CHAPTER 2

Descriptive Statistics

The Univariate Case

In this chapter we will consider descriptive techniques designed to summarize data on a single dependent variable. These techniques are often said to be **univariate** in nature because only one variable is involved. (In Chapter 3, we will look at several techniques designed for the **bivariate** case—that is, for situations where data have been collected on two dependent variables.)

We begin this chapter by looking at several ways data can be summarized using pictures. These so-called picture techniques include frequency distributions, stem-and-leaf displays, histograms, and bar graphs. Next, the topic of distributional shape is considered; here, you will learn what it means when a data set is said to be normal, skewed, bimodal, or rectangular. After that, we examine the concept of central tendency and various methods used to represent a data set's average score. We then turn our attention to how researchers usually summarize the variability, or spread, within their data sets; these techniques include four different kinds of range, the standard deviation, and the variance. Finally, we consider two kinds of standard scores: z and T.

Picture Techniques

In this section, we consider some techniques for summarizing data that produce a picture of the data. I use the term *picture* somewhat loosely, since the first technique really leads to a table of numbers. In any event, our discussion of descriptive statistics begins with a consideration of three kinds of frequency distributions.

Frequency Distributions

A **frequency distribution** shows how many people (or animals or objects) were similar in the sense that, measured on the dependent variable, they ended up in the same category or had the same score. Three kinds of frequency distributions are often seen in published journal articles: **simple** (or ungrouped), **grouped**, and **cumulative**.

In Excerpt 2.1, we see an example of a **simple frequency distribution,** also called an **ungrouped frequency distribution.** The numbers in the left-hand column are the scores that could be earned on a personality inventory that focused on people's sensation-seeking thoughts. The numbers in the middle column indicate how many of the study's participants got each possible score. (Thus, there were four people who ended up with a score of 0, four others who got a score of 1, and so on.) The numbers in the right column indicate the percent of the full group that ended up with each possible score.

EXCERPT 2.1 • *Simple Frequency Distribution*

TABLE 2 ***The Distribution of the Participants' Sensation-Seeking Scores (TAS)***

TAS	*Frequency*	*Percent*
0	4	5.1
1	4	5.1
2	3	3.8
3	7	9.0
4	7	9.0
5	6	7.7
6	17	21.8
7	6	7.7
8	7	9.0
9	6	7.7
10	11	14.1
Total	78	100.0

Source: Rosenbloom, T. (2006). Sensation seeking and pedestrian crossing compliance. *Social Behavior and Personality, 34*(2), p. 116.

Two features of Excerpt 2.1 are worth noting. First, the authors used the word *frequency* to label the middle column of numbers. Sometimes you will see the word *number* or the abbreviation *f* used instead. Second, the number at the bottom of the middle column is sometimes labeled *N*, to represent the number of individuals in the group.

Excerpt 2.2 shows how a frequency distribution can help us understand the characteristics of a group relative to some categorical (rather than numerical) variable of interest. In the study associated with this excerpt, the researchers compared two groups of people—abstinent drug abusers and nonuser controls—in terms of a variety of measures while people in the two groups performed several different tasks. The frequency distribution shown here was included in the research report to show the "drug history" of the individuals in the abstinent group.

EXCERPT 2.2 • *Simple Frequency Distribution for a Qualitative Variable*

TABLE 2 ***History of Substance Abuse***

	Drug Abusers	
Abused Substance	***No.***	***%***
Marijuana	6	29.0
Cocaine only	1	4.8
Marijuana/cocaine only	4	24.0
Multiple drugs, not including heroin	3	15.0
Multiple drugs, including heroin	6	29.0

Source: Fishbein, D., Eldreth, D., Matochik, J., Isenberg, N., Hyde, C., London, E. D., Ernst, M., and Steckley, S. (2005). Cognitive performance and autonomic reactivity in abstinent drug abusers and nonusers. *Experimental and Clinical Psychopharmacology, 13*(1), p. 28.

In Excerpt 2.3, we see an example of a **grouped frequency distribution.** This frequency distribution deals with the complaints that college students had when they used a web-based triage system. There were 1,290 contacts over a four-month time period, with records kept as to the chief complaint of each contact. (The most frequent chief complaint was a sore throat.) The information in Excerpt 2.3 was included in the research report to show how long the college students said they had lived with their problems before using the web-based triage system.

The table in Excerpt 2.3 is a *grouped* frequency distribution because the far left-hand column has, on each row, a group of possible durations. This grouping of the chief complaints into the duration period—into what are technically called *class intervals*—allows the data to be summarized in a more compact fashion. If the data in this excerpt had been presented in an ungrouped frequency distribution with

EXCERPT 2.3 • *Grouped Frequency Distribution*

TABLE 3 ***Duration of Chief Complaint***

Duration	*f*	*%*
<1 d	184	14.3
1–3 d	413	32.0
4.7 d	209	16.2
>1 w	484	37.5
Total	1290	100.0

Source: Sole, M. L., Stuart, P. L., and Deichen, M. (2005). Web-based triage in a college health setting. *Journal of American College Health, 54*(5), p. 292.

the duration column set up to reflect single 24-hour periods, at least 9 rows would have been needed, and probably many more than that. (I'm guessing that some of the college students waited well over a week before making contact.)

In addition to simple and grouped frequency distributions, **cumulative frequency distributions** sometimes appear in journal articles. With this kind of summarizing technique, a researcher tells us, through an additional column of numbers labeled *cumulative frequency* or *cumulative percentage,* how many measured objects ended up with any given score and all other lower scores (or how many scores ended up in a given score interval and all other lower intervals). This kind of frequency distribution is shown in Excerpt 2.4. Notice how the cumulative percentage of 55.1 can be obtained by (1) adding together 1, 2, 14, and 42, (2) dividing that sum by the sum of the frequencies, 107, and (3) multiplying by 100. Or, you can obtain the same cumulative percentage by adding together the top four numbers in the percent column.

EXCERPT 2.4 • *Cumulative Frequency Distribution*

TABLE 1 *Frequency Distribution of SMOG Scores*

SMOG Score	*Frequency*	*%*	*Cumulative %*
7th grade	1	0.9	0.9
8th grade	2	1.9	2.8
9th grade	14	13.1	15.9
10th grade	42	39.3	55.1
11th grade	39	36.4	91.6
12th grade	9	8.4	100.0

Source: Wegner, M. V., and Girasek, D. (2003). How readable are child safety seat installation instructions? *Pediatrics, 111*(3), p. 589.

Stem-and-Leaf Displays

Although a grouped frequency distribution provides information about the scores in a data set, it carries with it the limitation of "loss of information." The frequencies tell us how many data points fell into each interval of the score continuum but they do not indicate, within any interval, how large or small the scores were. Hence, when researchers summarize their data by moving from a set of raw scores to a grouped frequency distribution, the precision of the original scores is lost.

A **stem-and-leaf display** is like a grouped frequency distribution that contains no loss of information. To achieve this objective, the researcher first sets up score intervals on the left side of an imaginary vertical line. These intervals, collectively called the *stem,* are presented in a coded fashion by showing the lowest score of each interval. Then, to the right of the vertical line, the final digit is given for each observed score that fell into the interval being focused on. An example of a stem-and-leaf display is presented in Excerpt 2.5.

EXCERPT 2.5 • *Stem-and-Leaf Display*

TABLE 2 ***Stem-and-Leaf Display of Compliance on a Per-Participant Basis for Study 1***

Stem	*Leaf*
100	000000000000
90	8888887766666555411000
80	8764
70	
60	8
50	
40	0
30	
20	40

Note. Leaf values correspond to the units digit of each participant's compliance score for the given stem (e.g., there were 6 individuals with a compliance score of 98%).

Source: Green, A. S., Rafaeli, E., Bolger, N., Shrout, P. E., and Reis, H. T. (2006). Paper or plastic? Data equivalence in paper and electronic diaries. *Psychological Methods, 11*(1), p. 90.

In the fifth row of this stem-and-leaf display, there is a 60 on the left (stem) side and an 8 on the right (leaf) side. This indicates that there was one score, a 68, within the interval represented by this row of the display (60–69 percent). The third row has four digits on the leaf side, and this indicates that four scores fell into this row's interval (80–89 percent). Using both stem and leaf from this row, we see that those four scores were 88, 87, 86, and 84. All other rows of this stem-and-leaf display are interpreted in the same way.

Notice that the 42 actual compliance scores in Excerpt 2.5 show up in the stem-and-leaf display. There is, therefore, no loss of information. Take another look at Excerpt 2.3, where a grouped frequency distribution was presented. Because of the loss of information associated with grouped frequency distributions, you cannot tell what the highest and lowest earned scores were, what specific scores fell into any interval, or whether gaps exist inside any intervals (as was the case in Excerpt 2.5 because no compliance scores fell between 70–79 percent, for instance).

Histograms and Bar Graphs

In a **histogram,** vertical columns (or thin lines) are used to indicate how many times any given score appears in the data set. With this picture technique, the baseline (that is, the horizontal axis) is labeled to correspond with observed scores on the dependent variable while the vertical axis is labeled with frequencies.[1] Then,

[1]Technically speaking, the horizontal and vertical axes of any graph are called the **abscissa** and **ordinate,** respectively.

columns (or lines) are positioned above each baseline value to indicate how often each of these scores was observed. Whereas a tall bar indicates a high frequency of occurrence, a short bar indicates that the baseline score turned up infrequently.

A **bar graph** is almost identical to a histogram in both form and purpose. The only difference between these two techniques for summarizing data concerns the nature of the dependent variable that defines the baseline. In a histogram, the horizontal axis is labeled with numerical values that represent a quantitative variable. In contrast, the horizontal axis of a bar graph represents different categories of a qualitative variable. In a bar graph, the ordering of the columns is quite arbitrary, whereas the ordering of the columns in a histogram must be numerically logical.

In Excerpt 2.6, we see an example of a histogram. Notice how this graph allows us to quickly discern the Rotter scores for the individuals included in the researchers' sample. Also notice that the columns had to be arranged as they were because the variable on the baseline was clearly quantitative in nature.

EXCERPT 2.6 • *Histogram*

A histogram shows the distribution of Rotter scale scores within the final sample (Figure 1). Recall that lower scores are associated with an internal orientation and higher scores are indicative of external orientation. The sample distribution of scores was roughly bell shaped but with a fatter lower tail and a relatively thinner upper tail, suggesting that the majority of students lay somewhere in between the two extremes and that internally oriented students were represented slightly more in the sample relative to their externally oriented cohorts.

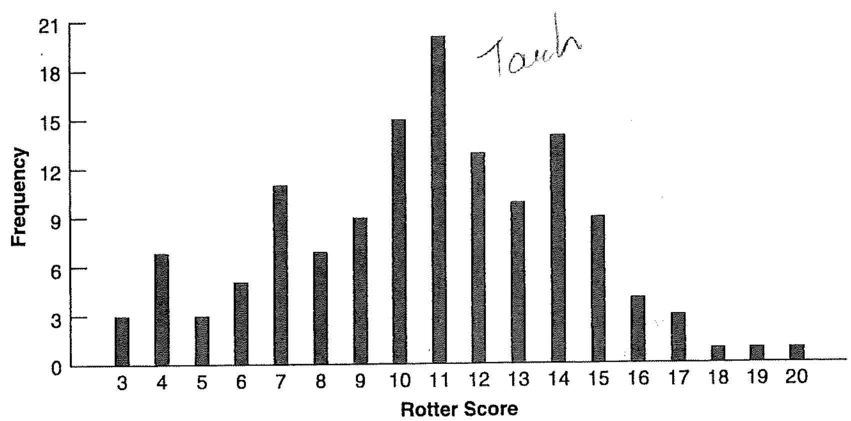

FIGURE 1 ***Histogram of Rotter scores for final sample (N = 136).***

Source: Grimes, P. W., Millea, M. J., and Woodruff, T. W. (2004). Grades—Who's to blame? Student evaluation of teaching and locus of control. *Journal of Economic Education, 35*(2), pp. 135, 137.

EXCERPT 2.7 • *Bar Graph*

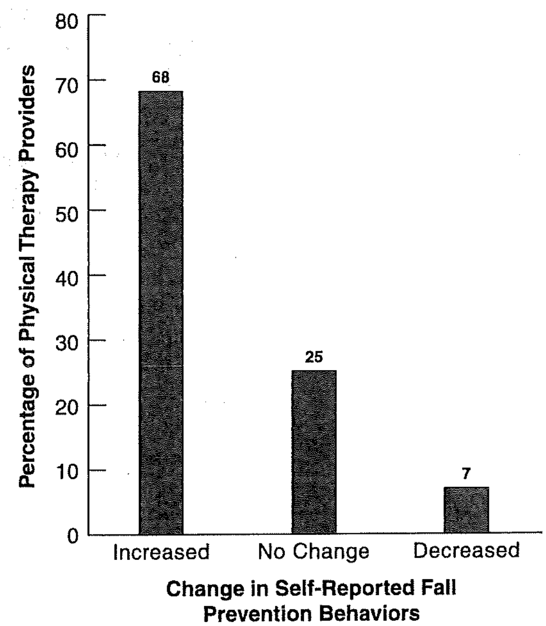

FIGURE 3 ***Distribution of change in self-reporte strategies with patients before and after Connecticu Prevention (CCFP) program (N = 94).***

Source: Brown, C. J., Gottschalk, M., Van Ness, P. H., F (2005). Changes in physical therapy providers' use of fall multicomponent behavioral change intervention. *Physical*

An example of a bar graph is presented in Ex bars is completely arbitrary. The short bars could h with the taller bars positioned on the right. Or, the alphabetically based on the labels beneath the bars.

Frequency Distributions in Words

Researchers sometimes present the information of a rather than in a picture. Excerpt 2.8 illustrates how frequency distribution.

In describing the age of their patients or researc summarize their data in one of three ways: (1) by rep

EXCERPT 2.8 • *A Frequency Distribution in Words*

The patients were 36 to 67 years old: 1 patient was between 30 and 39 years old, 10 were between 40 and 49 years old, 6 were between 50 and 59 years old, and 3 were between 60 and 69 years old.

Source: Blissitt, P. A., Mitchell, P. H., Newael, D. W., Woods, S. L., and Belza, B. (2006). Cerebrovascular dynamics with head-of-bed elevation in patients with mild or moderate vasospasm after aneurysmal subarachnoid hemorrhage. *American Journal of Critical Care, 15*(2), p. 210.

the mean and standard deviation, or (3) by reporting the range and the mean and the standard deviation.[2] That's also the case when researchers describe other characteristics of the people or animals from whom data were gathered. Few researchers provide a summary like the researchers did in Excerpt 2.8. This is unfortunate because such descriptions, as you can see, do not take up much space and yet allow us to get a good "feel" for the data the researcher is describing.

Distributional Shape

If researchers always summarized their quantitative data using one of the picture techniques just covered, then you could *see* whether the observed scores tended to congregate at one (or more) point along the score continuum. Moreover, a frequency distribution, a stem-and-leaf display, a histogram, or a bar graph would allow you to tell whether a researcher's data were symmetrical. To illustrate this nice feature of the picture techniques we have discussed, take another look at Excerpt 2.1. The frequency distribution for the 78 pedestrians' sensation-seeking scores shows nicely that (1) the TAS scores were spread out along the full range of the score continuum, and (2) the most frequently earned score was a 6.

Unfortunately, pictures of data sets do not appear in journal articles very often because they are costly to prepare and because they take up lots of space. By using some verbal descriptors, however, researchers can tell their readers what their data sets look like. To decipher such messages, you must understand the meaning of a few terms that researchers use to describe the **distributional shape** of their data.

If the scores in a data set approximate the shape of a **normal distribution,** most of the scores will be clustered near the middle of the continuum of observed scores, and there will be a gradual and symmetrical decrease in frequency in both directions away from the middle area of scores. Data sets that are normally

[2]We will consider the *mean* and the *standard deviation* later in this chapter.

distributed are said to resemble a bell-shaped curve, since a side drawing of a bell will start out low on either side and then bulge upward in the center. In Excerpt 2.6, we saw a histogram that resembles a normal distribution.

Research Navigator.com
Positively skewed
Negatively skewed

In **skewed distributions,** most of the scores end up being high (or low), with a small percentage of scores strung out in one direction away from the majority. Skewed distributions, consequently, are not symmetrical. If the tail of the distribution (formed by the small percentage of scores that is strung out in one direction) points toward the upper end of the score continuum, the distribution is said to be **positively skewed;** if the tail points toward the lower end of the score continuum, the term **negatively skewed** applies. In Excerpt 2.9, we see an example of a positively skewed distribution.

EXCERPT 2.9 • *A Positively Skewed Distribution*

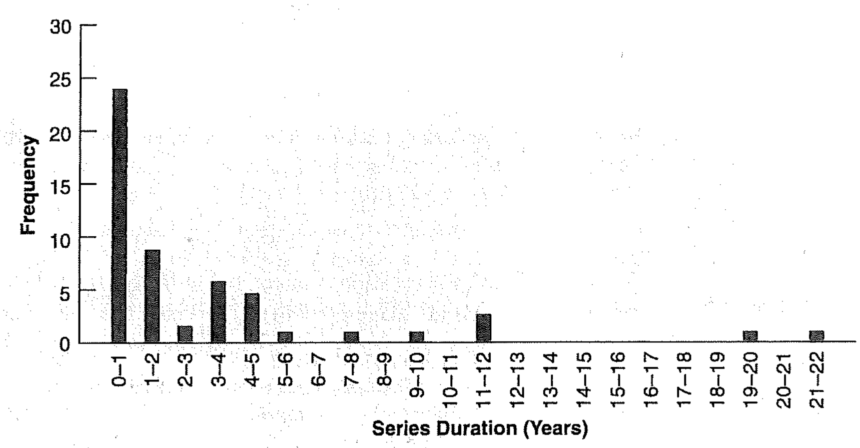

FIGURE 1 ***Duration of murder series (n = 51).***

Source: Snook, B., Cullen, R. M., Mokros, A., and Horbort, S. (2005). Serial murderers' spatial decisions: Factors that influence crime location choice. *Journal of Investigative Psychology and Offender Profiling,* 2, p. 153.

If the scores tend to congregate around more than one point along the score continuum, the distribution is said to be **multimodal** in nature. If there are two such places where scores are grouped together, we could be more specific and say that the data are distributed in a **bimodal** fashion. If scores are congregated at three distinct points, the term **trimodal** would come into play.[3]

[3]Distributions having just one "hump" are said to be **unimodal** in nature.

If scores are fairly evenly distributed along the score continuum without any clustering at all, the data set is said to be **rectangular.** Such a distributional shape would probably show up if someone (1) asked each person in a large group to indicate his or her birth month, and (2) created a histogram with 12 bars, beginning with January, arranged on the baseline. The bars making up this histogram would probably be approximately the same height. Looked at collectively, the bars making up this histogram would resemble a rectangle.

In Excerpts 2.10 through 2.12, we see a few examples of how researchers will sometimes go out of their way to describe the distributional shape of their data sets. Such researchers should be commended for indicating what their data sets looked like because these descriptions help others to understand the nature of the data that have been collected.

EXCERPTS 2.10–2.12 • *References to Different Distributional Shapes*

Scores on the internalizing difficulties scales were normally distributed at both 2 and 4 years of age.

Source: Bayer, Jordana K., Sanson, Ann V., Hemphill, Sheryl A. (2006). Children's moods, fears, and worries: Development of an early childhood parent questionnaire. *Journal of Emotional & Behavioral Disorders, 14*(1), p. 45.

The FNQ data for proportion of met need in the outpatient group were found to be positively skewed (as many participants reported low proportions of met need).

Source: Smith, M. J., Vaugban, F. L., Cox, L. J., McConville, H., Roberts, M., Stoddart, S., and Lew, A. R. (2006). The impact of community rehabilitation for acquired brain injury on carer burden: An exploratory study. *Journal of Head Trauma Rehabilitation, 21*(1), p. 80.

The total months of co-op per student ranged from three to 27; this distribution was bimodal, with 23 [of the 85] students completing nine months and 28 of the students completing 21 months of co-op.

Source: Hoffart, N., Diani, J. A., Connors, M., and Moynihan, P. (2006). Outcomes of cooperative education in a baccalaureate program IN NURSING. *Nursing Education Perspectives, 27*(3), p. 140.

As we have seen, two features of distributional shape are modality and skewness. A third feature is related to the concept of **kurtosis.** This third way of looking at distributional shape deals with the possibility that a set of scores can be nonnormal even though there is only one mode and even though there is no skewness in the data. This is possible because there may be an unusually large number of scores at the center of the distribution, thus causing the distribution to be overly peaked. Or, the hump in the middle of the distribution may be smaller than is the case in normal distributions, with both tails being thicker than in the famous bell-shaped curve.

When the concept of kurtosis is discussed in research reports, you may encounter the terms **leptokurtic** and **platykurtic.** These terms denote distributional shapes that are more peaked and less peaked (as compared with the normal distribution), respectively. The term **mesokurtic** signifies a distributional shape that is neither overly peaked nor overly flat.

As illustrated in Excerpts 2.9 through 2.12, researchers can communicate information about distributional shape via a picture or a label. They can also compute numerical indices that assess the degree of skewness and kurtosis present in their data. In Excerpt 2.13, we see a case in which a group of researchers presented such indices in an effort to help their readers understand what kind of distributional shape was created by each set of scores that had been gathered.

EXCERPT 2.13 • *Quantifying Skewness and Kurtosis*

To test whether the distribution of the PPVT–III scores within the African American sample deviated from normal, and to determine whether there was a floor effect, skewness and kurtosis values were examined. The skewness value of −.07 . . . and kurtosis value of −.14 . . . indicated a relatively normal distribution.

Source: Huaqing Qi, C., Kaiser, A. P., Milan, S., and Hancock, T. (2006). Language performance of low-income African American and European American preschool children on the PPVT-III. *Language, Speech, & Hearing Services in Schools, 37*(1), p. 9.

To properly interpret coefficients of skewness and kurtosis, keep in mind three things. First, both indices will turn out equal to zero for a normal distribution.[4] Second, a skewness value lower than zero indicates that a distribution is negatively skewed, whereas a value larger than zero indicates that a distribution is positively skewed; a kurtosis value less than zero indicates that a distribution is platykurtic, whereas a value greater than zero indicates that the distribution is leptokurtic. Finally, although there are no clear-cut guidelines for interpreting measures of skewness and kurtosis (mainly because there are different ways to compute such indices), most researchers consider data to be approximately normal in shape if the skewness and kurtosis values turn out to be anywhere from −1.0 to +1.0.

Depending on the objectives of the data analysis, a researcher should examine coefficients of skewness and kurtosis before deciding how to further analyze the

[4]Some formulas for computing skewness and kurtosis indices yield a value of +3 for a perfectly normal distribution. Most researchers, however, use the formulas that give values of zero for both skewness and kurtosis.

data. If a data set is found to be grossly nonnormal, the researcher may opt to do further analysis of the data using statistical procedures created for the nonnormal case. Or, the data can be "normalized" by means of a formula that revises the value of each score such that the revised data set represents a closer approximation to the normal.

Measures of Central Tendency

To help readers get a feel for the data that have been collected, researchers almost always say something about the typical or representative score in the group. They do this by computing and reporting one or more **measures of central tendency.** There are three such measures that are frequently seen in the published literature, each of which provides a numerical index of the **average** score in the distribution.

The Mode, Median, and Mean

The **mode** is simply the most frequently occurring score. For example, given the nine scores 6, 2, 5, 1, 2, 9, 3, 6, and 2, the mode is equal to 2. The **median** is the number that lies at the midpoint of the distribution of earned scores; it divides the distribution into two equally large parts. For the set of nine scores just presented, the median is equal to 3. Four of the nine scores are smaller than 3; four are larger.[5] The **mean** is the point that minimizes the collective distances of scores from that point. It is found by dividing the sum of the scores by the number of scores in the data set. Thus for the group of nine scores presented here, the mean is equal to 4.

In journal articles, authors sometimes use abbreviations or symbols when referring to their measure(s) of central tendency. The abbreviations *Mo* and *Mdn,* of course, correspond to the mode and median, respectively. The letter *M* always stands for the mean, even though all three measures of central tendency begin with this letter. The mean is also symbolized by $\bar{X}$ and μ.

In many research reports, the numerical value of only one measure of central tendency is provided. (That was the case with the model journal article presented in Chapter 1; take a look at Excerpt 1.9 to see which one was used.) Because it is not unusual for a real data set to be like our sample set of nine scores in that the mode, median, and mean assume different numerical values, researchers sometimes compute and report two measures of central tendency, or all three, so as to help readers better understand the data being summarized.

[5] When there is an even number of scores, the median is a number halfway between the two middle scores (once the scores are ordered from low to high). For example, if 9 is omitted from our sample set of scores, the median for the remaining eight scores would be 2.5—that is, the number halfway between 2 and 3.

In Excerpt 2.14, we see a case where two measures of central tendency were reported for the same data set. Excerpt 2.15 contains an example where all three averages—the mode, the median, and the mean—were provided.

EXCERPTS 2.14–2.15 • *Reporting Multiple Measures of Central Tendency*

The participating [nursing home] facilities varied in size from 7 to 164 beds, with a mean of 42 and a median of 30 beds.

Source: Charach, A., Hongmei, C., Schachar, R., and To, T. (2006). Correlates of methylphenidate use in Canadian children: A cross-sectional study. *Canadian Journal of Psychiatry, 51*(1), p. 21.

The number of PT visits varied among the 66 patients (mean = 11.4, median = 9, mode = 6).

Source: Harp, S. S. (2004). The measurement of performance in a physical therapy clinical program: A ROI approach. *Health Care Manager, 23*(2), p. 117.

The Relative Position of the Mode, Median, and Mean

In a true normal distribution (or in any unimodal distribution that is perfectly symmetrical), the values of the mode, median, and mean will be identical. Such distributions are rarely seen, however. In the data sets typically found in applied research studies, these three measures of central tendency assume different values. As a reader of research reports, you should know not only that this happens but also how the distributional shape of the data affects the relative position of the mode, median, and mean.

In a positively skewed distribution, a few scores are strung out toward the high end of the score continuum, thus forming a tail that points to the right. In this kind of distribution, the modal score ends up being the lowest (that is, positioned farthest to the left along the horizontal axis) while the mean ends up assuming the highest value (that is, positioned farthest to the right). In negatively skewed distributions, just the opposite happens; the mode ends up being located farthest to the right along the baseline while the mean assumes the lowest value. In Figure 2.1, we see a picture showing where these three measures of central tendency are positioned in skewed distributions.

After you examine Figure 2.1, return to Excerpt 2.6 and look at the histogram that summarizes the data. Because the distribution is not skewed very much, we should expect the mean, the median, and the mode to end up being similar. The actual values for these three measures of central tendency are 10.7, 10.8, and 11.0, respectively.

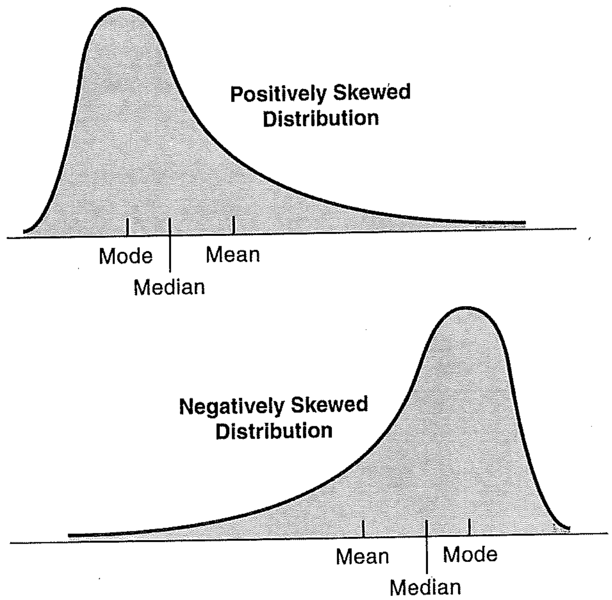

FIGURE 2.1 ***Location of the Mean, Median, and Mode in Skewed Distributions***

To see a case where the computed measures of central tendency turned out to be quite dissimilar, thus implying skewed data, consider Excerpt 2.16. With the median and mean equal to 2 and 9.5, respectively, and with the lowest possible value being 0, we should have been able to guess that the distribution was skewed even if the researchers had not mentioned this. To see if you can determine the nature of skewness from reported values of central tendency, take another look at Excerpt 2.15. For the 66 men in that study, were the number of visits to physical therapy positively skewed or negatively skewed?

EXCERPT 2.16 • *The Mean and Median in a Skewed Distribution*

At subsequent assessments, youths were asked about the number of individuals to whom the youth had disclosed during the past six months (i.e., since the last assessment). The number of individuals reported was used as the indicator of self-disclosure to others [and] the follow-up data were positively skewed [as] most youths reported very few new disclosures in the past 6 months; for example, median = 2.0, $M = 9.5$. . . .

Source: Rosario, M., Schrimshaw, E. W., Hunter, J., and Braun, L. (2006). Sexual identity development among lesbian, gay, and bisexual youths: Consistency and change over time. *Journal of Sex Research, 43*(1), p. 50.

In a bimodal distribution, there will be two points along the score continuum where scores tend to "pile up." If the distribution is symmetrical, the mean and median will be located halfway between the two modes. In a symmetrical trimodal distribution, the median and mean will assume a value equal to the middle of the three modes. Real data sets, however, rarely produce symmetrical bimodal or trimodal distributions. Any asymmetry (that is, skewness) will cause the median to be pulled off center toward the side of the distribution that has the longer tail—and the mean will be pulled even farther in that direction.

With full-fledged rectangular distributions, the mean and median will assume a value halfway between the high and low data points. In such distributions, there is no mode because all earned scores occur with equal frequency. If the distribution turns out to be only roughly rectangular, the median and mean will be located close together (and close to the halfway point between the high and low scores), but the mode could end up anywhere.

Other Measures of Central Tendency

Although the mode, median, and mean are the most popular measures of central tendency, there are other techniques for summarizing the average score in a data set. (Examples include the geometric mean and the harmonic mean.) Because these indices are rarely seen in research reports, they will not be discussed here. If you take an advanced course in statistics, however, you will encounter these alternative methods for computing an average score.

Measures of Variability

Descriptions of a data set's distributional shape and reports as to the central tendency value(s) help us to better understand the nature of data collected by a researcher. Although terms (e.g., *roughly normal*) and numbers (e.g., $M = 67.1$) help, they are not sufficient. To get a true feel for the data that have been collected, we also need to be told something about the variability among the scores. Let us consider now the standard ways that researchers summarize this aspect of their data sets.

The Meaning of Variability

Most groups of scores possess some degree of variability. That is, at least some of the scores differ (vary) from one another. A **measure of variability** simply indicates the degree of this **dispersion** among the scores. If the scores are very similar, there is little dispersion and little variability. If the scores are very dissimilar, there is a high degree of dispersion (variability). In short, **a measure of variability does nothing more than indicate how spread out the scores are.**

The term *variability* can also be used to pinpoint where a group of scores might fall on an imaginary homogeneous–heterogeneous continuum. If the scores are similar, they are **homogeneous** (and have low variability). If the scores are dissimilar, they are **heterogeneous** (and have high variability).

Even though a measure of central tendency provides a numerical index of the average score in a group, we need to know the variability of the scores to better understand the entire group of scores. For example, consider the following two groups of IQ scores:

Group I	*Group II*
102	128
99	78
103	93
96	101

In both groups the mean IQ is equal to 100. Although the two groups have the same mean score, their variability is obviously different. While the scores in the first group are very homogeneous (low variability), the scores in the second group are far more heterogeneous (high variability).

The specific measures of variability that we will now consider are similar in that the numerical index will be zero if all of the scores in the data set are identical, a small positive number if the scores vary to a small degree, or a large positive number if there is a great deal of dispersion among the scores. (No measure of variability, no matter how computed, can ever turn out equal to a negative value.)

The Range, Interquartile Range, Semi-Interquartile Range, and Box Plot

The **range** is the simplest measure of variability. It is the difference between the lowest and highest scores. For example, in Group I of the example just considered, the range is equal to 103–96, or 7. The range is usually reported by citing the extreme scores, but sometimes it is reported as the difference between the high and low scores. When providing information about the range to their readers, authors normally will write out the word *range.* Occasionally, however, this first measure of variability is abbreviated as *R*.

To see how the range can be helpful when we try to understand a researcher's data, consider Excerpts 2.17 and 2.18. Notice in Excerpt 2.17 how information concerning the range allows us to sense that the patients in this study were quite heterogeneous in terms of age. In contrast, the presentation of just the mean in Excerpt 2.18 puts us in the position of not knowing anything about how much variability existed among the patients' ages. Perhaps it was a very homogeneous group, with everyone in their early 60s. Or maybe the group was bimodal, with half the patients in their 50s and half in their 70s. Unless the range (or some other measure of variability) is provided, we are completely in the dark as to how similar or different the patients were in terms of their age.

EXCERPTS 2.17–2.18 • *Summarizing Data with and without the Range*

The mean age of the sample was 48.65 years (range = 23–64 years).

Source: Hacker, E. D., Ferrans, C., Verlen, E., Ravandi, F., van Besien, K., Gelms, J., and Dieterle, N. (2006). Fatigue and physical activity in patients undergoing hematopoietic stem cell transplant. *Oncology Nursing Forum, 33*(3), p. 618.

Patients had a mean age of 62.5 years.

Source: Gates, R., Cookson, T., Ito, M., Marcus, D., Gifford, A., Le, T. N., and Canh-Nhut, N. (2006). Therapeutic conversion from fosinopril to benazepril at a Veterans Affairs medical center. *American Journal of Health-System Pharmacy, 63*(11), p. 1067.

Research Navigator.com
Interquartile range

Whereas the range provides an index of dispersion among the full group of scores, the **interquartile range** indicates how much spread exists among the middle 50 percent of the scores. Like the range, the interquartile range is defined as the distance between a low score and a high score; these two indices of dispersion differ, however, in that the former is based on the high and low scores within the full group of data points whereas the latter is based on only *half* of the data—the middle half.

In any group of scores, the numerical value that separates the top 25 percent scores from the bottom 75 percent scores is the **upper quartile** (symbolized by Q_3). Conversely, the numerical value that separates the bottom 25 percent scores from the top 75 percent scores is the **lower quartile** (Q_1).[6] The interquartile range is simply the distance between Q_3 and Q_1. Stated differently, the interquartile range is the distance between the 75th and 25th percentile points.

In Excerpts 2.19 and 2.20, we see two cases in which the upper and lower quartiles were presented. In both of these excerpts, the values of Q_1 and Q_3 give us

EXCERPTS 2.19–2.20 • *Quartiles and the Interquartile Range*

The range of [scores] was 0.00 to 92.30, with a mean of 47.20. Scores of 24 and 64 were at the first and third quartiles of the sample, respectively.

Source: Kimonis, E. R., Frick, P. J., and Barry, C. T. (2004). Callous-unemotional traits and delinquent peer affiliation. *Journal of Consulting and Clinical Psychology, 72*(6), p. 958.

Interquartile range (IQR) = $34,140–$67,506.

Source: Boudreaux, E. D., Kim, S., Hohrmann, J. L., Clark, S., and Camargo, C. A. (2005). Interest in smoking cessation among emergency department patients. *Health Psychology, 24*(2), p. 222.

[6]The middle quartile, Q_2, divides any group of scores into upper and lower halves. Accordingly, Q_2 is always equal to the median.

information as to the dispersion among the middle 50 percent of the scores. In Excerpt 2.19, for example, the scores for the middle half of the tested individuals extended from 24 to 64, with the top one-fourth of the study's participants having scores between 64 and 92.30, and the bottom fourth of the participants having scores between 0 and 24. In Excerpt 2.20, the interquartile range regarding family income—used as a measure of socioeconomic status—helps us understand the children who were the focus of this investigation.

Sometimes, a researcher will compute the **semi-interquartile range** to index the amount of dispersion among a group of scores. As you would guess on the basis of its name, this measure of variability is simply equal to one-half the size of the interquartile range. In other words, the semi-interquartile range is nothing more than $(Q_3 - Q_1)/2$.

Research Navigator.com
Box plot

With a **box-and-whisker plot,** the degree of variability within a data set is summarized with a picture. To accomplish this objective, a rectangle (box) is drawn to the right of a vertical line labeled so as to correspond with scores on the dependent variable. The positions of the top and bottom sides of the rectangle are determined by Q_3 and Q_1, the upper and lower quartile points. On the outside of the rectangle, two vertical lines—called the *whiskers*—are drawn. Researchers use different rules for drawing the whiskers. Sometimes the whiskers extend up to the highest observed score and down to the lowest observed score. Other researchers use a rule that says that neither whisker should be longer than 1.5 times the height of the rectangle. If any scores are further out than this, they are considered to be outliers, and their positions are indicated by small circles or asterisks. Other researchers draw the whiskers so they extend out to points that represent the 5th and 95th percentiles.

In Excerpt 2.21, we see a case in which box-and-whisker plots were used to show how different sets of data compared with each other in a fully descriptive sense. Note the differences in the heights of the boxes and the length of the whiskers.

Although box-and-whisker plots are designed to communicate information about variability, they also reveal things about central tendency and distributional shape. Within the rectangle, a horizontal line is positioned so as to correspond to Q_2, the median. If this median line appears in the center of the box and if the whiskers are of equal lengths, then we can infer that the distribution of scores is probably symmetrical. On the other hand, the median will end up off-center and the whiskers will be of unequal lengths in skewed distributions. (If the median is on the lower side of the box while the top whisker is longer, the distribution is positively skewed; conversely, negatively skewed distributions cause the median line to be on the upper side of the box and the bottom whisker to be longer.)

Research Navigator.com
Standard deviation
Variance

Standard Deviation and Variance

Two additional indices of dispersion, the **standard deviation** and the **variance,** are usually better indices of dispersion than are the first three measures of variability

EXCERPT 2.21 • *Box-and-Whisker Plot*

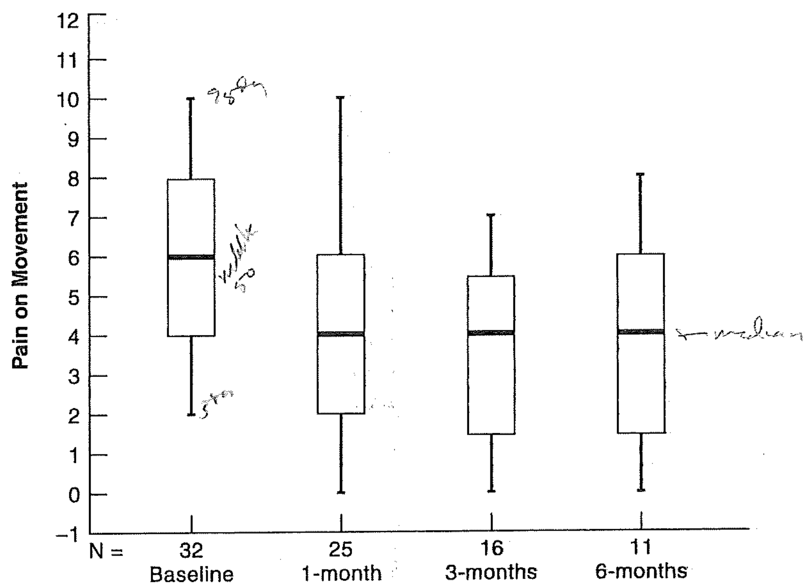

FIGURE 2 ***Boxplot for median pain score for worst pain on movement over time.***

Source: Clare, C., Royle, D., Saharia, K., Pearse, H., Oxberry, S., Oakley, K., Allsopp, L., Rigby, A. S., and Johnson, M. J. (2005). Painful bone metastases: A prospective observational cohort study. *Palliative Medicine, 19*(7), p. 523.

that we have considered. This is due to the fact that the standard deviation and variance are each based on all of the scores in a group (and not just the high and low scores or the upper and lower quartile points). The standard deviation is determined by (1) figuring how much each score deviates from the mean and (2) putting these deviation scores into a computational formula. The variance is found by squaring the value of the standard deviation.

In reporting their standard deviations, authors may employ the abbreviation *SD*, utilize the symbol s or σ, or simply write out the word **sigma.** Occasionally, authors will report the standard deviation using a plus-and-minus format—for example, 14.83 ± 2.51, where the first number (14.83) stands for the mean and the second number (2.51) stands for the standard deviation. The variance, being the square of the standard deviation, is symbolized as s^2 or σ^2.

Excerpts 2.22 through 2.25 illustrate four of the ways researchers indicate the numerical value of the standard deviation. In the first of these, the abbreviation *SD* was used. In the second, the "plus-and-minus" format was employed. In the third

EXCERPTS 2.22–2.25 • *Reporting on the Standard Deviation*

The age of the mothers ranged from 28 to 55 years old, with an average age of 40.56 years ($SD = 5.05$).

Source: Villar, P., Luengo, M. Á., Gómez-Fraguela, J. A., and Romero, E. (2006). Assessment of the validity of parenting constructs using the multitrait-multimethod model. *European Journal of Psychological Assessment, 22*(1), p. 61.

The mean age ($\pm SD$) of participants was 51.8 ± 13.4 years, and 63.5% were men.

Source: Sadri, H., MacKeigan, L. D., Leiter, L. A., and Einarson, T. R. (2005). Willingness to pay for inhaled insulin: A contingent valuation approach. *PharmacoEconomics, 23*(12), p. 1220.

The participants were 30 university students and 30 community-dwelling older adults. The mean age of the young adults was 19.7 (2.2) years while the older adults was 70.6 (4.7) years.

Source: Gilmore, G. C., Spinks, R. A., and Thomas, C. W. (2006). Age effects in coding tasks: Componential analysis and test of the sensory deficit hypothesis. *Psychology and Aging, 21*(1), p. 9.

Lucid dream recall frequency was found to increase from a mean of .33, ($s = .55$) dreams per week before the program started to a mean of .74, ($s = 1.09$) after the program terminated. . . .

Source: Paulsson, T., and Parker, A. (2006). The effects of a two-week reflection-intention training program on lucid dream recall. *Dreaming, 16*(1), p. 25.

---

excerpt, the researchers put values for the standard deviation in parentheses after each mean, without a ± symbol. And in the last of these four excerpts, we see the single letter *s* used to represent the standard deviation.

Excerpt 2.26 shows how information on the standard deviation can be included in a table. In this excerpt, each row of numbers corresponds to a different measuring instrument used within the researchers' study. The data in the first four rows came from scoring and then summarizing tests that were administered to the study's 79 children. The information in the fifth and sixth rows of the excerpt came from summarizing data collected from the children's parents on two social competence measures.

Although the standard deviation appears in research reports far more often than does any other measure of variability, a few researchers choose to describe the dispersion in their data sets by reporting the variance. Excerpt 2.27 is a case in

EXCERPT 2.26 • *Reporting the Standard Deviation in a Table*

TABLE 1 ***Descriptive Statistics for the Measure of Cognitive Ability and Social Competence***

Measures	*M*	*SD*	*Range*
Reading Span	2.84	0.80	2–5
Digit Span	5.80	1.98	2–11
Inhibition	50.94	6.95	33–76
Verbal Ability	109.46	12.86	65–134
MCB	1.71	0.21	1.02–1.95
SSRS	1.46	0.27	0.71–1.89

Note: MCB = My Child's Behavior social competence rating scale; SSRS = Social Skills Rating System.

Source: Tsethlikai, M., and Greenhoot, A. F. (2006). The influence of another's perspective on children's recall of previously misconstrued events. *Developmental Psychology, 42*(4), p. 737.

EXCERPT 2.27 • *Using the Variance to Measure Dispersion*

In addition, the $\varepsilon 4+$ group had a larger variance ($\sigma^2 = .32$) relative to the $\varepsilon 4-$ group ($\sigma^2 = .12$).

Source: Jacobson, M. W., Delis, D. C., Lansing, A., Houston, W., Olsen, R., Wetter, S., Bondi, M. W., and Salmon, D. P. (2005). Asymmetries in global-local processing ability in elderly people with the Apolipoprotein E-ε4 Allele. *Neuropsychology, 19*(6), p. 825.

point. Note that the researchers associated with this excerpt used the symbol σ^2 to represent this measure of variability.

Before concluding our discussion of the standard deviation and variance, I would like to offer a helpful hint concerning how to make sense out of these two indices of variability. Simply stated, I suggest using an article's reported standard deviation (or variance) to estimate what the range of scores probably was. Because the range is such a simple concept, the standard deviation or variance can be demystified by converting it into an estimated range.

To make a standard deviation interpretable, just multiply the reported value of this measure of variability by about 4 to obtain your guess as to what the range of the scores most likely was. Using 4 as the multiplier, this rule of thumb would tell you to guess that the range is equal to 20 for a set of scores in which the standard deviation is equal to 5. (If the research report were to indicate that the variance is equal to 9, you would first take the square root of 9 to get the standard deviation, and then you would multiply by 4 to arrive at a guess that the range was equal to 12.)

When giving you this rule of thumb, I have said that you should multiply the standard deviation by "about 4." To guess more accurately what the range most likely was in a researcher's data set, your multiplier will sometimes need to be a bit smaller or larger than 4. That's because the multiplier number needs to be adjusted on the basis of the number of scores on which the standard deviation is based. If there are 25 or so scores, use 4. If *N* is near 100, multiply the standard deviation by 5. And if *N* is gigantic, multiply by 6. With small *N*s, use a multiplier that is smaller than 4. With 10–20 scores in the group, multiplying by 3 works fairly well; when *N* is smaller than 10, setting the multiplier equal to 2 usually produces a good guess as to range.

It may strike you as somewhat silly to be guessing the range based on the standard deviation. If researchers regularly included the values of the standard deviation and the range when summarizing their data (as was done in Excerpt 2.26), there would be no need to make a guess as to the size of *R*. Unfortunately, most researchers present only the standard deviation—and by itself, a standard deviation provides little insight into the degree of variability within a set of scores.

One final comment is in order regarding this technique of using *SD* to guess *R*. What you will get is nothing more than a rough approximation, and you should not expect your guess of *R* to "hit the nail on the head." Using the standard deviation and range presented in Excerpt 2.26 (and using a multiplier of 5 because the *n* was 79), we see that our guess of *R* is never perfect for any of the six rows of table in the excerpt. But each of our six guesses turns out to approximate well the actual range, and it would help us understand how much spread was in the data if only the standard deviation is presented.

Other Measures of Variability

Of the five measures of variability discussed so far, you will encounter the range and the standard deviation most often when reading researcher-based journal articles. Occasionally, you will come across examples of the interquartile range, the semi-interquartile range, and the variance. And once in a great while, you will encounter some other measure of variability.

In Excerpt 2.28, we see a case where the *coefficient of variation* was used. As indicated within this excerpt, this measure of dispersion is nothing more than the standard deviation divided by the mean.

The coefficient of variation is useful when comparing the variability in two groups of scores where the means are known to be different. For example, suppose we wanted to determine which of two workers has the more consistent commuting time driving to work in the morning. If one of these workers lives five miles from work whereas the second lives 25 miles from work, a direct comparison of their

EXCERPT 2.28 • *Coefficient of Variation*

Variability was measured using a coefficient of variation (standard deviation/mean) to remove the effect of the magnitude of the data from the description of error. . . . As expected, the variability of articulator movement and of VOT, as measured by the coefficient of variation, was greater in the children as compared with the adults.

Source: Grigos, M. I., Saxman, J. H., and Gordon, A. M. (2005). Speech motor development during acquisition of the voicing contrast. *Journal of Speech, Language & Hearing Research, 48*(4), p. 743.

standard deviations (each based on 100 days of commuting to work) would not yield a fair comparison because the worker with the longer commute would be expected to have more variability. What *would* be fair would be to divide each commuter's standard deviation by his or her mean. Such a measure of variability is called the coefficient of variation.

Standard Scores

All of the techniques covered thus far in this chapter describe features of the entire data set. In other words, the focus of attention is on all *N* scores whenever a researcher summarizes a group of numbers by using one of the available picture techniques, a word or number that reveals the distributional shape, a numerical index of central tendency, or a quantification of the amount of dispersion that exists among the scores. Sometimes, however, researchers want to focus their attention on a single score within the group rather than on the full data set. When they do this, they usually convert the raw score being examined into a **standard score.**

Research Navigator.com
z-scores
T-scores

Although many different kinds of standard scores have been developed over the years, the ones used most frequently in research studies are called ***z*-scores** and ***T*-scores.** These two standard scores are identical in that each one indicates how many standard deviations a particular raw score lies above or below the group mean. In other words, the numerical value of the standard deviation is first looked upon as defining the length of an imaginary yardstick, with that yardstick then used to measure the distance between the group mean and the individual score being considered. For example, if you and several other people took a test that produced scores having a mean of 40 and a standard deviation of 8, and if your score on this test happened to be 52, you would be one and one-half yardsticks above the mean.

The two standard scores used most by researchers—*z*-scores and *T*-scores—perform exactly the same function. The only difference between them concerns the arbitrary values given to the new mean score and the length of the yardstick within the revised data set following conversion of one or more raw scores into standard scores. With *z*-scores, the mean is fixed at zero and the yardstick's length is set equal to 1. As a consequence, a z-score directly provides an answer to the question, "How many *SD*s is a given score above or below the mean?" Thus a z-score of $+2.0$ indicates that the person being focused on was 2 standard deviations above the group mean. Likewise, a z-score of -1.2 for someone else indicates that this person scored 1.2 standard deviations below the mean. A z-score close to 0, of course, would indicate that the original raw score was near the group mean.

With T-scores, the original raw score mean and standard deviation are converted to 50 and 10, respectively. Thus a person whose raw score positioned him or her two standard deviations above the mean would receive a T-score of 70. Someone else positioned 1.2 standard deviations below the mean would end up with a T-score of 38. And someone whose raw score was near the group mean would have a T-score near 50.

Although researchers typically apply their statistical procedures to the raw scores that have been collected, they occasionally will convert the original scores into z-scores or T-scores. Excerpts 2.29 and 2.30 provide evidence that these two standard scores are sometimes referred to in research summaries.

EXCERPTS 2.29–2.30 • *Standard Scores (z and T)*

Scores for each scale were transformed to z scores to facilitate comparison and analyses.

Source: Reynolds, S. J. (2006). Moral awareness and ethical predispositions: Investigating the role of individual differences in the recognition of moral issues. *Journal of Applied Psychology, 91*(1), p. 236.

Child behavior problems were assessed using the Child Behavior Checklist (CBCL), a 113-item, caregiver-report measure of child behavior problems [that] generates T scores ($M = 50$; $SD = 10$) for 3 summary scales (including total, internalizing, and externalizing behavioral problems) and 8 subscales (including attention problems, anxiety/depression, and withdrawal).

Source: Wade, S. L., Michaud, L., and Brown, T. M. (2006). Putting the pieces together: Preliminary efficacy of a family problem-solving intervention for children with traumatic brain injury. *Journal of Head Trauma Rehabilitation, 21*(1), p. 61.

A Few Cautions

Before concluding this chapter, I want to alert you to the fact that two of the terms discussed earlier are occasionally used by researchers who define them differently than I have. These two terms are *skewed* and *quartile.* I want to prepare you for the alternative meanings associated with these two concepts.

Regarding the term *skewed,* a few researchers use this word to describe a complete data set that is out of the ordinary. Used in this way, the term has nothing to do with the notion of distributional shape but instead is synonymous to the term *atypical.* In Excerpt 2.31, we see an example of how the word *skewed* was used in this fashion.

EXCERPT 2.31 • *Use of the Term Skewed to Mean Unusual or Atypical*

This study is not without limitations. The low response rate and hence small sample size of adolescents with AS significantly limits the ability to make generalizations. Sample selection was another limitation. The AS group was informed about the nature of the study, which might have skewed the sample toward adolescents already suspected of having anxiety.

Source: Farrugia, S., and Hudson, J. (2006). Anxiety in adolescents with Asperger Syndrome: Negative thoughts, behavior problems, and life interference. *Focus on Autism and Other Developmental Disabilities, 21*(1), p. 33.

The formal, statistical definition of *quartile* is "one of three points that divide a group of scores into four subgroups, each of which contains 25 percent of the full group." Certain researchers use the term *quartile* to designate the subgroups themselves. In this usage there are four quartiles (not three), with scores falling in the quartiles. Excerpt 2.32 provides an example of *quartile* being used in this fashion.

EXCERPT 2.32 • *Use of the Term Quartile to Designate Four Subgroups*

For the principal analyses reported here, we compared participants in the lowest quartile of frequency of exercise (<3 times/week) with those in the top 3 quartiles.

Source: Larson, E. B., Li, W., Bowen, J. D., McCormick, W. C., Teri, L., Crane, P., and Kukull, W. (2006). Exercise is associated with reduced risk for incident dementia among persons 65 years of age and older. *Annals of Internal Medicine, 144*(2), p. 76.

My second warning concerns the use of the term *average.* In elementary school, students are taught that (1) the average score is the mean score and (2) the median and the mode are *not* conceptually the same as the average. Unfortunately, you will have to undo your earlier learning if you're still under the impression that the words *average* and *mean* are synonymous.

In statistics, the term *average* is synonymous with the phrase "measure of central tendency," and either is nothing more than a generic label for *any* of several techniques that attempt to describe the typical or center score in a data set. Hence, if a researcher gives us information as to the "average score," we cannot be absolutely sure which average is being presented. It might be the mode, it might be the median, or it might be any of the many other kinds of average that can be computed. Nevertheless, you won't be wrong very often when you see the word "average" if you guess that reference is being made to the arithmetic mean. In Excerpt 2.33, for example, the average score is most likely a mean.

EXCERPT 2.33 • *Use of the Term Average*

Participants were 1,400 North American women, with an average age of 19.5 (SD = 5.87).

Source: Williams, M. T., and Bonner, L. (2006). Sex education attitudes and outcomes among North American women. *Adolescence, 41*(161), p. 3.

Research Navigator.com
Outliers

My final comment of the chapter concerns scores in a data set that lie far away from the rest of the scores. Such scores are called **outliers,** and they can come about because someone doesn't try when taking a test, doesn't understand the instructions, or consciously attempts to sabotage the researcher's investigation. Accordingly, researchers should (1) inspect their data sets to see if any outliers are present and (2) either discard such data points before performing any statistical analyses or perform analyses in two ways: with the outlier(s) included and with the outlier(s) excluded. In Excerpts 2.34 and 2.35, we see two cases in which data were examined for possible outliers. Notice how the researchers associated with these excerpts explained the rules they used to determine how deviant a score needed to be before it was tagged as an outlier. Also notice how these rules differed.

EXCERPTS 2.34–2.35 • *Dealing with Outliers*

Of returned surveys, three were deleted due to being outliers—the mean of their total scores was at least three standard deviations above or below the mean; two were upper outliers and one was a lower outlier.

Source: Johnson, H. L., and Fullwood, H. L. (2006). Disturbing behaviors in the secondary classroom: How do general educators perceive problem behaviors? *Journal of Instructional Psychology, 33*(1), p. 23.

Boxplots were used to identify outliers, defined as values >1.5 times the interquartile range away from the median. Identified outliers were removed from the data before statistical analysis of the differences between groups.

Source: Milner, C. E., Ferber, R., Pollard, C. D., Hamill, J., and Davis, I. S. (2006). Biomechanical factors associated with tibial stress fracture in female runners. *Medicine and Science in Sports and Exercise, 38*(2), p. 326.

If allowed to remain in a data set, outliers can create skewness and in other ways create problems for the researcher. Accordingly, the researchers who conducted the studies that appear in Excerpts 2.34 and 2.35 deserve credit for taking extra time to look for outliers before conducting any additional data analyses.

I should point out, however, that outliers potentially can be of legitimate interest in and of themselves. Instead of quickly tossing aside any outliers, researchers would be well advised to investigate any "weird cases" within their data sets. Even if the identified outliers have come about because of poorly understood directions, erratic measuring devices, low motivation, or effort to disrupt the study, researchers in these situations might ask the simple question, "Why did this occur?" More importantly, outliers that exist for other reasons have the potential, if considered thoughtfully, to provide insights into the genetic, psychological, and/or environmental factors that stand behind extremely high or low scores.

One Final Excerpt

As we finish this chapter, I want you to look at one final excerpt. Although it is quite short and despite the fact that it contains no tables or pictures, I think this excerpt stands as a good example of how researchers should describe their data. Judge for yourself. Read Excerpt 2.36 and then ask yourself this simple question: Can you imagine what the data looked like?

EXCERPT 2.36 • *A Good Descriptive Summary*

Mean scores were 23.5 ($SD = 6.6$) and 23.1 ($SD = 6.6$) on the MLQ Presence (MLQ–P) and Search (MLQ–S) subscales, respectively. Scores were slightly above but close to the midpoint of the scale (20). The shape of the distributions approximated normality, and scores were variable, as demonstrated by their standard deviations.

Source: Steger, M. F., Frazier, P., Oishi, S., and Kaler, M. (2006). The Meaning in Life Questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology, 53*(1), p. 84.

Review Terms

Average
Bar graph
Bimodal
Bivariate
Box-and-whisker plot
Cumulative frequency distribution
Dispersion
Distributional shape
Grouped frequency distribution
Heterogeneous
Histogram
Homogeneous
Interquartile range
Kurtosis