EXPLORING STATISTICS

TALES OF DISTRIBUTIONS

12TH EDITION

CHRIS SPATZ

12th Edition

Exploring Statistics
Tales of Distributions

Chris Spatz

Outcrop Publishers Conway, Arkansas

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

Exploring Statistics: Tales of Distributions
12th Edition
Chris Spatz

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

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

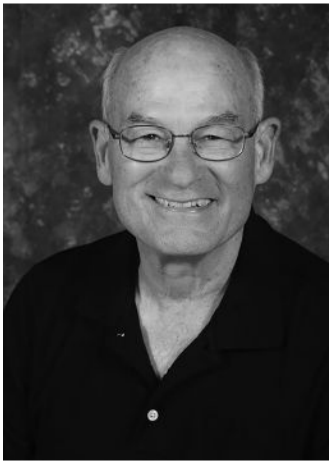
Examination copies are provided to academics and professionals to consider for adoption as a course textbook. Examination copies may not be sold or transferred to a third party. If you adopt this textbook, please accept it as your complimentary desk copy.

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 4

Exploring Data: Variability

OBJECTIVES FOR CHAPTER 4

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

1. Explain the concept of variability
2. Find and interpret the range of a distribution
3. Find and interpret the interquartile range of a distribution
4. Distinguish among the standard deviation of a population, the standard deviation of a sample used to estimate a population standard deviation, and the standard deviation used to describe a sample
5. Know the meaning of σ, $\hat{s}$, and S
6. For grouped and ungrouped data, calculate a standard deviation and interpret it
7. Calculate the variance of a distribution

LET'S LEAVE THE classroom, climb a mountain, and pose a question to the statistics guru at the top. "What's your best advice for those confronted with data?" Without pause, the guru answers, "Find out about the **variability** of the scores."

Variability
Having a range of values.

You studied *form* (frequency distributions and graphs) in Chapter 2 and *central tendency* in Chapter 3. Measures of *variability* tell a separate, independent story, and they deserve more attention than they get. If you are to understand a set of data, you must have a feel for form, central tendency, and especially, variability.

The value of knowing about variability is illustrated by a story of two brothers who, on a dare, went water skiing on Christmas Day (the temperature was about 35°F). On the *average*, each skier finished his turn 5 feet from the shoreline (where one may step off the ski into only 1 foot of very cold water). This bland average of the two actual stopping places, however, does not convey the excitement of the day.

The first brother, determined to avoid the cold water, held on to the towrope too long. Scraped and bruised, he finally stopped rolling at a spot 35 feet up on the rocky shore. The second brother, determined not to share the same fate, released the towrope too soon. Although he swam the 45 feet to shore very quickly, his lips were very blue. No, to hear that the average stopping place was 5 feet from the shoreline doesn't capture the excitement of the day.

This story reveals a cognitive habit that's been drummed into our heads since elementary school: Let the typical case (5 feet from shoreline) stand for all the cases. To get the full story, you have to ask about variability. Here are other situations in which variability is important.

1. You are a parent of a wonderful 8-month-old daughter. You read that at 8 months babies can pull up to a standing position. Your daughter cannot do this. Worse, typical babies crawl at 7 months of age. Your daughter does not crawl. Worry consumes you.

Caution: *Children vary. There is lots of variability about both 8-month and 7-month milestones. Knowing the range of months for normally developing babies can relieve much of your anxiety.*

2. Suppose that your temperature, taken with a thermometer under your tongue, is 97.5°F. You begin to worry. This is below even the average of 98.2°F that you learned in the previous chapter (based on Mackowiak, Wasserman, & Levine, 1992).

Caution: *There is variability around that mean of 98.2°F. Is 97.5°F below the mean by just a little or by a lot? Knowing about variability is necessary if you are to answer this question.*

3. Having graduated from college, you are considering two offers of employment, one in sales and the other in management. The pay is about the same for both. Using the library to check out the statistics for salespeople and managers, you find that those who have been working for 5 years in each type of job also have similar averages. You conclude that the pay for the two occupations is equal.

Caution: *Pay is more variable for those in sales than for those in management. Some in sales make much more than the average and some make much less, whereas the pay of those in management is clustered together. Your reaction to this difference in variability might help you choose.*

Central tendency statistics alone are almost never sufficient. Given a mean, nod appreciatively and ask about variability. Measures of variability provide important information that is completely independent of information provided by the mean, median, and mode. **Table 4.1** shows three very different distributions, each with a mean of 15. In every case, the 15 gives you a bit of information about the scores but leaves you completely in the dark about variability. Neglecting variability is a terrible flaw in many people's everyday thinking.

TABLE 4.1 Illustration of three different distributions that have the same mean

	X_1	X_2	X_3
	25	17	90
	20	16	30
	15	15	15
	10	14	0
	5	13	-60
Mean	15	15	15

This chapter is about four statistics and parameters that measure variability. The range is easy to calculate and gives a quick estimate. The interquartile range shows the middle 50% of the scores. The standard deviation provides an average of the distance that scores are from the mean. Finally, the variance gives another average distance of the scores from the mean. Most of this chapter is about the standard deviation.

Range

The **range** of a quantitative variable or an ordered categorical variable is the highest score minus the lowest score.

Range
Highest score minus lowest score.

$$\text{Range} = X_H - X_L$$

where X_H = highest score

X_L = lowest score

The range of the Satisfaction With Life Scale scores you worked with in previous chapters was 30 (see Table 2.2). The highest score was 35; the lowest was 5. Thus, the range is 35 – 5 = 30. In Chapter 2, you worked with oral body temperatures. (See Problem 2.3 and its answer.) The highest temperature was 99.5°F; the lowest was 96.4°F. The range is 3.1°F. Knowing that the range of normal body temperature is more than 3 degrees tends to soften a strict interpretation of a particular temperature.

In manufacturing, the range is used in some quality-control procedures. From a small sample of whatever is being manufactured, inspectors calculate a range and compare it to an expected figure. A large range means there is too much variability in the process; adjustments are called for.

The range is a quickly calculated, easy-to-understand statistic that conveys clearly how variable the data are. Though valuable, the range is often not sufficient. You can probably imagine two very different distributions of scores that have the same mean and the same range. (Go ahead, try.) If you are able to dream up two such distributions, it follows that additional measures of variability are needed if you are to distinguish among different distributions using just one measure of central tendency and one measure of variability.

Interquartile Range

The next measure of variability, the **interquartile range**, gives the range of scores that contain the middle 50% of the distribution. It is an important element of *boxplots*, which you will study in Chapter 5. (A boxplot shows several characteristics of a data set on one graph.)

Interquartile Range
Range of scores that contains the middle 50% of a distribution.

Percentile
Point below which a specified percentage of the distribution falls.

To find the interquartile range, you must have two different percentile scores. You may already be familiar with the concept of **percentile** scores. The 10th percentile score has 10% of the distribution below it; it is near the bottom. The 95th percentile score is near the top; 95% of the scores in the distribution are smaller. The 50th percentile divides the distribution into equal halves. For the interquartile range, you need the 25th and 75th percentile scores. Calculating percentiles is similar to finding the median. The median, the point that divides a distribution into equal halves, is the 50th percentile. Finding the 25th and 75th percentile scores involves the same kind of reasoning that you used to find the median.

The 25th percentile score has 25% of the scores below it. Look at **Table 4.2**, which shows a frequency distribution of 40 scores. To find the 25th percentile score, multiply 0.25 by N. Thus, 0.25 x 40 = 10. When $N = 40$, the 10th score up from the bottom is the 25th percentile score. You can see in Table 4.2 that there are five scores of 13 or lower. The 10th score is among the seven scores of 14. Thus, the 25th percentile is a score of 14.

TABLE 4.2 Finding the 25th and 75th percentiles

Score	f	
21	1	} 4 scores
20	3	
19	6	
18	3	
17	5	
16	4	
15	6	
14	7	
13	3	} 5 scores
12	2	
	$N = 40$	

The 75th percentile score has 75% of the scores below it. The easiest way to find it is to work from the top, using the same multiplication procedure, 0.25 times N (0.25 x 40 = 10). The 75th percentile score is the 10th score down from the top of the distribution. In Table 4.2, there are four scores 20 or larger. The 10th score is among the six scores of 19, so the 75th percentile score is 19. The interquartile range (IQR) is the 75th percentile minus the 25th percentile:

$$\text{IQR} = \text{75th percentile} - \text{25th percentile}$$

Thus, for the distribution in Table 4.2, IQR = 19 – 14 = 5. One interpretation is that about half the scores are between 14 and 19.

PROBLEMS

4.1. Find the range for each of the two distributions.

a. 17, 5, 1, 1

b. 0.45, 0.30, 0.30

***4.2.** Find the interquartile range of the Satisfaction With Life Scale scores in Table 3.3 (p. 50). Write a sentence of interpretation.

***4.3.** Find the interquartile range of the heights of both groups of those 20- to 29-year-old Americans. Use the frequency distributions you constructed for Problem 2.1. (These distributions may be in your files; they are also in Appendix G.)

Standard Deviation

The most popular measure of variability is the **standard deviation**. It is widely used because it provides both a measure of the width of a distribution and, given what you will learn in Chapter 7, the proportions of the distribution near the mean and far from the mean. Once you understand standard deviations, you can express quantitatively the difference between the two distributions you imagined previously in the section on the range. (You did do the imagining, didn't you?)

Standard deviation
Descriptive measure of the dispersion of scores around the mean.

The standard deviation (or its close relatives) turn up in every chapter after this one. Of course, in order to understand, you'll have to read all the material, figure out the problems, and do the interpretations. Lots of work. In return, though, you'll understand the most popular yardstick of variability—understanding that will last you a lifetime (if you get refreshers along the way).

Your principal task in this section is to learn the distinctions among three different standard deviations. Which standard deviation you use in a particular situation will be determined by your purpose. **Table 4.3** lists symbols, purposes, and descriptions. It will be worth your time to study Table 4.3 thoroughly now.

TABLE 4.3 Symbols, purposes, and descriptions of three standard deviations

Symbol	Purpose	Description
σ	Measure population variability	Lowercase Greek sigma. A parameter. Describes variability when a population of data is available
$\hat{s}$	Estimate population variability	Lowercase $\hat{s}$ (s-hat). A statistic. An estimate of a σ (in the same way that $\overline{X}$ is an estimate of μ). The variability statistic you will use most often in this book.
S	Measure sample variability	Capital S. A statistic. Describes the variability of a sample when there is no interest in estimating σ.

Standard Deviation as a Descriptive Index of Variability

Both σ and S are used to *describe* the variability of a set of data. σ is a parameter of a population; S is a statistic of a sample. The formulas for the two are almost identical, but the arithmetic that produces an answer can be arranged several ways.

The *deviation-score formula* shows clearly how each score contributes to the final result. The *raw-score formula* takes less time with a calculator, and its format turns up again in six more chapters. My suggestion is that you practice both methods. After that, you can revert to a calculator button or software to do calculations for you. The deviation-score formula requires an explanation of deviation scores.

Deviation Scores

A **deviation score** is a raw score minus the mean of the distribution, whether the distribution is a sample or a population.

$$\text{Deviation Score} = X - \bar{X} \text{ or } X - \mu$$

Deviation score
Raw score minus the mean of its distribution.

Raw scores that are greater than the mean produce positive deviation scores, raw scores that are less than the mean have negative deviation scores, and raw scores that are equal to the mean produce deviation scores of zero.

Table 4.4 provides an illustration of how to compute deviation scores for a small sample of data. From the sum of the scores in Table 4.4, I computed the mean, 8, and then subtracted it from each score. The result is deviation scores, which appear in the right-hand column.

TABLE 4.4 Computation of deviation scores from raw scores

Name	Score	Deviation score
Alex	14	6
Samantha	10	2
Luke	8	0
Zachary	5	–3
Stephen	3	-5
	$\sum X = 40$	$\sum(X - \bar{X}) = 0$
$\bar{X} = \frac{\sum X}{N} = \frac{40}{5} = 8$		

A deviation score tells you the number of points that a particular score deviates from, or differs from, the mean. In Table 4.4, the $X\text{-}\overline{X}$ value for Alex, 6, tells you that he scored six points above the mean. Luke scored at the mean, so his deviation score is 0, and Stephen, –5, scored five points below the mean.

error detection

Notice that the sum of the deviation scores is always zero. Add the deviation scores; if the sum is not zero, you made an error. You studied this concept before. In Chapter 3, you learned that $\sum(X\text{-}\overline{X}) = 0$.

Computing *S* and σ Using Deviation Scores

The deviation-score formula for computing the standard deviation as a descriptive index is

$$S = \sqrt{\frac{\sum(X - \overline{X})^2}{N}} \quad \text{or} \quad \sigma = \sqrt{\frac{\sum(X - \mu)^2}{N}}$$

where S = standard deviation of a sample

σ = standard deviation of a population

N = number of scores (same as the number of deviations)

The numerator of standard deviation formulas, $\sum(X\text{-}\overline{X})^2$, is shorthand notation that tells you to find the deviation score for each raw score, square the deviation score, and then add all the squares together. This sequence illustrates one of the rules for working with summation notation: Perform the operations within the parentheses first.

How can a standard deviation add to your understanding of a phenomenon? Let's take the sale of Girl Scout cookies. Table 4.5 presents some imaginary (but true-to-life) data on boxes of cookies sold by six Girl Scouts. I'll use these data to illustrate the calculation of the standard deviation, S. If these data were a population, the value of σ would be identical.

The numbers of boxes sold are listed in the X column of **Table 4.5**. The rest of the arithmetic needed for calculating S is also given. To compute S by the deviation-score formula, first find the mean. Obtain a deviation score for each raw score by subtracting the mean from the raw score. Square each deviation score and sum the squares to obtain $\sum(X\text{-}\overline{X})^2$. Divide $\sum(X\text{-}\overline{X})^2$ by N. Take the square root. The result is $S = 8.66$ boxes.

For convenience, I selected raw scores that produce an integer mean and, thus, integer deviation scores. If you are confronted with decimal deviation scores, it usually works to carry three decimals in the problem and then round the final answer to two decimal places.

TABLE 4.5 Cookie sales and computation of S by the deviation-score formula

Girl Scout	Boxes of cookies X	Deviation scores $X - \bar{X}$	$(X - \bar{X})^2$
Samantha	28	18	324
Selene	11	1	1
Caitlin	10	0	0
Ann	5	–5	25
Beth	4	–6	36
Sydney	2	–8	64
	$\sum X = 60$	$\sum(X - \bar{X}) = 0$	$\sum(X - \bar{X})^2 = 450$

$$\bar{X} = \frac{\sum X}{N} = \frac{60}{6} = 10.00 \text{ boxes} \qquad S = \sqrt{\frac{\sum(X - \bar{X})^2}{N}} = \sqrt{\frac{450}{6}} = 8.66 \text{ boxes}$$

Now, what does $S = 8.66$ boxes mean? How does it help your understanding? If S was zero, it would mean that each girl sold the same number of boxes. The closer S is to zero, the more confidence you can have in predicting that the number of boxes any girl sold was equal to the mean of the group. Conversely, the further S is from zero, the less confidence you have. With $S = 8.66$ and $\bar{X} = 10.00$, you can conclude that the girls varied a great deal in cookie sales.

I realize that, so far, my interpretation of the standard deviation has not given you any more information than the range does. (A range of zero means that each girl sold the same number of boxes, and so forth.) The range, however, has no additional information to give you—the standard deviation does, as you will see in Chapter 7.

Now look again at **Table 4.5** and the formula for S. Notice what is happening. The mean is subtracted from each score. This difference, whether positive or negative, is squared and these squared differences are added together. This sum is divided by N and the square root is found. Every score in the distribution contributes to the final answer, but they don't all contribute equally. A score such as 28, which is far from the mean, makes a large contribution to $\sum(X\text{-}\bar{X})^2$. This makes sense because the standard deviation is a yardstick of variability. Scores that are far from the mean cause the standard deviation to be larger. Take a moment to think through the contribution to the standard deviation made by a score near the mean.[1]

error detection

All standard deviations are positive numbers because variability varies from zero upward. If you find yourself trying to take the square root of a negative number, you've made an error.

[1] If you play with a formula, you will become more comfortable with it and understand it better. Make up a small set of numbers and calculate a standard deviation. Change one of the numbers, or add a number, or leave out a number. See what happens.

PROBLEMS

4.4. Give the symbol and purpose of each of the three standard deviations.

4.5. Using the deviation-score method, compute S for the three sets of scores.

- ***a.** 7, 6, 5, 2
- **b.** 14, 11, 10, 8, 8
- ***c.** 107, 106, 105, 102

4.6. Compare the standard deviation of Problem 4.5a with that of Problem 4.5c. What conclusion can you draw about the effect of the size of the scores on the following?

- **a.** standard deviation
- **b.** mean

***4.7.** The temperatures listed are averages for March, June, September, and December. Calculate the mean and standard deviation for each city. Summarize your results in a sentence.

San Francisco, CA	54°F	59°F	62°F	52°F
Albuquerque, NM	46°F	75°F	70°F	36°F

4.8. No computation is needed here; just eyeball the scores in set I and set II and determine which set is more variable or whether the two sets are equally variable.

a. x: 1, 2, 4, 1, 3
y: 9, 7, 3, 1, 10

b. x: 9, 10, 12, 11
y: 4, 5, 7, 6

c. x: 1, 3, 9, 6, 7
y: 14, 15, 14, 13, 14

d. x: 8, 4, 6, 3, 5
y: 4, 5, 7, 6, 15

e. x: 114, 113, 114, 112, 113
y: 14, 13, 14, 12, 13

Computing S and σ With the Raw-Score Formula

The deviation-score formula helps you understand what is actually going on when you calculate a standard deviation. Unfortunately, except in textbook examples, the deviation-score formula almost always has you working with decimal values. The raw-score formula, which is algebraically equivalent, involves far fewer decimals. It also produces answers more quickly.

The raw-score formula is

$$S \text{ or } \sigma = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}}$$

where $\sum X^2$ = sum of the squared scores

$(\sum X)^2$ = square of the sum of the raw scores

N = number of scores

TABLE 4.6 Cookie sales and computation of *S* by the raw-score formula

Girl Scout	Boxes of cookies X	X^2
Samantha	28	784
Selene	11	121
Caitlin	10	100
Ann	5	25
Beth	4	16
Sydney	2	4
	$\sum X = 60$	$\sum X^2 = 1050$
	$(\Sigma X)^2 = 3600$	

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N}} = \sqrt{\frac{1050 - \frac{60^2}{6}}{6}} = \sqrt{\frac{450}{6}} = 8.66 \text{ boxes}$$

Table 4.6 shows the steps for calculating S or σ by the raw-score formula. The data are those for boxes of cookies sold by six Girl Scouts. The arithmetic of Table 4.6 can be expressed in words: Square the sum of the values in the X column and divide the total by N. Subtract the result from the sum of the values in the X^2 column. Divide this difference by N and find the square root. The result is S or σ. Notice that the value of S in Table 4.6 is the same as the one you calculated in Table 4.5. In this case, the mean is an integer, so the deviation scores introduced no rounding errors.[2]

$\sum X^2$ and $(\sum X)^2$: Did you notice the difference in these two terms when you were working with the data in Table 4.6? You cannot calculate a standard deviation correctly unless you recognize the difference. Reexamine Table 4.6 if you aren't sure of the difference between $\sum X^2$ and $(\sum X)^2$. Be alert for these two sums in the problems coming up.

[2]Some textbooks and statisticians prefer the algebraically equivalent formula

$$\sqrt{\frac{N\sum X^2 - (\sum X)^2}{N^2}}$$

I am using the formula in the text because the same formula, or parts of it, will be used in other procedures. Yet another formula is often used in the field of testing. To use this formula, you must first calculate the mean:

$$S = \sqrt{\frac{\sum X^2}{N} - \bar{X}^2} \quad \text{and} \quad \sigma = \sqrt{\frac{\sum X^2}{N} - \mu^2}$$

All these arrangements of the arithmetic are algebraically equivalent.

error detection

The range is usually two to five times greater than the standard deviation when $N = 100$ or less. The comparison (which can be calculated quickly) will tell you if you made any large errors in calculating a standard deviation.

PROBLEMS

***4.9.** Look at distributions **a** and **b**. Just by looking at the data, decide which one has the larger standard deviation and estimate its size. Finally, make a choice between the deviation-score and the raw-score formulas and compute S for each distribution. Compare your computation with your estimate.

a. 5, 4, 3, 2, 1, 0

b. 5, 5, 5, 0, 0, 0

4.10. For each of the distributions in Problem 4.9, divide the range by the standard deviation. Is the result between 2 and 5?

4.11. Look at distributions **a** and **b**. Note that **a** is more variable than **b** due to the difference in the lowest score. Calculate σ for each distribution using the raw-score formula. Notice the difference produced by changing one score.

a. 9, 8, 8, 7, 2

b. 9, 8, 8, 7, 6

$\hat{s}$ as an Estimate of σ

Remember that $\hat{s}$ is the principal statistic in this chapter. It will be used again and again throughout this text. A "hat" on a symbol indicates that something is being estimated. If you have sample data and you want an estimate of σ, calculate $\hat{s}$. Note that the difference between $\hat{s}$ and σ is that $\hat{s}$ has $N - 1$ in the denominator, whereas σ has N.

$$\hat{s} = \sqrt{\frac{\sum(X - \bar{X})^2}{N - 1}}$$

The issue of dividing by N or by $N - 1$ sometimes leaves students shrugging their shoulders and muttering, "OK, I'll memorize it and do it however you want." I would like to explain, however, why you use $N - 1$ in the denominator when you have sample data and want to estimate σ.

Because the formula for σ is

$$\sigma = \sqrt{\frac{\sum(X - \mu)^2}{N}}$$

it would seem logical just to calculate $\overline{X}$ from the sample data, substitute $\overline{X}$ for μ in the numerator of the formula, and calculate an answer. This solution will, more often than not, produce a numerator that is *too small* [as compared to the value of $\sum (X - \mu)^2$].

To explain this surprising state of affairs, remember (page 47) a characteristic of the mean: For any set of scores, the expression $\sum (X - \overline{X})^2$ is minimized. That is, for any set of scores, this sum is smaller when $\overline{X}$ is used than it is if some other number (either larger or smaller) is used in place of $\overline{X}$.

Thus, for a sample of scores, substituting $\overline{X}$ for μ gives you a numerator that is minimized. However, what you want is a value that is the same as $\sum (X - \mu)^2$. Now, if the value of $\overline{X}$ is at all different from μ, then the minimized value you get using $\sum (X - \overline{X})^2$ will be too small.

The solution that statisticians adopted for this underestimation problem is to use $\overline{X}$ and then to divide the too-small numerator by a smaller denominator—namely, $N - 1$. This results in a statistic that is a much better estimator of σ.[3]

You just finished four dense paragraphs and a long footnote—lots of ideas per square inch. You may understand it already, but if you don't, take 10 or 15 minutes to reread, do the exercise in footnote 3, and think.

Note also that as N gets larger, the subtraction of 1 from N has less and less effect on the size of the estimate of variability. This makes sense because the larger the sample size is, the closer $\overline{X}$ will be to μ, on average.

One other task comes with the introduction of $\hat{s}$: the decision whether to calculate σ, S, or $\hat{s}$ for a given set of data. Your choice will depend on your purpose. If your purpose is to estimate the variability of a population using data from a sample, calculate $\hat{s}$. (This purpose is common in inferential statistics.) If your purpose is to describe the variability of a sample or a population, use S or σ, respectively.

Calculating $\hat{s}$

To calculate $\hat{s}$ from raw scores, I recommend this formula:

$$\hat{s} = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N - 1}}$$

[3] To illustrate this issue for yourself, use a small population of scores and do some calculations. For a population with scores of 1, 2, and 3, σ = 0.82. Three different samples with $N = 2$ are possible in the population. For each sample, calculate the standard deviation using $\overline{X}$ and N in the denominator. Find the mean of these three standard deviations. Now, for each of the three samples, calculate the standard deviation using $N - 1$ in the denominator and find the mean of these three. Compare the two means to the σ that you want to estimate.

Mathematical statisticians use the term *unbiased estimator* for statistics whose average value (based on many samples) is exactly equal to the parameter of the population the samples came from. Unfortunately, even with $N - 1$ in the denominator, $\hat{s}$ is not an unbiased estimator of σ, although the bias is not very serious. There is, however, an unbiased measure of variability called the *variance*. The sample variance, $\hat{s}^2$, is an unbiased estimator of the population variance, σ^2. (See the Variance section that follows.)

This formula is the same as the raw-score formula for σ except for $N - 1$ in the denominator. This raw-score formula is the one you will probably use for your own data.[4]

Sometimes, you may need to calculate a standard deviation for data already arranged in a frequency distribution. For a simple frequency distribution or a grouped frequency distribution, the formula is

$$\hat{s} = \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{N}}{N - 1}}$$

where f is the frequency of scores in an interval.

Here are some data to illustrate the calculation of $\hat{s}$ both for ungrouped raw scores and for a frequency distribution. Consider puberty. As you know, females reach puberty earlier than males (about 2 years earlier on the average). Is there any difference between the genders in the *variability* of reaching this developmental milestone? Comparing standard deviations will give you an answer.

If you have only a sample of ages for each sex and your interest is in all females and all males, $\hat{s}$ is the appropriate standard deviation. **Table 4.7** shows the calculation of $\hat{s}$ for eight females. Work through the calculations. The standard deviation is 2.19 years.

TABLE 4.7 **Age of puberty and calculation of $\hat{s}$ (ungrouped raw scores)**

Age (X)	X^2
17	289
15	225
13	169
12	144
12	144
11	121
11	121
11	121
$\sum X = 102$	$\sum X^2 = 1334$

$\bar{X} = 12.75$ $\qquad$ $(\Sigma X)^2 = 10{,}404$

$$\hat{s} = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}} = \sqrt{\frac{1334 - \frac{(102)^2}{8}}{7}} = \sqrt{\frac{33.50}{7}} = 2.19 \text{ years}$$

[4] Calculators with standard deviation functions differ. Some use N in the denominator, some use $N - 1$, and some have both standard deviations. You will have to check yours to see how it is programmed.

To move from ungrouped raw scores to a simple frequency distribution, I created the data in **Table 4.8**, which shows the age of puberty of 16 males. Work through the accompanying calculations. For males, $\hat{s} = 1.44$ years.

TABLE 4.8 **Age of puberty and calculation of $\hat{s}$ for males (simple frequency distribution)**

Age (X)	f	fX	X^2	fX^2
18	1	18	324	324
17	1	17	289	289
16	2	32	256	512
15	4	60	225	900
14	5	70	196	980
13	3	39	169	507
	$N = 16$	$\sum fX = 236$		$\sum fX^2 = 3512$

$\bar{X} = 14.75$ years

$$\hat{s} = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}} = \sqrt{\frac{3512 - \frac{(236)^2}{16}}{15}} = \sqrt{\frac{31}{15}} = 1.44 \text{ years}$$

Thus, based on sample data, you can conclude that there is more variability among females in the age of reaching puberty than there is among males. (Incidentally, you would be correct in your conclusion—I chose the numbers so they would produce results that are similar to population figures.)

Here are three final points about simple and grouped frequency distributions.

- $\sum fX^2$ is found by squaring X, multiplying by f, and then summing.
- $(\sum fX)^2$ is found by multiplying f by X, summing, and then squaring.
- For grouped frequency distributions, X is the midpoint of a class interval.

error detection

$\sum f X^2$ and $(\sum f X)^2$ are similar in appearance, but they tell you to do different operations. The different operations produce different results.

For both grouped and simple frequency distributions, most calculators give you $\sum fX^2$ and $(\sum fX)^2$ if you properly key in X and f for each line of the distribution. Procedures differ depending on the brand. The time you invest in learning will be repaid several times over in future chapters (not to mention the satisfying feeling you will get).

clue to the future

You will be glad to know that in your efforts to calculate a standard deviation, you have produced two other useful statistics along the way. The number you took the square root of to get the standard deviation is the variance (see below). The expression in the numerator of the standard deviation, $\sum(X\text{-}\overline{X})^2$, is the sum of squares, which is prominent in Chapters 11–13.

Graphing Standard Deviations

The information on variability that a standard deviation conveys can often be added to a graph. **Figure 4.1** is a double-duty bar graph of the puberty data that shows standard deviations as well as means. The lines extend one standard deviation above and below the mean. From this graph, you can see at a glance that the mean age of puberty for females is younger than for males and that they are more variable in reaching puberty. One caution: Other measures of variability besides standard deviations are presented as an extended line. The figure caption or the legend tells you the meaning of the line.

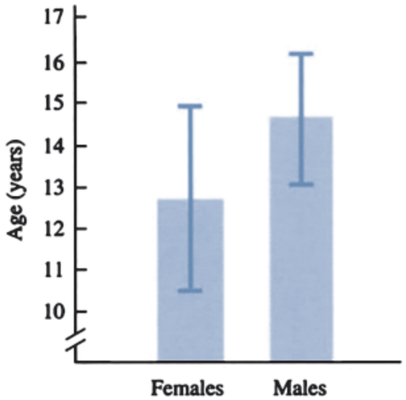


FIGURE 4.1 Bar graph of mean age of onset puberty for females and males (error bars = I standard deviation)

Variance

The last step in calculating a standard deviation is to find a square root. The number you take the square root of is the **variance**. The symbols for the variance are σ^2 (population variance) and $\hat{s}^2$ (sample variance used to estimate the population variance). By formula,

Variance
Square of the standard deviation.

$$\sigma^2 = \frac{\sum(X-\mu)^2}{N} \quad \text{and} \quad \hat{s}^2 = \frac{\sum(X-\bar{X})^2}{N-1} \quad \text{or} \quad \hat{s}^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}$$

The difference between σ^2 and $\hat{s}^2$ is the term in the denominator. The population variance uses N, and the estimate of the population variance uses $N - 1$. The variance is not very useful as a descriptive statistic. It is, however, of enormous importance in inferential statistics, especially in a technique called the *analysis of variance* (Chapters 11, 12, and 13).

Statistical Software Programs

There are many computer software programs that calculate statistics. One of the most widely used is called IBM SPSS. At several places in this book, I have included look-alike tables from IBM SPSS. **Table 4.9** has IBM SPSS output of measures of central tendency and variability of the Satisfaction With Life Scale (SWLS) scores you have been working with since Chapter 2. In IBM SPSS, *Std. Deviation* is calculated with $N - 1$ in the denominator.

TABLE 4.9 IBM SPSS output of descriptive statistics for the SWLS scores

Statistics		
SWLSscores		
N	Valid	100
	Missing	0
Mean		24.0000
Median		25.0000
Mode		27.00
Std. Deviation		6.41809
Variance		41.192
Range		30.00
Percentiles	25	21.0000
	50	25.0000
	75	28.0000

PROBLEMS

4.12. Make a freehand bar graph that shows the means and standard deviations of the Problem 4.7 temperature data for San Francisco and Albuquerque. Include a caption.

4.13. Recall the student who spent an average of \$2.90 a day at the Student Center. Suppose the student wanted to reduce Student Center spending. Write a sentence of advice if $\hat{s}$ = \$0.02 and a second sentence if $\hat{s}$ = \$2.50.

4.14. Describe in words the relationship between the variance and the standard deviation.

4.15. A researcher had a sample of freshman class scores on attitudes toward authority. She wished to estimate the standard deviation of the scores of the entire freshman class. (She had data from 20 years earlier and believed that current students were more homogeneous than students in the past.) Calculate the proper standard deviation and variance.

$$N = 21 \qquad \sum X = 304 \qquad \sum X^2 = 5064$$

4.16. Here are those data on the heights of the Americans in their 20s that you have been working with. For each group, calculate $\hat{s}$. Write a sentence of interpretation.

Women		Men	
Height (in.)	*f*	Height (in.)	*f*
72	1	77	1
71	2	76	1
69	3	75	2
68	3	74	1
67	5	73	4
66	6	72	7
65	9	71	6
64	7	70	7
63	6	69	8
62	4	68	4
61	1	67	2
60	2	66	2
59	1	65	3
		64	1
		62	1

4.17. A high school English teacher measured the attitudes of 11th-grade students toward poetry. After a 9-week unit on poetry, she measured the students' attitudes again. She was disappointed to find that the mean change was zero. Following are some representative scores. (High scores indicate favorable attitudes.) Calculate $\hat{s}$ for before scores and after scores, and write a conclusion based on the standard deviations.

Before	7	5	3	5	5	4	5	6
After	9	8	2	1	8	9	1	2

4.18. Two widely recognized personality characteristics are agreeableness (high scorers are sympathetic and altruistic; low scorers are egocentric, competitive, and antagonistic) and conscientiousness (high scorers are purposeful, self-controlled, and reliable; low scorers are lackadaisical and hedonistic). The samples that follow approximate the norms for college-age individuals (McCrae & Costa, 2010). Find means and standard deviations. Write an interpretation.

Agreeableness	122	106	92	136	97	129	116
Conscientiousness	137	93	123	143	88	105	109

4.19. In manufacturing, engineers strive for consistency. The following data are the errors, in millimeters, in giant gizmos manufactured by two different processes. Choose S or $\hat{s}$ and determine which process produces the more consistent gizmos.

Process A	0	1	–2	0	–2	3
Process B	1	–2	–1	1	–1	2

***4.20.** Find the interquartile range for normal oral body temperature and write a sentence of interpretation. For data, use your answer to Problem 2.17a.

4.21. Estimate the population standard deviation for oral body temperature using data you worked with in Problem 2.17a. You may find $\sum fX$ and $\sum fX^2$ by working from the answer to Problem 2.17a, or, if you understand what you need to do to find these values from that table, you can use $\sum fX = 3928.0$ and $\sum fX^2 = 385{,}749.24$.

CP **4.22.** Reread the description of the classic study by Bransford and Franks (Problem 2.4, p. 34). Using the frequency distribution you compiled for that question (or the answer in Appendix G), calculate the mean, median, and mode. Find the range and the interquartile range. After considering why Bransford and Franks would gather such data, calculate an appropriate standard deviation and variance. Write a paragraph explaining what your calculations show.

4.23. Return to the objectives at the beginning of the chapter. Can you do each one?

KEY TERMS

Deviation score (p. 64)
Interquartile range (p. 61)
Percentile (p. 62)
Range (p. 61)
Standard deviation (p. 63)
Variability (p. 59)
Variance (p. 73)

CHAPTER

5

Other Descriptive Statistics

OBJECTIVES FOR CHAPTER 5

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

1. Use z scores to compare two scores in one distribution
2. Use z scores to compare scores in one distribution with scores in a second distribution
3. Construct and interpret boxplots
4. Identify outliers in a distribution
5. Calculate an effect size index and interpret it
6. Compile descriptive statistics and an explanation into a Descriptive Statistics Report

THIS CHAPTER INTRODUCES four statistical techniques that have two things in common. First, the four techniques are descriptive statistics. Second, each one combines two or more statistical elements. Fortunately, you studied all the elements in the preceding two chapters. The techniques in this chapter should prove quite helpful as you improve your ability to explore data, understand it, and convey your understanding to others.

The first statistic, the z score, tells you the relative standing of a raw score in its distribution. The formula for a z score combines a raw score with its distribution's mean and standard deviation. A z-score description of a raw score works regardless of the kind of raw scores or the shape of the distribution. This first section also defines *outliers*, which are extreme scores in a distribution. Suggestions of what to do about them are offered.

The second section covers boxplots. A boxplot is a graphic that displays the scores of one variable, much like a frequency polygon, but it conveys a lot more information than a polygon. With just one picture, a boxplot gives you the mean, median, range, interquartile range, and skew of the distribution.

This chapter introduces d, which is an *effect size index*. Effect size indexes show the size of a difference between two distributions. The two distributions often are two levels of an independent variable. If the difference is so small as to be of no consequence, investigation stops. Large differences invite further research. Effect size indexes are prominent in the "new statistics." Many journals require them.

The final section of this chapter has no new statistics. It shows you how to put together a Descriptive Statistics Report, which is an organized collection of descriptive statistics that helps a reader understand a set of data.

Probably all college students are familiar with measures of central tendency; most are familiar with measures of variability. However, only those with a good education in quantitative thinking are familiar with all the techniques presented in this chapter. So, learn this material. You will then understand more than most and you will be better equipped to explain what you understand.

Describing Individual Scores

Suppose one of your friends says he got a 95 on a math exam. What does that tell you about his mathematical ability? From your previous experience with tests, 95 may seem like a pretty good score. This conclusion, however, depends on a couple of assumptions, and unless those assumptions are correct, a score of 95 tells you very little. Let's return to the conversation with your friend.

After you say, "95! Congratulations," suppose he tells you that 200 points were possible. Now a score of 95 seems like something to hide. "My condolences," you say. But then he tells you that the highest score on that difficult exam was 105. Now 95 has regained respectability and you chortle, "Well, all right!" In response, he shakes his head and tells you that the mean score was 100. The 95 takes a nose dive. As a final blow, you find out that 95 was the lowest score, that nobody scored worse than your friend. With your hand on his shoulder, you cut off further discussion of the test with, "Come on, I'll buy you an ice cream cone."

This example illustrates that the meaning of a score of 95 depends on the rest of the test scores. Fortunately, there are several ways to convert a raw score into a measure that signals its place among the array of its fellow scores. Percentiles are one well-known example. I'll explain two others that are important for statistical analyses: *z* scores and outliers.

The *z* score

***z* score**
Score expressed in standard deviation units.

After being converted to a *z* **score**, a raw score's place among its fellow scores is revealed.

$$z = \frac{X - \overline{X}}{S}$$

The formula is a deviation score ($X - \overline{X}$) divided by a standard deviation (S). A positive deviation score tells you the raw score is above average; negative means the score is below average. However, deviation scores don't tell you how far above or below. For example, with a mean of 50, a deviation score of +5 is above the mean, but you have no idea how much above average. If the distribution has a range of 10 units, 55 is probably a top score. If the range is 100 units, +5 is just barely above average. Look at **Figure 5.1**, which is a picture of the ideas in this paragraph.

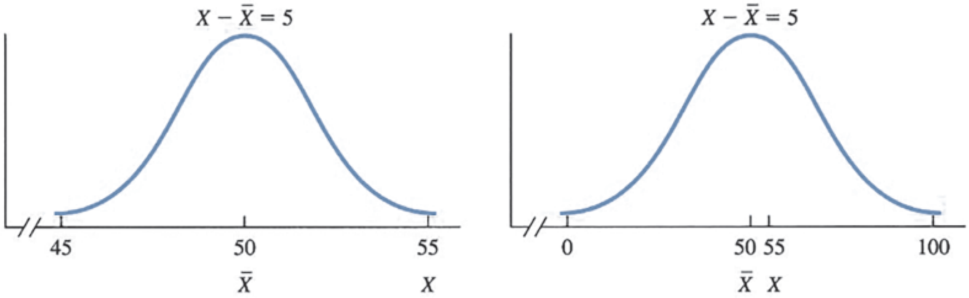


FIGURE 5.1 A comparison of the distance bewteen $\bar{X}$ and X for a distribution with a small standard deviation (left) and a large standard deviation (right)

Standard score
Score expressed in standard deviation units; *z* is one example.

To know a score's position in a distribution, the variability of the distribution must be taken into account. The way to do this is to divide X- $\bar{X}$ by a unit that measures variability, the standard deviation. The result is a deviation score per unit of standard deviation.[1] A *z* score is also referred to as a **standard score** because it is a deviation score expressed in standard deviation units.

Any distribution of raw scores can be converted into a distribution of *z* scores. The mean of a distribution of *z* scores is 0; its standard deviation is 1. A *z* score tells you the number of standard deviations a raw score is from the mean and whether it is above or below it. Thus, a *z* score of –3.0 represents a raw score far below the mean. (Three standard deviations below the mean is near the lowest point of a distribution.) If two raw scores in a distribution are converted to *z* scores, their relationship reveals their relative positions in the distribution. Finally, *z* scores are also used to compare two scores from different distributions, even if the scores are measuring different things. (If this seems like trying to compare apples and oranges, see Problem 5.5.)

In the general psychology course I took as a freshman, the professor returned tests with a *z* score rather than a percentage score. This *z* score was the key to figuring out your grade. A *z* score of +1.50 or higher was an A, and –1.50 or lower was an F. (*z* scores between +0.50 and +1.50 received Bs. If you assume the professor practiced symmetry, you can figure out the rest of the grading scale.)

Table 5.1 lists the raw scores (percentage correct) and z scores for four of the many students who took two tests in that class. Begin by noting that for Test 1, $\bar{X} = 54$ and $S = 10$. For Test 2, $\bar{X} = 86$ and $S = 6$. Well, so what? To answer this question, examine the scores of the four individuals. Consider the first student, Kris, who scored 76 on both tests. The two 76s appear to be the same, but the *z* scores show that they are not. The first 76 was a high A, and the second 76 was an F.

[1] This technique is based on the same idea that percent is based on. For example, 24 events at one time and 24 events a second time appear to be the same. But don't stop with appearances. Ask for the additional information that you need, which is, "24 events out of how many chances?" An answer of "50 chances the first time and 200 chances the second" allows you to convert both 24s to "per centum" (per one hundred). Clearly, 48 percent and 12 percent are different, even though both are based on 24 events. Other examples of this technique of dividing to make raw scores comparable include miles per gallon, per capita income, bushels per acre, and points per game.

The second student, Robin, appears to have improved on the second test, going from 54 to 86, but Robin's performance, relative to the class, was the same on both tests as revealed by z scores (z = 0, the class average). Marty also appears to have improved if you examine only the raw scores. However, the z scores reveal that Marty did *worse* on the second test. Finally, comparing Terry's and Robin's raw scores, you can see that although Terry scored four points higher than Robin on each test, the z scores show that Terry's improvement on the second test was greater than Robin's.

TABLE 5.1 Raw scores and *z* scores of selected students on two 100-point tests in general psychology

	Test 1		Test 2	
Student	Raw score	z score	Raw score	z score
Kris	76	+2.20	76	−1.67
Robin	54	.00	86	.00
Marty	58	+.40	82	−.67
Terry	58	+.40	90	+.67
	Test 1: $\bar{X}$ = 54		Test 2: $\bar{X}$ = 86	
	S = 10		S = 6	

The reason for these surprising comparisons is that the means and standard deviations were so different for the two tests. Perhaps the second test was easier, or the material was more motivating to students, or the students studied more. Maybe the teacher prepared better. Perhaps all of these reasons were true.

To summarize, z scores give you a way to compare raw scores. The basis of the comparison is the distribution itself rather than some external standard (such as a grading scale of 90%–80%–70%–60% for As, Bs, and so on).

A word of caution: the letter z is used as both a descriptive statistic and an inferential statistic. As a descriptive statistic, its range is limited. For a distribution of 100 or so scores, the z scores might range from approximately –3 to +3. For many distributions, especially when N is small, the range is less.

As an inferential statistic, however, z values are not limited to ±3. For example, z-score tests are used in Chapter 15 to help decide whether two populations are different. The value of z depends heavily on how different the two populations actually are. When z is used as an inferential statistic, values much greater than 3 can occur.

clue to the future

> z scores will turn up often as you study statistics. They are prominent in this book in Chapters 5, 7, 8, and 15.

Outliers

Outlier
An extreme score separated from the others and at least 1.5 × IQR beyond the 25th or 75th percentile.

Outliers are scores in a distribution that are unusually small or unusually large. They have a disproportionate influence, compared to any of the other scores, on the mean, standard deviation, and other statistical measures. They can certainly affect the outcome of a statistical analysis, and they appear to be common (Wilcox, 2005a).

Although there is no general agreement on how to identify outliers, Hogan and Evalenko (2006) found that the most common definition in statistics textbooks is

Lower outlier = 25th percentile – (1.5 × IQR)
Upper outlier = 75th percentile + (1.5 × IQR)

Using these definitions and the heights of 20- to 29-year-old women and men, we can determine heights that qualify as outliers.

For women:
Lower outlier = 25th percentile – (1.5 IQR) = 63 – 1.5(4) = 57 inches or shorter
Upper outlier = 75th percentile + (1.5 IQR) = 67 + 1.5(4) = 73 inches or taller

For men:
Lower outlier = 25th percentile – (1.5 IQR) = 68 – 1.5(4) = 62 inches or shorter
Upper outlier = 75th percentile + (1.5 IQR) = 72 + 1.5(4) = 78 inches or taller

What should you do if you detect an outlier in your data? Answer: *think*. Could the outlier score be a recording error? Is there a way to check? The outlier score may not be an error, of course. Each of us probably knows someone who is taller or shorter than the outlier heights identified above. Nevertheless, outliers distort means, standard deviations, and other statistics. Fortunately, mathematical statisticians have developed statistical techniques for data with outliers (Wilcox, 2005b), but these are typically covered in advanced courses.

PROBLEMS

5.1. The mean of any distribution has a z score equal to what value?

5.2. What conclusion can you reach about $\sum z$?

5.3. Under what conditions would you prefer that your personal z score be negative rather than positive?

5.4. Jayla and Ayana, twin sisters, were intense competitors, but they never competed against each other. Jayla specialized in long-distance running and Ayana was an excellent sprint swimmer. As you can see from the distributions in the accompanying table, each was the best in her event. Take the analysis one step further and use z scores to determine who is the more outstanding twin. You might start by looking at the data and making an estimate.

10k Runners	Time (min.)	50m Swimmers	Time (sec.)
Jayla	37	Ayana	24
Dott	39	Ta-Li	26
Laqueta	40	Deb	27
Marette	42	Aisha	28

5.5. Tobe grows apples and Zeke grows oranges. In the local orchards, the mean weight of apples is 5 ounces, with S = 1.0 ounce. For oranges, the mean weight is 6 ounces, with S = 1.2 ounces. At harvest time, each entered his largest specimen in the Warwick County Fair. Tobe's apple weighed 9 ounces and Zeke's orange weighed 10 ounces. This particular year, Tobe was ill on the day of judgment, so he sent his friend Hamlet to inquire who had won. Adopt the role of judge and use z scores to determine the winner. Hamlet's query to you is: "Tobe, or not Tobe; that is the question."

5.6. Taz's anthropology professor drops the poorest exam grade in the term. Taz scored 79 on the first exam. The mean was 67 and the standard deviation, 4. On the second exam, Taz made 125. The class mean was 105 and the standard deviation, 15. On the third exam, the mean was 45 and the standard deviation, 3. Taz got 51. Which test should be dropped?

5.7. Using your answer to Problem 4.20, determine which of the following temperatures qualifies as an outlier.

a. 98.6° F   d. 96.6° F
b. 99.9° F   e. 100.5° F
c. 96.0° F

Boxplots

Boxplot
Graph that shows a distribution's range, interquartile range, skew, median, and sometimes other statistics.

At the beginning of Chapter 3, I said that to understand a distribution, you need information about central tendency, variability, and form. A **boxplot** provides all three. In fact, a typical boxplot gives you two measures of central tendency, two measures of variability, and two ways to estimate skewness. Skew, of course, is a description of the form of a distribution.[2] Because boxplots are so informative, they are helpful during the initial exploratory stages of data analysis and also later when explanations are given to others.

Boxplots can be oriented horizontally or vertically. **Figure 5.2** shows both versions for the Satisfaction With Life Scale (SWLS) scores you first encountered in Chapter 2. Vertical versions put the scores of the variable on the upright axis; horizontal versions display scores horizontally. An older name for this graphic is "box and whisker plot," and as you can see, it consists of a *box* and *whiskers* (and, within the box, a *line* and a *dot*).

[2] Boxplots were dreamed up by John Tukey (1915–2000), who invented several statistical techniques that facilitate the exploration of data. See Lovie's (2005) entry on Exploratory Data Analysis.

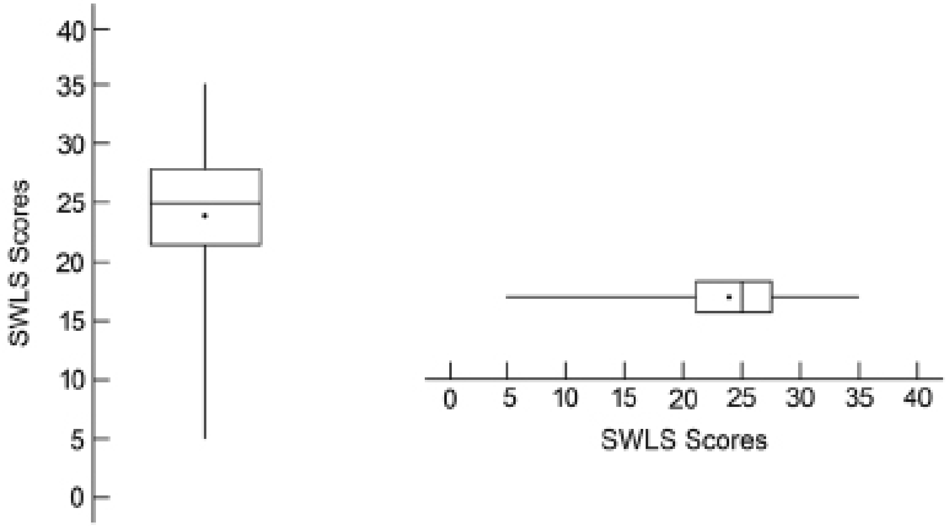


FIGURE 5.2 **Vertical and horizontal boxplots of Satisfaction With Life Scale scores**

Interpreting Boxplots

Central tendency A boxplot gives two measures of central tendency. The line inside the box is at the median. The dot indicates the mean. In either panel of **Figure 5.2**, you can estimate the median as 25 and the mean as slightly less (it is actually 24).

Variability Both the box and the whiskers in a boxplot tell you about variability. The box covers the interquartile range, which you studied in Chapter 4. The bottom of the box (vertical orientation) and the left end of the box (horizontal orientation) align with the 25th percentile score. The top of the box and the right end correspond to the 75th percentile score. In **Figure 5.2**, you can estimate these scores as 21 and 28, respectively. Thus, IQR = 7.

The *whiskers* extend from the box to the most extreme scores in the distribution. Thus, the whisker tips tell you the range. Reading from the scales in Figure 5.2, you can see that the highest score is 35 and the lowest is 5.

Skew Both the relationship of the mean to the median and any difference in the lengths of the whiskers help you determine the skew of the distribution. You already know that, in general, when the mean is less than the median, the skew is negative; when the mean is greater than the median, the skew is positive. The relationship of the mean to the median is readily apparent in a boxplot.

Often, but not always, skew is indicated by whiskers of different length. If the longer whisker corresponds to the lower scores, the skew is negative; if the longer whisker goes with the higher scores, the skew is positive. Given these two rules of thumb to determine skew, you can conclude that the distribution of SWLS scores in Figure 5.2 is negatively skewed.

Variations in boxplots Boxplots are versatile; they are easily modified to present more information or less information. Outliers can be indicated with marks or asterisks beyond the end of the whiskers. If the data do not permit the calculation of a mean, eliminate the dot. Perhaps most importantly, several groups can be shown on one graphic in an uncluttered way (as compared to several frequency polygons on one axis).

Boxplot questions Look at **Figure 5.3**, which shows the boxplots of four distributions. Seven questions follow. See how many you get right. If you get all of them right, consider skipping the explanations that follow the answers.

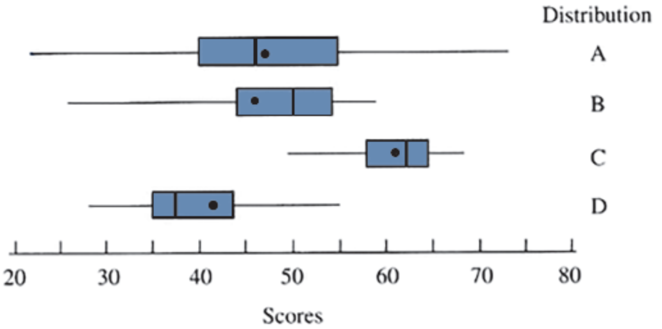


FIGURE 5.3 **Boxplots of four distributions**

Questions

1. Which distribution has the greatest positive skew?
2. Which distribution is the most compact?
3. Which distribution has a mean closest to 40?
4. Which distribution is most symmetrical?
5. Which distribution has a median closest to 50?
6. Which distribution is most negatively skewed?
7. Which distribution has the greatest range?

Answers

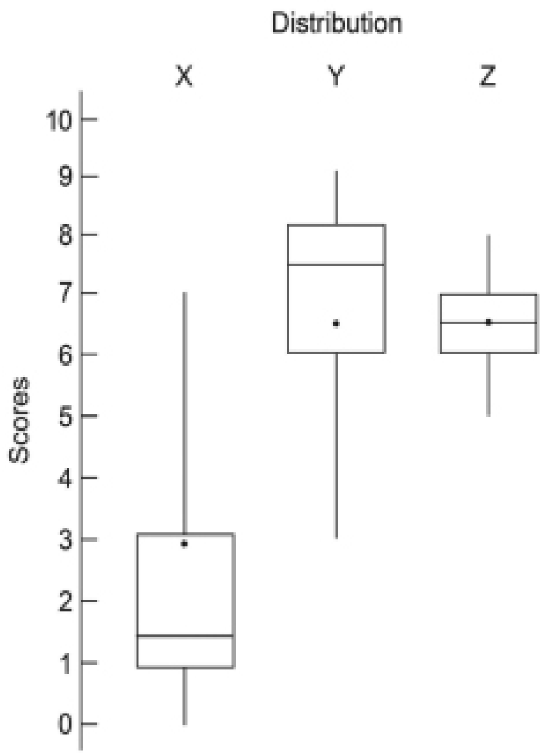
1. *Positive skew*: Distribution D. The mean is greater than the median, and the high-score whisker is longer than the low-score whisker.
2. *Most compact:* Distribution C. The range is smaller than other distributions.
3. *Mean closest to 40:* Distribution D.
4. *Most symmetrical:* Distribution A. The mean and median are about the same, and the whiskers are about the same length.
5. *Median closest to 50:* Distribution B.
6. *Most negative skew:* Distribution B. The difference between the mean and median is greater in B than in C, and the difference in whisker length is greater in B than in C.
7. *Greatest range:* Distribution A.

error detection

A computer-generated boxplot and familiarity with boxplot interpretation can reveal gross errors such as impossible or improbable scores.

PROBLEMS

5.8. Tell the story (mean, median, range, interquartile range, and form) of each of the three boxplots in the figure that follows.



5.9. You have already found the elements needed for boxplots of the heights of 20- to 29-year-old women and men. (See Problems 2.1, 3.4, and 4.3.) Draw boxplots of the two groups using one horizontal axis.

***5.10.** Create a horizontal boxplot (but without the mean) of the oral body temperature data based on Mackowiak et al. (1992). Find the statistics you need from your answers to Problems 2.17 and 4.20 (Appendix G).

Effect Size Index

- Men are taller than women.
- First-born children score higher than second-born children on tests of cognitive ability.
- Interrupted tasks are remembered better than uninterrupted tasks.

These three statements follow a pattern that is common when quantitative variables are compared. The pattern is, “The average *X* is greater than the average *Y*.” It happens that the three statements are true on average, but they leave an important question unanswered. How much taller, higher, or better is the first group than the second? This is a question of *effect size*. There are several ways to convey effect size. Some are everyday expressions, and some are statistics calculated from data.

For the heights of men and women, a satisfactory indication of effect size is that men, on the average, are about 5 inches taller than women. But how satisfactory is it to know that the mean difference in cognitive ability scores of first-born and second-born children is 13 points or that the difference in recall of interrupted and uninterrupted tasks is three tasks? What is needed is a statistic that works when the measurement scale is unfamiliar. Such statistics are called **effect size indexes**. Kirk (2005) describes several.

Effect size index
Amount or degree of separation between two distributions.

The Effect Size Index, *d*

Probably the most common effect size index is *d*, where

$$d = \frac{\mu_1 - \mu_2}{\sigma}$$

Thus, *d* is a difference between means per standard deviation unit. To calculate *d*, you must estimate the parameters with statistics. Samples from the μ_1 population and the μ_2 population produce $\overline{X}_1$ and $\overline{X}_2$. Calculating an estimate of σ requires knowing about degrees of freedom, a concept that is better introduced in connection with hypothesis testing (Chapter 8 and chapters that follow). So, at this point you will have to be content to have σ given to you.

You can easily see that if the difference between means is zero, then $d = 0$. Also, depending on σ, a given difference between two means might produce a small *d* or a large *d.* The sign of *d* depends on which group is assigned 1 and which is assigned 2. If this decision is arbitrary, the sign of *d* is unimportant. However, in a comparison of an experimental group and a control group, it is conventional to designate the experimental group as Group 1.

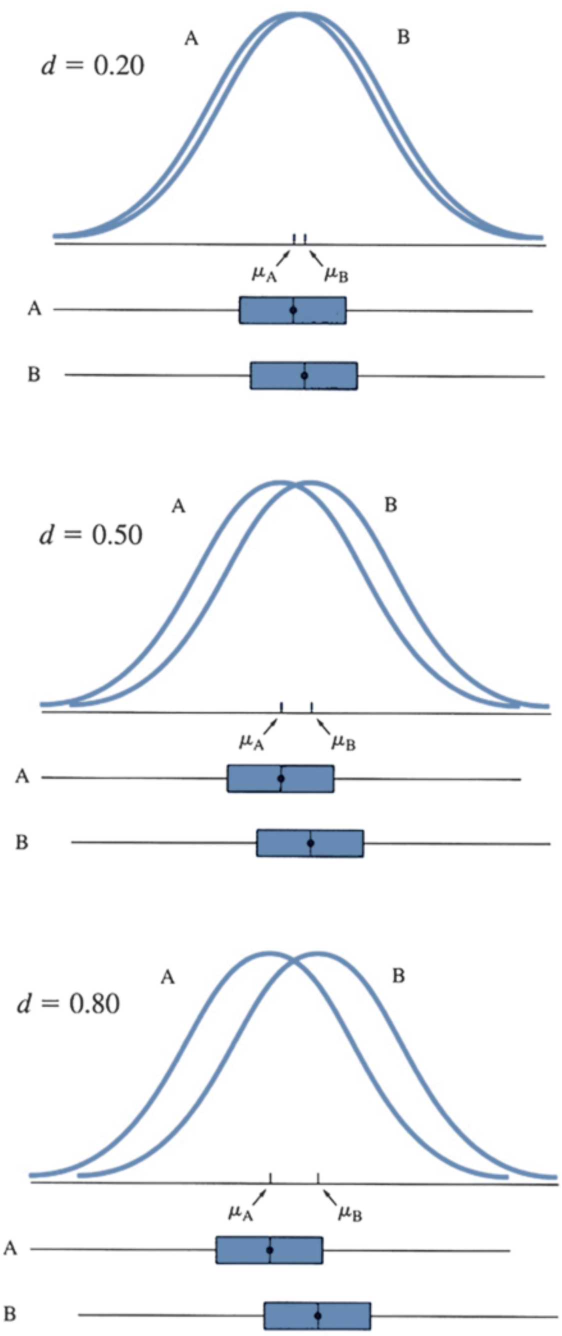


FIGURE 5.4 **Frequency ploygons and boxplots of two populations that differ by small (d=0.20), medium (d=0.50) and large (d=0.80) amounts**

The Interpretation of *d*

A widely accepted convention of what constitutes small, medium, and large effect sizes was proposed by Jacob Cohen (1969)[3], who was on our exploration tour in Chapter 1.

Small effect	$d = 0.20$
Medium effect	$d = 0.50$
Large effect	$d = 0.80$

To get a visual idea of these d values, look at **Figure 5.4**. The three tiers illustrate small, medium, and large values of d for both frequency polygons and boxplots. In the top panel, the mean of Distribution B is two-tenths of a standard deviation unit greater than the mean of Distribution A (d = 0.20). You can see that there is a great deal of overlap between the two distributions. Study the other two panels of Figure 5.4, examining the amount of overlap for $d = 0.50$ and $d = 0.80$.

To illustrate further, let's take the heights of women and men. Just intuitively, what adjective would you use to describe the difference in the heights of women and men? A small difference? A large difference?

[3] For an easily accessible source of Cohen's reasoning in proposing these conventions, see Cohen (1992), p. 99.

Well, gender height differences are certainly obvious, ones that everyone sees. Let's find the effect size index for the difference in heights of women and men. You already have estimates of μ_{women} and μ_{men} from your work on Problem 3.6: $\bar{X}_{\text{women}} = 65.1$ inches and $\bar{X}_{\text{men}} = 70.0$ inches. For this problem, $\sigma = 2.8$ inches. Thus,

$$d = \frac{\mu_1 - \mu_2}{\sigma} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma} = \frac{65.1 - 70.0}{2.8} = -1.75$$

The interpretation of $d = -1.75$ is that the difference in heights of women and men is just huge, more than twice the size of a difference that would be designated "large." So, here is a reference point for effect size indexes. If a difference is so great that everyone is aware of it, then the effect size index, d, will be greater than large.

Let's take another example. Women, on average, have higher verbal scores than men do. What is the effect size index for this difference? To answer this question, I consulted Hedges and Nowell (1995). Their analysis of six studies with very large representative samples and a total $N = 150{,}000$ revealed that $\bar{X}_{\text{women}} = 513$, $\bar{X}_{\text{men}} = 503$, and $\sigma = 110$.[4] Thus,

$$d = \frac{\mu_1 - \mu_2}{\sigma} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma} = \frac{513 - 503}{110} = 0.09$$

A d value of 0.09 is less than half the size of the value considered "small." Thus, you can say that although the average verbal ability of women is better than that of men, the difference is very small.

Figure 5.5 shows overlapping frequency polygons with d values of 1.75 and 0.09, the d values found in the two examples in this section.

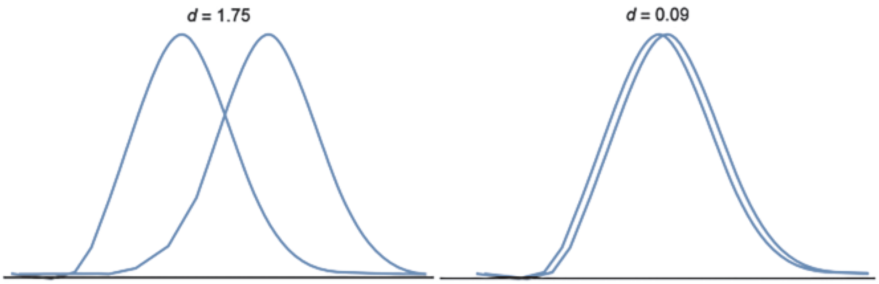


FIGURE 5.5 Frequency polygons that show effect size indexes of 1.75 (heights of men and women) and 0.09 (verbal scores of men and women)

Cohen's rule-of-thumb conventions for d are common fare in statistics textbooks, which present problems and exercises singly and not with other studies on the same topic. Actual research projects, however, are embedded in a literature of other studies on similar topics. In those cases, it is always more informative to interpret d by comparing it to other d values rather than to rely on rule-of-thumb adjectives.

[4] I created the means and standard deviation, which are similar to SAT scores, so the result would mirror the conclusions of Hedges and Nowell (1995).

clue to the future

An effect size index turns up again in Chapters 6, 9, 10, 11, 13, and 14. Effect size indexes are an increasingly important addition to the toolkit of researchers. They are usually required in journals that publish quantitative data.

The Descriptive Statistics Report

Techniques from this chapter and the previous two can be used to compile a Descriptive Statistics Report, which gives you a fairly complete story for a set of data.[5] The most interesting Descriptive Statistics Reports are those that compare two or more distributions of scores. To compile a Descriptive Statistics Report for two groups, you should (a) construct boxplots, (b) find the effect size index, and (c) tell the story that the data reveal. As for telling the story, cover the following points, arranging them so that your story is told well.

- Form of the distributions
- Central tendency
- Overlap of the two distributions
- Interpretation of the effect size index

To illustrate a Descriptive Statistics Report, I'll use the heights of the men and women that you began working with in Chapter 2. The first tasks are to assemble the statistics needed for boxplots and to calculate an effect size index. Look at **Table 5.2**, which shows these statistics. The next step is to construct boxplots (your answer to Problem 5.9). The final task is to write a paragraph of interpretation. To write a paragraph, I recommend that you make notes and then organize your points, selecting the most important one to lead with. Write a rough draft. Revise the draft until you are satisfied with it.[6] My final version is **Table 5.3**, which is a Descriptive Statistics Report of the heights of women and men.

TABLE 5.2 **Descriptive statistics for a Descriptive Statistics Report of the heights of women and men**

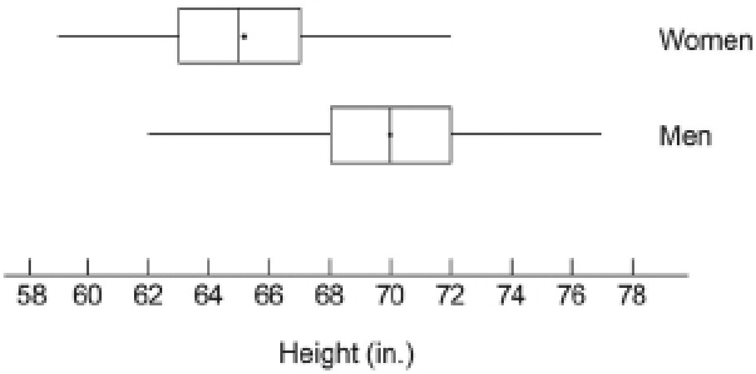
	Heights of 20- to 29-year-old Americans	
	Women (in.)	Men (in.)
Mean	65.1	70.0
Median	65	70
Minimum	59	62
Maximum	72	77
25th percentile score	63	68
75th percentile score	67	72
Effect size index	1.75	

[5] A more complete report contains inferential statistics and their interpretation.

[6] For me, several revisions are needed. This paragraph got 10.

TABLE 5.3 A Descriptive Statistics Report on the heights of women and men

The graph shows boxplots of heights of American women and men, aged 20-29. The difference in means produces an effect size index of $d = 1.75$.



The mean height of women is 65.1 inches; the median is 65 inches. The mean height of men is 70.0 inches; the median is also 70 inches. Men are about 5 inches taller than women, on average. Although the two distributions overlap, the designation "67 inches or taller" applies to more than 75% of the men but only about 25% of the women. This difference in the two distributions is reflected by an effect size index (d) of 1.75, a very large value. (A value of 0.80 is traditionally considered large.) The heights of women and men are distributed fairly symmetrically.

I will stop with just one example of a Descriptive Statistics Report. The best way to learn and understand is to create reports of your own. Thus, problems follow shortly.

For the first (but not the last) time, I want to call your attention to the subtitle of this book: *Tales of Distributions*. A Descriptive Statistics Report is a tale of distributions of scores. What on earth would be the purpose of such stories?

The purpose might be to better understand a scientific phenomenon that you are intensely interested in. The purpose might be to explain to your boss the changes that are taking place in your industry; perhaps it is to convince a quantitative-minded customer to place a big order with you. Your purpose might be to better understand a set of reports on a troubled child (perhaps your own). At this point in your efforts to educate yourself in statistics, you have a good start toward being able to tell the tale of a distribution of data. Congratulations! And, oh yes, here are some more problems so you can get better.

PROBLEMS

5.11. Find the effect size indexes for the three data sets in the table. Write an interpretation for each d value, using Cohen's guidelines.

	Group 1 mean	Group 2 mean	Standard deviation
Set a.	14	12	4
Set b.	10	19	10
Set c.	10	19	30

5.12. In hundreds of studies with many thousands of participants, first-born children scored better than second-born children on tests of cognitive ability. Data based on Zajonc (2001) and Zajonc and Bargh (1980) provide mean scores of 469 for first-born and 456 for second-born. For this measure of cognitive ability, $\sigma = 110$. Find *d* and write a sentence of interpretation.

5.13. HAVING THE BIGGEST is a game played far and wide for fun, fortune, and fame (such as being listed in Guinness World Records). For example, the biggest cabbage was grown in 2012 by Scott Robb in Alaska. It weighed 138 pounds. The biggest pumpkin (2625 pounds) was grown by Mathias Willemijns of Belgium in 2016. Calculate $\overline{X}$ and S from the three scores below, which are representative of contest cabbages and pumpkins. Use z scores to determine the BIG winner between Robb and Willemijns.

Cabbages	Pumpkins
117	2200
100	1900
83	1600

5.14. Among legendary cattle ranches in America, some storytellers include the ephemeral *Stats Bar-X Ranch*. After round-up one year, several outfits competed. The Bar-X cowboys were best at knife throwing, and nobody beat the Dragging y at lassoing, but who was "best of all"? Being modern cowboys, they used statistics rather than six-shooters to settle the question. Analyze the accuracy scores that follow with z scores to determine which outfit was more outstanding.

	Knife throwing	Lassoing
Stats Bar-X	75	70
Dragging *y*	60	75
O RO Ranch	50	65
Babbitt Ranches	55	60
Four Sixes	40	65

5.15. For the Satisfaction With Life Score data, determine the highest score and the lowest score that qualify as outliers. What scores (if any) in Table 2.3 are outliers? See Problem 4.2 for the percentiles you need.

5.16. Is psychotherapy effective? This first-class question has been investigated by many researchers. The data that follow were constructed to mirror the classic findings of Smith and Glass (1977), who analyzed 375 studies.

The psychotherapy group received treatment during the study; the control group did not. A participant's psychological health score at the beginning of the study was subtracted from the same participant's score at the end of the study, giving the score listed in the table. Thus, each score provides a measure of the change in psychological health for an individual. (A negative score means that the person was worse at the end of the study.) For these change scores, $\sigma = 10$. Create a Descriptive Statistics Report.

Psychotherapy	Control	Psychotherapy	Control
7	9	13	–3
11	3	–15	22
0	13	10	–7
13	1	5	10
–5	4	9	–12
25	3	15	21
–10	18	10	–2
34	–22	28	5
7	0	–2	2
18	–9	23	4

KEY TERMS

Boxplot (p. 82)
Descriptive Statistics Report (p. 89)
Effect size index (p. 86)
Interquartile range (p. 81)
Outliers (p. 81)
Skew (p. 83)
Standard score (p. 79)
z score (p. 78)

Transition Passage

To Bivariate Statistics

So far in this exploration of statistics, the raw data have appeared as a single string of measurements of one variable. Heights, dollars, Satisfaction With Life Scale scores, and number of boxes of Girl Scout cookies were all analyzed, but in every case (except for line graphs), the statistics were calculated on the scores of just one variable. These data are called *univariate distributions*.

The chapter that follows is about statistics that are calculated for a distribution of two variables, which are called *bivariate distributions*. In a bivariate distribution, each score on one variable is paired with a score on the other variable. Analyses of bivariate distributions reveal answers to questions about the relationship between the two variables. For example, the questions might be

- What is the relationship between a person's verbal ability and mathematical ability?
- Knowing a person's verbal aptitude score, what should we predict as his or her freshman grade point average?

Other pairs of variables that might be related include the following:

- Height of daughters and height of their fathers
- Stress and infectious diseases
- Size of groups taking college entrance examinations and scores received

By the time you finish Chapter 6, you will know whether or not these pairs of variables are related and, if so, the direction and degree of the relationship.

Chapter 6, "Correlation and Regression," explains two statistical methods. *Correlation* is a method used to determine the direction and degree of relationship between two variables. *Regression* is a method used to predict scores for one variable when you have measurements on a second variable.

CHAPTER

6

Correlation and Regression

OBJECTIVES FOR CHAPTER 6

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

1. Explain the difference between univariate and bivariate distributions
2. Explain the concept of correlation and the difference between positive and negative correlation
3. Draw scatterplots
4. Compute a Pearson product-moment correlation coefficient, *r*
5. Discuss the effect size index for *r*
6. Calculate and discuss common variance
7. Recognize correlation coefficients that indicate a reliable test
8. Discuss the relationship of correlation to cause and effect
9. Identify situations in which a Pearson *r* does not accurately reflect the degree of relationship
10. Name and explain the elements of the regression equation
11. Compute regression coefficients and fit a regression line to a set of data
12. Interpret the appearance of a regression line
13. Predict scores on one variable based on scores from another variable

CORRELATION AND REGRESSION: My guess is that you have some understanding of the concept of correlation and that you are not as comfortable with the word *regression*. Speculation aside, correlation is simpler. Correlation is a statistical technique that describes the direction and degree of relationship between two variables.

Regression is more complex. In this chapter, you will use the regression technique to accomplish two tasks, *drawing* the line that best fits the data and *predicting* a person's score on one variable when you know that person's score on a second, correlated variable. Regression has other uses, but you will have to put those off until you study more advanced statistics.

The ideas identified by the terms *correlation* and *regression* were developed by Sir Francis Galton in England well over 100 years ago. Galton was a genius (he could read at age 3) who had an amazing variety of interests, many of which he actively pursued during his 89 years. He once listed his occupation as "private gentleman," which meant that he had inherited money and did not have to work at a job. Lazy, however, he was not. Galton traveled widely and wrote prodigiously (17 books and more than 200 articles).

From an early age, Galton was enchanted with counting and quantification. Among the many things he tried to quantify were weather, individuals, beauty, characteristics of criminals,