PRESENTING DATA

Descriptive Statistics

6

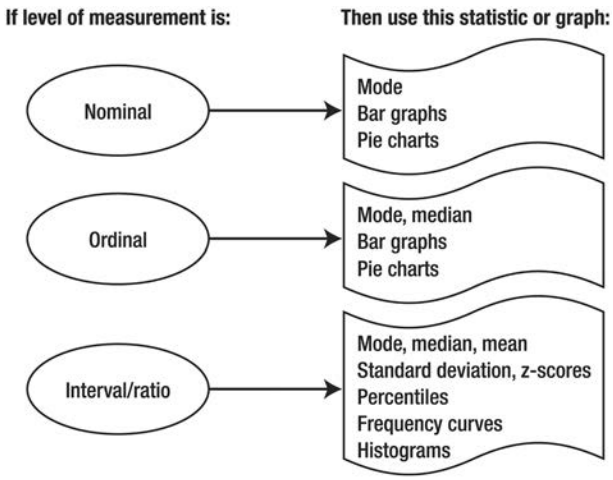
Just think of how stupid the average person is, and then realize that half of them are even stupider!

—George Carlin, comedian

LEARNING GOALS

Understanding how to describe your findings with graphs, tables, and statistics is the focus of this chapter. By the end of the chapter, you should be able to decide when to use the mean, median, mode, and standard deviation (see Figure 6.1). You should also understand the concept of the normal curve and z-scores. In addition, you will learn the concepts of probability and statistical significance.

Figure 6.1 Statistical Decision Steps (also see Statistical Analysis Decision Tree in Appendix)



Finally comes the big moment: What did you discover from your surveys? You eagerly await the outcome after weeks, maybe months, of designing a questionnaire, developing a sampling strategy, distributing surveys, begging people to complete them, and coding and downloading data. Now what do you do? Where do you even begin? How are you going to make sense of all these responses to the questionnaires? Presenting data, analyzing relationships among the variables, and applying statistical tests to the findings are the focus of the next four chapters. This chapter discusses the first steps in describing the data and introduces some basic theoretical concepts about the normal curve and significance level.

PRESENTING UNIVARIATE DATA

If we want to describe, explain, explore, or predict some phenomenon, we must first be sure that the variables in the study are actually variables. Imagine a situation in which the only people to have completed your survey were all women (and you weren't purposely choosing such a sample). The sex variable is now a constant and you can no longer use sex as a variable in explaining or predicting any other variables in the study. Hence, before any further data analysis is accomplished, it is important to do some *univariate* (one variable at a time) *analysis* and look at every item in the study to get a description of the variability of responses. This is necessary in order to decide whether the variables can be used for additional statistical analysis.

In the process of doing a descriptive analysis of the variables, you get a demographic profile of the respondents and an overview of all the behaviors and attitudes measured. These data may be all that are needed for a descriptive or exploratory study but just the start for other kinds of research. There are several ways of presenting univariate information about the variables in your study, including frequency distributions, graphs, and statistical measures.

Frequency Tables

A *frequency table* or *distribution* shows how often each response (a *value*) was given by the respondents to each item (a *variable*). Frequency tables are especially useful when a variable has a limited number of values, such as with nominal or ordinal measures. It is less useful when an interval/ratio variable has many values: For example, when age varies from 18 to 89 in your study, the output could have over 70 rows of numbers, making the table all but unreadable.

The frequency for each value is listed in absolute raw numbers of occurrence and in percentages relative to the number of total responses. Percentages can be presented in terms of both the total number of questionnaires coded and the total number of those actually responding to the question (sometimes called the *valid percent*). A

percent is the proportion of responses standardized on the basis of 100. "Per cent" means "per 100" from the Latin *per centum* for "by hundred."

A proportion is calculated by dividing the number of responses given for a particular value by the total number of responses for the variable; then that proportion (usually somewhere between 0.0 and 1.0) is multiplied by 100 to get the percent. Sometimes if occurrences in a population are small, such as crime rates, numbers are presented "per 1,000" or "per 10,000" instead of "per 100." For example, the U.S. Bureau of Justice Statistics reports that in 2015 there were almost 111 property crimes (such as auto thefts and burglaries) per 1,000 households. Note that the unit of analysis in this case is a household, not a person. This means that for every 1,000 households in the United States, 111 experienced some property crime. This would be the same as saying 11.1 for every 100 households, or simply 11.1 percent.

Let's review a frequency table, as presented by output from the SPSS software program (Table 6.1). (Note that other data analysis programs may have different formats and information than the SPSS examples presented in this book, but the concepts should be applicable in most situations.) Three people of 154 in a survey of college students taking a research methods course did not report their political party affiliation. Therefore, 88 respondents said they were Democrats, representing 57.1 percent of the total number of people who completed the questionnaire (154), but 58.3 percent of those who actually answered the question (151); this is the valid percent. The cumulative percent is useful only for ordinal or interval/ratio measures since it requires that the values accumulate in some order.

We could conclude that there seems to be a variable here; it is not a constant. How do you decide if this is the case? What if 90 percent said they were Democrats, or 80

Table 6.1 Frequency Table, SPSS

		POLITICAL PARTY			
		FREQUENCY	PERCENT	VALID PERCENT	CUMULATIVE PERCENT
Valid	Democrats	88	57.1	58.3	58.3
	Republicans	23	14.9	15.2	73.5
	Independents	31	20.1	20.5	94.0
	Other	9	5.8	6.0	100.0
	Total	151	98.1	100.0	
Missing		3	1.9		
	Total	3	1.9		
Total		154	100.0*		

Note: Total percentages in tables may reflect rounding errors.

percent, or 75 percent? When is variability evident? There is no set rule: You have to decide if there are enough respondents in each of the categories (values) of the variable to do further data analysis. Clearly in this case, only nine people have another political identity than the three most common ones, and "other" would not be a useful category for further analyses, although the variable itself can still be used. Perhaps the "other" category should be combined with the "Independents" by recoding the data.

Charts and Graphs

In addition to a table of numbers, you can represent your findings visually with a *graph* or *chart*. If the variable has a limited number of discrete values, as with nominal or ordinal measures, then select a *bar graph* or *pie chart* to illustrate what you found. These graphic representations give a quick visual description of your variables. Figure 6.2a shows a pie chart for religion. Figure 6.2b shows the same information with a bar graph.

If your data are continuous or interval/ratio measures, histograms and *frequency curves* (sometimes called *frequency polygons*) are better ways of visually presenting univariate information. *Histograms* (Figure 6.3) are similar to bar charts, but the bars are adjacent and touching each other to indicate the continuous nature of the measure. Their width and height communicate the number of responses grouped within some interval. The intervals of the values for the variable are placed along the horizontal or *x-axis*, and the frequency range, in raw numbers or in percentages, is designated along the vertical or *y-axis* of the graph.

A frequency polygon (Figure 6.4) is the result of connecting the midpoints of each of the intervals (bars) in the histogram with a line. Other visual ways of displaying information, such as other line graphs, stem-and-leaf displays, box plots, and stacked bar graphs, are described in more advanced statistics books and online; these graphs are available in most statistical computer packages. Line graphs are especially useful for depicting changes over time.

The Normal Curve. One of the most important frequency polygons is the normal curve; it is at the core of most social science statistics and methodologies. Whether or not a variable is normally distributed in a population can affect the interpretations made about the results. By definition, a *normal curve* is bell-shaped (statistically measured by something called *kurtosis*) and symmetrical (statistically measured by *skewness*); that is, if you cut the curve in half, the right and left sides have the same shape. Kurtosis indicates how peaked or flat the bell-shaped curve is. The normally skewed curve in Figure 6.5 would be called "mesokurtic" in shape and would have a statistical value of zero kurtosis and zero skewness.

Figure 6.2 Examples of Pie Chart (a) and Bar Graph (b)

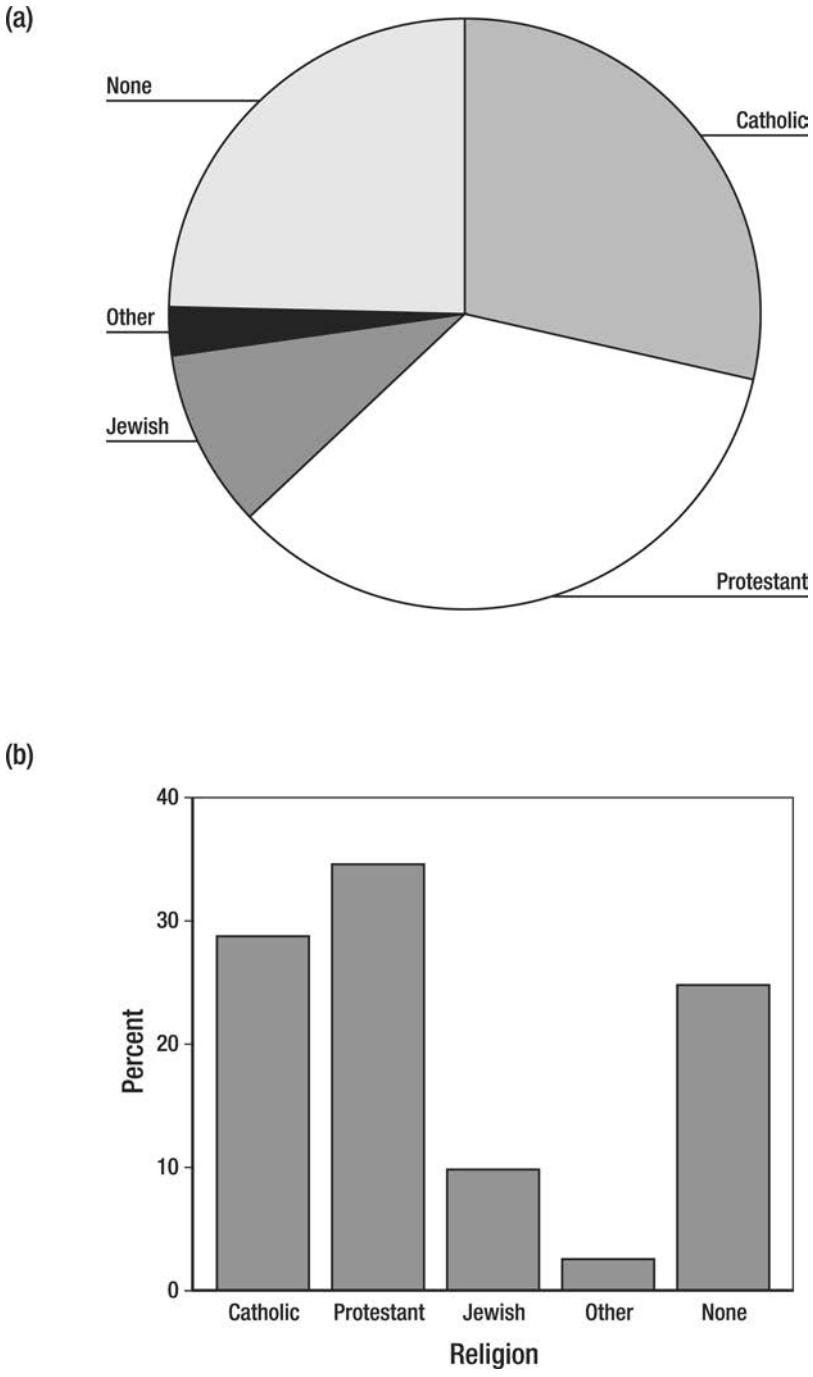


Figure 6.3 Example of a Histogram

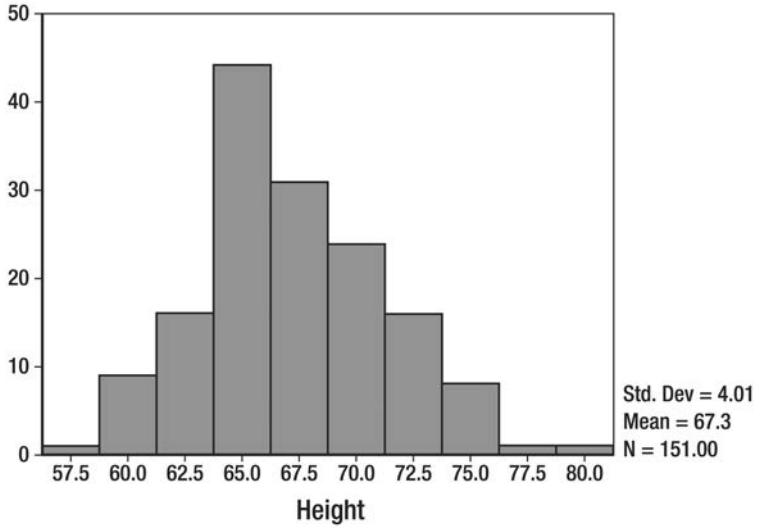


Figure 6.4 Example of a Frequency Polygon

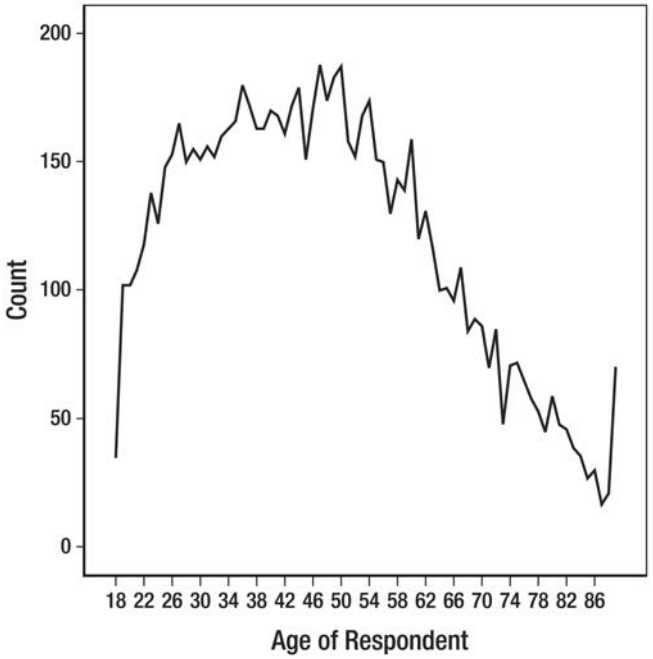
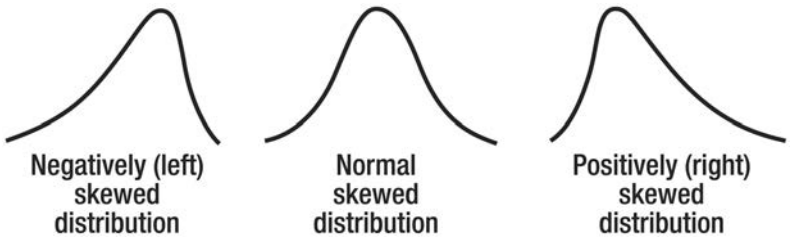


Figure 6.5 Examples of Skewed and Normal Distributions



If there are more scores bunched together on the left side or "negative" tail of the distribution and fewer on the right or "positive" side, then the curve is said to be *positively skewed*. A few high scores skew, or distort, the sample in favor of the positive end of the distribution. A *negative skew* is when few of the scores are at the left side or negative tail of the distribution and most are bunched at the higher end. A few extremely low scores distort the results in favor of the negative side of the distribution. This seems to go against intuition or the words we use when we talk about a "skewed" sample in everyday language, but think of a perfect bell-shaped curve as a piece of string, and whichever end you pull and stretch out to distort the curve a little tells you whether it's negatively or positively skewed. If you tug at the right side to include a few extreme high scores, you are skewing the sample in the positive direction.

Univariate Statistics

Generating statistical information about each variable in a study is another way to find out what you have and to understand more about the distribution of the variables in a sample. Of most importance is a *measure of central tendency*, which provides a quick summary of where the responses are clustered. Depending on whether the variable is nominal, ordinal, or interval/ratio, a mode, median, or mean is used. These measures can also tell you something about the distribution of a variable's values. When all three central tendency measures are equal, there is a perfect normal curve; all are in the peaked center of the distribution. When the mean is higher than the median, there is a positive skew, because a few high scores distort the mean away from the median; a negative skew is indicated by a mean lower than the median, since a few low scores pull the mean down. Let's look at each of these measures (see Box 6.1 for examples of many descriptive statistics).

BOX 6.1
BASIC DESCRIPTIVE MEASURES

Table 6.2 shows some SPSS output using data from General Social Survey (GSS) interviews.

Table 6.2 Descriptive Statistics

STATISTICS		
Age of Respondent		
N	Valid	1,401
	Missing	3
Mean		45.56
Median		42.00
Mode		34
Standard deviation		16.914
Variance		286.083
Range		71
Minimum		18
Maximum		89
Percentiles	20	30.00
	25	33.00
	50	42.00
	75	56.00
	80	61.00

Only three people out of 1,404 declined to reveal their age in the interviews. For those 1,401 valid responses, the following descriptive statistics were calculated. Because age is an interval/ratio measure, the mean and standard deviation are used. These data tell us much about one characteristic of the sample: 45.56 is the average age, 42 is the age half of the respondents are above and half are below (the median and fiftieth percentile), and the most frequently occurring age is 34. However, a frequency table needs to be viewed in order to find out exactly what percentage of the total are 34 (it's 3.1 percent of the 1,401 respondents).

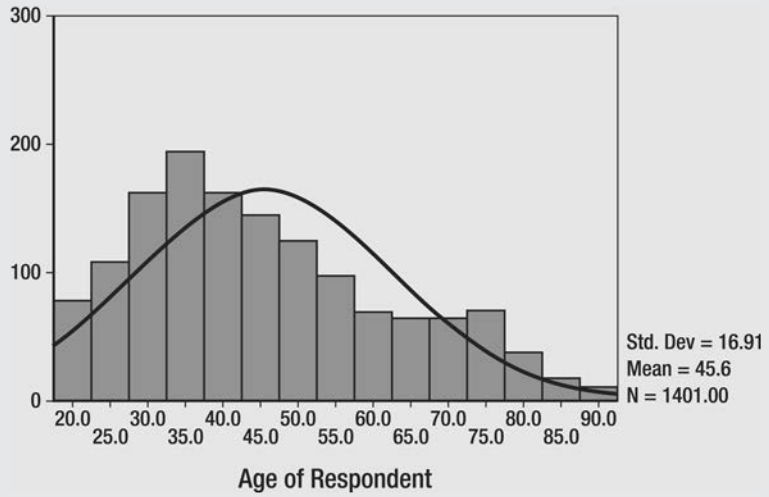
The data also show that the youngest person in the survey is 18 and the oldest is 89, for a range of 71 years. To calculate the interquartile range, simply subtract the age or value at the twenty-fifth percentile from the value at the seventy-fifth percentile to get 23 years. Other percentiles can be used, for example, to show that 20 percent of the sample is over 61 or 80 percent are under 61 (the value at the eightieth percentile), whereas 20 percent are under the age of 30. The standard deviation is 16.9 years; note that this is the square root of the variance, 286.08. These numbers are meaningful primarily in comparison to another sample where age is used. If another survey reported that its mean age is also 45.56, but with a standard deviation of 20.3,

BOX 6.1 CONTINUED

then you would conclude that although it has the same average, its respondents are more widely dispersed around that mean. The ranges indicate this wider dispersion as well.

With these data, you could tell if the distribution of ages formed a normal curve by looking at the mean, median, and mode. If all are approximately the same number, then you have a normal distribution. Although the mean and median are within four years here, note that the mode is much lower. Because the mean is higher than the median (suggesting that there are some very elderly people pulling the mean higher from the median), and given that there is a clustering of respondents around a younger age (as the mode tells us), then it's likely there is a positive skew rather than a normal distribution of age. A positive skew occurs when most values cluster at the low end and a few values at the high or positive end pull the distribution in that direction. A histogram (shown in Figure 6.6), appropriate for continuous interval/ratio measures, illustrates the skew.

Figure 6.6 Histogram With Normal Curve



The Mode. For nominal data, the *mode* is used as a measure of central tendency. It is obtained by finding the most frequently selected value for a variable. A simple look at a frequency table, bar graph, or pie chart to see which value has the largest raw frequency or percentage of occurrence is all it takes to "calculate" this measure. For example, the modal political party in Table 6.1 is Democrat because it has the largest frequency count and valid percent. In Figure 6.2, you can also see in the pie chart and bar graph that Protestant is the mode for the variable religion, even though it is not the religion of the majority of the respondents. Do not confuse the mode with the "majority" answer, which is a response that is more than 50 percent. The most frequently occurring value could have been selected by fewer than 50 percent of the respondents and still qualify as the modal response, but a majority response is always the mode. When there are two values that are selected with equal frequency, the result is a *bimodal* distribution.

The Median. If the values for the variable are ranked or ordered categories (ordinal data), a *median* is the ideal measure of central tendency to report in addition to the mode. The median—like the median that runs down the middle of a highway—is the halfway point, the value above which half the values fall, and below which the other half fall. It has virtually nothing to do with the actual values, just the number of values.

For example, imagine you asked five people to state their birth orders; you first list their responses in order: 1, 1, 2, 3, 3. The halfway point is the third response because this would place two below and two above that response. For these findings, the median birth order is 2. And if the fifth person instead said she was number 7 in her family of 8, the distribution would look like this: 1, 1, 2, 3, 7. Notice what happens: The median birth order remains 2. It is not affected by the size of the value, as the mean is. This is why it is a better statistic to use when you have a skewed sample of values that include extremely low or high scores like income. Not everyone would appreciate being told that the average yearly income of people using Facebook includes Mark Zuckerberg's earnings!

When you have an even number of responses—for example, only four people are surveyed and you found their birth order to be 1, 2, 3, 7—then you take the halfway point between the two middle values, in this case, between 2 and 3, which results in a median birth order of 2.5. How to calculate the median when the data are grouped—that is, when the ordered categories represent ranges (such as when value 1 represents ages 10 to 20, 2 equals 21 to 30, and so on)—is discussed in more detail in statistics books and available with an online search of the Internet.

Percentiles. The median is also called the fiftieth percentile. A *percentile* tells you the percentage of responses that fall above and below a particular point. So when you receive test scores on some national exam, like the SAT or ACT, and the results are at the eightieth percentile, it means that you scored higher than 80 percent of those taking it and lower than 20 percent.

Percentiles can be used to show the dispersion of scores for ordinal or interval/ratio data by finding the value at the twenty-fifth percentile (called the first quartile) and subtracting it from the value at the seventy-fifth percentile (the third quartile), resulting in the *interquartile range*. Percentiles can be broken down into any number of categories, such as deciles to get the scores at every tenth percentile. A *range* indicates the spread of scores by subtracting the highest and lowest values. However, the range is affected by extreme scores unlike interquartile ranges, which are better suited when there are some very high and very low values in the distribution. Many college guidebooks present comparisons of SAT scores using the interquartile range.

The Mean. The most sophisticated measure of central tendency, and one that forms the basis of advanced statistics, is the arithmetic *mean*. You have calculated this many times, and it follows you around school as the grade point average (GPA). The mean

is the sum of the values divided by the number of values and is most suitable for interval/ratio variables. $\bar{X}$ (pronounced X-bar) is used for samples, and mu (μ) is used for means when you have surveyed every element in the population.

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

Calculating the mean for some ordinal scales (such as Likert ones) that look like equal-appearing interval scales is acceptable. When the interval/ratio measure is discrete rather than continuous, the mean often produces a peculiar number, such as the 2.3 mean number of children in families, or the 38.5 average number of books borrowed at the library per day (can you really borrow just half a book?) (see Box 6.2 for an example using the mean and median).

BOX 6.2
DESCRIPTIVE STATISTICS IN A PUBLISHED REPORT

The Pew Research Center (2016) surveyed 1,520 Americans 18 and older about the number of books they read in a given year. Table 6.3 shows what they found:

Table 6.3 Number of Books Read in the Previous 12 Months

	MEAN	MEDIAN
All respondents	4	12
Gender		
Men	3	9
Women	5	15
Age		
18 to 29	5	12
30 to 49	4	12
50 to 64	3	11
65+	3	13
Education level		
Less than high school	0	3
High school diploma	2	9
Some college	4	12
College+	7	17

N = 1,520

BOX 6.2 CONTINUED

Remember, the mean is a mathematical calculation affected by extreme scores, while the median tells us simply what number is at the fiftieth percentile, that is, where half the respondents are above and below that point. So the univariate statistics tell us that half these respondents read fewer than 12 books a year, and half read more than 12 a year. Yet these same people read four books a year on average. What does this tell you? With the mean much smaller than the median, the findings suggest that there are people in the sample reading an extremely lower number of books, which pulls the results in a lower direction.

We can also see by these data how other differences begin to emerge when we look at bivariate results (discussed in more detail in Chapter 7). For example, men and women seem to read different numbers of books per year on average. A statistical analysis, like a t-test, would tell us whether such a difference in means (3 for men, 5 for women) is significant. Chapter 8 discusses how to compare means.

Notice the big difference by education: Americans with a college degree or higher read substantially more books in the past 12 months, especially compared with respondents who have a high school degree or less.

Using the median along with the mean gives you more information than either does alone. What is missing from these data that would give us even more information about the distributions for each subgroup? It would help to have, along with the means, the standard deviations, a very useful statistic for interval/ratio data, to let us know how the results are distributed around these means.

Standard Deviation. Besides describing the dispersion of values using a range, such as the interquartile range, a more powerful measure of dispersion, the *standard deviation*, is available for interval/ratio data. Think of it as the average variation of all the values from the mean. With the standard deviation, we can compare similar variables in different samples or in the same sample at different points in time. The larger it is from zero, the more dispersed the scores are in the sample for that particular variable.

The standard deviation number itself is in the units of the values of the variable—for example, scores on a 100-point reading test—and cannot be compared to a standard deviation calculated on a different variable, like inches for height or number of books read in a year. They are most useful when comparing similar measures between two (or more) different groups, or two (or more) sets of similar measures for the same group. For a quick sense of whether a particular interval/ratio variable in your study is indeed a variable, this statistic provides that information. The further away the number is from no deviation of zero, the more dispersed the scores are for that variable in your sample. Box 6.3 illustrates how the standard deviation is used and calculated.

BOX 6.3
CALCULATING THE STANDARD DEVIATION

The standard deviation (*s* for sample statistics, σ [sigma] for population parameters) is calculated by taking each score (*X*) and subtracting (deviating) it from the mean. Because the mean is the perfect mathematical middle of a set of values, when you deviate the scores from the mean you notice that, when added together, these deviations always equal zero. If one score is −2 below the mean, then another one is +2 above the mean, and so on for every value in the distribution. This creates problems because to calculate a mean you have to add up the scores and divide by the number of scores. If the sum of those scores equals zero, good luck in doing any division!

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

The calculation for the standard deviation requires that the deviations be squared to eliminate the negative numbers; then they can be added and divided by the number of scores minus one (sometimes called *df*, or the *degrees of freedom*, a concept explained in Chapter 7). The resulting number is called the *variance* (s^2), a core concept of statistical analysis. A good deal of the time, we want to understand why data vary in our sample. We usually want to explain the variance in our dependent variables in terms of the variance in our independent variables. We have to remember, though, that we sort of arbitrarily squared all those differences in order to calculate the variance, so just to "undo" what we did, we take the square root of the variance and the resulting number is called the standard deviation.

Let's take a teacher who found that the 30 students in her third-grade class had an average reading score of 75 and another teacher discovered that his 30 third graders also had an average reading score of 75. These findings wouldn't tell you much more than that both classes seemed to be at the same level of reading ability. Yet every student in the first class might have scored exactly 75 to achieve an average of 75 (add up 30 scores of 75 and divide by 30), while in the second class 15 of the students may have scored 60 and another 15 may have scored 90, also resulting in an average of 75. The second class has a much larger dispersion of scores: The students range from 60 to 90. Comparing the two means by themselves wouldn't uncover this very unusual and important difference.

What we need then is the standard deviation to give us this information. In the first class, the standard deviation is zero. This is obtained by subtracting all 30 scores of 75 from the mean of 75, resulting in lots of zeros, and when zero is squared, you still get zero. The sum of zero divided by 30 remains zero even after you take the square root of zero. See what happens when you use zeros in multiplication or division!

On the other hand, in the second class, when you subtract 15 scores of 60 from 75, you get many scores of −15, and when you subtract the other 15 scores of 90 from 75 you see many scores of +15. Square all those numbers and you now have 30 values of 225 to add together. Divide that total by 29 (*N*−1) and take the square root of the new number to get the standard deviation of 15.26. This is a much larger number than zero.

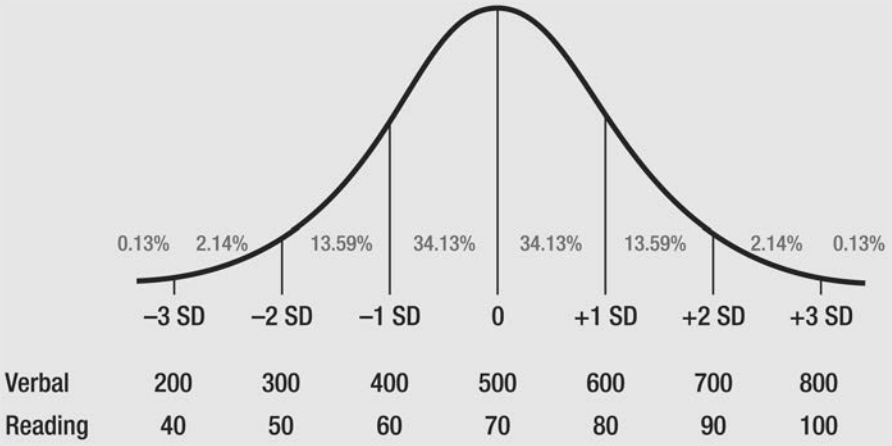
When comparing the two classes' reading scores, the standard deviation tells us that the students in the second class are more widely dispersed in their reading ability than the students in the first class. For the first class, reading scores are a constant, not a variable, as a frequency table or histogram would have also illustrated.

THE NORMAL CURVE AND Z-SCORES

Standard deviations are also important for figuring out where exactly a particular value or a sample of scores is relative to the mean. If the distribution of scores form a normal curve—indicated by the frequency curve or by looking to see if the mean, median, and mode are the same number, as they are when the distribution is normal—you can compare any score with others by standardizing them using *z-scores*. These are used to locate a particular value in a distribution of values, to get its percentile rank, or to compare it with a score measured in different units. Let's say you want to compare students' grades on a classroom reading test with their verbal aptitude scores on a national test. Because they are measured in different units, you must first translate the values to z-scores. Box 6.4 illustrates how this works.

BOX 6.4
CALCULATING Z-SCORES

Figure 6.7 Normal Curve Distribution



Imagine that a student scored 600 on a national SAT verbal aptitude test that goes from 200 to 800 points. Let's assume the mean for that test is 500 and the standard deviation is 100. Now let's also imagine that a student has a score of 80 on a classroom reading test where 70 is the mean and the standard deviation is 10. Assume that both distributions form a normal curve (Figure 6.7). Is the student consistent in how she stands within each of these normal distributions? Is a score of 600 equivalent to a score of 80? To answer that question, z-scores are calculated. A z-score is obtained by subtracting the mean of the distribution from a score and dividing the result by the standard deviation, in order to "standardize" it.

BOX 6.4 CONTINUED

For the verbal aptitude score, 600–500 = 100, then 100/100 = 1. This student has a z-score of 1 for a test score of 600. For the reading test, 80–70 = 10, then 10/10 = 1. We now know that a verbal score of 600 and a grade of 80 on the reading test are comparable standardized scores, even though they are in different units of measurement. Both are one standard deviation unit above the mean. If it were a minus result, it would tell us the score is below the mean.

A z-score also gives us the *percentile* for the score. Based on the mathematics in the formula for the normal curve (available in advanced statistics books and on the Internet), a normal distribution of values has a mean of zero and a standard deviation of 1. Approximately 99.7 percent of all the scores in a distribution (the area under the normal curve contains all the scores) fall within three standard deviations above and three standard deviations below the mean; approximately 68 percent of all scores fall within plus or minus one standard deviation of the mean; and 95 percent fall within two standard deviation units of the mean. Using this information and a table of z-scores and percentages (also found in most statistics books and on the Internet), we can figure out the exact percentile of each score in the distribution.

For example, 34.13 percent of all the scores fall between the mean and one standard deviation above or below the mean. Since the curve is symmetrical, we also know that 50 percent are below the mean; therefore, a z-score of 1.0 tells us that 84.13 percent of all the scores are below that one (50 + 34.13). In other words, a verbal aptitude score of 600 has a percentile rank of 84.13. Conversely, 15.87 percent of all the scores are higher than 600. We can also say that approximately 68 percent of all the scores fall between 400 and 600, since 34 percent fall between the mean (in this case, 500) and one standard deviation unit (in this case, 100 points) below the mean and another 34 percent fall within one unit above the mean.

When a score is exactly the same as the mean of the distribution—say, a score of 500 on the verbal test—then the z-score would be zero. In this case, 50 percent of the scores are above and another 50 percent are below 500. You can see that the mean is also the median in the normal curve.

Statistical Significance, Confidence Intervals, and the Central Limit Theorem

Z-scores are used for calculating the percentiles of individual scores. Therapists, guidance counselors, and others working with data about individuals can see if a client, for example, is significantly different from the norm on a psychological test. You could also see how you stand in comparison with others who took the same test, especially if your teacher is "grading on the curve," that is, using a normal curve to determine grades.

Besides giving us a percentile, z-scores assist in determining confidence intervals for understanding probability levels of significance and in determining sampling error. A z-score tells us the probability of obtaining a score by chance. If 50 percent of the scores are above the mean in a normal curve distribution, then the probability of finding someone with a verbal aptitude score above a mean of 500 is 50 percent or, conversely, for finding someone below the mean of 500. If the percentile is 84 percent,

then the odds of finding someone with a score above 600 is 16 percent, 84 percent for finding someone below 600, and so on, all based on the concept of the normal curve.

Statistical Significance. One of the conventions in social science research is to declare that a finding is *statistically significant* if the probability of obtaining a statistic by chance alone is less than 5 percent. This can be either 5 percent at one end, or tail, of the normal distribution for one-directional tests, or 2.5 percent at the low or negative end *plus* 2.5 percent at the high or positive tail of the distribution for two-directional tests. A *two-tailed* or *two-directional test of significance* exists when we allow two chances in our hypothesis for the outcome to be different: either differently high or differently low from what is expected. For a score to be significant, then, it must be lower or higher than the z-score at which 2.5 percent of the scores fall above or below. The top 2.5 percent of values in a normal distribution have a z-score higher than +1.96 and 2.5 percent would be lower than −1.96, or above a z-score of approximately 2 and below a z-score of −2.

If we set our significance level (often called the *alpha level*) to the probability of obtaining a statistic by chance 1 percent of the time, that is, in the top .5 percent or the bottom .5 percent for two-directional hypotheses (.5 + .5 = 1), then we would use a 2.58 z-score (plus or minus) as the point above or below which significance would occur. Consider a distribution of verbal scores where 500 is the mean and 100 is the standard deviation: A verbal score of 758 or higher would be in the top .5 percent. Each z-score unit in this distribution is "worth" 100 standard deviation points, so 2.58 times 100 results in 258 points above the mean of 500. A score of 242 or lower would be in the bottom .5 percent (500 minus 258) of the distribution of all verbal scores. The probability of selecting someone randomly from a normally distributed population whose verbal score is above 758 *or* below 242 would be less than 1 percent; selecting someone with those scores could then be called a statistically significant outcome. Notice that we get two chances to be significantly different from the norm: either very high scores or very low ones. That's what is meant by a two-tailed test of significance (see Box 6.5 for an explanation about probabilities and significance levels).

Alpha significance levels are usually represented as $p < .05$ or $p < .01$ or $p < .001$, meaning the probability of obtaining that statistic (or score) by chance is less than 5 in 100 (5 percent), 1 in 100 (1 percent), or 1 in 1,000 (.1 percent), respectively. A single asterisk (*) typically appears in tables of data to signify that the .05 significance level has been reached; two asterisks are used for .01, and three for the .001 level.

If we want to see whether a particular score is significantly *higher* than the average—that is, it would be significantly different only if the score were higher but not lower than the average—then we would apply a *one-tailed* or *one-directional test* of

significance. In this case, to be statistically significant, the score should be higher than 95 percent of the values and in the top 5 percent with a z-score of 1.64, or −1.64 if looking for scores significantly lower than the average and in the bottom 5 percent. In short, when the probability of obtaining a particular statistic by chance is less than the cutoff point established, we can state that the hypothesis or research question being tested is statistically significant. Call the media to announce your significant finding and tweet your excitement to everyone!

BOX 6.5
UNDERSTANDING PROBABILITY AND SIGNIFICANCE LEVELS

When calculating chance outcomes and using "or" as the criterion, you *add* the probabilities. When you use "and" as the criterion, you *multiply* the probabilities. For example, if you were asked to pick a red card from a deck of cards (26 chances out of 52 or 26 ÷ 52 = .50) *or* a black card (also .50), then the probability of being successful (significant) is .50 + .50 = 1.00. In short, you have a 100 percent chance of picking a red *or* black card. This should come as no surprise!

But if you asked what the chances are of picking a black card (.50) *and* an ace (4 chances out of 52 = .077), then the probability of picking a black ace is .077 × .50 = .038. There is a 3.8 percent probability you will pick a black ace from a deck of cards. Another way of figuring this is to calculate the chances of getting the ace of spades *or* the ace of clubs using the additive rule. The probability of picking the ace of spades is 1 out of 52, or .019. The probability of picking the ace of clubs is also .019. Therefore the chances of picking either the ace of spades *or* the ace of clubs is .019 + .019 = .038. In short, you have two chances out of 52 to pick a black ace, or 2 ÷ 52 = .038.

The key point to remember here is that when we are testing to see what the probability is of obtaining a statistic or value by chance that is significantly higher *or* lower than the mean at the .05 alpha level (for a two-tailed hypothesis), then we are looking for a value that is either in the top .025 area of the normal curve (the top 2.5 percent) *or* in the lower .025 area in order to get an added probability of .05, that is, .025 + .025 = .05. When we do, we can declare that the statistic or value is a significant finding, just as we would in randomly pulling a black ace from a deck of cards, because the odds of doing so are certainly less than .05. As we just saw, the probability is .038 and therefore $p < .05$.

Here's another easy way of remembering significance levels: Think about 1,000 people standing around flipping coins. Let's say you ask them to flip a coin ten times. At what point do you accuse some of them of using a phony coin? If they get five heads or tails? Nah, that's a pretty likely outcome. When they get three or four heads or tails, or six or seven or what? Those seem fairly reasonable outcomes as well. However, the odds of getting ten heads or ten tails by chance when you flip a coin ten times is .001. That is, only one person in that group of 1,000 is likely to do so by chance alone. Any more and you might be suspicious that some people have magic coins.

BOX 6.5 CONTINUED

The probability of getting nine heads or tails is around .01, or 1 in every 100 people flipping a coin ten times. You wouldn't expect many more than ten people in your sample of 1,000 to get nine heads or tails. And the chance of flipping eight heads or tails is approximately .044, or less than .05 using our criterion of significance.

In other words, you would say that it would be a statistically significant event if someone got eight ($p < .05$), nine ($p < .01$), or ten ($p < .001$) heads flipping a coin ten times since the probabilities are so low that these could occur by accident. So when you see significance levels, translate them into coin flips and ask yourself if the statistical findings are as rare as these coin flip results. If they are, you have a statistically significant finding.

By the way, in everyday language, we often ask what the "odds" are when we actually mean what the "probability" is. A probability is a ratio calculated by dividing a desired outcome by all possible outcomes (resulting in a number between 0 and 1), like selecting any ace from a deck of 52 playing cards (4/52 = 0.077, or 7.7 percent). Odds are a ratio of the likelihood of an outcome happening to the likelihood of it not happening. Choosing any ace from a deck of cards would be 4 to 48, represented with a colon, 4:48 (or simplifying, 1:12). You would say or write this as "the odds of picking any ace from a deck of cards is 1 to 12" while the "the probability of picking any ace is 7.7 percent."

Type I and Type II Errors. Because a decision to declare statistical significance is based on probability, we could be wrong some of the time. We might incorrectly conclude there is a relationship between an independent and dependent variable in the population from which the sample was chosen when there really isn't one. Finding a relationship when there isn't one could occur the same percentage of time established by the alpha level of significance. For example, if we wanted to declare statistical significance at the .05 level, then the probability of the statistic occurring by chance is less than 5 percent or, put another way, we are 95 percent confident it is not due to chance but instead to a real finding.

However, this could be one of those chance times that a statistic of the magnitude calculated happened accidentally and was not a result of an actual finding and relationship in the population. The likelihood, then, of declaring a relationship statistically significant when it is not is the value of alpha; this is called a *Type I error*. In the words of the null hypothesis, a Type I error is rejecting a null hypothesis that is truly null, when we should have accepted it. There is no relationship between the independent and dependent variables, but we rejected that by mistake and declared there is a relationship. Normally, when a statistic's significance level is less than .05, or whatever other number was selected, we say that we have a statistically significant finding and we reject the null hypothesis. That is, we reject that there is no relationship and declare the alternative hypothesis that there is a relationship. In this case, we should have accepted the null hypothesis of no relationship.

We hold a press conference to announce our findings, yet we should be aware that our outcome could be the result of chance and not because of the independent variable, as the statistical calculations have led us to believe. By setting a probability strict enough to limit the number of Type I errors, we become more reassured that our results are genuine and not likely to be due to chance alone. Thus, many researchers choose .01 or .001 as stricter levels of significance that are more difficult to achieve. The probability of making a Type I error and declaring a relationship when there really isn't any, in these cases, is less than 1 in 100 (1 percent) or 1 in 1,000 (.1 percent).

On the other hand, we might be making a *Type II error* instead and accept a null hypothesis that should be rejected, that is, declaring that there is no relationship between our variables when in fact there really is a statistically significant one. These two errors are intertwined: Decreasing a Type I error increases the possibility of making a Type II error. The goal of testing hypotheses is to minimize errors when making inferences about a population from a sample, but convention encourages us to use at least .05 as the level of significance ($p < .05$) when testing statistics, depending on what is being tested and for what reasons. Exploratory research may allow a lower standard, while predictive research looking at the outcomes of a new kind of life-saving drug or public policy program, for example, might require a more stringent test, such as $p < .001$.

The key point to remember is that any results and statistics never prove a hypothesis or research question. Statistics suggest a finding that is tentative; we just never know for sure whether what we found occurred by chance or is the result of an actual influence of the independent variable on the dependent variable. All we can say is that the probability of a relationship happening by accident is less than a certain standard we set.

Central Limit Theorem. Making inferences about a population and establishing probabilities assume a normal distribution. What happens, though, when the distribution of a variable is not normal, as is usually the case? There are other distribution shapes and appropriate statistics that can be used for other models, but these are beyond the scope of this book. Luckily, though, the distribution of statistics (like means) is assumed to be normally distributed. If you were to plot on a graph the reading test means of all possible samples of people taken from a population, the distribution would approach a normal curve, especially the larger the sample size is. This is something in reality you wouldn't do, because if you had the time and money to gather all possible samples of a certain size, you would end up with a survey of the entire population. If the population were small enough for you to survey, then you wouldn't have a need to sample in the first place.

What we are talking about is a theoretical idea known as the *central limit theorem* and something useful in understanding sample size, sampling error, and confidence limits. It states that a distribution of sample means—again, not to be confused with

the distribution of individual scores from one sample—will approach a normal curve, the larger the sample size and the larger the number of samples taken. And the mean and standard deviation (here called the *standard error of the mean*) of all the sample means will be the true population mean (μ) and population standard deviation (σ).

Consider a situation in which we happen to know what the population parameters are for a group of people. Imagine that the entire population of some small village is 1,495 people and has an average age of $\bar{X}$ = 46.23 and a standard deviation of s = 17.42. As illustrated previously, the distribution of age is not normally distributed; it has a slight positive skew. Now, let's take random samples from that population of size five (see Table 6.4). If we were able to take all possible combinations of samples of size five, calculate the mean age for each of those samples, plot them on a graph, and then calculate a mean of all those means, the graph would begin to look like a normal curve and the mean of all those means would be equal to 46.23, the true population mean.

Because the sample size is so small (only five out of 1,495), many of the samples are way off from the true mean of 46.23. Yet even without showing all possible samples of size five from the population, when we calculate a mean for just five of these sample means (as shown in Table 6.4), we get 45.96, which is pretty close to the true population parameter of 46.23. What if the one random sample of five we actually generated turned out to be, say, sample three? The probability is great that, with small sample sizes, the sample statistics (in this case, the mean of 56 and standard deviation of 16.51) will be further away from the true population parameters and we are more likely to end up with a sample that does not represent the population. There would be greater sampling error, hence the justification to gather samples of larger sizes (and good-quality ones using random probability sampling methods).

The central limit theorem argues that larger sample sizes will result in a distribution of sample statistics that is closer to a normal curve with most of the values close to the mean (a curve that is narrow and peaked), and therefore we would be more confident in the accuracy of predicting the population information and make fewer errors

Table 6.4 Samples of Five

	SAMPLE SIZE (N)	MEAN ($\bar{X}$)	STANDARD DEVIATION (S)
Sample 1	5	44.8	19.15
Sample 2	5	35.2	17.21
Sample 3	5	56.0	16.51
Sample 4	5	40.0	10.07
Sample 5	5	53.8	15.51

doing so. Recall the M&M candy demonstration in Chapter 5, which also suggested larger sample sizes reflected the population distribution of colors more accurately.

Now consider random sample sizes of 50 from the same population in Table 6.5. Here, the mean of the five means is 46.44, almost exactly the true population mean of 46.23. And that is with only five samples. In other words, any one of the Table 6.5 samples would be a much better estimate of the true population mean than the samples that had only five people in it (Table 6.4). Still not convinced? Then take a look at random sample sizes of 100 in Table 6.6.

The mean of the sample means is 46.76, pretty close again to the actual population mean of 46.23. Also note how close the standard deviations are to the true population standard deviation of 17.42. Any one of these samples would provide a more accurate estimate of the population parameters than any of the samples with sizes of five. Remember, in actual research we survey only one sample. Notice that the increase in accuracy from sample sizes of 50 to 100 may not be worth the extra time and money; doubling the sample size did not increase the number of samples that were closer to the true population. In general, as we increase the random sample size, more of the samples' means tend to be closer to one another and the true mean, resulting in fewer sampling errors.

Table 6.5 Samples of 50

	SAMPLE SIZE (*N*)	MEAN ($\bar{X}$)	STANDARD DEVIATION (*S*)
Sample 1	50	46.0	16.37
Sample 2	50	46.8	16.79
Sample 3	50	46.5	16.38
Sample 4	50	44.6	13.80
Sample 5	50	48.3	17.16

Table 6.6 Samples of 100

	SAMPLE SIZE (*N*)	MEAN ($\bar{X}$)	STANDARD DEVIATION (*S*)
Sample 1	100	46.4	17.14
Sample 2	100	46.5	17.59
Sample 3	100	47.2	17.70
Sample 4	100	44.6	16.05
Sample 5	100	49.1	17.67

In Chapter 5, we discussed how large a sample should be for a good study. One of the points made was that the larger the random sample size, the more likely it is to capture the diversity that exists in the population. This idea is based primarily on the concepts just described about the central limit theorem. Similar calculations are made by professional research organizations when determining sample size. For example, one formula used to determine sample size takes sampling error into account: margin of error equals 1 divided by the square root of the sample size. Choosing a sample size depends on the population size, its heterogeneity on a variety of key characteristics being studied, the amount of sampling error you're willing to tolerate, and the amount of money and time available to do the research.

Confidence Intervals. The central limit theorem is a theoretical concept because in reality no one takes all possible samples from a population. However, it does help in determining whether the one sample we actually have is representative of the population, that is, whether the sample mean and standard deviation are fairly accurate estimates of the true population mean and standard deviation. Using *inferential statistics*, we infer (arrive at some conclusion about) the population parameters from the sample statistics with some degree of confidence. We can do this since the normal curve tells us with z-scores and percentiles how far off the statistics are from the mean. Not unlike political polls that present results in terms of plus or minus four or five percentage points of the true population percentage, what we also do is determine how confident we are that the true mean of the population (or other statistic) is within a range of plus or minus a certain number of points.

This range is referred to as the *confidence interval*, and the numbers at the beginning and end of the interval are called the *confidence limits*. You see this range in the output results performed by most data analysis software programs. If we use the traditional cutoff point for significance of .05, we can state that we are 95 percent confident that the true population mean—or whichever other statistic from the sample we are using to infer the population parameter—is within this confidence interval range. It states that if we were to get 100 samples of the same size as the one we just did and calculate the confidence limits for each one of them, 95 of those confidence limits would contain the true actual population mean. We could also set the alpha level at .01, or 1 percent, and use a z-score of plus or minus 2.58 to calculate the confidence interval if we want to be 99 percent confident that the true population mean is within that range. In any case, we will never know the true population parameter unless we do a survey of every element in the population, so our results are always estimates and subject to some sampling error.

What the central limit theorem provides us with, then, is proof that even when an individual variable is not normally distributed in a sample, if large-enough sample

sizes are taken, the distribution of statistics (like the mean) from those samples will be normal. This fact allows us to use probabilities in making conclusions and inferences about the population from which the one sample we actually have comes. Our inferences have fewer errors, the more confident we are that the true number is within a particular range. And we make many more errors, the smaller the sample size is, because the distribution of statistics from small random samples has a standard deviation (standard error) that is larger than the true population's dispersion.

The central limit theorem tells us that larger sample sizes produce distributions that have more limited dispersions, ones in which the normal curve is narrow and peaked. We will make fewer errors in estimating the population parameters if we have a sample from a theoretical distribution of all samples that is normally distributed this way since we know that the probability of obtaining a sample that is statistically very different from the norm (the mean, for example) are 5 in 100, 1 in 100, 1 in 1,000, or whatever alpha level we set as our standard of statistical significance.

Considerations in Descriptive Analysis

Practically speaking, what we need to be aware of when reviewing the output from various statistical procedures is the information about probability or significance levels, confidence intervals, standard error, and sample size, as described in the next three chapters. These numbers are our guide to making more accurate inferences or generalizations about the characteristics of the population from which we drew our random sample. Without a random sample, however, generalizations about a population are not possible, and inferential statistics cannot be calculated. Yet we can still use many of these statistics and graphs to describe what we have in our study.

However, it's critically important to remember that research ethics play a central role at this stage of the research process. Making conclusions about a population using data collected from a convenience and nonrepresentative sample demonstrates not only a lack of understanding about sampling but also unprofessional ethical choices. A researcher must always report the findings accurately, use the appropriate statistics, and present them correctly in graphs and tables. Numbers can easily be manipulated by using exaggerated scales on a graph (beginning the y-axis units at 50 and increasing with units of 5, for example, as opposed to starting at 0 and increasing with units of 10), by selecting a mean when a median should be used, or by presenting the results as if they applied to an entire population rather than just the sample. The ethical use of statistics and the accurate interpretation of them must be kept in mind as you learn in the next three chapters how to do more complex data analysis.

Displaying and describing the basic statistics for each of our variables is the first step in analyzing data. These univariate statistics help us determine if the items from

our questionnaire are actually variables and useful for later data analysis. They also allow us to describe the characteristics of the sample through displays of the demographic findings. Once we determine which items are variable enough for further analysis, we can begin to evaluate the research questions and hypotheses we developed using concepts of inferential statistics and probability levels. To do this, we need to explore relationships between two or more variables at a time. The next chapter discusses the many ways to assess whether variables are correlated.

REVIEW: WHAT DO THESE KEY TERMS MEAN?

Alpha levels
Bar graphs
Central limit theorem
Confidence limits and interval
Frequency curves or polygons
Frequency tables or distribution
Graphs and charts
Histograms
Inferential statistics
Interquartile range
Kurtosis
Mean
Measures of central tendency
Median
Mode
Normal curve
One- and two-tailed tests
Percent and valid percent
Percentiles
Pie charts
Positive and negative skews
Range
Standard deviation
Standard error of the mean
Statistical significance
Type I and Type II errors
Univariate analysis
Values and variables
Z-scores

TEST YOURSELF

1. For each of the following variables in a study, list the one best measure of central tendency you can use to describe the data and a graph to depict the distribution of values (pick any one if more than one can be used).

	Levels of Measurement	Graph/Chart
a. Number of text messages sent per day		
b. Race/ethnicity		
c. Skewed number of hours studied in the past week		
d. Type of car owned		

2. Here are some results using data from 151 students who reported their height (in inches):
   Mean: 67.3 Median: 67 Mode: 68 Standard deviation: 4
   a. How can you tell if this is approximately a normal curve?
   b. Using the standard deviation, calculate the following:
      i. What percentage of the sample is between 63.3 and 71.3 inches?
      ii. What height would put you in the shortest 16 percent of the class?
      iii. What percentage of the respondents are taller than 67 inches?
      iv. If you set the significance level at 5 percent ($p < .05$), and you were to randomly select one respondent from the sample to find someone "significantly different from the mean," approximately what height would it take for this respondent to qualify as statistically taller or statistically shorter than the mean? Is this a two-tailed or one-tailed research question?
      v. Assume this is a random sample from a population: Between approximately what heights would you be 95 percent confident that the true population mean is?

INTERPRET: WHAT DO THESE REAL EXAMPLES TELL US?

1. In a published academic article, Elias et al. (2017) reported on a study analyzing racial differences in attitudes toward homosexuality in the United States. In addition to race/ethnicity, the researchers collected information on sexual orientation, religion, education, and age.

   Here are the demographic characteristics for age and race/ethnicity of the survey sample respondents:

	MEAN	STANDARD DEVIATION
Black	39.89	11.79
Hispanic	37.10	13.71
White	55.33	14.91

   a. What do the means tell us in terms of the age of the various racial/ethnic respondents? How would you put into words what this table of means tells us?
   b. What do the standard deviations say? Which group has more or less diversity in terms of age? How would you put into words what this table of standard deviations tells us?
2. For a study on marital satisfaction and online infidelity-related behaviors on social media sites, McDaniel et al. (2017: 90) presented these demographic results: "The mean age of wives was 31.59 years old (SD = 4.44; Range = 20 to 42), and the mean

age of husbands was 33.26 (SD = 5.05; Range = 22 to 52). Participants self-reported their yearly household income, with the median income being $69,500; Range = $0 to $250,000."

a. Why did they use the median for income and not the mean?
b. Put into words what a median income of $69,500 tells us.
c. Explain what the "Range" is, what information it communicates, and how it might be related to the choice of using the median for income.

CONSULT: WHAT COULD BE DONE?

You've been hired to do the data analysis for a survey conducted at a local business. The questionnaire contains the following variables, and you need to advise the business which ones are good for further analysis in its study.

1. What do you do first? Describe the steps needed to be taken at the start of data analysis to determine which are useful variables for further analysis.
2. Which measures of central tendency and what kinds of graphs would be best to use to describe the following variables?
   a. Type of job performed in the organization
   b. Hours worked per week
   c. Seniority ranking (oldest employee to newest)
   d. Monthly salary
   e. A checklist of employee benefits (health plan, dental insurance, childcare, etc.).

DECIDE: WHAT DO YOU DO NEXT?

For your study on how people develop and maintain diverse friendships, especially on social media, respond to the following items:

1. For each variable in the questionnaire you developed in the previous chapter, describe which measures of central tendency, other statistical descriptions, and graphs would be best to use during the first phase of data analysis.
2. If you have actually collected data, now is the time to code and enter the data into a statistical program and begin data analysis with descriptive statistics.