$62,456

$54,365

$45,500

$37,668

There are five values. The middle-most value is $54,365, and that's the median.

Now, what if the number of values is even? An even number of scores means there is no middle value. Let's add a value ($64,500) to the list so there are six income levels. Here they are sorted with the largest value first:

$135,456

$64,500

$62,456

$54,365

$45,500

$37,668

When there is an even number of values, the median is simply the mean of the two middle values. In this case, the middle two cases are $54,365 and $62,456. The mean of those two values is $58,410.50. That's the median for that set of six values.

What if the two middle-most values are the same, such as in the following set of data?

$45,678

$25,567

$25,567

$13,234

Then the median is same as both of those middle-most values. In this example, it's $25,567.

If we had a series of values that was the number of days spent in rehabilitation for a sports-related injury for seven different patients, the numbers might look like this:

43

34

32

12

51

6

27

As we did before, we can order the values (51, 43, 34, 32, 27, 12, 6) and then select the middle value as the median, which in this case is 32. So, the median number of days spent rehabilitating an injury is 32.

If you know about medians, you should also know about . Percentile ranks are used to define the percentage of cases equal to or below a certain point in a distribution or set of scores. For example, if a score is "at the

75th percentile," it means that the score is at or above 75% of the other scores in the distribution. The median is also known as the 50th percentile, because it's the point at or below which 50% of the cases in the distribution fall. Other percentiles are useful as well, such as the 25th percentile, often called Q_1, and the 75th percentile, referred to as Q_3. So what's Q_2? The median, of course.

Here comes the answer to the question you've probably had in the back of your mind since we started talking about the median. Why use the median instead of the mean? For one very good reason. The median is insensitive to extreme scores, whereas the mean is not. When you have a set of scores in which one or more scores are extreme, the median better represents the centermost value of that set of scores than any other measure of central tendency. Yes, even better than the mean.

What do we mean by *extreme*? It's probably easiest to think of an extreme score as one that is very different from the group to which it belongs. For example, consider the list of five incomes that we worked with earlier (shown again here):

$135,456

$54,365

$37,668

$32,456

$25,500

The value $135,456 is more different from the other four than is any other value in the set. We would consider that an extreme score.

The best way to illustrate how useful the median is as a measure of central tendency is to compute both the mean and the median for a set of data that contains one or more extreme scores and then compare them to see which one best represents the group. Here goes.

The mean of the set of five scores you see above is the sum of the set of five divided by 5, which turns out to be $57,089. On the other hand, the median for this set of five scores is $37,668. Which is more representative of the group? The value $37,668, because it clearly lies closer to most of the scores, and we like to think about "the average" (in this case, we are using the median as a measure of average) as being representative or assuming a central position. In fact, the mean value of $57,089 falls above the fourth highest value ($54,365) and is not very central or representative of the distribution.

It's for this reason that certain social and economic indicators (often involving income) are reported using a median as a measure of central tendency—"The

median income of the average American family is . . ."—rather than using the mean to summarize the values. There are just too many extreme scores that would **skew**, or significantly distort, what is actually a central point in the set or distribution of scores. As an example, the median annual family income in the United States for 2014 (the most recent year we could find data for) was about $54,000, while the mean annual family income was about $73,000. Which is closer to your family's income?

You learned earlier that sometimes the mean is represented by the capital letter *M* instead of $\bar{X}$. Well, other symbols are used for the median as well. We like the letter *M*, but some people confuse it with the mean, so they use *Med* or *Mdn* for median. Don't let that throw you—just remember what the median is and what it represents, and you'll have no trouble adapting to different symbols.

Here are some interesting and important things to remember about the median:

- We use the word *median* to describe other middle things, like the median on a highway, that stripe down the middle of a road.
- Because the median is based on how many cases there are, and not the values of those cases, extreme scores (sometimes called **outliers**) only count a little.

COMPUTING THE MODE

The third and last measure of central tendency that we'll cover, the mode, is the most general and least precise measure of central tendency, but it plays a very important part in understanding the characteristics of a sample of scores. The **mode** is the value that occurs most frequently. Like the median, there is no formula for computing the mode.

To compute the mode, follow these steps:

List all the values in a distribution but list each value only once.

Tally the number of times that each value occurs.

The value that occurs most often is the mode.

For example, an examination of the political party affiliation of 300 people might result in the following distribution of scores: