EXPLORING STATISTICS

TALES OF DISTRIBUTIONS

12TH EDITION

CHRIS SPATZ

12th Edition

Exploring Statistics

Tales of Distributions

Chris Spatz

Outcrop Publishers Conway, Arkansas

Exploring Statistics: Tales of Distributions
12th Edition
Chris Spatz

Online study guide available at
http://exploringstatistics.com/studyguide.php

Cover design: Grace Oxley
Answer Key: Jill Schmidlkofer
Webmaster & Ebook: Fingertek Web Design, Tina Haggard
Managers: Justin Murdock, Kevin Spatz

Copyright © 2019 by Outcrop Publishers, LLC
All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the publisher, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law. For permission requests, contact info@outcroppublishers.com or write to the publisher at the address below.

Outcrop Publishers
615 Davis Street
Conway, AR 72034
Email: info@outcroppublishers.com
Website: outcroppublishers.com
Library of Congress Control Number: [Applied for]

ISBN-13 (hardcover): 978-0-9963392-2-3
ISBN-13 (ebook): 978-0-9963392-3-0
ISBN-13 (study guide): 978-0-9963392-4-7

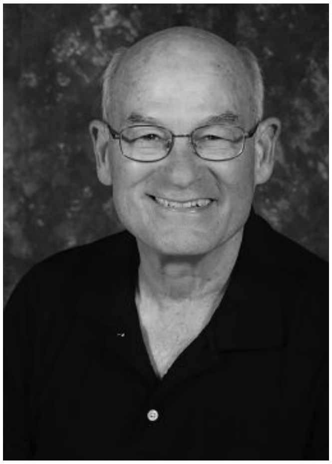
Examination copies are provided to academics and professionals to consider for adoption as a course textbook. Examination copies may not be sold or transferred to a third party. If you adopt this textbook, please accept it as your complimentary desk copy.

Ordering information:
Students and professors – visit exploringstatistics.com
Bookstores – email info@outcroppublishers.com

Photo Credits – Chapter 1
Karl Pearson – Courtesy of Wellcomeimages.org
Ronald A. Fisher – R.A. Fisher portrait, 0006973, Special Collections Research Center, North Carolina State University Libraries, Raleigh, North Carolina
Jerzy Neyman – Paul R. Halmos Photograph Collection, e_ph 0223_01, Dolph Briscoe Center for American History, The University of Texas at Austin
Jacob Cohen – New York University Archives, Records of the NYU Photo Bureau

Printed in the United States of America by Walsworth ®
2 3 4 5 6 7 24 23 22 21 20

About The Author



Chris Spatz is at Hendrix College where he twice served as chair of the Psychology Department. Dr. Spatz's undergraduate education was at Hendrix, and his PhD in experimental psychology is from Tulane University in New Orleans. He subsequently completed postdoctoral fellowships in animal behavior at the University of California, Berkeley, and the University of Michigan. Before returning to Hendrix to teach, Spatz held positions at The University of the South and the University of Arkansas at Monticello.

Spatz served as a reviewer for the journal Teaching of Psychology for more than 20 years. He co-authored a research methods textbook, wrote several chapters for edited books, and was a section editor for the *Encyclopedia of Statistics in Behavioral Science.*

In addition to writing and publishing, Dr. Spatz enjoys the outdoors, especially canoeing, camping, and gardening. He swims several times a week (mode = 3). Spatz has been an opponent of high textbook prices for years, and he is happy to be part of a new wave of authors who provide high-quality textbooks to students at affordable prices.

With love and affection,

this textbook is dedicated to

Thea Siria Spatz, Ed.D., CHES

Brief Contents

Contents

Preface

Even if our statistical appetite is far from keen, we all of us should like to know enough to understand, or to withstand, the statistics that are constantly being thrown at us in print or conversation—much of it pretty bad statistics. The only cure for bad statistics is apparently more and better statistics. All in all, it certainly appears that the rudiments of sound statistical sense are coming to be an essential of a liberal education.

– Robert Sessions Woodworth

Exploring Statistics: Tales of Distributions (12th edition) is a textbook for a one-term statistics course in the social or behavioral sciences, education, or an allied health/nursing field. Its focus is conceptualization, understanding, and interpretation, rather than computation. Designed to be comprehensible and complete for students who take only one statistics course, it also includes elements that prepare students for additional statistics courses. For example, basic experimental design terms such as independent and dependent variables are explained so students can be expected to write fairly complete interpretations of their analyses. In many places, the student is invited to stop and think or do a thought exercise. Some problems ask the student to decide which statistical technique is appropriate. In sum, this book's approach is in tune with instructors who emphasize critical thinking in their course.

This textbook has been remarkably successful for more than 40 years. Students, professors, and reviewers have praised it. A common refrain is that the book has a conversational, narrative style that is engaging, especially for a statistics text. Other features that distinguish this textbook from others include the following:

- Data sets are approached with an attitude of exploration.
- Changes in statistical practice over the years are acknowledged, especially the recent emphasis on effect sizes and confidence intervals.
- Criticism of null hypothesis significance testing (NHST) is explained.
- Examples and problems represent a variety of disciplines and everyday life.
- Most problems are based on actual studies rather than fabricated scenarios.
- Interpretation is emphasized throughout.
- Problems are interspersed within a chapter, not grouped at the end.
- Answers to all problems are included.
- Answers are comprehensively explained—over 50 pages of detail.
- A final chapter, *Choosing Tests and Writing Interpretations*, requires active responses to comprehensive questions.

- Effect size indexes are treated as important descriptive statistics, not add-ons to NHST.
- Important words and phrases are defined in the margin when they first occur.
- *Objectives*, which open each chapter, serve first for orientation and later as review items.
- *Key Terms* are identified for each chapter.
- *Clues to the Future* alert students to concepts that come up again.
- *Error Detection* boxes tell ways to detect mistakes or prevent them.
- *Transition Passages* alert students to a change in focus in chapters that follow.
- *Comprehensive Problems* encompass all (or most) of the techniques in a chapter.
- *What Would You Recommend*? problems require choices from among techniques in several chapters.

For this 12th edition, I increased the emphasis on effect sizes and confidence intervals, moving them to the front of Chapter 9 and Chapter 10. The controversy over NHST is addressed more thoroughly. Power gets additional attention. Of course, examples and problems based on contemporary data are updated, and there are a few new problems. In addition, a helpful *Study Guide to Accompany Exploring Statistics* (12th edition) was written by Lindsay Kennedy, Jennifer Peszka, and Leslie Zorwick, all of Hendrix College. The study guide is available online at exploringstatistics.com.

Students who engage in this book and their course can expect to:

- Solve statistical problems
- Understand and explain statistical reasoning
- Choose appropriate statistical techniques for common research designs
- Write explanations that are congruent with statistical analyses

After many editions with a conventional publisher, *Exploring Statistics: Tales of Distributions* is now published by Outcrop Publishers. As a result, the price of the print edition is about one-fourth that of the 10th edition. Nevertheless, the authorship and quality of earlier editions continue as before.

Acknowledgments

The person I acknowledge first is the person who most deserves acknowledgment. And for the 11th and 12th editions, she is especially deserving. This book and its accompanying publishing company, Outcrop Publishers, would not exist except for Thea Siria Spatz, encourager, supporter, proofreader, and cheer captain. This edition, like all its predecessors, is dedicated to her.

Kevin Spatz, manager of Outcrop Publishers, directed the distribution of the 11th edition, advised, week by week, and suggested the cover design for the 12th edition. Justin Murdock now serves as manager, continuing the tradition that Kevin started. Tina Haggard of Fingertek Web Design created the book's website, the text's ebook, and the online study guide. She provided advice and solutions for many problems. Thanks to Jill Schmidlkofer, who edited the extensive answer section again for this edition. Emily Jones Spatz created new drawings for the text. I'm particularly grateful to Grace Oxley for a cover design that conveys exploration, and to Liann Lech, who copyedited for clarity and consistency. Walsworth® turned a messy collection of files into a handsome book—thank you Nathan Stufflebean and Dennis Paalhar. Others who were instrumental in this edition or its predecessors include Jon Arms, Ellen Bruce, Mary Kay Dunaway, Bob Eslinger, James O. Johnston, Roger E. Kirk, Rob Nichols, Jennifer Peszka, Mark Spatz, and Selene Spatz. I am especially grateful to Hendrix College and my Hendrix colleagues for their support over many years, and in particular, to Lindsay Kennedy, Jennifer Peszka, and Leslie Zorwick, who wrote the study guide that accompanies the text.

This textbook has benefited from perceptive reviews and significant suggestions by some 90 statistics teachers over the years. For this 12th edition, I particularly thank

Jessica Alexander, Centenary College
Lindsay Kennedy, Hendrix College
Se-Kang Kim, Fordham University
Roger E. Kirk, Baylor University
Kristi Lekies, The Ohio State University
Jennifer Peszka, Hendrix College
Robert Rosenthal, University of California, Riverside

I've always had a touch of the teacher in me—as an older sibling, a parent, a professor, and now a grandfather. Education is a first-class task, in my opinion. I hope this book conveys my enthusiasm for it. (By the way, if you are a student who is so thorough as to read even the acknowledgments, you should know that I included phrases and examples in a number of places that reward your kind of diligence.)

If you find errors in this book, please report them to me at spatz@hendrix.edu. I will post corrections at the book's website: exploringstatistics.com.

CHAPTER 1

Introduction

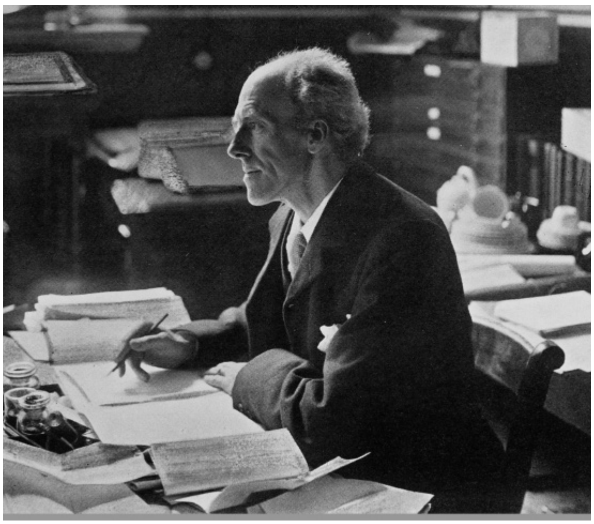
OBJECTIVES FOR CHAPTER 1

After studying the text and working the problems in this chapter, you should be able to:

1. Distinguish between descriptive and inferential statistics
2. Define population, sample, parameter, statistic, and variable as they are used in statistics
3. Distinguish between quantitative and categorical variables
4. Distinguish between continuous and discrete variables
5. Identify the lower and upper limits of a continuous variable
6. Identify four scales of measurement and distinguish among them
7. Distinguish between statistics and experimental design
8. Define independent variable, dependent variable, and extraneous variable and identify them in experiments
9. Describe statistics' place in epistemology
10. List actions to take to analyze a data set
11. Identify a few events in the history of statistics

WE BEGIN OUR exploration of statistics with a trip to London. The year is 1900.

Walking into an office at University College London, we meet a tall, well-dressed man about 40 years old. He is Karl Pearson, Professor of Applied Mathematics and Mechanics. I ask him to tell us a little about himself and why he is an important person. He seems authoritative, glad to talk about himself. As a young man, he says, he wrote essays, a play, and a novel, and he also worked for women's suffrage. These days, he is excited about this new branch of biology called genetics. He says he supervises lots of data gathering.



Karl Pearson

Pearson, warming to our group, lectures us about the major problem in science—there is no agreement on how to decide among competing theories. Fortunately, he just published a new statistical method that provides an objective way to decide among competing theories, regardless of the discipline. The method is called chi square.[1] Pearson says, "Now, arguments will be much fewer. Gather a thousand data points and calculate a chi square test. The result gives everyone an objective way to determine whether or not the data fit the theory."

Exploration Notes from a student: Exploration off to good start. Hit on a nice, easy-to-remember date to start with, visited a founder of statistics, and had a statistic called chi square described as a big deal.

Ronald A. Fisher

Our next stop is Rothamsted Experiment Station just north of London. Now the year is 1925. There are fields all around the agricultural research facility, each divided into many smaller plots. The growth in the fields seems quite variable.

Arriving at the office, the atmosphere is congenial. The staff is having tea. There are two topics—a new baby and a new book. We get introduced to Ronald Fisher, the chief statistician. Fisher is a small man with thick glasses and red hair.

He tells us about his new child[2] and then motions to a book on the table. Sneaking a peek, we read the title: *Statistical Methods for Research Workers*. Fisher becomes focused on his book, holding forth in an authoritative way.

He says the book explains how to conduct experiments and that an experiment is just a comparison of two or more conditions. He tells us we don't need a thousand data points. He says that small samples, randomly selected, are the way for science to progress. "With an experiment and my technique of analysis of variance," he exclaims, "you can determine why that field out there"—here he waves toward the window—"is so variable. We can find out what makes some plots lush and some mimsy." *Analysis of variance*,[3] he says, works in any discipline, not just agriculture.

Exploration Notes: Looks like statistics had some controversy in it.[4] Also looks like progress. Statistics is used for experiments, too, and not just for testing theories. And Fisher says experiments can be used to compare anything. If that's right, I can use statistics no matter what I major in.

---

[1] Chi square, which is explained in this book in Chapter 14, has been called one of the 20 most important inventions in the 20th century (Hacking, 1984).
[2] (in what will become a family with eight children).
[3] explained in Chapters 11-13
[4] The slight sniping I've built into this story is just a hint of the strong animosity between Fisher and Pearson.


Jerzy Neyman

Next we go to Poland to visit Jerzy Neyman at his office at the University of Warsaw. It is 1933. As we walk in, he smiles, seems happy we've arrived, and makes us feel completely welcome.

Motioning to an envelope on his desk, he tells us it holds a manuscript that he and Egon Pearson[5] wrote. "The problem with Fisher's analysis of variance test is that it focuses exclusively on finding a difference between groups. Suppose the statistical test doesn't detect a difference. Does that prove there is no difference? No, of course not. It may be that the test was just not sensitive enough to detect the difference. Right?"

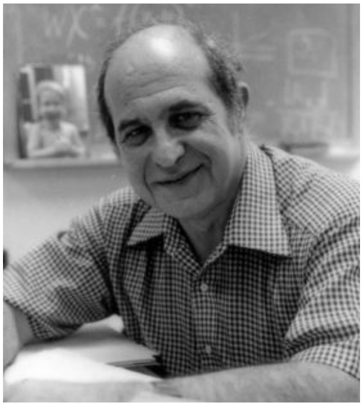
At his question, a few of us nod in agreement. Seeing uncertainty, he notes, "Maybe a larger sample is needed to find the difference, you see? Anyway, what we've done is expand statistics to cover not just finding a difference, but also what it means when the test doesn't find a difference. Our approach is what you people in your time will call *null hypothesis significance testing.*"

Exploration Notes: Statistics seems like a work in progress. Changing. Now it is not just about finding a difference but also about what it means not to find a difference. Also, looks like null hypothesis significance testing is a phrase that might turn up on tests.

Our next trip is to libraries, say, anytime between 1940 and 2000. For this exploration, the task is to examine articles in professional journals published in various disciplines. The disciplines include anthropology, biology, chemistry, defense strategy, education, forestry, geology, health, immunology, jurisprudence, manufacturing, medicine, neurology, ophthalmology, political science, psychology, sociology, zoology, and others. I'm sure you get the idea—the whole range of disciplines that use quantitative measures in their research. What this exploration produces is the discovery that all of these disciplines rely on a data analysis technique called *null hypothesis significance testing* (NHST).[6] Many different statistical tests are employed. However, for all the tests in all the disciplines, the phrase, "$p < .05$" turns up frequently.

Exploration Notes: It seems that all that earlier controversy has subsided and scientists in all sorts of disciplines have agreed that NHST is the way to analyze quantitative data. All of them seem to think that if there is a comparison to be made, applying NHST is a necessary step to get correct conclusions. All of them use "$p < .05$," so I'll have to be sure to find out exactly what that means.

[5] Egon Pearson was Karl Pearson's son.
[6] Null hypothesis significance testing is first explained in Chapters 9 and 10.



Jacob Cohen

Our next excursion is a 1962 visit with Jacob Cohen at New York University in New York City. He is holding his article about studies published in the *Journal of Abnormal and Social Psychology*, a leading psychology journal. He tells us that the NHST technique has problems. Also, he says we should be calculating an effect size statistic, which will show whether the differences observed in our experiments are large or small.

Exploration Notes: The idea of an effect size index makes a lot of sense. Just knowing there is a difference isn't enough. How big is the difference? Wonder what "problems with NHST" is all about.

Back to the library for a final excursion to check out recent events. We come across a 2014 article by Geoff Cumming on the "new statistics." We find things like, "avoid NHST and use better techniques" (p. 26) and "we should not trust any p value" (p. 13). This seems like awfully strong advice. Are researchers taking this advice? Looking through more of today's research in journals in several fields, we find that most statistical analyses use NHST and there are many instances of "$p < .05$."

Exploration Notes, Conclusion: These days, it looks like statistics is in transition again. There's a lot of controversy out there about how to analyze data from experiments. The NHST approach is still very common, though, so it's clear I must learn it. But I want to be prepared for changes. I hope knowing NHST will be helpful for the future.[7]

Welcome to statistics at a time when the discipline is once again in transition. A well-established tradition (null hypothesis significance testing) has been in place for almost a century but is now under attack. New ways of thinking about data analysis are emerging, and along with them, a collection of statistics that do not include the traditional NHST approach. As for the immediate future, though, NHST remains the method most widely used by researchers in many fields. In addition, much of the thinking required for NHST is required for other approaches.

Our exploration tour is over, so I'll quit supplying notes; they are your responsibility now. As your own experience probably shows, making up your own summary notes improves retention of what you read. In addition, I have a suggestion. Adopt a mindset that thinks *growth*. A student with a growth mindset expects to learn new things. When challenges arise, as they

[7] Not only helpful, but necessary, I would say.

inevitably do, acknowledge them and figure out how to meet the challenge. A growth mindset treats ability as something to be developed (see Dweck, 2016). If you engage yourself in this course, you can expect to use what you learn for the rest of your life.

The main title of this book is “Exploring Statistics.” *Exploring* conveys the idea of uncovering something that was not apparent before. An attitude of *searching, wondering, checking*, and so forth is what I want to encourage. (Those who object to traditional NHST procedures are driven by this exploration motivation.) As for this book’s subtitle, “Tales of Distributions,” I’ll have more to say about it as we go along.

Disciplines that Use Quantitative Data

Which disciplines use quantitative data? The list is long and more variable than the list I gave earlier. The examples and problems in this textbook, however, come from psychology, biology, sociology, education, medicine, politics, business, economics, forestry, and everyday life. Statistics is a powerful method for getting answers from data, and this makes it popular with investigators in a wide variety of fields.

Statistics is used in areas that might surprise you. As examples, statistics has been used to determine the effect of cigarette taxes on smoking among teenagers, the safety of a new surgical anesthetic, and the memory of young school-age children for pictures (which is as good as that of college students). Statistics show which diseases have an inheritance factor, how to improve short-term weather forecasts, and why giving intentional walks in baseball is a poor strategy. All these examples come from Statistics: *A Guide to the Unknown*, a book edited by Judith M. Tanur and others (1989). Written for those “without special knowledge of statistics,” this book has 29 essays on topics as varied as those above.

In American history, the authorship of 12 of *The Federalist* papers was disputed for a number of years. (*The Federalist* papers were 85 short essays written under the pseudonym “Publius” and published in New York City newspapers in 1787 and 1788. Written by James Madison, Alexander Hamilton, and John Jay, the essays were designed to persuade the people of the state of New York to ratify the Constitution of the United States.) To determine authorship of the 12 disputed papers, each was graded with a quantitative *value analysis* in which the importance of such values as national security, a comfortable life, justice, and equality was assessed. The value analysis scores were compared with value analysis scores of papers known to have been written by Madison and Hamilton (Rokeach, Homant, & Penner, 1970). Another study, by Mosteller and Wallace, analyzed *The Federalist* papers using the frequency of words such as *by* and *to* (reported in Tanur et al., 1989). Both studies concluded that Madison wrote all 12 essays.

Here is an example from law. Rodrigo Partida was convicted of burglary in Hidalgo County, a border county in southern Texas. A grand jury rejected his motion for a new trial. Partida’s attorney filed suit, claiming that the grand jury selection process discriminated against Mexican-Americans. In the end (*Castaneda v. Partida*, 430 U.S. 482 [1976]), Justice Harry

Blackmun of the U.S. Supreme Court wrote, regarding the number of Mexican-Americans on grand juries, "If the difference between the expected and the observed number is greater than two or three standard deviations, then the hypothesis that the jury drawing was random (is) suspect." In Partida's case, the difference was approximately 12 standard deviations, and the Supreme Court ruled that Partida's attorney had presented *prima facie* evidence. (*Prima facie* evidence is so good that one side wins the case unless the other side rebuts the evidence, which in this case did not happen.) *Statistics: A Guide to the Unknown* includes two essays on the use of statistics by lawyers.

Gigerenzer et al. (2007), in their public interest article on health statistics, point out that lack of statistical literacy among both patients and physicians undermines the information exchange necessary for informed consent and shared decision making. The result is anxiety, confusion, and undue enthusiasm for testing and treatment.

Whatever your current interests or thoughts about your future as a statistician, I believe you will benefit from this course. A successful statistics course teaches you to identify questions a set of data can answer; determine the statistical procedures that will provide the answers; carry out the procedures; and then, using plain English and graphs, tell the story the data reveal.

The best way for you to acquire all these skills (especially the part about telling the story) is to *engage* statistics. Engaged students are easily recognized; they are prepared for exams, are not easily distracted while studying, and generally finish assignments on time. *Becoming* an engaged student may not be so easy, but many have achieved it. Here are my recommendations. Read with the goal of understanding. Attend class. Do all the assignments (on time). Write down questions. Ask for explanations. Expect to understand. (Disclaimer: I'm not suggesting that you marry statistics, but just engage for this one course.)

Are you uncertain about whether your background skills are adequate for a statistics course? For most students, this is an unfounded worry. Appendix A, Getting Started, should help relieve your concerns.

What Do You Mean, "Statistics"?

Descriptive statistic
A number that conveys a particular characteristic of a set of data.

Mean
Arithmetic average; sum of scores divided by number of scores.

Inferential statistics
Method that uses sample evidence and probability to reach conclusions about unmeasurable populations.

The *Oxford English Dictionary* says that the word *statistics* came into use almost 250 years ago. At that time, *statistics* referred to a country's quantifiable political characteristics—characteristics such as population, taxes, and area. Statistics meant "state numbers." Tables and charts of those numbers turned out to be a very satisfactory way to compare different countries and to make projections about the future. Later, tables and charts proved useful to people studying trade (economics) and natural phenomena (science). Statistical thinking spread because it helped. Today, two different techniques are called *statistics*.

Descriptive statistics[8] produce a number or a figure that summarizes or describes a set of data. You are already familiar with some descriptive statistics. For example, you know about the arithmetic average, called the **mean**. You have probably known how to compute a mean since elementary school—just add up the numbers and divide the total by the number of entries. As you already know, the mean describes the central tendency of a set of numbers. The basic idea of descriptive statistics is simple: They summarize a set of data with one number or graph. This book covers about a dozen descriptive statistics.

The other statistical technique is **inferential statistics**. Inferential statistics use measurements from a sample to reach conclusions about a larger, *unmeasured* population. There is, of course, a problem with samples.

Samples always depend *partly* on the luck of the draw; chance helps determine the particular measurements you get.

If you have the measurements for the entire population, chance doesn't play a part—all the variation in the numbers is "true" variation. But with samples, some of the variation is the true variation in the population and some is just the chance ups and downs that go with a sample. Inferential statistics was developed as a way to account for the effects of chance that come with sampling. This book will cover about a dozen and a half inferential statistics.

Here is a textbook definition: Inferential statistics is a method that takes chance factors into account when samples are used to reach conclusions about populations. Like most textbook definitions, this one condenses many elements into a short sentence. Because the idea of using samples to understand populations is perhaps the *most important concept* in this course, please pay careful attention when elements of inferential statistics are explained.

Inferential statistics has proved to be a very useful method in scientific disciplines. Many other fields use inferential statistics, too, so I selected examples and problems from a variety of disciplines for this text and its auxiliary materials. *Null hypothesis significance testing*, which had a prominent place in our exploration tour, is an inferential statistics technique.

Here is an example from psychology that uses the NHST technique. Today, there is a lot of evidence that people remember the tasks they fail to complete better than the tasks they complete. This is known as the *Zeigarnik effect*. Bluma Zeigarnik asked participants in her experiment to do about 20 tasks, such as work a puzzle, make a clay figure, and construct a box from cardboard.[9] For each participant, half the tasks were interrupted before completion. Later, when the participants were asked to recall the tasks they worked on, they listed more of the interrupted tasks (average about 7) than the completed tasks (about 4).

One good question to start with is, "Did interrupting make a big difference or a small difference?" In this case, interruption produced about three additional memory items compared to the completion condition. This is a 75% difference, which seems like a big change, given our experience with tests of memory. The question of "How big is the difference?" can often be answered by calculating an *effect size index*.

---

[8] Boldface words and phrases are defined in the margin and also in Appendix D, Glossary of Words.

[9] A summary of this study can be found in Ellis (1938). The complete reference and all others in the text are listed in the References section at the back of the book.

clue to the future

> Calculating effect size indexes is first addressed in Chapter 5. It is also a topic in Chapters 9-14.

So, should you conclude that interruption improves memory? Not yet. It might be that interruption actually has no effect but that several *chance factors* happened to favor the interrupted tasks in Zeigarnik's particular experiment. One way to meet this objection is to conduct the experiment again. Similar results would lend support to the conclusion that interruption improves memory. A less expensive way to meet the objection is to use inferential statistics such as NHST.

NHST begins with the actual data from the experiment. It ends with a probability—the probability of obtaining data like those actually obtained if it is true that interruption has *no* effect on memory. If the probability is very small, you can conclude that interruption *does* affect memory. For Zeigarnik's data, the probability was tiny.

Now for the conclusion. One version might be, "After completing about 20 tasks, memory for interrupted tasks (average about 7) was greater than memory for completed tasks (average about 4). The approximate 75% difference cannot be attributed to chance because chance by itself would rarely produce a difference between two samples as large as this one." The words *chance* and *rarely* tell you that probability is an important element of inferential statistics.

My more complete answer to what I mean by "statistics" is Chapter 6 in *21st Century Psychology: A Reference Handbook* (Spatz, 2008). This 8-page chapter summarizes in words (no formulas) the statistical concepts usually covered in statistics courses. This chapter can orient you as you begin your study of statistics and later provide a review after you finish your course.

clue to the future

> The first part of this book is devoted to descriptive statistics (Chapters 2–6) and the second part to inferential statistics (Chapters 7–15). Inferential statistics is the more comprehensive of the two because it combines descriptive statistics, probability, and logic.

Statistics: A Dynamic Discipline

Many people continue to think of statistics as a collection of techniques that were developed long ago, that have not changed, and that will be the same in the future. That view is mistaken. Statistics is a dynamic discipline characterized by more than a little controversy. New techniques in both descriptive and inferential statistics continue to be developed. Controversy continues too, as you saw at the end of our exploration tour. To get a feel for the issues when the controversy entered the mainstream, see Dillon (1999) or Spatz (2000) for nontechnical summaries. For more technical explanations, see Nickerson (2000). To read about current approaches, see Erceg-Hurn and Mirosevich (2008), Kline (2013), or Cumming (2014).

In addition to controversy over techniques, attitudes toward data analysis shifted in recent years. The shift has been toward the idea of exploring data to see what it reveals and away from using statistical analyses to nail down a conclusion. This shift owes much of its impetus to John Tukey (1915–2000), who promoted Exploratory Data Analysis (Lovie, 2005). Tukey invented techniques such as the boxplot (Chapter 5) that reveal several characteristics of a data set simultaneously.

Today, statistics is used in a wide variety of fields. Researchers start with a phenomenon, event, or process that they want to understand better. They make measurements that produce numbers. The numbers are manipulated according to the rules and conventions of statistics. Based on the outcome of the statistical analysis, researchers draw conclusions and then write the story of their new understanding of the phenomenon, event, or process. Statistics is just one tool that researchers use, but it is often an essential tool.

Some Terminology

Like most courses, statistics introduces you to many new words. In statistics, most of the terms are used over and over again. Your best move, when introduced to a new term, is to *stop, read* the definition carefully, and *memorize* it. As the term continues to be used, you will become more and more comfortable with it. Making notes is helpful.

Populations and Samples

A **population** consists of all the scores of some specified group. A **sample** is a subset of a population. The population is the thing of interest. It is defined by the investigator and includes all cases. The following are some populations:

Family incomes of college students in the fall of 2017
Weights of crackers eaten by obese male students
Depression scores of Alaskans
Gestation times for human beings
Memory scores of human beings[10]

Population
All measurements of a specified group.

Sample
Measurements of a subset of a population.

[10] I didn't pull these populations out of thin air; they are all populations that researchers have gathered data on. Studies of these populations will be described in this book.

Investigators are always interested in populations. However, as you can determine from these examples, populations can be so large that not all the members can be studied. The investigator must often resort to measuring a sample that is small enough to be manageable. A sample taken from the population of incomes of families of college students might include only 40 students. From the last population on the list, Zeigarnik used a sample of 164.

Most authors of research articles carefully explain the characteristics of the samples they use. Often, however, they do not identify the *population*, leaving that task to the reader.

The answer to the question "What is the population?" depends on the specifics of a research area, but many researchers generalize generously. For example, for some topics it is reasonable to generalize from the results of a study on rats to "all mammals." In all cases, however, the reason for gathering data from a sample is to generalize the results to a larger population even though sampling introduces some uncertainty into the conclusions.

Parameters and Statistics

Parameter
Numerical or nominal characteristic of a population.

Statistic
Numerical or nominal characteristic of a sample.

A **parameter** is some numerical (number) or nominal (name) characteristic of a population. An example is the mean reading readiness score of all first-grade pupils in the United States. A **statistic** is some numerical or nominal characteristic of a sample. The mean reading readiness score of 50 first-graders is a statistic, and so is the observation that 45% are girls. A parameter is constant; it does not change unless the population itself changes. The mean of a population is exactly one number. Unfortunately, the parameter often cannot be computed because the population is unmeasurable. So, a statistic is used as an estimate of the parameter, although, as suggested before, statistics tend to differ from one sample to another. If you have five samples from the same population, you will probably have five different sample means. In sum, parameters are constant; statistics are variable.

Variables

Variable
Something that exists in more than one amount or in more than one form.

A **variable** is something that exists in more than one amount or in more than one form. Height and eye color are both variables. The notation *67 inches* is a numerical way to identify a group of persons who are similar in height. Of course, there are many other groups, each with an identifying number. Blue and brown are common eye colors, which might be assigned the numbers *0* and *1*. All participants represented by *0* have the same eye color. I will often refer to numbers like *67* and *0* as *scores* or *test scores*. A score is simply the result of measuring a variable.

Quantitative Variables

Scores on **quantitative variables** tell you the *degree* or *amount* of the thing being measured. At the very least, a larger score indicates more of the variable than a smaller score does.

Quantitative variable
Variable whose scores indicate different amounts.

Continuous Variables. **Continuous variables** are quantitative variables whose scores can be any value or intermediate value over the variable's possible range. The continuous memory scores in Zeigarnik's experiment make up a quantitative, continuous variable. *Number of tasks recalled* scores come in whole numbers such as 4 or 7, but it seems reasonable to assume that the thing being measured, memory, is a continuous variable. Thus, of two participants who both scored 7, one just barely got 7 and the other almost scored 8. Picture the continuous variable, *recall*, as **Figure 1.1**.

Continuous variable
A quantitative variable whose scores can be any amount.

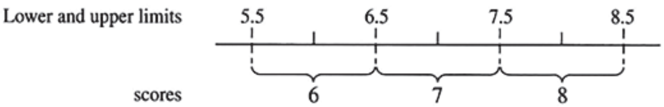


FIGURE 1.1 **The lower and upper limits of recall scores of 6, 7, and 8**

Figure 1.1 shows that a score of 7 is used for a range of possible recall values—the range from 6.5 to 7.5. The number 6.5 is the **lower limit** and 7.5 is the **upper limit** of the score of 7. The idea is that recall can be any value between 6.5 and 7.5, but that all the recall values in this range are expressed as 7. In a similar way, a charge indicator value of 62% on your cell phone stands for all the power values between 61.5% (the lower limit) and 62.5% (the upper limit).

Lower limit
Bottom of the range of possible values that a measurement on a continuous variable can have.

Upper limit
Top of the range of possible values that a measurement on a continuous variable can have.

Sometimes scores are expressed in tenths, hundredths, or thousandths. Like integers, these scores have lower and upper limits that extend halfway to the next value on the quantitative scale.

Discrete Variables. Some quantitative variables are classified as **discrete variables** because intermediate values are not possible. The number of siblings you have, the number of times you've been hospitalized, and how many pairs of shoes you have are examples. Intermediate scores such as 2½ just don't make sense.

Discrete variable
Variable for which intermediate values between scores are not meaningful.

Categorical Variables

Categorical variable
Variable whose scores differ in kind, not amount.

Categorical variables (also called *qualitative variables*) produce scores that differ in *kind* and not *amount*. Eye color is a categorical variable. Scores might be expressed as *blue* and *brown* or as *0* and *1*, but substituting a number for a name does not make eye color a quantitative variable. American political affiliation is a categorical variable with values of Democrat, Republican, Independent, and Other. College major is another categorical variable.

Some categorical variables have the characteristic of *order*. College standing has ordered measurements of senior, junior, sophomore, and freshman. Military rank is a categorical variable with scores such as sergeant, corporal, and private. Categorical variables such as color and gender do not have an inherent order. All categorical variables produce discrete scores, but not all discrete scores are from a categorical variable.

Problems and Answers

At the beginning of this chapter, I urged you to engage statistics. Have you? For example, did you read the footnotes? Have you looked up any words you weren't sure of? (How near are you to dictionary definitions when you study?) Have you read a paragraph a second time, wrinkled your brow in concentration, made notes in the book margin, or promised yourself to ask your instructor or another student about something you aren't sure of? *Engagement* shows up as activity. Best of all, the activity at times is a nod to yourself and a satisfied, "Now I understand."

From time to time, I will use my best engagement tactic: I'll give you a set of problems so that you can practice what you have just been reading about. Working these problems correctly is additional evidence that you have been engaged. You will find the answers at the end of the book in Appendix G. Here are some suggestions for *efficient* learning.

1. Buy yourself a notebook or establish a file for statistics. Save your work there. When you make an error, don't remove it—note the error and rework the problem correctly. Seeing your error later serves as a reminder of what not to do on a test. If you find that I have made an error, write to me with a reminder of what not to do in the next edition.
2. Never, never look at an answer before you have worked the problem (or at least tried twice to work the problem).
3. For each set of problems, work the first one and then immediately check your answer against the answer in the book. If you make an error, find out why you made it—faulty understanding, arithmetic error, or whatever.
4. Don't be satisfied with just doing the math. If a problem asks for an interpretation, write out your interpretation.
5. When you finish a chapter, go back over the problems immediately, reminding yourself of the various techniques you have learned.
6. Use any blank spaces near the end of the book for your special notes and insights.

Now, here is an opportunity to see how actively you have been reading.

PROBLEMS

1.1. The history-of-statistics tour began with what easy-to-remember date?
1.2. The dominant approach to inferential statistics that is under attack is called ___________.
1.3. Identify each number below as coming from a quantitative variable or a categorical variable.
 a. 65 – seconds to work a puzzle
 b. 319 – identification number for intellectual disability in the American Psychiatric Association manual
 c. 3 – group identification for small-cup daffodils
 d. 4 – score on a high school advanced placement exam
 e. 81 – milligrams of aspirin
1.4. Place lower and upper limits beside the continuous variables. Write *discrete* beside the others.
 a. ______________________ 20, seconds to work a puzzle
 b. ______________________ 14, number of concerts attended
 c. ______________________ 3, birth order
 d. ______________________ 10, speed in miles per hour
1.5. Write a paragraph that gives the definitions of *population, sample, parameter,* and *statistic* and the relationships among them.
1.6. Two kinds of statistics are ____________ statistics and ____________ statistics. Fill each blank with the correct adjective.
 a. To reach a conclusion about an unmeasured population, use ___________ statistics.
 b. ____________ statistics take chance into account to reach a conclusion.
 c. ____________ statistics are numbers or graphs that summarize a set of data.

Scales of Measurement

Numbers mean different things in different situations. Consider three answers that appear to be identical but are not:

What number were you wearing in the race?	"5"
What place did you finish in?	"5"
How many minutes did it take you to finish?	"5"

The three 5s all look the same. However, the three variables (identification number, finish place, and time) are quite different. Because of the difference in what the variables measure, each 5 has a different interpretation.

To illustrate this difference, consider another person whose answers to the same three questions were 10, 10, and 10. If you take the first question by itself and know that the two people had scores of 5 and 10, what can you say? You can say that the first runner was different from the second, but *that is all*. (Think about this until you agree.) On the second question, with scores of 5 and 10, what can you say? You can say that the first runner was faster than the second and, of course, that they are different.

Comparing the 5 and 10 on the third question, you can say that the first runner was twice as fast as the second runner (and, of course, was faster and different).

The point of this discussion is to draw the distinction between the *thing* you are interested in and the *number* that stands for the thing. Much of your experience with numbers has been with pure numbers or quantitative measures such as time, length, and amount. Four and two have a relationship of *twice as much* and *half as much*. And, for distance and seconds, four *is* twice two; for amounts, two is half of four. But these relationships do not hold when numbers are used to measure some things. For example, for political race finishes, *twice* and *half* are not helpful. Second place is not half or twice anything compared to fourth place.

S. S. Stevens (1946) identified four different *scales of measurement*, each of which carries a different set of information. Each scale uses numbers, but the information that can be inferred from the numbers differs. The four scales are *nominal*, *ordinal, interval,* and *ratio*.

In the **nominal scale**, numbers are used simply as names and have no real quantitative value. Numerals on sports uniforms are an example. Thus, 45 is *different* from 32, but that is all you can say. The person represented by 45 is not "more than" the person represented by 32, and certainly it would be meaningless to calculate the mean of the two numbers. Examples of nominal variables include psychological diagnoses, personality types, and political parties. Psychological diagnoses, like other nominal variables, consist of a set of categories. People are assessed and then classified into one of the categories. The categories have both a name (such as posttraumatic stress disorder or autism spectrum disorder) and a number (309.81 and 299.00, respectively). On a nominal scale, the numbers mean only that the categories are different. In fact, for a nominal scale variable, the numbers could be assigned to categories at random. Of course, all things that are alike must have the same number.

Nominal scale
Measurement scale in which numbers serve only as labels and do not indicate any quantitative relationship.

The **ordinal scale** has the characteristic of the nominal scale (different numbers mean different things) plus the characteristic of indicating *greater than* or *less than*. In the ordinal scale, the object with the number 3 has less or more of something than the object with the number 5. Finish places in a race are an example of an ordinal scale. The runners finish in rank order, with 1 assigned to the winner, 2 to the runner-up, and so on. Here, 1 means less time than 2. Judgments about anxiety, quality, and recovery often correspond to an ordinal scale. "Much improved," "improved," "no change," and "worse" are levels of an ordinal recovery variable. Ordinal scales are characterized by *rank order*.

Ordinal scale
Measurement scale in which numbers are ranks; equal differences between numbers do not represent equal differences between the things measured.

The third kind of scale is the **interval scale**, which has the properties of both the nominal and ordinal scales plus the additional property that *intervals between the numbers are equal.* "Equal interval" means that the distance between the things represented by 2 and 3 is the same as the distance between the things represented by 3 and 4. Temperature is measured on an interval scale. The difference in temperature between 10°C and 20°C is the same as the difference between 40°C and 50°C. The Celsius thermometer, like all interval scales, has an arbitrary zero point. On the Celsius thermometer, this zero point is the freezing point of water at sea level. Zero degrees on this scale does not mean the complete absence of heat; it is simply a convenient starting point. With interval data, there is one restriction: You may not make simple ratio statements. You may not say that 100° is twice as hot as 50° or that a person with an IQ of 60 is half as intelligent as a person with an IQ of 120.[11]

Interval scale
Measurement scale in which equal differences between numbers represent equal differences in the thing measured. The zero point is arbitrarily defined.

The fourth kind of scale, the **ratio scale**, has all the characteristics of the nominal, ordinal, and interval scales plus one other: It has a *true zero point*, which indicates a complete absence of the thing measured. On a ratio scale, zero means "none." Height, weight, and time are measured with ratio scales. Zero height, zero weight, and zero time mean that no amount of these variables is present. With a true zero point, you can make ratio statements such as 16 kilograms is four times heavier than 4 kilograms.[12] **Table 1.1** summarizes the major differences among the four scales of measurement.

Ratio scale
Measurement scale with characteristics of interval scale; also, zero means that none of the thing measured is present.

TABLE 1.1 Characteristics of the four scales of measurement

	Scale characteristics			
Scale of measurement	Different numbers for different things	Numbers convey greater than and less than	Equal differences mean equal amounts	Zero means none of what was measured was detected
Nominal	Yes	No	No	No
Ordinal	Yes	Yes	No	No
Interval	Yes	Yes	Yes	No
Ratio	Yes	Yes	Yes	Yes

[11] Convert 100°C and 50°C to Fahrenheit (F = 1.8C + 32) and suddenly the "twice as much" relationship disappears.

[12] Convert 16 kilograms and 4 kilograms to pounds (1 kg = 2.2 lbs) and the "four times heavier" relationship is maintained.

Knowing the distinctions among the four scales of measurement will help you in two tasks in this course. The kind of *descriptive statistics* you can compute from numbers depends, in part, on the scale of measurement the numbers represent. For example, it is senseless to compute a mean of numbers on a nominal scale. Calculating a mean Social Security number, a mean telephone number, or a mean psychological diagnosis is either a joke or evidence of misunderstanding numbers.

Understanding scales of measurement is sometimes important in choosing the kind of *inferential statistic* that is appropriate for a set of data. If the dependent variable (see next section) is a nominal variable, then a chi square analysis is appropriate (Chapter 14). If the dependent variable is a set of ranks (ordinal data), then a nonparametric statistic is required (Chapter 15). Most of the data analyzed with the techniques described in Chapters 7–13 are interval and ratio scale data.

The topic of scales of measurement is controversial among statisticians. Part of the controversy involves viewpoints about the underlying thing you are interested in and the number that represents the thing (Wuensch, 2005). In addition, it is sometimes difficult to classify some of the variables used in the social and behavioral sciences. Often they appear to fall between the ordinal scale and the interval scale. For example, a score may provide more information than simply rank, but equal intervals cannot be proven. Examples include aptitude and ability tests, personality measures, and intelligence tests. In such cases, researchers generally treat the scores as if they were interval scale data.

Statistics and Experimental Design

Here is a story that will help you distinguish between statistics (applying straight logic) and experimental design (observing what actually happens). This is an excerpt from a delightful book by E. B. White, *The Trumpet of the Swan* (1970, pp. 63–64).

> The fifth-graders were having a lesson in arithmetic, and their teacher, Miss Annie Snug, greeted Sam with a question.
>
> "Sam, if a man can walk three miles in one hour, how many miles can he walk in four hours?" "It would depend on how tired he got after the first hour," replied Sam. The other pupils roared. Miss Snug rapped for order.
>
> "Sam is quite right," she said. "I never looked at the problem that way before. I always supposed that man could walk twelve miles in four hours, but Sam may be right: that man may not feel so spunky after the first hour. He may drag his feet. He may slow up."
>
> Albert Bigelow raised his hand. "My father knew a man who tried to walk twelve miles, and he died of heart failure," said Albert.
>
> "Goodness!" said the teacher. "I suppose *that* could happen, too."

"Anything can happen in four hours," said Sam. "A man might develop a blister on his heel. Or he might find some berries growing along the road and stop to pick them. That would slow him up even if he wasn't tired or didn't have a blister."

"It would indeed," agreed the teacher. "Well, children, I think we have all learned a great deal about arithmetic this morning, thanks to Sam Beaver."

Everyone had learned how careful you have to be when dealing with figures.

Statistics involves the manipulation of numbers and the conclusions based on those manipulations (Miss Snug). Experimental design (also called research methods) deals with all the things that influence the numbers you get (Sam and Albert). **Figure 1.2** illustrates these two approaches to getting an answer. This text could have been a "pure" statistics book, from which you would learn to analyze numbers without knowing where they came from or what they referred to. You would learn about statistics, but such a book would be dull, dull, dull. On the other hand, to describe procedures for collecting numbers is to teach experimental design—and this book is for a statistics course. My solution to this conflict is generally to side with Miss Snug but to include some aspects of experimental design throughout the book. Knowing experimental design issues is especially important when it comes time to interpret a statistical analysis. Here's a start on experimental design.

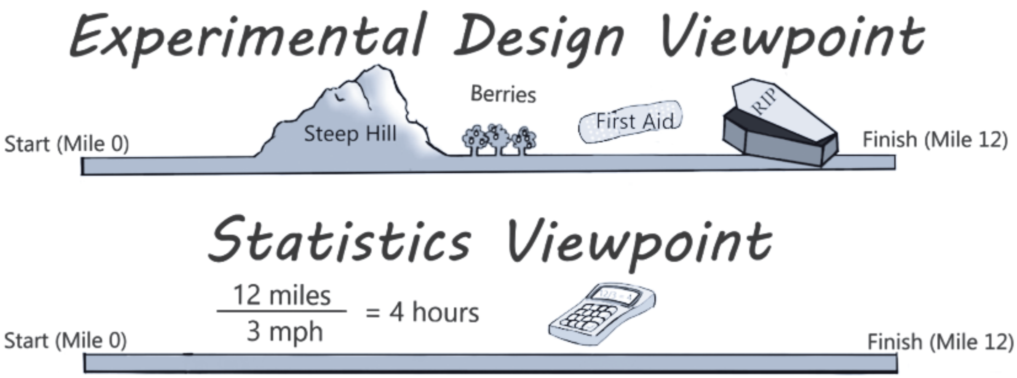


FIGURE 1.2 **Travel time from an experimental design viewpoint and a statistical viewpoint**

Experimental Design Variables

The overall task of an experimenter is to discover relationships among variables. Variables are things that vary, and researchers have studied personality, health, gender, anger, caffeine, memory, beliefs, age, skill. . . . (I'm sure you get the picture—almost anything can be a variable.)

Independent and Dependent Variables

Independent variable
Variable controlled by the researcher; changes in this variable may produce changes in the dependent variable.

Dependent variable
Observed variable that is expected to change as a result of changes in the independent variable in an experiment.

Level
One value of the independent variable.

Treatment
One value (or level) of the independent variable.

A simple experiment has two major variables, the **independent variable** and the **dependent variable**. In the simplest experiment, the researcher selects two values of the independent variable for investigation. Values of the independent variable are usually called **levels** and sometimes called **treatments**.

The basic idea is that the researcher finds or creates two groups of participants that are similar except for the independent variable. These individuals are measured on the dependent variable. The question is whether the data will allow the experimenter to claim that the values on the dependent variable *depend* on the level of the independent variable.

The values of the dependent variable are found by *measuring* or *observing* participants in the investigation. The dependent variable might be scores on a personality test, number of items remembered, or whether or not a passerby offered assistance. For the independent variable, the two groups might have been selected because they were already different—in age, gender, personality, and so forth. Alternatively, the experimenter might have produced the difference in the two groups by an experimental manipulation such as creating different amounts of anxiety or providing different levels of practice.

An example might help. Suppose for a moment that as a budding gourmet cook you want to improve your spaghetti sauce. One of your buddies suggests adding marjoram. To investigate, you serve spaghetti sauce at two different gatherings. For one group of guests, the sauce is spiced with marjoram; for the other it is not. At both gatherings, you count the number of favorable comments about the spaghetti sauce. *Stop reading; identify the independent and the dependent variables.*

The dependent variable is the number of favorable comments, which is a measure of the taste of the sauce. The independent variable is marjoram, which has two levels: present and absent.

Extraneous Variables

Extraneous variable
Variable other than the independent variable that may affect the dependent variable.

One of the pitfalls of experiments is that every situation has other variables besides the independent variable that might possibly be responsible for the changes in the dependent variable. These other variables are called **extraneous variables**. In the story, Sam and Albert noted several extraneous variables that could influence the time to walk 12 miles.

Are there any extraneous variables in the spaghetti sauce example? Oh yes, there are many, and just one is enough to raise suspicion about a conclusion that relates the taste of spaghetti sauce to marjoram. Extraneous variables include the amount and quality of the other ingredients in the sauce, the spaghetti itself, the "party moods" of the two groups, and how hungry everyone was. If any of these extraneous variables was actually operating, it weakens the claim that a difference in the comments about the sauce is the result of the presence or absence of marjoram.

The simplest way to remove an extraneous variable is to be sure that all participants are equal on that variable. For example, you can ensure that the sauces are the same except for marjoram by mixing up the ingredients, dividing it into two batches, adding marjoram to one batch but not the other, and then cooking. The "party moods" variable can be controlled (equalized) by conducting the taste test in a laboratory. Controlling extraneous variables is a complex topic covered in courses that focus on research methods and experimental design.

In many experiments, it is impossible or impractical to control all the extraneous variables. Sometimes researchers think they have controlled them all, only to find that they did not. The effect of an uncontrolled extraneous variable is to prevent a simple cause-and-effect conclusion. Even so, if the dependent variable changes when the independent variable changes, something is going on. In this case, researchers can say that the two variables are *related*, but that other variables may play a part, too.

At this point, you can test your understanding by engaging yourself with these questions: What were the independent and dependent variables in the Zeigarnik experiment? How many levels of the independent variable were there?[13]

How well did Zeigarnik control extraneous variables? For one thing, each participant was tested at both levels of the independent variable. That is, the recall of each participant was measured for interrupted tasks and for completed tasks. One advantage of this technique is that it naturally controls many extraneous variables. Thus, extraneous variables such as age and motivation were exactly the same for tasks that were interrupted as for tasks that were not because the same people contributed scores to both levels.

At various places in the following chapters, I will explain experiments and the statistical analyses using the terms *independent* and *dependent variables*. These explanations usually assume that all extraneous variables were controlled; that is, you may assume that the experimenter knew how to design the experiment so that changes in the dependent variable could be attributed correctly to changes in the independent variable. However, I present a few investigations (like the spaghetti sauce example) that I hope you recognize as being so poorly designed that conclusions cannot be drawn about the relationship between the independent variable and dependent variable. Be alert.

---

[13] Try for answers. Then, if need be, here's a hint: First, identify the dependent variable; for the dependent variable, you don't know values until data are gathered. Next, identify the independent variable; you can tell what the values of the independent variable are just from the description of the design.

Here's my summary of the relationship between statistics and experimental design. Researchers suspect that there is a relationship between two variables. They design and conduct an experiment; that is, they choose the levels of the independent variable (treatments), control the extraneous variables, and then measure the participants on the dependent variable. The measurements (data) are analyzed using statistical procedures. Finally, the researcher tells a story that is consistent with the results obtained and the procedures used.

Statistics and Philosophy

The two previous sections directed your attention to the relationship between statistics and experimental design; this section will direct your thoughts to the place of statistics in the grand scheme of things.

Epistemology
The study or theory of the nature of knowledge.

Explaining the grand scheme of things is the task of philosophy and, over the years, many schemes have been advanced. For a scheme to be considered a grand one, it has to address **epistemology**—that is, to propose answers to the question: How do we acquire knowledge?

Both *reason* and *experience* have been popular answers among philosophers.[14] For those who emphasize the importance of reason, mathematics has been a continuing source of inspiration. Classical mathematics starts with axioms that are assumed to be true. Theorems are thought up and are then proved by giving axioms as reasons. Once a theorem is proved, it can be used as a reason in a proof of other theorems.

Statistics has its foundations in mathematics and thus, a statistical analysis is based on reason. As you go about the task of calculating $\overline{X}$ or $\hat{s}$, finding confidence intervals, and telling the story of what they mean, know deep down that you are engaged in logical reasoning. (Experimental design is more complex; it includes experience and observation as well as reasoning.)

In the 19th century, science concentrated on observation and description. The variation that always accompanied a set of observations was thought to be due to imprecise observing, imprecise instruments, or failure of nature to "hit the mark." During the 20th century, however, statistical methods such as NHST revolutionized the philosophy of science by focusing on the variation that was always present in data (Salsburg, 2001; Gould, 1996). Focusing on variation allowed changes in the data to be associated with particular causes.

As the 21st century approached, flaws in the logic of NHST statistics began to be recognized. In addition to logical flaws, the practices of researchers and journal editors (such as requiring a statistical analysis to show that $p < .05$) came under scrutiny. This concern with how science is conducted has led to changes in how data are analyzed and how information is shared. The practice of statistics is in transition.

[14] In philosophy, those who emphasize reason are rationalists and those who emphasize experience are empiricists.

Let's move from formal descriptions of philosophy to a more informal one. A very common task of most human beings can be described as *trying to understand*. Statistics has helped many in their search for better understanding, and it is such people who have recommended (or demanded) that statistics be taught in school. A reasonable expectation is that you, too, will find statistics useful in your future efforts to understand and persuade.

Speaking of persuasion, you have probably heard it said, "You can prove anything with statistics." The implied message is that a conclusion based on statistics is suspect because statistical methods are unreliable. Well, it just isn't true that statistical methods are unreliable, but it is true that people can misuse statistics (just as any tool can be misused). One of the great advantages of studying statistics is that you get better at recognizing statistics that are used improperly.

Statistics: Then and Now

Statistics began with counting, which, of course, was prehistory. The origin of the mean is almost as obscure. It was in use by the early 1700s, but no one is credited with its discovery. Graphs, however, began when J. H. Lambert, a Swiss-German scientist and mathematician, and William Playfair, an English political economist, invented and improved graphs in the period 1765 to 1800 (Tufte, 2001).

The Royal Statistical Society was established in 1834 by a group of Englishmen in London. Just 5 years later, on November 27, 1839, at 15 Cornhill in Boston, a group of Americans founded the American Statistical Society. Less than 3 months later, for a reason that you can probably figure out, the group changed its name to the American Statistical Association, which continues today (*www.amstat.org*).

According to Walker (1929), the first university course in statistics in the United States was probably "Social Science and Statistics," taught at Columbia University in 1880. The professor was a political scientist, and the course was offered in the economics department. In 1887, at the University of Pennsylvania, separate courses in statistics were offered by the departments of psychology and economics. By 1891, Clark University, the University of Michigan, and Yale had been added to the list of schools that taught statistics, and anthropology had been added to the list of departments. Biology was added in 1899 (Harvard) and education in 1900 (Columbia).

You might be interested in when statistics was first taught at your school and in what department. College catalogs are probably your most accessible source of information.

This course provides you with the opportunity to improve your ability to understand and use statistics. Kirk (2008) identifies four levels of statistical sophistication:

Category 1—those who understand statistical presentations
Category 2—those who understand, select, and apply statistical procedures
Category 3—applied statisticians who help others use statistics
Category 4—mathematical statisticians who develop new statistical techniques and discover new characteristics of old techniques

I hope that by the end of your statistics course, you will be well along the path to becoming a Category 2 statistician.

How to Analyze a Data Set

The end point of analyzing a data set is a story that explains the relationships among the variables in the data set. I recommend that you analyze a data set in three steps. The first step is exploratory. Read all the information and examine the data. Calculate descriptive statistics and focus on the differences that are revealed. In this textbook, descriptive statistics are emphasized in Chapters 2 through 6 and include graphs, means, and effect size indexes. Calculating descriptive statistics helps you develop preliminary ideas for your story (Step 3). The second step is to answer the question, What are the effects that chance could have on the descriptive statistics I calculated? An answer requires inferential statistics (Chapter 7 through Chapter 15). The third step is to write the story the data reveal. Incorporate the descriptive and inferential statistics to support the conclusions in the story. Of course, the skills you've learned and taught yourself about composition will be helpful as you compose and write your story. Don't worry about length; most good statistical stories about simple data sets can be told in one paragraph.

Write your story using journal style, which is quite different from textbook style. Textbook style, at least this textbook, is chatty, redundant, and laced with footnotes.[15] Journal style, on the other hand, is terse, formal, and devoid of footnotes. Paragraphs labeled Interpretation in Appendix G, this textbook's answer section, are examples of journal style. And for guidance in writing up an entire study, see Appelbaum et al. (2018).

Helpful Features of This Book

At various points in this chapter, I encouraged your engagement in statistics. Your active participation is *necessary* if you are to learn statistics. For my part, I worked to organize this book and write it in a way that encourages active participation. Here are some of the features you should find helpful.

Objectives

Each chapter begins with a list of skills the chapter is designed to help you acquire. Read this list of objectives first to find out what you are to learn to do. Then thumb through the chapter and read the headings. Next, study the chapter, working all the problems *as you come to them*. Finally, reread the objectives. Can you meet each one? If so, put a check mark beside that objective.

---

[15] You are reading the footnotes, aren't you? Your answer — "Well, yes, it seems I am."

Problems and Answers

The problems in this text are in small groups within the chapter rather than clumped together at the end. This encourages you to read a little and work problems, followed by more reading and problems. Psychologists call this pattern spaced practice. Spaced practice patterns lead to better performance than massed practice patterns. The problems come from a variety of disciplines; the answers are in Appendix G.

Some problems are conceptual and do not require any arithmetic. Think these through and write your answers. Being able to calculate a statistic is almost worthless if you cannot explain in English what it means. Writing reveals how thoroughly you understand. To emphasize the importance of explanations, I highlited **Interpretation** in the answers in Appendix G. On occasion, problems or data sets are used again, either later in that chapter or in another. If you do not work the problem when it is first presented, you are likely to be frustrated when it appears again. To alert you, I have put an asterisk (*) beside problems that are used again.

At the end of many chapters, comprehensive problems are marked with a CP Working these problems requires knowing most of the material in the chapter. For most students, it is best to work all the problems, but be sure you can work those marked with a CP

clue to the future

Often a concept is presented that will be used again in later chapters. These ideas are separated from the rest of the text in a box labeled "Clue to the Future." You have already seen two of these "Clues" in this chapter. Attention to these concepts will pay dividends later in the course.

error detection

I have boxed in, at various points in the book, ways to detect errors. Some of these "Error Detection" tips will also help you better understand the concept. Because many of these checks can be made early, they can prevent the frustrating experience of getting an impossible answer when the error could have been caught in Step 2.

Figure and Table References

Sometimes the words Figure and Table are in boldface print. This means that you should examine the figure or table at that point. Afterward, it will be easy for you to return to your place in the text—just find the boldface type.

Transition Passages

At six places in this book, there are major differences between the material you just finished and the material in the next section. "Transition Passages," which describe the differences, separate these sections.

Glossaries

This book has three separate glossaries of words, symbols, and formulas.

1. *Words*. The first time an important word is used in the text, it appears in boldface type accompanied by a definition in the margin. In later chapters, the word may be boldfaced again, but margin definitions are not repeated. Appendix D is a complete glossary of words (p. 401). I suggest you mark this appendix.
2. *Symbols*. Statistical symbols are defined in Appendix E (p. 405). Mark it too.
3. *Formulas*. Formulas for all the statistical techniques used in the text are printed in Appendix F (p. 407), in alphabetical order according to the name of the technique.

Computers, Calculators, and Pencils

Computer software, calculators, and pencils with erasers are all tools used at one time or another by statisticians. Any or all of these devices may be part of the course you are taking. Regardless of the calculating aids that you use, however, your task is the same:

- Read a problem.
- Decide what statistical procedure to use.
- Apply that procedure using the tools available to you.
- Write an interpretation of the results.

Pencils, calculators, and software represent, in ascending order, tools that are more and more error-free. People who routinely use statistics routinely use computers. You may or may not use one at this point. Remember, though, whether you are using a software program or not, your principal task is to understand and describe.

For many of the worked examples in this book, I included the output of a popular statistical software program, IBM SPSS.[16] If your course includes IBM SPSS, these tables should help familiarize you with the program.

[16] The original name of the program was Statistical Package for the Social Sciences.

Concluding Thoughts

The first leg of your exploration of statistics is complete. Some of the ideas along the path are familiar and perhaps a few are new or newly engaging. As your course progresses, you will come to understand what is going on in many statistical analyses, and you will learn paths to follow if you analyze data that you collect yourself.

This book is a fairly complete introduction to elementary statistics. Of course, there is lots more to statistics, but there is a limit to what you can do in one term. Even so, exploration of paths not covered in a textbook can be fun. Encyclopedias, both general and specialized, often reward such exploration. Try the *Encyclopedia of Statistics in Behavioral Sciences or the International Encyclopedia of the Social and Behavioral Sciences*. Also, when you finish this course (but before any final examination), I recommend Chapter 16, the last chapter in this book. It is an overview/integrative chapter.

Most students find that this book works well as a textbook in their statistics course. I recommend that you keep the book after the course is over to use as a *reference* book. In courses that follow statistics and even after leaving school, many find themselves looking up a definition or reviewing a procedure.[17] Thus, a familiar textbook becomes a valued guidebook that serves for years into the future. For me, exploring statistics and using them to understand the world is quite satisfying. I hope you have a similar experience.

PROBLEMS

1.7. Name the four scales of measurement identified by S. S. Stevens.

1.8. Give the properties of each of the scales of measurement.

1.9. Identify the scale of measurement in each of the following cases.

- **a.** Geologists have a "hardness scale" for identifying different rocks, called Mohs' scale. The hardest rock (diamond) has a value of 10 and will scratch all others. The second hardest will scratch all but the diamond, and so on. Talc, with a value of 1, can be scratched by every other rock. (A fingernail, a truly handy field-test instrument, has a value between 2 and 3.)
- **b.** The volumes of three different cubes are 40, 64, and 65 cubic inches.
- **c.** Three different highways are identified by their numbers: 40, 64, and 65.
- **d.** Republicans, Democrats, Independents, and Others are identified on the voters' list with the numbers 1, 2, 3, and 4, respectively.
- **e.** The winner of the Miss America contest was Miss New York; the two runners-up were Miss Ohio and Miss California.[18]
- **f.** The prices of the three items are $3.00, $10.00, and $12.00.
- **g.** She earned three degrees: B.A., M.S., and Ph.D.

---

[17] This book's index is unusually extensive. If you make margin notes, they will help too.

[18] Contest winners have come most frequently from these states, which have had six winners each.

1.10. Undergraduate students conducted the three studies that follow. For each study, identify the dependent variable, the independent variable, the number of levels of the independent variable, and the names of the levels of the independent variable.

a. Becca had students in a statistics class rate a résumé, telling them that the person had applied for a position that included teaching statistics at their college. The students rated the résumé on a scale of 1 (not qualified) to 10 (extremely qualified). All the students received identical résumés, except that the candidate's first name was Jane on half the résumés and John on the other half.

b. Michael's participants filled out the Selfism scale, which measures narcissism. (Narcissism is neurotic self-love.) In addition, students were classified as first-born, second-born, and later-born.

c. Johanna had participants read a description of a crime and "Mr. Anderson," the person convicted of the crime. For some participants, Mr. Anderson was described as a janitor. For others, he was described as a vice president of a large corporation. For still others, no occupation was given. After reading the description, participants recommended a jail sentence (in months) for Mr. Anderson.

1.11. Researchers who are now well known conducted the three classic studies that follow. For each study, identify the dependent variable, the independent variable, and the number and names of the levels of the independent variable. Complete items i and ii.

a. Theodore Barber hypnotized 25 people, giving each a series of suggestions. The suggestions included arm rigidity, hallucinations, color blindness, and enhanced memory. Barber counted the number of suggestions the participants complied with (the mean was 4.8). For another 25 people, he simply asked them to achieve the best score they could (but no hypnosis was used). This second group was given the same suggestions, and the number complied with was counted (the mean was 5.1). (See Barber, 1976.)

i. Identify a nominal variable and a statistic.

ii. In a sentence, describe what Barber's study shows.

b. Elizabeth Loftus had participants view a film clip of a car accident. Afterward, some were asked, "How fast was the car going?" and others were asked, "How fast was the car going when it passed the barn?" (There was no barn in the film.) A week later, Loftus asked the participants, "Did you see a barn?" If the barn had been mentioned earlier, 17% said yes; if it had not been mentioned, 3% said yes. (See Loftus, 1979.)

i. Identify a population and a parameter.

ii. In a sentence, describe what Loftus's study shows.

c. Stanley Schachter and Larry Gross gathered data from obese male students for about an hour in the afternoon. At the end of this time, a clock on the wall was correct (5:30 p.m.) for 20 participants, slow (5:00 p.m.) for 20 others, and fast (6:00 p.m.) for 20 more. The actual time, 5:30, was the usual dinnertime for these students. While participants filled out a final questionnaire, Wheat Thins® were freely available. The weight of the crackers each student consumed was measured. The means were as follows: 5:00 group—20 grams; 5:30 group—30 grams; 6:00 group—40 grams. (See Schachter and Gross, 1968.)

i. Identify a ratio scale variable.

ii. In a sentence, describe what this study shows.

1.12. There are uncontrolled extraneous variables in the study described here. Name as many as you can. Begin by identifying the dependent and independent variables.

An investigator concluded that Textbook A was better than Textbook B after comparing the exam scores of two statistics classes. One class met MWF at 10:00 a.m. for 50 minutes, used Textbook A, and was taught by Dr. X. The other class met for 2.5 hours on Wednesday evening, used Textbook B, and was taught by Dr. Y. All students took the same comprehensive test at the end of the term. The mean score for Textbook A students was higher than the mean score for Textbook B students.

1.13. In philosophy, the study of the nature of knowledge is called _____.

1.14. a. The two approaches to epistemology identified in the text are _____ and _____.

b. Statistics has its roots in _____.

1.15. Your textbook recommends a three-step approach to analyzing a data set. Summarize the steps.

1.16. Read the objectives at the beginning of this chapter. Responding to them will help you consolidate what you have learned.

KEY TERMS

Categorical variable (p. 12)
Continuous variable (p. 11)
Dependent variable (p. 18)
Descriptive statistics (p. 6)
Discrete variable (p. 11)
Epistemology (p. 20)
Extraneous variable (p. 18)
Independent variable (p. 18)
Inferential statistics (p. 6)
Interval scale (p. 15)
Level (p. 18)
Lower limit (p. 11)
Mean (p. 7)
Nominal scale (p. 14)
Ordinal scale (p. 14)
Parameter (p. 10)
Population (p. 9)
Qualitative variable (p. 12)
Quantitative variable (p. 11)
Ratio scale (p. 15)
Sample (p. 9)
Statistic (p. 10)
Treatment (p. 18)
Upper limit (p. 11)
Variable (p. 10)

The online *Study Guide for Exploring Statistics* (12th ed.) is available for sale at exploringstatistics.com

Transition Passage

To Descriptive Statistics

STATISTICAL TECHNIQUES ARE often categorized as descriptive statistics and inferential statistics. The next five chapters are about descriptive statistics. You are already familiar with some of these descriptive statistics, such as the mean, range, and bar graphs. Others may be less familiar—the correlation coefficient, effect size index, and boxplot. All of these and others that you study will be helpful in your efforts to understand data.

The phrase *Exploring Data* appears in three of the chapter titles that follow. This phrase is a reminder to approach a data set with the attitude of an explorer, an attitude of *What can I find here?* Descriptive statistics are especially valuable in the early stages of an analysis as you explore what the data have to say. Later, descriptive statistics are essential when you convey your story of the data to others. In addition, many descriptive statistics have important roles in the inferential statistical techniques that are covered in later chapters. Let's get started.

CHAPTER 2

Exploring Data: Frequency Distributions and Graphs

OBJECTIVES FOR CHAPTER 2

After studying the text and working the problems in this chapter, you should be able to:

1. Arrange a set of scores into a simple frequency distribution
2. Arrange a set of scores into a grouped frequency distribution if you are given the class intervals
3. Describe the characteristics of frequency polygons, histograms, and bar graphs and explain the information each one provides
4. Name certain distributions by looking at their shapes
5. Describe the characteristics of a line graph
6. Comment on the recent use of graphics

YOU HAVE PROBABLY heard the expression *the data speak* for themselves. The implication is that the meaning of the data is plainly there in the numbers. This implication is mistaken. It is the researcher's explanation that speaks for the data. After that, you and I and others give our interpretation. We draw conclusions and then cite our interpretation as support. Because many decisions are based on data, studying statistics puts you in a much better position to interpret data, draw appropriate conclusions, and decide what to do.

The chapters that follow immediately are about descriptive statistics. To get an overview, carefully read this chapter title and those of the next four chapters.

You'll begin your study of descriptive statistics with a group of raw scores. **Raw scores**, or raw data, are available from many sources. As one example, if you administer a questionnaire to a group of college students, the scores from the questionnaire are raw scores. I will illustrate several of the concepts in this chapter and the next three with raw scores of 100 representative college students.

Raw Score
Score obtained by observation or from an experiment.

If you would like to engage fully in this chapter, take 3 minutes to read, fill out the five-item questionnaire in **Figure 2.1**, and calculate your score. Do this now, before you read further. (If you read the five chapter titles, I'd bet that you answered the five questions. Engagement always helps.)

Instructions: Five statements follow. Use the scale below (1-7) to indicate your level of agreement or disagreement with each statement. Please be open and honest in your response.

1 = Strongly disagree	5 = Slightly agree
2 = Disagree	6 = Agree
3 = Slightly disagree	7 = Strongly agree
4 = Neither agree nor disagree	

_____ 1. In most ways my life is close to my ideal.
_____ 2. The conditions in my life are excellent.
_____ 3. I am satisfied with my life.
_____ 4. So far I have gotten the important things I want in life.
_____ 5. If I could live my life over, I would change almost nothing.

Scoring instructions: Add the numbers in the blanks to determine your score.

FIGURE 2.1 Questionnaire

Subjective well-being is a person's evaluation of his or her life. This increasingly hot topic has implications beyond personal well-being. For example, Diener and Seligman (2004) marshaled evidence that governments would be better served to focus on policies that increase subjective well-being rather than continuing with their traditional concern, which is policies of economic well-being.

For a person, subjective well-being has emotional components and cognitive components. An important cognitive component is satisfaction with life. To measure this cognitive component, Ed Diener and his colleagues (Diener et al., 1985) developed the Satisfaction With Life Scale (SWLS). The SWLS is the short, five-item questionnaire in Figure 2.1. It is a reliable, valid measure of a person's global satisfaction with life. **Table 2.1** is an unorganized collection of SWLS scores collected from 100 representative college students.

TABLE 2.1 Representative scores of 100 college students on the Satisfaction With Life Scale

15	31	22	26	19	27	33	24	27	25
20	26	29	32	21	13	35	9	25	25
28	23	17	27	30	16	11	29	26	20
23	30	16	24	27	29	26	10	23	34
5	19	28	29	27	30	32	22	17	13
35	28	27	25	26	25	23	21	29	28
20	27	30	22	22	12	32	25	24	23
20	24	26	26	29	33	29	24	20	25
19	25	9	21	32	30	27	24	10	5
22	26	26	28	23	27	25	28	27	31

Simple Frequency Distributions

Table 2.1 is just a jumble of numbers, not very interesting or informative. (My guess is that you glanced at it and went quickly on.) A much more informative presentation of the 100 scores is an arrangement called a **simple frequency distribution**. Table 2.2 is a simple frequency distribution—an ordered arrangement that shows the frequency of each score. Look at **Table 2.2**. (I would guess that you spent more time on Table 2.2 than on Table 2.1 and that you got more information from it.)

Simple frequency distribution
Scores arranged from highest to lowest, each with its frequency of occurrence.

TABLE 2.2 Rough draft of a simple frequency distribution of Satisfaction With Life Scale scores for a representative sample of 100 college students

SWLS score (X)	Tally marks	Frequency (f)	SWLS score (X)	Tally marks	Frequency (f)
35	II	2	19	III	3
34	I	1	18		0
33	II	2	17	II	2
32	IIII	4	16	II	2
31	II	2	15	I	1
30	~~IIII~~	5	14		0
29	~~IIII~~ II	7	13	II	2
28	~~IIII~~ I	6	12	I	1
27	~~IIII~~ ~~IIII~~	10	11	I	1
26	~~IIII~~ IIII	9	10	II	2
25	~~IIII~~ IIII	9	9	II	2
24	~~IIII~~ I	6	8		0
23	~~IIII~~ I	6	7		0
22	~~IIII~~	5	6		0
21	III	3	5	II	2
20	~~IIII~~	5			N = 100
	(continued above)				

The SWLS score column in Table 2.2 tells the name of the variable that is being measured. The generic name for all variables is *X*, which is the symbol used in formulas. The Frequency *(f)* column shows how frequently a score occurred. The tally marks are used when you construct a rough-draft version of a table and are not usually included in the final form. *N* is the number of scores and is found by summing the numbers in the *f* column. You will have a lot of opportunities to construct simple frequency distributions, so here are the steps. In each step, the first sentence is a general instruction; the italicized sentence refers to the data in Table 2.1.

1. Find the highest and lowest scores. *Highest score is 35; lowest is 5*.
2. In column form, write in descending order all the numbers. *35 to 5*.
3. At the top of the column, name the variable being measured. *Satisfaction With Life Scale scores*.

4. Start with the number in the upper left-hand corner of the scores, draw a line under it, and place a tally mark beside that number in the column of numbers. *Underline 15 in Table 2.1, and place a tally mark beside 15 in Table 2.2*.
5. Continue underlining and tallying for all the unorganized scores.
6. Add a column labeled *f* (frequency).
7. Count the number of tallies by each score and enter the count in the *f* column. *2, 1, 2, 4, . . . , 0, 2.*
8. Add the numbers in the *f* column. If the sum is equal to *N*, you haven't left out any scores. *Sum = 100*.

error detection

Underlining numbers is much better than crossing them out. Later, when you check your work or do additional analyses, numbers with lines through them can be difficult to read.

A simple frequency distribution is a useful way to explore a set of data because you pick up valuable information from just a glance. For example, the highest and lowest scores are readily apparent in any frequency distribution. In addition, after some practice with frequency distributions, you can ascertain the general shape of the distribution and make informed guesses about central tendency and variability, which are explained in Chapter 3 and Chapter 4.

Table 2.3 shows a formal presentation of the data from Table 2.1. Formal presentations are used to present data to colleagues, professors, supervisors, editors, and others. Formal presentations do not include tally marks and often do not include zero-frequency scores.

TABLE 2.3 Formal simple frequency distribution of Satisfaction With Life Scale scores for a representative sample of 100 college students

SWLS score (X)	Frequency (f)	SWLS score (X)	Frequency (f)
35	2	22	5
34	1	21	3
33	2	20	5
32	4	19	3
31	2	17	2
30	5	16	2
29	7	15	1
28	6	13	2
27	10	12	1
26	9	11	1
25	9	10	2
24	6	9	2
23	6	5	2
			N = 100
(continued above)			

Grouped Frequency Distributions

Many researchers would condense Table 2.3 even more. The result is a **grouped frequency distribution. Table 2.4** is a rough-draft example of such a distribution.[1] Raw data are usually condensed into a grouped frequency distribution when researchers want to present the data as a graph or as a table.

Grouped frequency distribution
Scores compiled into intervals of equal size. Includes the frequency of scores in each interval.

Class interval
A range of scores in a grouped frequency distribution.

In a grouped frequency distribution, scores are grouped into equal-sized ranges called **class intervals**. In Table 2.4, the entire range of scores, from 35 to 5, is reduced to 11 class intervals. Each interval covers three scores; the symbol *i* indicates the size of the interval. In Table 2.4, $i = 3$. The midpoint of the interval is noteworthy because it represents all the scores in that interval. For example, five students had scores of 15, 16, or 17. The midpoint of the class interval 15–17 is 16.[2] The midpoint, 16, represents all five scores. There are no scores in the interval 6–8, but zero-frequency intervals are included in formal grouped frequency distributions if they are within the range of the distribution.

TABLE 2.4 Rough draft of a grouped frequency distribution of Satisfaction With Life Scale scores (i = 3)

SWLS scores (class interval)	Midpoint (X)	Tally marks	(f)
33-35	34	~~IIII~~	5
30-32	31	~~IIII~~ ~~IIII~~ I	11
27-29	28	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ III	23
24-26	25	~~IIII~~ ~~IIII~~ ~~IIII~~ ~~IIII~~ IIII	24
21-23	22	~~IIII~~ ~~IIII~~ IIII	14
18-20	19	~~IIII~~ III	8
15-17	16	~~IIII~~	5
12-14	13	III	3
9-11	10	~~IIII~~	5
6-8	7		0
3-5	4	II	2
			$N = 100$

Class intervals have lower and upper limits, much like simple scores obtained by measuring a continuous variable. A class interval of 15–17 has a lower limit of 14.5 and an upper limit of 17.5.

The only difference between grouped frequency distributions and simple frequency distributions is class intervals. The details of establishing class intervals are described in Appendix B. For problems in this chapter, I will give you the class intervals to use.

[1] A formal grouped frequency distribution, like a formal simple frequency distribution, does not include tally marks.
[2] When the scores are whole numbers, make *i* an odd number. With *i* odd, the midpoint of the class interval is a whole number.

PROBLEMS

***2.1.** (This is the first of many problems with an asterisk. An asterisk means the information in the problem is used in other problems later. If you collect all your answers in a notebook or file, you'll have a handy reference when a problem turns up again.)

The following numbers are the heights in inches of two groups of Americans in their 20s. Choose the group that interests you more and organize the 50 numbers into a simple frequency distribution using the rough-draft form. For the group you choose, your result will be fairly representative of the whole population of 20- to 29-year-olds (*Statistical Abstract of the United States:* 2013, 2012).

Women					*Men*				
68	67	63	65	71	72	65	72	68	70
66	66	62	65	65	69	73	71	69	67
60	72	64	61	65	77	67	72	73	70
69	64	64	66	60	73	64	72	69	69
65	67	65	62	68	70	71	71	70	75
69	65	63	64	71	72	68	62	68	74
66	64	59	63	65	66	70	72	66	75
63	63	66	64	68	69	71	68	73	69
66	67	62	62	67	71	69	69	65	76
67	69	64	63	65	71	72	65	70	70

***2.2.** The frequency distribution you will construct for this problem is a little different. The "scores" are names and thus nominal data.

A political science student traveled on every street in a voting precinct on election day morning, recording the yard signs (by initial) for the five candidates. The five candidates were Attila (A), Bolivar (B), Gandhi (G), Lenin (L), and Mao (M). Construct an appropriate frequency distribution from her observations. (She hoped to find the relationship between yard signs and actual votes.)

G A M M M G G L A G B A A G G B L M M
A G G B G L M A A L M G G M G L G A A
B L G G A G A M L M G B A G L G M A

***2.3.** You may have heard or read that the normal body temperature (oral) for humans is 98.6°F. The numbers that follow are temperature readings from healthy adults, aged 18–40. Arrange the data into a grouped frequency distribution, using 99.3–99.5 as the highest class interval and 96.3–96.5 as the lowest. (Based on Mackowiak, Wasserman, & Levine, 1992.)

98.1	97.5	97.8	96.4	96.9	98.9	99.5	98.6	98.2	98.3
97.9	98.0	97.2	99.1	98.4	98.5	97.4	98.0	97.9	98.3
98.8	99.5	98.7	97.9	97.7	99.2	98.0	98.2	98.3	97.0
99.4	98.9	97.9	97.4	97.8	98.6	98.7	97.9	98.4	98.8

***2.4.** An experimenter read 60 related statements to a class. He then asked the students to indicate which of the next 20 statements were among the first 60. Due to the relationships among the concepts in the sentences, many seemed familiar but, in fact, none of the 20 had been read before. The following scores indicate the number (out of 20) that each student had "heard earlier." (See the classic experiment by Bransford & Franks, 1971.) Arrange the scores into a simple frequency distribution. Based on this description and your frequency distribution, write a sentence of interpretation.

14	11	10	8	12	13	11	10	16	11
11	9	9	7	14	12	9	10	11	6
13	8	11	11	9	8	13	16	10	11
9	9	8	12	11	10	9	7	10	

***2.5.** This problem is based on a study by Chen, Chavez, Ong, and Gunderson (2017). Students in statistics classes identified resources that could be helpful for an upcoming test and how each would be utilized. Examples included lecture notes, textbook chapters, study partners, and instructor office hours. The numbers below represent the additional points each student earned on the 100 point exam. Organize the scores into a simple frequency distribution. Comment on the results.

3	7	0	5	3	6	1	2	4	4
9	0	7	3	1	6	9	3	6	1
1	3	4	4	3	2	2	5	5	4
2	5	3	3	2					

Graphs of Frequency Distributions

You have no doubt heard the saying *A picture is worth a thousand words*. When it comes to numbers, a not-yet-well-known saying is *A graph is better than a thousand numbers*. Actually, as long as I am rewriting sayings, I would also like to say *The more numbers you have, the more valuable graphics are*. There are many kinds of graphics, and they are becoming more and more important for data analysis and persuasion.

Abscissa
The horizontal axis of a graph; *x*-axis.

Ordinate
The vertical axis of a graph; *y*-axis.

Pictures that present statistical data are called *graphics*. The most common graphic is a graph composed of a horizontal axis (variously called the baseline, *x*-axis, or **abscissa**) and a vertical axis (called the *y*-axis, or **ordinate**). To the right and upward are both positive directions; to the left and downward are both negative directions. The axes cross at the origin. **Figure 2.2** is a picture of these words.

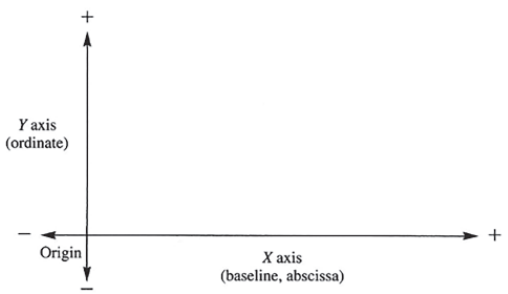


FIGURE 2.2 The elements of a graph

This section presents three kinds of graphs that are used to present frequency distributions—*frequency polygons, histograms, and bar graphs*. Frequency distribution graphs present an entire set of observations of a sample or a population. If the variable being graphed is a continuous variable, use a frequency polygon or a histogram; if the variable is categorical or discrete, use a bar graph. (Think about the SWLS scores in Table 2.4 and the yard sign data in Problem 2.2. Which categories do these variables fall in?)

Frequency Polygon

Frequency polygon
Graph of a frequency distribution of a continuous variable; points are connected by lines.

A **frequency polygon** is used to graph continuous variables. The SWLS scores in Table 2.4 are continuous data, so a frequency polygon is appropriate. The result is **Figure 2.3**. An explanation of Figure 2.3 will serve as an introduction to constructing polygons.

FIGURE 2.3 Frequency polygon of Satisfaction With Life Scale scores of 100 college students

Each point of the frequency polygon represents two numbers: the class midpoint directly below it on the *x*-axis and the frequency of that class directly across from it on the *y*-axis. By looking at the data points in **Figure 2.3**, you can see that the five students who scored 33, 34, or 35 are represented by the midpoint 34, zero students by the midpoint 7, and so on.

The name of the variable being graphed goes on the *x*-axis (Satisfaction With Life Scale scores). The *x*-axis is marked off in equal intervals, with each tick mark indicating the midpoint of a class interval. Low scores are on the left. The lowest class interval midpoint, 1, and the highest, 37, have zero frequencies. Frequency polygons are closed at both ends. In cases where the lowest score in the distribution is well above zero, it is conventional to replace numbers smaller than the lowest score on the *x*-axis with a slash mark, which indicates that the scale to the origin is not continuous. (See **Figure 4.1**, p. 73)

Histogram

A **histogram** is another graphing technique that is appropriate for continuous variables. **Figure 2.4** shows the SWLS data of Table 2.4 graphed as a histogram. A histogram is constructed by raising bars from the *x*-axis to the appropriate frequencies on the *y*-axis. The lines that separate the bars intersect the *x*-axis at the lower and upper limits of the class intervals.

Histogram
Graph of a frequency distribution of a continuous variable; frequencies indicated by contiguous vertical bars.

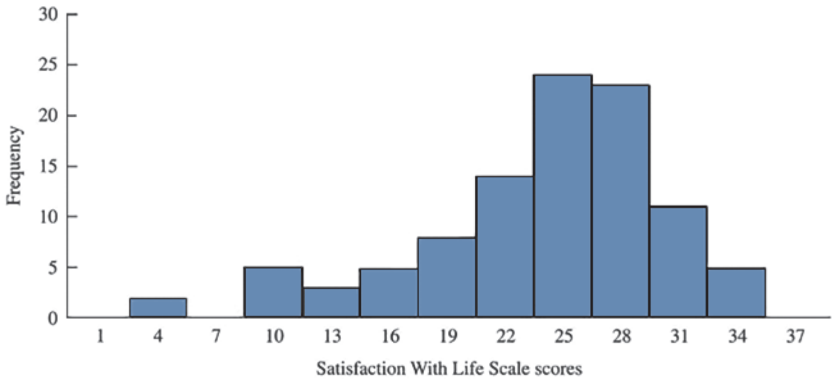


FIGURE 2.4 **Histogram of Satisfaction With Life Scale scores of 100 college students**

Here are some considerations for deciding whether to use a frequency polygon or a histogram. If you are displaying two overlapping distributions on the same axes, a frequency polygon is less cluttered and easier to comprehend than a histogram. However, it is easier to read frequencies from a histogram.

Bar Graph

A **bar graph** is used to present the frequencies of categorical variables and discrete variables. A conventional bar graph looks like a histogram except it has wider spaces between the bars. The space is a signal that the variable is not continuous. Conventionally, bar graphs have the name of the variable being graphed on the x-axis and frequency on the *y*-axis.

Bar graph
Graph of a frequency distribution of nominal or categorical data; bars are separated.

If an ordinal scale variable is being graphed, the order of the values on the *x*-axis follows the order of the variable. If, however, the variable is a nominal scale variable, then *any* order on the *x*-axis is permissible. Alphabetizing might be best. Other considerations may lead to some other order.

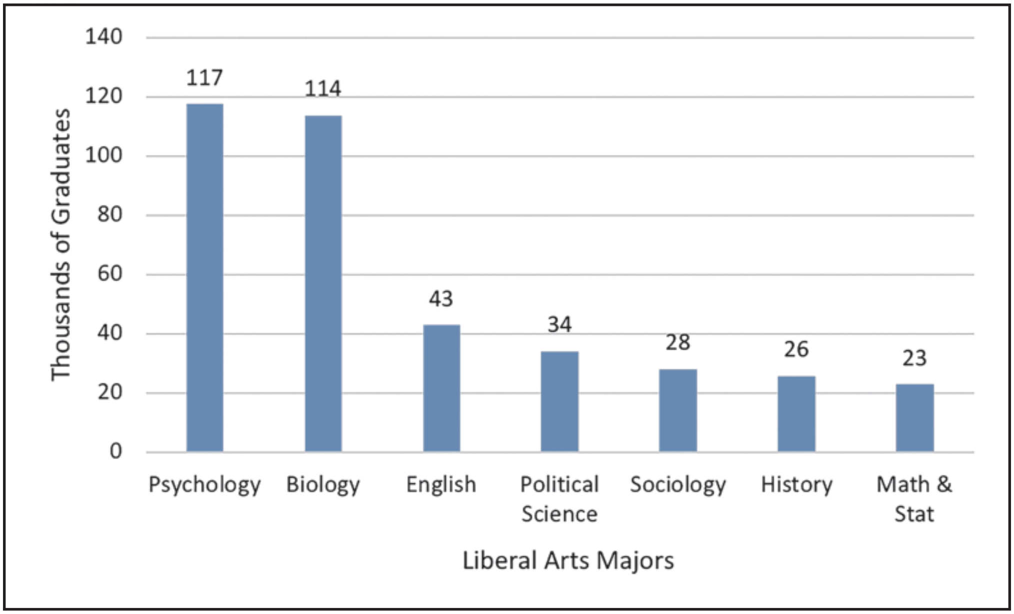


FIGURE 2.5 **Thousands of baccalaureate graduates for the seven most common liberal arts majors (academic year 2015-2016)**

Figure 2.5 is a bar graph of the seven most common liberal arts majors among U.S. baccalaureate graduates in 2015-2016. The majors are listed on the *x*-axis; the number of graduates is on the *y*-axis. I ordered the majors from most to fewest graduates, but other orders would be more appropriate in other circumstances. For example, placing English in the left-hand position might be better for a presentation focused on English majors.

PROBLEMS

2.6. For the height data in Problem 2.1, you constructed a frequency distribution. Graph it, being careful to include the zero-frequency scores.

2.7. The average work week for employees varies from country to country. In 2017, the most recently available figures for number of hours worked were Brazil, 39; France, 35; Germany, 35; Japan, 39; Turkey, 48; United Kingdom, 36; United States, 39; and Viet Nam, 47. What kind of graph is appropriate for these data? Design the graph and execute it.

***2.8**. Decide whether the following distributions should be graphed as frequency polygons or as bar graphs. Graph both distributions.

X Class interval	(*f*)
48-52	1
43-47	1
38-42	2
33-37	4
28-32	5
23-27	7
18-22	10
13-17	12
8-12	6
3-7	2

Y Class interval	(*f*)
54-56	3
51-53	7
48-50	15
45-47	14
42-44	11
39-41	8
36-38	7
33-35	4
30-32	5
27-29	2
24-26	0
21-23	0
18-20	1

2.9. Look at the frequency distribution that you constructed for the yard-sign observations in Problem 2.2. Which kind of graph should be used to display these data? Compose and graph the distribution.

Describing Distributions

Words, pictures, and mathematics are all used to describe a distribution's shape. This section uses words and pictures. One mathematical method, chi square, is covered in Chapter 14. A distribution's shape conveys the variability in the distribution and the likelihood of any particular score. For continuous variables and ordered categorical variables, a distribution's shape is meaningful. For unordered categorical distributions, however, whatever shape the distribution has is arbitrary.

Symmetrical Distributions

Symmetrical distributions have two halves that more or less mirror each other; the left half looks pretty much like the right half. In many cases, symmetrical distributions are bell-shaped; the highest frequencies are in the middle of the distribution, and scores on either side occur less and less frequently. The distribution of heights in Problem 2.6 is an example.

Normal distribution (normal curve)
A mathematically defined, theoretical distribution or a graph of observed scores with a particular shape.

There is a special case of a bell-shaped distribution that you will soon come to know very well. It is the **normal distribution** (or **normal curve**).

The left panel in **Figure 2.6** is an illustration of a normal distribution.

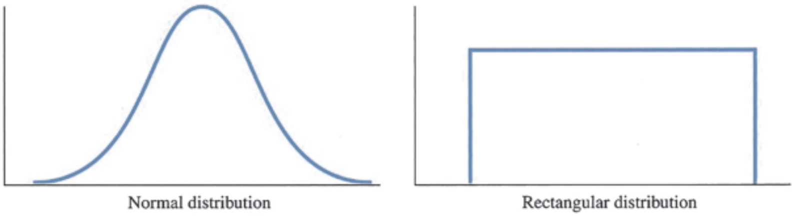


FIGURE 2.6 **A normal distribution and a rectangular distribution**

A rectangular distribution (also called a *uniform distribution*) is a symmetrical distribution that occurs when the frequency of each value on the *x*-axis is the same. The right panel of **Figure 2.6** is an example. You will see this distribution again in Chapter 7.

Skewed Distributions

Skewed distribution
Asymmetrical distribution; may be positive or negative.

Positive skew
Graph with a great preponderance of low scores.

Negative skew
Graph with a great preponderance of high scores.

In some distributions, the scores that occur most frequently are near one end of the scale, which leaves few scores at the other end. Such distributions are **skewed**. Skewed distributions, like a skewer, have one end that is thin and narrow. On graphs, if the thin point is to the right—the positive direction—the curve has a **positive skew**. If the thin point is to the left, the curve is **negatively skewed. Figure 2.7** shows a negatively skewed curve on the left and a positively skewed curve on the right. The data in Table 2.4 are negatively skewed.

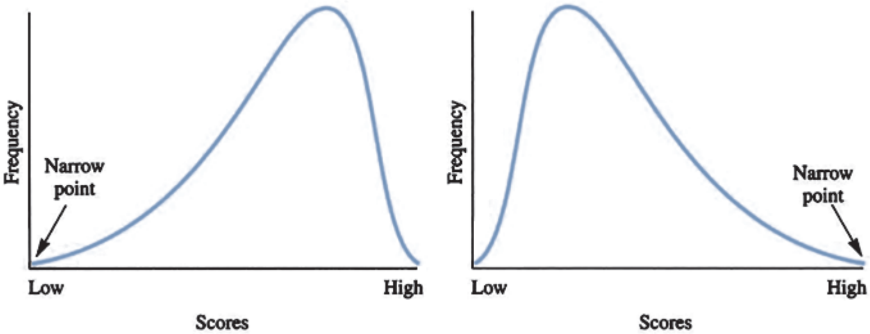


FIGURE 2.7 **A negatively skewed curve (left) and a positively skewed curve (right)**

Bimodal Distributions

A graph with two distinct humps is called a **bimodal distribution.**[3] Both distributions in **Figure 2.8** would be referred to as bimodal even though the humps in the distribution on the right aren't the same height. For either distributions or graphs, if two high-frequency scores are separated by scores with lower frequencies, the name *bimodal* is appropriate.

Bimodal distribution
Distribution with two modes.

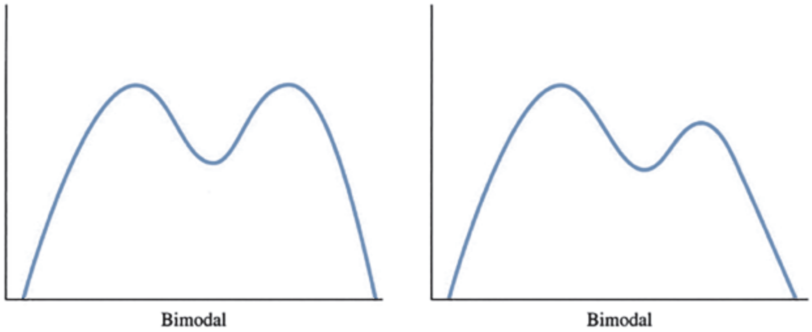


FIGURE 2.8 **Two bimodal curves**

Any set of measurements of a phenomenon can be arranged into a distribution; scientists and others find them most helpful. Sometimes measurements are presented in a frequency distribution table and sometimes in a graph (and sometimes both ways).

In the first chapter, I said that by learning statistics you could become more persuasive. Frequency distributions and graphs (but especially graphs) will be valuable to you as you face audiences large and small over the years.

The Line Graph

Perhaps the graph most frequently used by scientists is the **line graph**. A line graph is a picture of the relationship between *two* variables. A point on a line graph represents the value on the Y variable that goes with the corresponding value on the X variable. To understand a line graph, begin by reading the labels on the axes. Next, determine what the points on the graph represent. In social and biological science disciplines, a point is often an average (which might be of people, places, or things).

Line graph
Graph that uses lines to show the relationship between two variables.

[3] The mode of a distribution is the score that occurs most frequently.

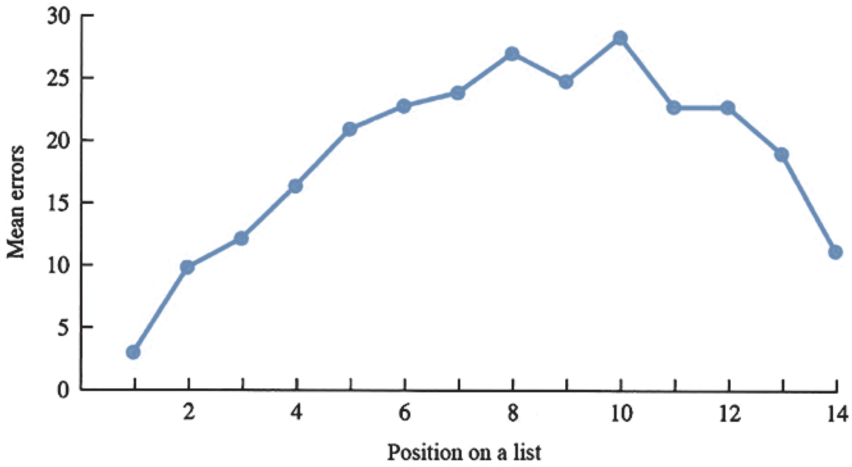


FIGURE 2.9 **Mean number of errors in memorizing a list of 14 items**

To give you an example, suppose students read a 14-item list, then repeat those items they remember. They do this until they repeat the list without error, which might take as many as 30 trials. Some items will be more difficult to remember than others. Which ones? **Figure 2.9** shows that items just past the middle produce the most errors. This phenomenon is called the *serial position effect*. So, if you find yourself memorizing a list, spend more time on items in the middle.

Line graphs typically do not connect to the x-axis. With frequency polygons, the connection to the *x*-axis is a data point—no one received that score. Connecting a line graph to the *x*-axis says there is a data point there. Don't do it unless the point actually represents data.

More on Graphics

Almost 100 graphs and figures are in the pages that follow. Some are frequency distributions and others are line graphs. In addition, you will find boxplots and scatterplots (with explanations). Although this chapter is the only one with *graphs* in the title, this textbook and scientific research in general are full of graphs. New designs are being created regularly.

Graphs are one of our most powerful ways of persuading others. When a person is confused or is of the opposite opinion, a graph can be convincing. One particular champion of graphics is Edward Tufte, an emeritus professor at Yale University. His books, *The Visual Display of Quantitative Information* (1983, 2001) and *Beautiful Evidence* (2006), celebrate and demonstrate the power of graphs, and both include a reprint of "(perhaps) the best statistical graphic ever drawn" (1983, p. 40).[4]

[4] I'll give you just a hint about this graphic. A French engineer drew it almost two centuries ago to illustrate a disastrous military campaign by Napoleon against Russia in 1812–13. This graphic was nominated by Wainer (1984) as the "World's Champion Graph."

Tufte says that regardless of your field, when you construct a quality graphic, it improves your understanding. So, if you find yourself somewhere on that path between confusion and understanding, you should try to construct a graph. A graph communicates information with the simultaneous presentation of words, numbers, and pictures. In the heartfelt words of a sophomore engineering student I know, "Graphs sure do help."

Designing an effective and pleasing graph requires planning and rough drafts. Sometimes conventional advice is helpful and sometimes not. For example, "Make the height 60 to 75 percent of the width" works for some data sets and not for others. Graphic designers should heed Tufte's (1983) admonition, "It is better to violate any principle than to place graceless or inelegant marks on paper" (p. 191).

But how can you learn to put graceful and elegant marks on paper? Tufte (2001, 2006) has advice. Nicol and Pexman (2010) provide a detailed guide that illustrates a variety of graphs and figures; they also cover poster presentations. Finally, revising your own productions before you submit them will move you toward this goal.

A Moment to Reflect

At this point in your statistics course, I have a question for you: How are you doing? I know that questions like this are better asked by a personable human than by an impersonal textbook. Nevertheless, please give your answer. I hope you can say "OK" or perhaps something better. If your answer isn't "OK" (or better), now is the time to make changes. Develop a plan. Share it with someone who will support you. Implement your plan.

Perhaps you have a test coming up soon? Here's an exercise I developed that is based on Chen et al. (2017). In their study, statistics students who completed a planning exercise improved their test scores by 1/3 of a grade. To use this exercise yourself, go to "free materials" at exploringstatistics.com well before the date of your exam.

PROBLEMS

***2.10.** Determine the direction of the skew for the two curves in Problem 2.8 by examining the curves or the frequency distributions (or both).

2.11. Describe how a line graph is different from a frequency polygon.

2.12. From the data for the five U.S. cities that follow, construct a line graph that shows the relationship between elevation above sea level and mean January temperature. The latitudes of the five cities are almost equal; all are within one degree of 35°N latitude. Write a sentence of interpretation.

City	Elevation (feet)	Mean January temperature (°F)
Albuquerque, NM	5000	34
Amarillo, TX	3700	35
Flagstaff, AZ	6900	29
Little Rock, AR	350	39
Oklahoma City, OK	1200	36

2.13. Without looking at Figures 2.6 and 2.8, sketch a normal distribution, a rectangular distribution, and a bimodal distribution.

2.14. Is the narrow point of a positively skewed distribution directed toward the right or the left?

2.15. For each frequency distribution listed, tell whether it is positively skewed, negatively skewed, or approximately symmetrical (bell-shaped or rectangular).

- **a.** Age of all people alive today
- **b.** Age in months of all first-graders
- **c.** Number of children in families
- **d.** Wages in a large manufacturing plant
- **e.** Age at death of everyone who died last year
- **f.** Shoe size

2.16. Write a paragraph on graphs.

2.17. **a.** Problem 2.3 presented data on oral body temperature. Use those data to construct a simple frequency distribution.

b. Using the grouped frequency distribution you created for your answer to Problem 2.3, construct an appropriate graph, and describe its form.

2.18. Read and respond to the six objectives at the beginning of the chapter. Engaged responding will help you remember what you learned.

KEY TERMS

Abscissa (p. 35)
Bar graph (p. 37)
Bimodal distribution (p. 41)
Class intervals (p. 33)
Frequency (p. 31)
Frequency polygon (p. 36)
Grouped frequency distribution (p. 33)
Histogram (p. 37)
Line graph (p. 41)
Negative skew (p. 40)
Normal distribution (normal curve) (p. 40)
Ordinate (p. 35)
Positive skew (p. 40)
Raw scores (p. 29)
Rectangular distribution (p. 40)
Simple frequency distribution (p. 31)
Skewed distribution (p. 40)
Symmetrical distribution (p. 39)

CHAPTER 3

Exploring Data: Central Tendency

OBJECTIVES FOR CHAPTER 3

After studying the text and working the problems in this chapter, you should be able to:

1. Find the mean, median, and mode of a simple frequency distribution
2. Determine whether a measure of central tendency is a statistic or a parameter
3. Identify three characteristics of the mean
4. Detect bimodal distributions
5. Determine the central tendency measure that is most appropriate for a set of data
6. Estimate the direction of skew of a frequency distribution from the relationship between the mean and median
7. Calculate a weighted mean

EVERY DISTRIBUTION HAS three features—*form, central tendency,* and *variability.* These features are independent of each other and if you are familiar with all three, you'll have a fairly complete understanding of the distribution. In the previous chapter, you learned about frequency distributions and graphs, two descriptive methods that show the *form* of a distribution.

This chapter is about **central tendency**; four measures are covered. All are single numbers or descriptors that indicate a location for the bulk of a distribution's scores. Often defined as a *typical score* or a *representative score*, measures of central tendency are almost a necessity for exploring a set of data. Used by themselves, however, they do not provide any information about form or variability.

Central tendency
Descriptive statistics that indicate a typical or representative score.

clue to the future

The distributions you work with in this chapter are empirical distributions based on observed scores. This chapter and the next three are about these empirical frequency distributions. Starting with Chapter 7 and throughout the rest of the book, you will also use theoretical distributions—distributions that come from mathematical formulas and logic rather than from actual observations.

Measures of Central Tendency

There are a number of measures of central tendency; the most popular are the mean, median, and mode.

The Mean

Mean
The arithmetic average; the sum of the scores divided by the number of scores.

The symbol for the **mean** of a sample is $\overline{X}$ (pronounced "mean" or "X-bar"). The symbol for the mean of a population is μ (a Greek letter, mu, pronounced "mew"). Of course, an $\overline{X}$ is only one of many possible means from a population. Because other samples from that same population produce somewhat different $\overline{X}$s, a degree of uncertainty goes with $\overline{X}$.[1]

If you had an entire population of scores, you could calculate μ and it would carry no uncertainty with it. Most of the time, however, the population is not available and you must make do with a sample. The difference in $\overline{X}$ and μ, then, is in the interpretation. $\overline{X}$ carries some uncertainty with it; μ does not.

Here's a simple example of a sample mean. Suppose a college freshman arrives at school in the fall with a promise of a monthly allowance for spending money. Sure enough, on the first of each month, there is money to spend. However, 3 months into the school term, our student discovers a recurring problem: too much month left at the end of the money.

Unable to secure a calendar compressor, our student zeros in on money spent at the Student Center. For a 2-week period, he records everything bought at the center, a record that includes coffee, both regular and cappuccino Grande, bagels (with cream cheese), chips, soft drinks, ice cream, and the occasional banana. His data are presented in **Table 3.1.**[2] You already know how to compute the mean of these numbers, but before you do, eyeball the data and then write down your estimate of the mean in the space provided. The formula for the mean is

$$\overline{X} = \frac{\sum X}{N}$$

where $\overline{X}$ = the mean
Σ = an instruction to add (Σ is uppercase Greek sigma)
X = a score or observation; ΣX means to add all the Xs
N = number of scores or observations

For the data in Table 3.1,

$$\overline{X} = \frac{40.60}{14} = \$2.90$$

These data are for a 2-week period, but our freshman is interested in his expenditures for at least 1 month and, more likely, for many months. Thus, the result is a sample mean and the symbol $\overline{X}$ is appropriate. The amount, $2.90, is an estimate of the amount that our friend spends at the Student Center each day. $2.90 may seem unrealistic to you. If so, go back and explore the data; you can find an explanation.

[1] Many scientific journals use a capital *M* as the symbol of the mean.

[2] In formal writing, scientists and academics insist that *data* is a plural noun. Thus, *data* are and *a datum* is. Newspapers, magazines, and general conversation treat data as singular. I don't know when we will admit defeat on this issue.

TABLE 3.1 Student Center expenditures during a 2-week period

Day	Money Spent
1	$4.25
2	2.50
3	5.25
4	0.00
5	4.90
6	0.85
7	0.00
8	0.00
9	5.70
10	3.00
11	0.00
12	0.00
13	8.90
14	5.25
	$\sum$ = $40.60
Your estimate of the mean______	

Now we come to an important part of any statistical analysis, which is to answer the question, "So what?" Calculating numbers or drawing graphs is a part of almost every statistical problem, but unless you can tell the story of what the numbers and pictures mean, you won't find statistics worthwhile.

The first use you can make of Table 3.1 is to estimate the student's monthly Student Center expenses. This is easy to do. Thirty days times $2.90 is $87.00. Now, let's suppose our student decides that this $87.00 is an important part of the "monthly money problem." The student has three apparent options. The first is to get more money. The second is to spend less at the Student Center. The third is to justify leaving things as they are. For this third option, our student might perform an economic analysis to determine what he gets in return for his almost $90 a month. His list might be pretty impressive: lots of visits with friends; information about classes, courses, and professors; a borrowed book that was just super; thousands of calories; and more.

The point of all this is that part of the attack on the student's money problem involved calculating a mean. However, an answer of $2.90 doesn't have much meaning by itself. Interpretations and comparisons are called for.

Characteristics of the mean Three characteristics of the mean will be referred to again in this book. First, if the mean of a distribution is subtracted from each score in that distribution and the differences are added, the sum will be zero; that is, $\sum(X\text{-}\bar{X}) = 0$. The statistic, $X\text{-}\bar{X}$ is called a deviation score. To demonstrate to yourself that $\sum(X\text{-}\bar{X}) = 0$, you might pick a few numbers to play with (numbers 1, 2, 3, 4, and 5 are easy to work with).

Second, if those deviation scores are squared and then summed, which is expressed as $\sum(X\text{-}\bar{X})^2$, the result will be smaller than the sum you get if any number other than $\bar{X}$ is used. That is, using the mean minimizes the sum of squared deviations. Again, you can demonstrate this relationship for yourself by playing with a small set of scores.

Third, the formula $\Sigma X/N$ produces a sample mean that is an unbiased estimator of μ, the population mean. Some sample statistics, however, are not unbiased estimators of their population parameters. (See the next chapter on variability.)

clue to the future

In the next chapter, the first two characteristics of the mean will come up again explicitly. The third one will be important background information as you study the different formulas for the standard deviation.

The Median

Median
Point that divides a distribution of scores into equal halves.

The **median** is the *point* that divides a distribution of scores into two equal parts. To find the median of the Student Center expense data, begin by arranging the daily expenditures from highest to lowest. The result is **Table 3.2**, which is called an *array*. Because there are 14 scores, the halfway point, or median, will have seven scores above it and seven scores below it. The seventh score from the bottom is \$2.50. The seventh score from the top is \$3.00. The median, then, is halfway between these two scores, or \$2.75. (The halfway point between two numbers is the mean of the two numbers. Thus, [\$2.50 + \$3.00]/2 = \$2.75.) The median is a hypothetical *point* in the distribution; it may or may not be an actual score.

Table 3.2 Data of Table 3.1 arrayed in descending order

X	
\$8.90	7 scores
5.70	
5.25	
5.25	
4.90	
4.25	
3.00	
	Median = \$2.75
2.50	7 scores
0.85	
0.00	
0.00	
0.00	
0.00	
0.00	

What is the interpretation of a median of \$2.75? The simplest interpretation is that on half the days our student spends less than \$2.75 in the Student Center and on the other half he spends more.

What if there had been an odd number of days in the sample? Suppose the student chose to sample half a month, or 15 days. Then the median would be the eighth score. The eighth score has seven scores above and seven below. For example, if an additional day was included, during which \$5.00 was spent, the median would be \$3.00. If the additional day's expenditure was zero, the median would be \$2.50.

The reasoning you just went through can be facilitated by using a formula that locates the median in an array. The formula for finding the location of the median is

$$\text{Median location} = \frac{N + 1}{2}$$

The location may be at an actual score (as in the second example) or a point between two scores (the first example).

The Mode

The third central tendency statistic is the **mode**. The mode is the most frequently occurring score—the score observed most often. For the Student Center expense data, the mode is $0.00. **Table 3.2** shows the mode most clearly. The zero amount occurred five times, $5.25 twice, and all other amounts once.

Mode
Score that occurs most frequently in a distribution.

When a mode is reported, it often helps to tell the percentage of times it occurred. "The mode was $0.00, which occurred on 36% of the days" is certainly more informative than "The mode was $0.00."

PROBLEMS

3.1. Find the mean and median of the following sets of scores.
- **a.** 2, 5, 15, 3, 9
- **b.** 9, 13, 16, 20, 12, 11
- **c.** 8, 11, 11, 8, 11, 8

3.2. Find the mode of each set of scores.
- **a.** 0, 4, 4, 5, 6
- **b.** 12, 7, 10, 12, 11, 9

3.3. What three mathematical characteristics of the mean were described in this section?

Finding Central Tendency of Simple Frequency Distributions

Mean

Table 3.3 is an expanded version of Table 2.3, the frequency distribution of Satisfaction With Life Scale (SWLS) scores in Chapter 2. The steps for finding the mean from a simple frequency distribution are in the text, but first estimate the mean of the SWLS scores in the space at the bottom of **Table 3.3,**[3] ignoring the summary numbers.

The first step in calculating the mean from a simple frequency distribution is to multiply each score in the *X* column by its corresponding *f* value, so that all the people who make a particular score are included. Next, sum the *fX* values and divide the total by *N*. (*N* is the sum of the *f* values.) The result is the mean. In terms of a formula,

$$\mu \text{ or } \overline{X} = \frac{\sum fX}{N}$$

[3] If the data are arranged in a frequency distribution, you can estimate the mean by selecting the most frequent score (the mode) or by selecting the score in the middle of the list.

TABLE 3.3 Simple frequency distribution of Satisfaction With Life Scale scores including *fX* calculations

SWLS score (X)	(f)	fX	SWLS score (X)	(f)	fX
35	2	70	22	5	110
34	1	34	21	3	63
33	2	66	20	5	100
32	4	128	19	3	57
31	2	62	17	2	34
30	5	150	16	2	32
29	7	203	15	1	15
28	6	168	13	2	26
27	10	270	12	1	12
26	9	234	11	1	11
25	9	225	10	2	20
24	6	144	9	2	18
23	6	138	5	2	10
	(continued above)			Σ = 100	2400

Your estimate of the mean______

For the data in Table 3.3,

$$\mu \text{ or } \overline{X} = \frac{\sum fX}{N} = \frac{2400}{100} = 24.00$$

How does 24.00 compare to your estimate?

Two questions are common at this point. "How many decimal places?" and "Is 24.00 a $\overline{X}$ or a μ?" If the data are whole numbers, two decimal places is usually sufficient. The second question requires a follow-up question, "Is there any interest in a group larger than these 100?" If the answer is no, the 100 scores are a population and 24.00 = μ. If the answer is yes, the 100 scores are a sample and 24.00 = $\overline{X}$.

Median

The formula for finding the location of the median that you used earlier works for a simple frequency distribution, too.

$$\text{Median location} = \frac{N+1}{2}$$

Thus, for the scores in Table 3.3,

$$\text{Median location} = \frac{N+1}{2} = \frac{100+1}{2} = 50.5$$

To find the 50.5th position, begin adding the frequencies in Table 3.3 from the bottom (2 + 2 + 2 + 1 + . . .). The total is 43 by the time you include the score of 24. Including 25 would make the total 52—more than you need. So the 50.5th score is among those nine scores of 25. The median is 25.

Suppose you start the quest for the median at the top of the array rather than at the bottom. Again, the location of the median is at the 50.5th position in the distribution. To get to 50.5, add the frequencies from the top (2 + 1 + 2 + . . .). The sum of the frequencies of the scores from 35 down to and including 26 is 48. The next score, 25, has a frequency of 9. Thus, the 50.5th position is among the scores of 25. The median is 25.

error detection

Calculating the median by starting from the top of the distribution produces the same result as calculating the median by starting from the bottom.

Mode

It is easy to find the mode from a simple frequency distribution. In **Table 3.3**, the score with the highest frequency, which is 10, is the mode. So, mode = 27.

In Chapter 2, you learned to recognize curves with two distinct humps as bimodal. Such curves have distributions with two high-frequency scores (modes) separated by one or more low frequency scores.

PROBLEMS

3.4. Which of the following distributions is bimodal?

a. 10, 12, 9, 11, 14, 9, 16, 9, 13, 20
b. 21, 17, 6, 19, 23, 19, 12, 19, 16, 7
c. 14, 18, 16, 28, 14, 14, 17, 18, 18, 6

***3.5.** Refer to Problem 2.2, the political-scientist student's yard-sign data.

a. Decide which measure of central tendency is appropriate and find it.
b. Is this central tendency measure a statistic or a parameter?
c. Write a sentence of interpretation.

***3.6.** Refer to Problem 2.1. Find the mean, median, and mode of the heights of both groups of Americans in their 20s. Work from the Appendix G answers to Problem 2.1.

3.7. Find the median of **a**, starting from the bottom. Find the median of **b**, starting at the top. The median of distribution **c** can probably be solved by inspection.

a.

X	f
15	4
14	3
13	5
12	4
11	2
10	1

b. 4, 0, -1, 2, 1, 0, 3, 1, -2, -2, -1, 2, 1, 3, 0, -2, 1, 2, 0, 1
c. 28, 27, 26, 21, 18, 10

error detection

Estimating (also called eyeballing) is a valuable way to avoid big mistakes. Begin work by quickly making an estimate of the answers. If your calculated answers differ from your estimates, wisdom dictates that you reconcile the difference. You have either overlooked something when estimating or made a computation mistake.

When to Use the Mean, Median, and Mode

A common question is, "Which measure of central tendency should I use?" The general answer is, given a choice, use the mean. Sometimes, however, the data limit your choice. Here are three considerations.

Scale of Measurement

A mean is appropriate for ratio or interval scale data, but not for ordinal or nominal distributions. A median is appropriate for ratio, interval, and ordinal scale data, but not for nominal data. The mode is appropriate for any of the four scales of measurement.

You have already thought through part of this issue in working Problem 3.5. In it, you found that the yard sign names (very literally, a nominal variable) could be characterized with a mode, but it would be impossible to try to add up the names and divide by N or to find the median of the names.

For an ordinal scale such as class standing in college, either the median or the mode makes sense. The median would probably be sophomore, and the mode would be freshman.

Skewed Distributions

Even if you have interval or ratio data, the mean may be a misleading choice if the distribution is severely skewed. Here's a story to illustrate.

The developer of Swampy Acres Retirement Homesites is attempting to sell building lots in a southern "paradise" to out-of-state buyers. The marks express concern about flooding. The developer reassures them: "The average elevation of the lots is 78.5 feet and the water level in this area has never, ever exceeded 25 feet." The developer tells the truth, but this average truth is misleading. The actual lay of the land is shown in **Figure 3.1**. Now look at **Table 3.4**, which shows the elevations of the 100 lots arranged in a grouped frequency distribution. (Grouped frequency distributions are explained in Appendix B.

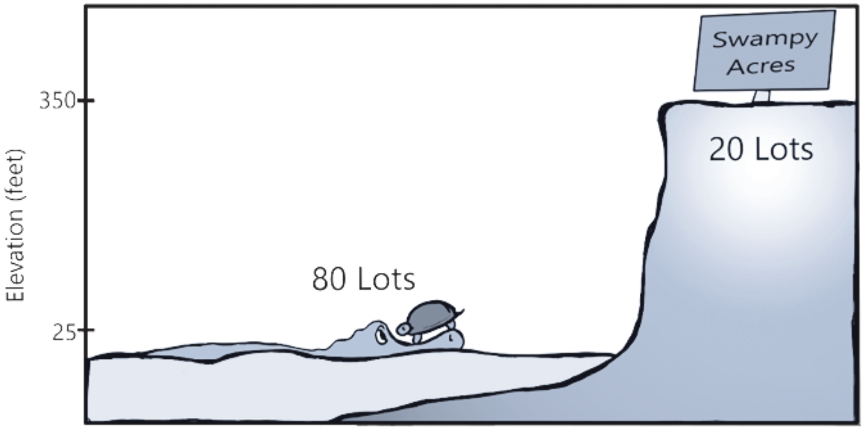


FIGURE 3.1 **Elevation of Swampy Acres**

TABLE 3.4 **Frequency distribution of lot elevations at Swampy Acres**

Elevation in feet	Midpoint X	Number of Lots (f)	fX
348-352	350	20	7000
13-17	15	30	450
8-12	10	30	300
3-7	5	20	100
		$\sum = 100$	7850

To calculate the mean of a grouped frequency distribution, multiply the midpoint of each interval by its frequency, add the products, and divide by the total frequency. Thus, in Table 3.4,

$$\mu = \frac{\sum fX}{N} = \frac{7850}{100} = 78.5 \text{ feet}$$

The mean is 78.5 feet, exactly as the developer said. However, only the 20 lots on the bluff are out of the flood zone; the other 80 lots are, on the average, under water. The mean, in this case, is misleading. What about the median? The median of the distribution in Table 3.4 is 12.5 feet, well under the high-water mark, and a better overall descriptor of Swampy Acres Retirement Homesites. The distribution in **Table 3.4** is severely skewed, which leads to the big disparity between the mean and the median.

Income data present real-life examples of severely skewed distributions. For 2016, the U.S. Census Bureau reported that mean household income was $83,143. The median, however, was $59,039. For distributions that are severely or even moderately skewed, the median is often preferred because it is unaffected by extreme scores.

Open-Ended Class Intervals

Even if you have interval or ratio data and the distribution is fairly symmetrical, there is a situation for which you *cannot* calculate a mean. If the highest interval or the lowest interval of a grouped frequency distribution is open-ended, there is no midpoint and you cannot calculate a mean. Age data are sometimes reported with the oldest being "75 and over." The U.S. Census Bureau reports household income with the largest category as $200,000 and over. Because there is no midpoint to "75 and over" or to "$200,000 or more," you cannot calculate a mean. Medians and modes are appropriate measures of central tendency when one or both of the extreme class intervals are open ended.

In summary, use the mean if it is appropriate. To follow this advice, you must recognize data for which the mean is not appropriate. Perhaps **Table 3.5** will help.

TABLE 3.5 Data characteristics and recommended central tendency statistic

	Recommended statistic		
Data characteristic	Mean	Median	Mode
Nominal scale data	No	No	Yes
Ordinal scale data	No	Yes	Yes
Interval scale data	Yes	Yes	Yes
Ratio scale data	Yes	Yes	Yes
Open-ended category(ies)	No	Yes	Yes
Skewed distribution	No	Yes	Yes

Determining Skewness From the Mean and Median

You can determine skewness by comparing the mean and median (most of the time). If the mean is smaller than the median, expect negative skew. If the mean is larger than the median, expect positive skew. **Figure 3.2** shows the relationship of the mean to the median for a negatively skewed distribution (left) and a positively skewed distribution (right) of continuous data.

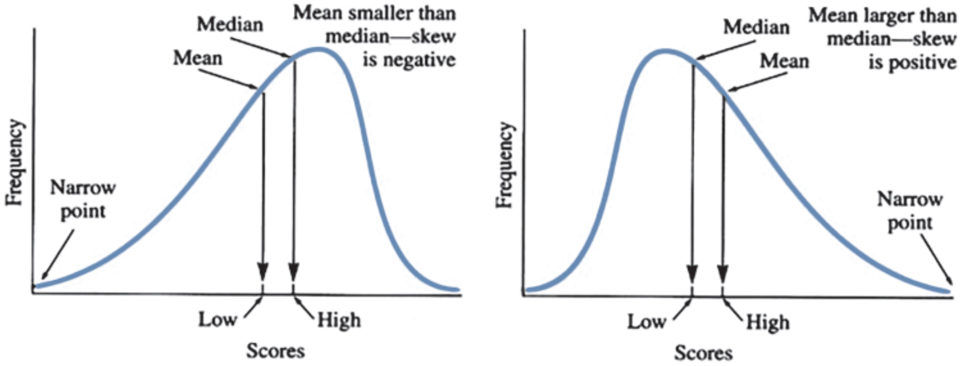


FIGURE 3.2 The effect of skewness on the relative position of the mean and median for continuous data

I'll illustrate by changing the slightly skewed data in Table 3.1 into more severely skewed data. The original expenditures in Table 3.1 have a mean of $2.90 and a median of $2.50, a difference of just 40 cents. If I add an expenditure of $100.00 to the 14 scores, the mean jumps to $9.37 and the median moves up to $3.00. The difference now is $6.37 rather than 40 cents. This example follows the general rule that the greater the difference between the mean and median, the greater the skew.[4] Note also that the mean of the new distribution ($9.37) is greater than every score in the distribution except for $100.00. For severely skewed distributions, the mean is not a typical score.

This rule about the mean/median relationship usually works for continuous data such as SWLS scores and dollars, but it is less trustworthy for discrete data such as the number of adult residents in U.S. households (von Hippel, 2005).

One response to severely skewed distributions is to use the median. Another solution, often found in inferential statistics, is the trimmed mean. A trimmed mean is calculated by excluding a certain percentage of the values from each tail of the distribution. Means trimmed by 10% and 20% are common.

The Weighted Mean

Sometimes, several sample means are available from the same or similar populations. In such cases, a weighted mean, $\bar{X}_W$, is the best estimate of the population parameter μ. If every sample has the same N, you can compute a weighted mean by adding the means and dividing by the number of means.

Weighted mean
Overall mean calculated from two or more samples with different *N*s.

If the sample means are based on *N*s of different sizes, however, you cannot use this procedure. Here is a story that illustrates the right way and the wrong way to calculate a weighted mean.

At my undergraduate college, a student with a cumulative grade point average (GPA) of 3.25 in the middle of the junior year was eligible to enter a program to "graduate with honors." [In those days (1960), usually fewer than 10% of a class had GPAs greater than 3.25.] Discovering this rule after my sophomore year, I decided to figure out if I had a chance to qualify.

Calculating a cumulative GPA seemed easy enough to do: Four semesters had produced GPAs of 3.41, 3.63, 3.37, and 2.16. Given another GPA of 3.80, the sum of the five semesters would be 16.37, and dividing by 5 gave an average of 3.27, well above the required 3.25.

Graduating with honors seemed like a great ending for college, so I embarked on a goal-oriented semester—a GPA of 3.80 (a B in German and an A in everything else). And, at the end of the semester, I had accomplished the goal. Unfortunately, "graduating with honors" was not to be.

There was a flaw in my method of calculating my cumulative GPA. My method assumed that all of the semesters were equal in weight, that they had all been based on the same number of credit hours. My calculations based on this assumption are shown on the left side of **Table 3.6.**

Unfortunately, all five semesters were not the same; the semester with the GPA of 2.16 was based on 19 hours, rather than the usual 16 or so.[5] Thus, that semester should have been weighted more heavily than semesters with fewer hours.

[4] For a mathematical measure of skewness, see Kirk (2008), p. 129.
[5] That semester was more educational than a GPA of 2.16 would indicate. I read a great deal of American literature that spring, but unfortunately, I was not registered for any courses in American literature.

TABLE 3.6 Two methods of calculating a mean from five semesters' GPAs; the method on the left is correct only if all semesters have the same number of credit hours

Flawed method	Correct method		
Semester GPA	Semester GPA	Credit hours	GPA x hours
3.41	3.41	17	58
3.63	3.63	16	58
3.37	3.37	19	64
2.16	2.16	19	41
3.80	3.80	16	61
$\Sigma = 16.37$		$\Sigma = 87$	$\Sigma = 282$

$\overline{X} = \frac{16.37}{5} = 3.27$ $\qquad$ $\overline{X}_W = \frac{282}{87} = 3.24$

The formula for a weighted mean is

$$\overline{X}_W = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + \cdots + N_K\overline{X}_K}{N_1 + N_2 + \cdots + N_K}$$

where $\overline{X}_W$ = the weighted mean

$\overline{X}_1, \overline{X}_2, \overline{X}_K$ = sample means

N_1, N_2, N_K = sample sizes

K = number of samples

The right side of Table 3.6 shows the calculation of a weighted mean, which is required to correctly determine cumulative GPA. Each semester's GPA is multiplied by its number of credit hours. These products are summed and that total is divided by the sum of the hours. As you can see from the numbers on the right, the actual cumulative GPA was 3.24, not high enough to qualify for the honors program.

More generally, to find the mean of a set of means, multiply each separate mean by its N, add these products together, and divide the total by the sum of the Ns. As an example, three means of 2.0, 3.0, and 4.0, calculated from the scores of samples with Ns of 6, 3, and 2, produce a weighted mean, $\overline{X}_W$, of 2.64. Do you agree?

A final cautionary note about all measures of central tendency: As stand-ins or representatives of a data set, they can deceive. The variability of the scores and the form of the distribution have their say, too, and sometimes their say is, "This time, central tendency statistics give a false impression."

Estimating Answers

You may have noticed that I have (subtly?) worked in the advice: As your first step, estimate an answer. Here's why I think you should begin a problem by estimating the answer: (a) Taking a few seconds to estimate keeps you from plunging into the numbers before you fully understand the problem. (b) An estimate is especially helpful when a calculator or a computer does the bulk of your number crunching. Although these wonderful machines don't make errors, the people who enter the numbers or give instructions occasionally do. An initial estimate helps you catch these mistakes. (c) Noticing that an estimate differs from a calculated value gives you a chance to correct an error before anyone else sees it.

It can be exciting to look at an answer, whether it is on paper, a calculator display, or a computer screen, and say, "That can't be right!" and then to find out that, sure enough, the displayed answer is wrong. If you develop your ability to estimate answers, I promise that you will sometimes experience this excitement.

PROBLEMS

3.8. For each of the following situations, tell which measure of central tendency is appropriate and why.

- **a.** As part of a study on prestige, an investigator sat on a corner in a high-income residential area and classified passing automobiles according to color: black, gray, white, silver, green, and other.
- **b.** In a study of helping behavior, an investigator pretended to have locked himself out of his car. Passersby who stopped were classified on a scale of 1 to 5 as (1) very helpful, (2) helpful, (3) slightly helpful, (4) neutral, and (5) discourteous.
- **c.** In a study of household income in a city, the following income categories were established: \$0–\$20,000, \$20,001–\$40,000, \$40,001–\$60,000, \$60,001–\$80,000, \$80,001–\$100,000, and \$100,001 and more.
- **d.** In a study of per capita income in a city, the following income categories were established: \$0– \$20,000, \$20,001–\$40,000, \$40,001–\$60,000, \$60,001–\$80,000, \$80,001–\$100,000, \$100,001–\$120,000, \$120,001–\$140,000, \$140,001–\$160,000, \$160,001–\$180,000, and \$180,001–\$200,000.
- **e.** First admissions to a state mental hospital for 5 years were classified by disorder: schizophrenic, delusional, anxiety, dissociative, and other.
- **f.** A teacher gave her class an arithmetic test; most of the children scored in the range 70–79. A few scores were above this, and a few were below.

3.9. A senior psychology major conducted the same experiment three times with groups that consisted of 12, 31, and 17 participants. For the three sessions the mean percent correct was 74, 69, and 75, respectively. What is the mean score of the participants?

3.10.

- **a.** By inspection, determine the direction of skew of the expenditure data in Table 3.2 and the elevation data in Table 3.4. Verify your judgment by comparing the means and medians.
- **b.** For Problem 2.10, you determined by inspection the direction of skew of two distributions (X and Y). To verify your judgment, calculate and compare the mean and median of each distribution. The data are in Problem 2.8.

3.11. A 3-year veteran of the local baseball team was calculating his lifetime batting average. The first year he played for half the season and batted .350 (28 hits in 80 at-bats). The second year he had about twice the number of at-bats and his average was .325. The third year, although he played even more regularly, he was in a slump and batted only .275. Adding the three season averages and dividing the total by 3, he found his lifetime batting average to be .317. Is this correct? Explain.

CP **3.12.** The data in Table 3.1 (money spent at the Student Center) were arranged so I could illustrate characteristics of central tendency statistics. The data that follow are based on a study of actual expenditures for food by freshmen students at Central Michigan University. (See http://www.westga.edu/~bquest/2008/local08.pdf). Students who reported positive expenditures for a two-week period are shown. Find the mean, median, and mode. Determine the direction of skew from the relationship between the mean and median.

$20, $30, $9, $22, $18, $54, $24, $2, $81, $24, $20, $33, $13, $28, $20

KEY TERMS

Array (p. 48)
Bimodal (p. 51)
Central tendency (p. 46)
Empirical distribution (p. 45)
Estimating (p. 56)
Mean (p. 46)
Median (p. 48)
Mode (p. 49)
Open-ended class intervals (p. 54)
Scale of measurement (p. 52)
Skewed distributions (p. 52, 54)
Theoretical distribution (p. 45)
Trimmed mean (p. 55)
Weighted mean (p. 55)