

# 9

## TESTING THE DIFFERENCE BETWEEN TWO MEANS

### CHAPTER OUTLINE

- 9.1 An Example From the Research: You Can Just Wait
- 9.2 The Sampling Distribution of the Difference
  - Characteristics of the sampling distribution of the difference
- 9.3 Inferential Statistics: Testing the Difference Between Two Sample Means
  - State the null and alternative hypotheses ( $H_0$  and  $H_1$ )
  - Make a decision about the null hypothesis
    - Calculate the degrees of freedom ( $df$ )
    - Set alpha ( $\alpha$ ), identify the critical values, and state a decision rule
    - Calculate a statistic:  $t$ -test for independent means
    - Make a decision whether to reject the null hypothesis
    - Determine the level of significance
  - Draw a conclusion from the analysis
  - Relate the result of the analysis to the research hypothesis
  - Assumptions of the  $t$ -test for independent means
  - Summary
- 9.4 Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)
  - An example from the research: Kids' motor skills and fitness
  - Inferential statistics: Testing the difference between two sample means (unequal sample sizes)
    - State the null and alternative hypotheses ( $H_0$  and  $H_1$ )
    - Make a decision about the null hypothesis
    - Draw a conclusion from the analysis
    - Relate the result of the analysis to the research hypothesis
  - Assumptions of the  $t$ -test for independent means (unequal sample sizes)
  - Summary
- 9.5 Inferential Statistics: Testing the Difference Between Paired Means
  - Research situations appropriate for within-subjects research designs
  - An example from the research: Web-based family interventions
  - Inferential statistics: Testing the difference between paired means
    - Calculate the difference between the paired scores
    - State the null and alternative hypotheses ( $H_0$  and  $H_1$ )
    - Make a decision about the null hypothesis
    - Draw a conclusion from the analysis
    - Relate the result of the analysis to the research hypothesis

*(Continued)*

(Continued)

- Assumptions of the  $t$ -test for dependent means
  - Summary
- 9.6 Looking Ahead  
9.7 Summary
- 9.8 Important Terms  
9.9 Formulas Introduced in This Chapter  
9.10 Using SPSS  
9.11 Exercises

This chapter returns to a discussion of the process of calculating inferential statistics to test research hypotheses. Chapter 7 introduced this process with the simplest example, the test of one mean, which evaluated the difference between a sample mean ( $\bar{X}$ ) and a hypothesized population mean ( $\mu$ ). This chapter, on the other hand, discusses inferential statistics that evaluate the difference between the means of two samples drawn from two populations. Although the calculations in this chapter are slightly more complicated than those in Chapter 7, throughout this chapter you'll see many critical similarities between the test of one mean and the difference between two means.

## 9.1 AN EXAMPLE FROM THE RESEARCH: YOU CAN JUST WAIT

Like most college towns, Berkeley has an inordinate number of coffee houses. The author of this book has spent a great deal of time in these businesses preparing lectures, grading exams, and, of course, drinking a lot of coffee. It is for this reason that this book's author became introduced to a sign next to a restroom door:

Remember: How long a minute is depends on which side of the door you're on.

Does it ever seem to you that people move slower when they know you're waiting for them than when they don't know you're there? Well, perhaps it isn't your imagination.

Two sociologists at Pennsylvania State University, Barry Ruback and Daniel Juieng, were interested in studying territorial behavior, defined as "marking, occupying, or defending a location in order to indicate presumed rights to the particular place" (Ruback & Juieng, 1997, p. 821). Although you may think of territorial behavior as protecting our homes from burglars, the researchers studied this behavior in public places. For example, imagine you're in a busy library working on a photocopy machine when someone walks up and, without saying a word to you, makes it apparent he also wishes to use it. Do you (a) speed up to finish more quickly, (b) continue to work in the same manner as if no one were waiting, or (c) deliberately slow down and actually take longer than if no one were there?

The theory of territorial behavior states that people sometimes select choice (c). In describing this behavior, Ruback and Juieng (1997) proposed that when a person possesses a limited resource that is desired by others, the person will maintain possession of the resource to defend it from "intruders" and "would be territorial even when they had completed their task at the location and the territory no longer served any function to them" (p. 823).

The researchers chose to test their beliefs in a familiar setting: a shopping mall. Picture yourself driving in a crowded parking lot when you see someone leave the mall and walk to his car. You drive over and wait for him to leave. And you wait . . . and wait . . . and wait. Based on the theory of territorial behavior, the research hypothesis in the Ruback and Juieng (1997) study was that even if people no longer need a parking space, they will take longer to leave the space when another driver is waiting than when no such "intruder" is present.

To collect their data, the researchers watched drivers leave parking spaces at a large shopping mall. As such, this study, which we will refer to as the parking lot study, is an example of observational research. As defined in Chapter 1, observational research involves the systematic and objective observation of naturally occurring events, with little or no intervention on the part of the researcher.

The researchers in the parking lot study were interested in measuring the amount of time taken by drivers to leave a parking space. Using a stopwatch, they “started timing the moment the departing shopper opened the driver’s side car door and stopped timing when the car had completely left the parking space” (Ruback & Juieng, 1997, p. 823). They also recorded whether or not another driver was waiting to use the parking space. If there was another driver waiting, the departing driver was defined as having an “intruder”; if not, the departing driver had “no intruder” present. Therefore, this study involved two variables. Driver group, the independent variable, consisted of two groups: Intruder and No Intruder. Because the Driver group variable consists of distinct categories or groups, it is measured at the nominal level of measurement. Time, the dependent variable, was the amount of time in seconds taken to leave the parking space. Because Time is a numeric variable with a true zero point (0 seconds), it is measured at the ratio level of measurement.

The original study consisted of 200 drivers. To save space, the example in this chapter will use a smaller sample of 30 drivers, equally divided between the Intruder and No Intruder groups. (Although the data in this example differ from the original study, the results of the analysis mirror those reached by the researchers.) The time in seconds taken by the 15 drivers in each group to leave their parking spaces is presented in Table 9.1(a); these data have been organized into grouped frequency distribution tables (Table 9.1(b)). An examination of these tables shows that the departure times of those in the Intruder group are generally longer than those in the No Intruder group. For example, the modal interval for the Intruder group is 31 to 40 seconds as opposed to 21 to 30 seconds for the No Intruder group. However, the shape of the distribution for both groups is somewhat normal, with each group having a range of about 40 seconds from the quickest driver to the slowest.

The next step in analyzing the collected data is to calculate descriptive statistics of the dependent variable for each level of the independent variable (Table 9.2). This table includes additions and changes to the notational system used in this book. For example, subscripts have been added to the mathematical symbols to distinguish each group’s descriptive statistics—in the parking lot study, the sample sizes for the Intruder and No Intruder groups are symbolized by  $N_1$  and  $N_2$  rather than simply  $N$ . Also, the subscript  $i$  is used instead of a number to represent a group without specifying any particular one. For example, the symbol  $\bar{X}_i$  represents the mean of either of the two groups.

Previous chapters have created figures to illustrate the distribution of scores for a variable. For example, bar charts were used in Chapter 2 to display the frequencies for the values of a variable measured at the nominal or ordinal level of measurement, and histograms and frequency polygons were used for interval or ratio variables. This chapter introduces figures used to illustrate descriptive statistics such as measures of central tendency and variability. Just as there are different types of figures for variables, there are different ways of displaying descriptive statistics, the choice being a function of the nature of the variables.

When the independent variable is measured at the nominal or ordinal level of measurement and the dependent variable is measured at the interval or ratio level of measurement, descriptive statistics are typically displayed using a bar graph. A **bar graph** is a figure in which bars are used to represent the mean of the dependent variable for each level of the independent variable. For the parking lot study, the bar graph in Figure 9.1 displays the descriptive statistics for the dependent variable Time for the Intruder and No Intruder groups.

The height of the bars in the bar graph in Figure 9.1 represents the mean of the dependent variable (Time) for each of the two groups. The sample mean serves as an estimate of the mean of the population from which the sample was drawn. However, from our discussion of sampling error in Chapters 6 and 7, we know that the means of samples drawn from the population vary as the result of random, chance factors; this variability is estimated using a statistic known as the standard error of the mean. To illustrate the variability of sample means, the T-shaped lines extending above and below the mean in each bar in Figure 9.1 measure one standard error of the mean above and below the sample mean ( $\pm 1 s_{\bar{X}}$ ).

**TABLE 9.1** THE TIME (IN SECONDS) FOR DRIVERS IN THE INTRUDER AND NO INTRUDER GROUPS

(a) Raw Data

Intruder			
Driver	Time	Driver	Time
1	23	9	35
2	62	10	45
3	28	11	37
4	55	12	43
5	31	13	39
6	51	14	41
7	33	15	40
8	48		

No Intruder			
Driver	Time	Driver	Time
1	54	9	35
2	19	10	25
3	46	11	34
4	20	12	28
5	42	13	31
6	21	14	29
7	38	15	30
8	23		

(b) Grouped Frequency Distribution Tables

Intruder		
Time	f	%
> 60	1	7%
51-60	2	13%
41-50	4	27%
31-40	6	40%
21-30	2	13%
≤ 20	0	0%
Total	15	100%

No Intruder		
Time	f	%
> 60	0	0%
51-60	1	7%
41-50	2	13%
31-40	4	27%
21-30	6	40%
≤ 20	2	13%
Total	15	100%

Using Formula 7-5 from Chapter 7, the standard error of the mean  $s_{\bar{X}}$  for the two groups is calculated as follows:

$$\begin{aligned}
 \text{Intruder} \\
 s_{\bar{X}} &= \frac{s_1}{\sqrt{N_1}} \\
 &= \frac{10.42}{\sqrt{15}} = \frac{10.42}{3.87} \\
 &= 2.69
 \end{aligned}$$

$$\begin{aligned}
 \text{No Intruder} \\
 s_{\bar{X}} &= \frac{s_2}{\sqrt{N_2}} \\
 &= \frac{10.08}{\sqrt{15}} = \frac{10.08}{3.87} \\
 &= 2.60
 \end{aligned}$$

**TABLE 9.2** DESCRIPTIVE STATISTICS OF TIME FOR DRIVERS IN THE INTRUDER AND NO INTRUDER GROUPS

(a) Mean ( $\bar{X}_i$ )

Intruder	No Intruder
$\bar{X}_1 = \frac{\sum X}{N}$ $= \frac{23+62+\dots+41+40}{15} = \frac{611}{15}$ $= 40.73$	$\bar{X}_2 = \frac{\sum X}{N}$ $= \frac{54+19+\dots+29+30}{15} = \frac{475}{15}$ $= 31.67$

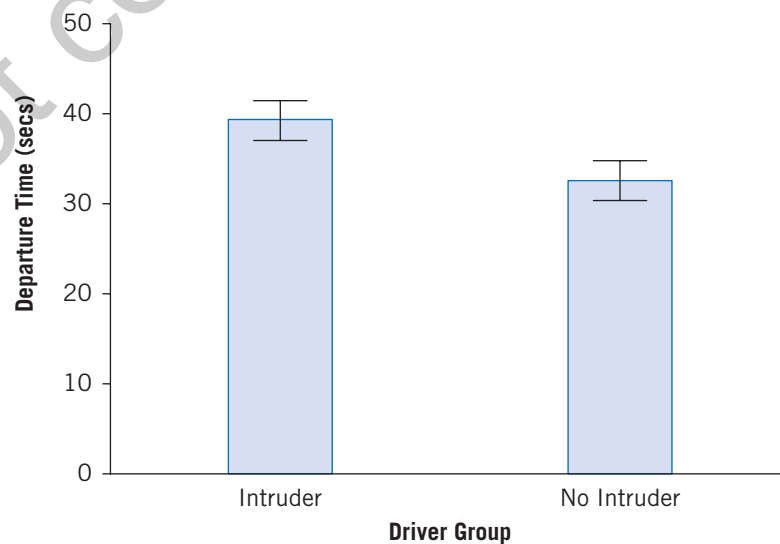
(b) Standard Deviation ( $s$ )

$s_1 = \sqrt{\frac{\Sigma(X - \bar{X})^2}{N - 1}}$ $= \sqrt{\frac{(23 - 40.73)^2 + \dots + (40 - 40.73)^2}{15 - 1}}$ $= \sqrt{\frac{314.47 + \dots + .54}{14}} = \sqrt{108.50}$ $= 10.42$	$s_2 = \sqrt{\frac{\Sigma(X - \bar{X})^2}{N - 1}}$ $= \sqrt{\frac{(54 - 31.67)^2 + \dots + (30 - 31.67)^2}{15 - 1}}$ $= \sqrt{\frac{498.78 + \dots + 2.78}{14}} = \sqrt{101.52}$ $= 10.08$
---	--

In Figure 9.1, for the Intruder group, the area covered by the T-shaped lines extends from  $40.73 \pm 2.69$ . From our discussion of confidence intervals in Chapter 8, we know that the range represented by the T-shaped line represents an interval or range with a stated probability of containing the mean of the population on the dependent variable.

Looking at Table 9.2 and Figure 9.1, we see that the mean time for the Intruder group ( $M = 40.73$ ) is 9.06 seconds greater than that for the No Intruder group ( $M = 31.67$ ). This difference in departure time provides initial support for the hypothesis that people will take longer to leave a parking space when an intruder is present than when there is no intruder. However, to formally test the study's research hypothesis, the next step is to calculate an inferential statistic to conclude the difference between the two sample means is not due to chance but instead is statistically significant.

FIGURE 9.1 ■ BAR GRAPH OF TIME FOR DRIVERS IN THE INTRUDER AND NO INTRUDER GROUPS



In Chapter 7, testing the difference between a sample mean and a population mean ( $\bar{X} - \mu$ ) involved determining the probability of obtaining the value of the sample mean. To address the question, “What is the probability of obtaining a sample mean of 150.35 WCPM assuming the mean in the population is 124.81 WCPM?”, we relied on the sampling distribution of the mean, defined as the distribution of all possible values of the sample mean when an infinite number of samples of size  $N$  are randomly selected from the population. However, as the goal in this chapter is to test the difference between two sample means ( $\bar{X}_1 - \bar{X}_2$ ), we now need to determine the probability of obtaining our particular difference between the two sample means. In the parking lot study, we need to answer the question, “What is the probability of obtaining our difference in departure time of 9.06 seconds between the Intruder and No Intruder groups?” To answer this question, we need a different type of sampling distribution: a distribution of differences between sample means. This distribution is introduced and illustrated in the next section.

## 9.2 THE SAMPLING DISTRIBUTION OF THE DIFFERENCE

The **sampling distribution of the difference** is the distribution of all possible values of the difference between two sample means when an infinite number of pairs of samples of size  $N$  are randomly selected from two populations. It, like the sampling distribution of the mean (Chapter 7), is an example of sampling distribution, which is a distribution of statistics for samples randomly drawn from populations. The sampling distribution of the difference is used to determine the probability of obtaining any particular difference between two sample means. As such, we will rely on this distribution to test the difference between the departure times of the Intruder and No Intruder groups in the parking lot study.

To illustrate the sampling distribution of the difference, imagine you have two populations that do *not* differ on some variable, such that the two population means  $\mu_1$  and  $\mu_2$  are equal. You randomly draw samples from these two populations, calculate the mean for each sample, and then calculate the difference between the two sample means ( $\bar{X}_1 - \bar{X}_2$ ). What should be the difference between the two means? Because we’ve stated that the two populations do not differ, we expect the difference between the two sample means to be zero. However, because the concept of sampling error implies there will be variability among sample means drawn from the populations, there will also be variability among differences between sample means. Consequently, the difference between the two sample means may not be equal to zero. Furthermore, if we were to repeat this process of randomly drawing samples and calculating the difference between sample means an infinite number of times, we could create a distribution of these differences. This distribution is known as the sampling distribution of the difference.

### Characteristics of the Sampling Distribution of the Difference

Like any other distribution, the sampling distribution of the difference may be characterized in terms of its modality, symmetry, and variability. In terms of its modality, the mean of the sampling distribution of the difference is equal to 0 under the assumption that if the two population means are equal, the average difference between sample means should be zero. Next, in terms of its symmetry, the sampling distribution of the difference is approximately normal, assuming the two samples are of sufficient size (typically defined as  $N \geq 30$ ); like the  $t$ -distribution discussed in Chapter 7, the shape of the distribution changes as a function of the size of the samples. Finally, the variability of the sampling distribution of the difference is measured by the **standard error of the difference** ( $s_{\bar{X}_1 - \bar{X}_2}$ ), defined as the standard deviation of the sampling distribution of the difference. The standard error of the difference represents the variability of differences between two sample means (the formula used to calculate the standard error of the difference will be presented later in this chapter).

The characteristics of the sampling distribution of the difference enable researchers to determine the probability of obtaining any particular difference between the means of two samples. For the parking lot study, determining the probability of obtaining our difference of 9.06 seconds between the Intruder ( $M = 40.73$ ) and No Intruder ( $M = 31.67$ ) groups will allow us to test the study's research hypothesis that people will take longer to leave a parking space when another driver is waiting than when no such "intruder" is present. In order to determine the probability of the difference between two sample means, it must be transformed into a statistic in a manner very similar to what we encountered in Chapter 7. The next section describes the process of calculating and evaluating an inferential statistic that tests the difference between two sample means.



### LEARNING CHECK 1

#### Reviewing What You've Learned So Far

1. Review questions
  - a. What changes to this book's notational system are introduced as a function of having two groups rather than one?
  - b. What are the differences between visual displays of variables such as bar charts and visual displays of descriptive statistics such as bar graphs?
  - c. What is the main purpose and characteristics of the sampling distribution of the difference?
  - d. What is the difference between the sampling distribution of the mean (Chapter 7) and the sampling distribution of the difference?
  - e. What is the difference between the standard error of the mean (Chapter 7) and the standard error of the difference?
2. Construct a bar graph for each of the following situations (assume the independent variable is Gender and the dependent variable is Score):
  - a. Males ( $N = 6, M = 3.00, s = 1.79$ ); Females ( $N = 6, M = 5.00, s = 1.41$ )
  - b. Males ( $N = 9, M = 13.00, s = 1.73$ ); Females ( $N = 9, M = 11.00, s = 2.06$ )
  - c. Males ( $N = 12, M = .64, s = .27$ ); Females ( $N = 12, M = .45, s = .25$ )
3. For each of the following situations, (a) calculate the mean, standard deviation, and standard error of the mean for each group and (b) construct a bar graph (assume the independent variable is Type of Pet and the dependent variable is Obedience).
  - a. Dog: 1, 7, 1, 0, 1  
Cat: 6, 10, 5, 1, 8
  - b. Dog: .76, .80, .67, .42, .56, .78, .49, .31, .24, .69  
Cat: .92, .51, .48, .65, .32, .71, .93, .26, .37, .14
  - c. Dog: 13, 7, 9, 11, 12, 7, 8, 10, 6, 7, 5, 10, 9, 8  
Cat: 5, 4, 8, 10, 7, 6, 8, 4, 6, 9, 3, 5, 6, 7

## 9.3 INFERENCE STATISTICS: TESTING THE DIFFERENCE BETWEEN TWO SAMPLE MEANS

This section describes the steps involved in calculating and evaluating an inferential statistic that tests the difference between two sample means. In the parking lot study, do drivers take longer to leave a parking space when another driver is waiting than when there is no intruder? To test this hypothesis, we'll follow the steps introduced in earlier chapters:

- state the null and alternative hypotheses ( $H_0$  and  $H_1$ ),
- make a decision about the null hypothesis,

- draw a conclusion from the analysis, and
- relate the result of the analysis to the research hypothesis.

In discussing each of these steps, we will note both similarities and differences between the research situations discussed in this chapter versus those in Chapter 7, which involved one sample mean rather than the difference between two sample means.

### State the Null and Alternative Hypotheses ( $H_0$ and $H_1$ )

The process of hypothesis testing begins by stating the null hypothesis ( $H_0$ ), which reflects the conclusion that no change, difference, or relationship exists among groups or variables in the population. When testing the difference between two sample means, the null hypothesis states that the means of the two populations are equal. For the parking lot study, the null hypothesis is that the mean amount of time taken to leave a parking space is the same for drivers with or without an intruder. This absence of difference between the two populations is reflected in the following null hypothesis:

$$H_0: \mu_{\text{Intruder}} = \mu_{\text{No Intruder}}$$

The alternative hypothesis ( $H_1$ ), which is mutually exclusive from the null hypothesis, states that in the population there does exist a change, difference, or relationship between groups or variables. For the parking lot study, one way to state the alternative hypothesis is to reflect the conclusion that the mean departure time in the two populations is *not* equal to one another. This conclusion is represented by the following alternative hypothesis:

$$H_1: \mu_{\text{Intruder}} \neq \mu_{\text{No Intruder}}$$

The use of the “not equals” ( $\neq$ ) symbol in the above alternative hypothesis implies that the mean departure time of drivers in the No Intruder group in the population may *either* be greater than *or* less than the mean departure time of drivers in the Intruder group in the population. As such, this alternative hypothesis is considered “nondirectional” or “two-tailed.” From the study’s research hypothesis, which predicts that people will take a *greater* amount of time to leave in the presence of an intruder than when there is no intruder, it may have been reasonable for the researchers to propose the directional (one-tailed) alternative hypothesis  $H_1: \mu_{\text{Intruder}} > \mu_{\text{No Intruder}}$ . However, they chose the traditional approach of a nondirectional alternative hypothesis to allow for the possibility of a statistically significant difference in the opposite direction of what was anticipated.

### Make a Decision About the Null Hypothesis

Once the null and alternative hypotheses have been stated, the next step is to make the decision whether to reject the null hypothesis. This involves completing the steps introduced in Chapter 7:

- calculate the degrees of freedom ( $df$ );
- set alpha ( $\alpha$ ), identify the critical values, and state a decision rule;
- calculate a statistic:  $t$ -test for independent means;
- make a decision whether to reject the null hypothesis; and
- determine the level of significance.

We complete each of these steps for the parking lot study below, highlighting changes new to this chapter.



### Calculate the Degrees of Freedom (*df*)

The degrees of freedom (*df*) is defined as the number of values or quantities that are free to vary when a statistic is used to estimate a parameter. In testing the mean of one sample in Chapter 7, the degrees of freedom was equal to  $N - 1$  (the number of scores in the sample minus 1). However, because samples have now been drawn from two populations rather than one, the number of degrees of freedom has changed. Formula 9-1 presents the formula for the degrees of freedom for the difference between two sample means:

$$df = (N_1 - 1) + (N_2 - 1) \quad (9-1)$$

where  $N_1$  and  $N_2$  are the sample sizes of the two groups.

For the parking lot study, where both groups have a sample size of 15 ( $N_i = 15$ ), the degrees of freedom are calculated as follows:

$$\begin{aligned} df &= (N_1 - 1) + (N_2 - 1) \\ &= (15 - 1) + (15 - 1) = 14 + 14 \\ &= 28 \end{aligned}$$

By combining the degrees of freedom for the two samples, there are a total of 28 degrees of freedom ( $df = 28$ ) for the data in the parking lot study.

### Set Alpha ( $\alpha$ ), Identify the Critical Values, and State a Decision Rule

The second step in making the decision to reject the null hypothesis consists of three parts. The first part is to set alpha ( $\alpha$ ), which is the probability of the inferential statistic needed to reject the null hypothesis. The second part is to use the degrees of freedom and stated value of alpha to identify the critical values of the statistic we are about to calculate; the critical values determine the values of the statistic that result in the decision to reject the null hypothesis. The third part is to state a decision rule that explicitly states the logic to be followed in making the decision whether to reject the null hypothesis.

In this chapter, alpha will be set to the traditional value of .05, implying the null hypothesis will be rejected if the probability of the calculated value of the statistic is less than .05. Furthermore, because the alternative hypothesis in the parking lot study ( $H_1: \mu_{\text{Intruder}} \neq \mu_{\text{No Intruder}}$ ) is nondirectional, alpha may be stated as “ $\alpha = .05$  (two-tailed).”

To identify the critical values of the statistic, we must first determine which particular inferential statistic will be calculated. In Chapter 7, when the population standard deviation for a variable was unknown, the sample mean was transformed into a  $t$ -statistic, which was part of the Student  $t$ -distribution. In this chapter, we have a similar situation in that the population standard deviations for the two groups are assumed to be unknown. As such, we will transform the difference between the two sample means into a  $t$ -statistic and again rely on the Student  $t$ -distribution.

To demonstrate how to identify the critical value of  $t$ , we return to the table of critical values for the  $t$ -distribution provided in Table 3 in the back of this book. Three pieces of information are needed to identify the critical value: the degrees of freedom ( $df$ ), alpha ( $\alpha$ ), and the directionality of the alternative hypothesis (one-tailed or two-tailed). For the parking lot study, because we've calculated the degrees of freedom to be 28 ( $df = 28$ ), we move down the  $df$  column of Table 3 until we reach the number 28. Next, because our alternative hypothesis is nondirectional, we move to the right to the set of columns labeled “Level of significance for two-tailed test.” Within this set of columns, we move to the column for the stated value of .05 for alpha ( $\alpha$ ). Using these three pieces of information, we find a critical  $t$  value of 2.048. Therefore, for the parking lot study, the critical values may be stated as the following:

$$\text{For } \alpha = .05 \text{ (two-tailed) and } df = 28, \text{ critical values} = \pm 2.048.$$

The critical values and regions of rejection and nonrejection for the parking lot study are illustrated in Figure 9.2. This figure demonstrates that values of the  $t$ -statistic that are either less than  $-2.048$  or greater than  $2.048$  have a “low” probability of occurring, that is, a probability less than  $.05$ .

Once the critical values have been identified, it is helpful to explicitly state a decision rule that specifies the logic to be followed in making the decision to reject or not reject the null hypothesis. For the parking lot study, the following decision rule is stated:

$$\text{If } t < -2.048 \text{ or } > 2.048, \text{ reject } H_0; \text{ otherwise, do not reject } H_0.$$

In this situation, the null hypothesis will be rejected if the value of  $t$  we calculate from our data is either less than  $-2.048$  or greater than  $2.048$  because such a value is located in the region of rejection, meaning it has a low ( $< .05$ ) probability of occurring.

### Calculate a Statistic: $t$ -Test for Independent Means

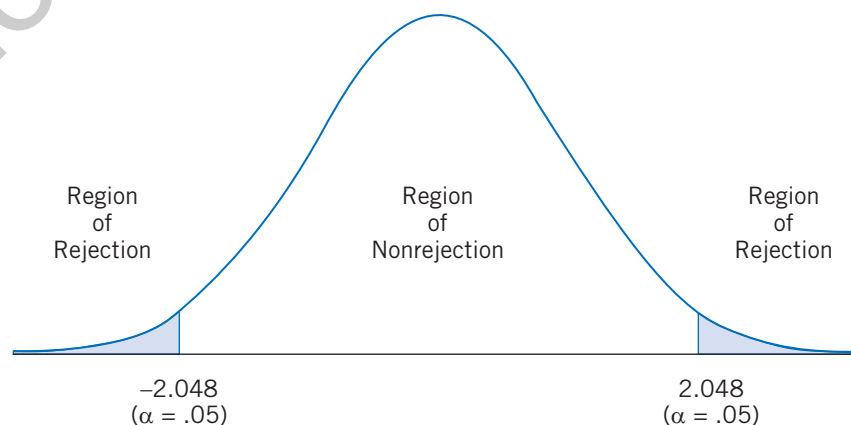
The next step in hypothesis testing is to calculate a value of an inferential statistic. As mentioned earlier, in testing the difference between two sample means, we will calculate a value of a  $t$ -statistic. Formula 9-2 provides the formula for the  **$t$ -test for independent means**, defined as an inferential statistic that tests the difference between the means of two samples drawn from two populations:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} \quad (9-2)$$

where  $\bar{X}_1$  and  $\bar{X}_2$  are the means for the two groups and  $(s_{\bar{X}_1 - \bar{X}_2})$  is the standard error of the difference. This statistic is called the  $t$ -test for “independent” means to indicate that the data were collected from two samples drawn from two different populations, and that scores in one sample are unrelated to scores in the other sample. (Later in this chapter, we’ll discuss research situations in which the two scores are related to each other because both are collected from the same participant.) For the parking lot study, “independence” implies that the No Intruder and Intruder drivers are from two different populations of drivers.

If you compare Formula 9-2 with Chapter 7’s Formula 7-4 (the  $t$ -test for one mean), you see the two formulas are very similar. First, the numerator of both formulas involves the difference between means, which in both cases is the primary issue of interest in the analysis. However, the numerator of Formula 9-2 includes

FIGURE 9.2 CRITICAL VALUES AND REGIONS OF REJECTION AND NONREJECTION FOR THE PARKING LOT STUDY



the difference between two sample means rather than the difference between a sample mean and a population mean. Second, although the denominator of both formulas contains a measure of variability, the denominator of Formula 9-2 measures the variability of differences between sample means rather than the variability of sample means.

The first step in calculating the  $t$ -statistic is to calculate the standard error of the difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ) using Formula 9-3:

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}} \quad (9-3)$$

where  $s_1$  and  $s_2$  are the standard deviations for the two groups and  $N_1$  and  $N_2$  are the sample sizes of the two groups. The formula for the standard error of the difference is very similar to the formula for the standard error of the mean ( $s_{\bar{X}}$ ) discussed in Chapter 7, the critical difference being that we must now take into account the variability in two samples rather than one.

Using the standard deviations calculated in Table 9.2, the standard error of the difference for the parking lot study is calculated as follows:

$$\begin{aligned} s_{\bar{X}_1 - \bar{X}_2} &= \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}} \\ &= \sqrt{\frac{(10.42)^2}{15} + \frac{(10.08)^2}{15}} = \sqrt{7.24 + 6.77} = \sqrt{14.01} \\ &= 3.74 \end{aligned}$$

In calculating the standard error of the difference, students sometimes get confused by the fact that the symbol  $s_{\bar{X}_1 - \bar{X}_2}$  includes a minus sign ( $-$ ) but the formula contains the plus sign ( $+$ ). The formula involves addition ( $+$ ) because in order to estimate the variability of differences between two sample means, we need to combine (add together) the variability of the two samples. We are evaluating the difference between sample means, *not* the difference between sample standard deviations.

Inserting our value of the standard error of the difference and the Intruder and No Intruder sample means into Formula 9-2, a value of the  $t$ -statistic may now be calculated:

$$\begin{aligned} t &= \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} \\ &= \frac{40.73 - 31.67}{3.74} = \frac{9.06}{3.74} \\ &= 2.42 \end{aligned}$$

As you can see from these calculations, we've transformed the difference between the two sample means ( $40.73 - 31.67$ ) into a  $t$ -statistic. Furthermore, the calculated value for the  $t$ -statistic ( $t = 2.42$ ) is a positive number. This is because the mean of the first group,  $M_{\text{Intruder}} = 40.73$ , is greater than the mean of the second group,  $M_{\text{No Intruder}} = 31.67$ . Because the designation of the first and second groups is arbitrary, the  $t$ -statistic would have been  $-2.42$  if the No Intruder group had been the first group and the Intruder group the second.

### Make a Decision Whether to Reject the Null Hypothesis

The next step is to make a decision whether to reject the null hypothesis by comparing the value of the  $t$ -statistic calculated from the data with the identified critical values. If the  $t$ -statistic exceeds one of the critical values, it falls in the region of rejection, and the decision is made to reject the null hypothesis. This decision implies that



## LEARNING CHECK 2

### Reviewing What You've Learned So Far

1. Review questions
  - a. In testing the difference between two sample means, what conclusions are represented by the null and alternative hypotheses ( $H_0$  and  $H_1$ )?
  - b. What is the difference between the degrees of freedom ( $df$ ) for the test of one sample mean (Chapter 7) and the difference between two sample means?
  - c. What are the main differences between the formulas for the  $t$ -test for one sample mean (Chapter 7) and the  $t$ -test for independent means?
  - d. What is the main difference between the formulas for the standard error of the mean (Chapter 7) and the standard error of the difference?
2. State the null and alternative hypotheses ( $H_0$  and  $H_1$ ) for the following research questions:
  - a. Do children from single-parent families have higher self-esteem than children from intact families?
  - b. Does tap water taste any different than bottled water?
  - c. Who is more likely to intervene when a crime takes place: someone walking alone or someone walking with other people?
3. For each of the following situations, calculate the degrees of freedom ( $df$ ) and identify the critical value of  $t$  (assume  $\alpha = .05$ ).
  - a.  $N_1 = 7, N_2 = 7, H_1: \mu_1 \neq \mu_2$
  - b.  $N_1 = 11, N_2 = 11, H_1: \mu_1 \neq \mu_2$
  - c.  $N_1 = 8, N_2 = 8, H_1: \mu_1 > \mu_2$
4. For each of the following situations, calculate the standard error of the difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ).
  - a.  $N_1 = 9, s_1 = 3.00, N_2 = 9, s_2 = 6.00$
  - b.  $N_1 = 16, s_1 = 8.00, N_2 = 16, s_2 = 12.00$
  - c.  $N_1 = 24, s_1 = 3.38, N_2 = 24, s_2 = 3.76$
5. For each of the following situations, calculate the standard error of the difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ) and the  $t$ -test for independent means.
  - a.  $N_1 = 6, \bar{X}_1 = 9.00, s_1 = 3.00, N_2 = 6, \bar{X}_2 = 6.00, s_2 = 2.00$
  - b.  $N_1 = 15, \bar{X}_1 = 53.62, s_1 = 13.54, N_2 = 15, \bar{X}_2 = 49.12, s_2 = 11.87$
  - c.  $N_1 = 36, \bar{X}_1 = 7.63, s_1 = 2.90, N_2 = 36, \bar{X}_2 = 10.50, s_2 = 3.38$

the difference between the means of the two groups is statistically significant, meaning it is unlikely to occur as the result of chance factors.

For the parking lot study, the decision about the null hypothesis may be stated the following way:

$$t = 2.42 > 2.048 \therefore \text{reject } H_0 (p < .05)$$

Because our calculated value of  $t$  of 2.42 is greater than the critical value 2.048, it falls in the region of rejection and the decision is made to reject the null hypothesis because its probability is less than .05; in other words,  $p < .05$ . As a result, we've decided that the difference between the two groups in the amount of time taken to leave their parking spaces ( $M_{\text{Intruder}} = 40.73$  and  $M_{\text{No Intruder}} = 31.67$ ) is statistically significant.

### Determine the Level of Significance

If and when the null hypothesis is rejected, it is appropriate to determine whether the probability of the  $t$ -statistic is not only less than .05 but also less than .01. In Chapter 7, we discussed that when researchers calculate a statistic using statistical software, they often report the exact probability of the statistic (i.e., " $p = .017$ ") rather

than use cutoffs such as “ $p < .05$ .” However, in this textbook, we include the step of determining whether the probability of a statistic meets the .01 level of significance because “ $p < .01$ ” is commonly included in tables and figures of statistics in journal articles.

To determine whether the probability of the  $t$ -statistic for the parking lot study is less than .01, we return to the table of critical values in Table 3. Moving down to the  $df = 28$  row, we move to the right until we are under the .01 column under “Level of significance for two-tailed test.” Here, we find the critical value 2.763. Comparing the calculated  $t$ -statistic of 2.42 with the  $\alpha = .01$  critical value of 2.763, we reach the following conclusion:

$$t = 2.42 < 2.763 \therefore p < .05 \text{ (but not } < .01)$$

Figure 9.3 illustrates the location of the value of  $t$  from this example relative to the .05 and .01 critical values. Because the calculated  $t$ -statistic of 2.42 was greater than the .05 critical value of 2.048, it falls in the region of rejection and we made the decision to reject the null hypothesis. However, because 2.42 is *less* than the .01 critical value of 2.763, its probability is less than .05 but *not* less than .01. Consequently, we would report “ $p < .05$ ” as the level of significance. As a reminder, when the value of a statistic doesn’t exceed the .01 critical value, this does *not* mean the null hypothesis is not rejected; that decision has already been made. It simply means that the probability of the statistic is not less than .01.

### Draw a Conclusion From the Analysis

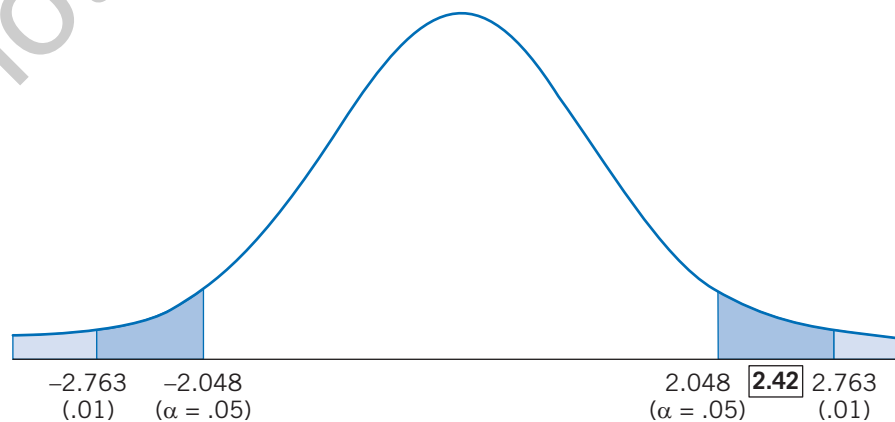
Given that we’ve made the decision to reject the null hypothesis, what conclusion can be drawn about the difference between the departure times of the Intruder and No Intruder groups? One way to report the results of the analysis is the following:

The mean departure time for the 15 drivers in the Intruder group ( $M = 40.73$  s) is significantly greater than the mean departure time for the 15 drivers in the No Intruder group ( $M = 31.67$  s),  $t(28) = 2.42$ ,  $p < .05$ .

Note that this single sentence provides the following information:

- the dependent variable (“The mean departure time”),
- the two samples (“15 drivers . . . 15 drivers”),
- the independent variable (“Intruder group . . . No Intruder group”),
- descriptive statistics (“ $M = 40.73$  s . . .  $M = 31.67$  s”),

FIGURE 9.3 ■ DETERMINING THE LEVEL OF SIGNIFICANCE FOR THE PARKING LOT STUDY



- the nature and direction of the findings (“significantly greater than”), and
- information about the inferential statistic (“ $t(28) = 2.42, p < .05$ ”), which indicates the inferential statistic calculated ( $t$ ), the degrees of freedom (28), the value of the statistic (2.42), and the level of significance ( $p < .05$ ).

### Relate the Result of the Analysis to the Research Hypothesis

It’s critical to relate the result of the statistical analysis back to the research hypothesis it was designed to test. In the parking lot study, does the finding that drivers took a significantly longer amount of time to leave a parking space in the presence of an intruder support or not support the study’s research hypothesis? Here is what the authors of the study had to say:

The present series of studies is consistent with prior findings that people display territorial defense in public territories. . . . What is new about the present research is that it suggests people sometimes display territorial behavior merely to keep others from possessing the space even when it no longer has any value to them. (Ruback & Juieng, 1997, p. 831)

Note that the researchers did not say that their study “proved” their hypothesis but rather that it “*suggests* people *sometimes* display territorial behavior.” As was discussed in Chapter 6, research hypotheses cannot be proven because data are collected from a sample rather than the entire population.

The process of testing the difference between two sample means is summarized in Box 9.1, using the parking lot study as an example. In closing, you may remember an old saying: “A watched pot never boils.” Research suggests this may reflect how people behave in social situations. Sometimes a ‘watched’ person really does take longer to move!

### Assumptions of the $t$ -Test for Independent Means

The goal of statistical procedures such as the  $t$ -test is to test hypotheses researchers have about populations by analyzing data collected from samples of these populations. These procedures make certain mathematical assumptions about the distribution of scores for variables in the population. To appropriately use these procedures, researchers must determine whether the data in their samples meet these assumptions. This section discusses assumptions related to the  $t$ -test for independent means and strategies available to researchers if their data do not meet these assumptions.

#### SUMMARY BOX 9.1 TESTING THE DIFFERENCE BETWEEN TWO SAMPLE MEANS (PARKING LOT STUDY EXAMPLE)

##### State the null and alternative hypotheses ( $H_0$ and $H_1$ )

$$H_0: \mu_{\text{Intruder}} = \mu_{\text{No Intruder}} \quad H_1: \mu_{\text{Intruder}} \neq \mu_{\text{No Intruder}}$$

##### Make a decision about the null hypothesis

Calculate the degrees of freedom ( $df$ )

$$df = (N_1 - 1) + (N_2 - 1) = (15 - 1) + (15 - 1) = 14 + 14 = 28$$

Set alpha ( $\alpha$ ), identify the critical values, and state a decision rule

$$\text{If } t < -2.048 \text{ or } > 2.048, \text{ reject } H_0; \text{ otherwise, do not reject } H_0$$

Calculate a statistic:  $t$ -test for independent means

Calculate the standard error of the difference ( $s_{\bar{x}_1 - \bar{x}_2}$ )

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}} = \sqrt{\frac{(10.42)^2}{15} + \frac{(10.08)^2}{15}} = \sqrt{14.01} = 3.74$$

Calculate the  $t$ -statistic ( $t$ )

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{40.73 - 31.67}{3.74} = \frac{9.06}{3.74} = 2.42$$

Make a decision whether to reject the null hypothesis

$$t = 2.42 > 2.048 \therefore \text{reject } H_0 (p < .05)$$

Determine the level of significance

$$t = 2.42 < 2.763 \therefore p < .05 \text{ (but not } < .01)$$

#### Draw a conclusion from the analysis

The mean departure time for the 15 drivers in the Intruder group ( $M = 40.73$  s) is significantly greater than the mean departure time for the 15 drivers in the No Intruder group ( $M = 31.67$  s),  $t(28) = 2.42$ ,  $p < .05$ .

#### Relate the result of the analysis to the research hypothesis

"The present series of studies is consistent with prior findings that people display territorial defense in public territories. . . . What is new about the present research is that it suggests people sometimes display territorial behavior merely to keep others from possessing the space even when it no longer has any value to them" (Ruback & Juieng, 1997, p. 831).

The first assumption related to the  $t$ -test for independent means is the assumption of normality, which is the assumption that scores for the dependent variable in each of the two populations are approximately normally distributed; that is, the shape of the two distributions resembles normal (bell-shaped) curves. The implication of this assumption is that distributions of scores for samples drawn from these populations are also expected to be approximately normal. The assumption of normality is common to many of the statistical procedures discussed in this book.

The second assumption applicable to the  $t$ -test for independent means is **homogeneity of variance**, which is the assumption that the variance of scores in the two populations is the same. Assuming this assumption is true, the variance in two samples drawn from the two populations should also be the same. Figure 9.4(a) illustrates two distributions that meet the assumptions of normality and homogeneity of variance.

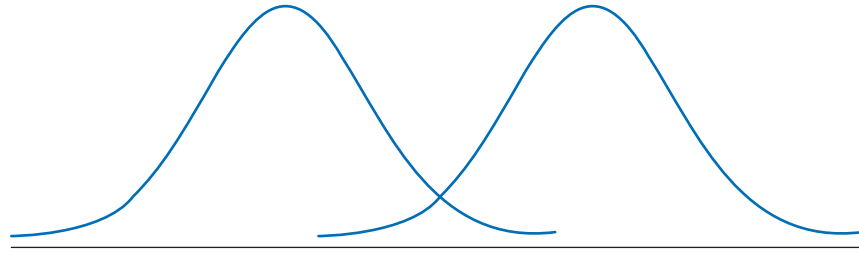
A violation of homogeneity of variance might take place if the variance in one sample is much less or greater than the variance in the other sample; this is illustrated in Figure 9.4(b). When the homogeneity of variance assumption is violated, it's possible that the two samples are not representative of their populations, which in turn raises doubt about the results of statistical analyses conducted on the samples.

Violating the assumptions of normality and homogeneity of variance increases the possibility of making the wrong decision regarding the null hypothesis for a statistic such as the  $t$ -test. For example, a researcher may decide to reject the null hypothesis and conclude the difference between the two sample means is significant when the groups do not actually differ in the populations from which the samples were drawn. (These types of errors in decision making are discussed in greater detail in the next chapter.)

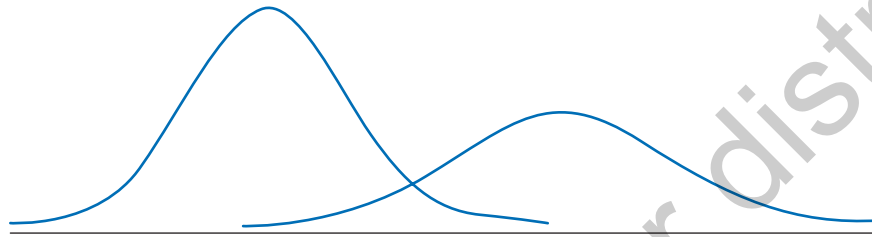
Statistical tests have been developed to determine whether the assumptions of normality and homogeneity of variance are met in a set of data. However, we will not discuss these tests for two reasons: They are

FIGURE 9.4 ■ ILLUSTRATION OF THE HOMOGENEITY OF VARIANCE ASSUMPTION

(a) Two Distributions That Meet the Assumption of Homogeneity of Variance



(b) Two Distributions That Possibly Violate the Assumption of Homogeneity of Variance



beyond the scope of this textbook, and the need to conduct these tests has been called into question (i.e., Zimmerman, 2004). As we mentioned in Chapter 7, research has found that statistics such as the  $t$ -statistic are “robust,” meaning they can withstand moderate violations of their mathematical assumptions. To withstand a violation of an assumption implies that, even if the data from a sample violate an assumption, the decision made regarding the null hypothesis is the same decision that would have been made had the assumption not been violated. For example, imagine that two populations do not differ on a variable that is normally distributed in the populations. Even if data from samples of these populations happen to violate the assumptions, a robust statistic such as the  $t$ -test will lead to the correct decision, which in this case is to not reject the null hypothesis. This is particularly true when the two samples are of equal size and are sufficiently large (i.e.,  $N_i \geq 30$ ).

Even when a statistic is considered robust, researchers may wish to address possible violations of assumptions by altering how they analyze their data. One way to do this is to use different formulas to calculate and evaluate the  $t$ -test (Welch, 1938); we will not present these formulas for the same reasons mentioned earlier. A second way to address extreme violations of assumptions is to employ an alternative statistical procedure, one that is not based on these assumptions. For example, rather than calculate the  $t$ -test for independent means, the data can be analyzed using a statistical procedure known as the Mann-Whitney  $U$ . Examples of these alternative types of statistical procedures are introduced in Chapter 14 of this textbook.

### Summary

The parking lot study compared the means of two groups calculated from samples of equal size ( $N_i = 15$ ). Having sample sizes equal to each other is preferable because it helps ensure that the two samples are treated equally in analyzing the data. However, as it is possible that the two groups may not have the same sample size, the next section describes how to test the difference between two sample means that are based on unequal sample sizes.





### LEARNING CHECK 3

#### Reviewing What You've Learned So Far

1. Review questions
  - a. What are the two mathematical assumptions related to the  $t$ -test for independent means?
  - b. What does it mean to violate the assumption of heterogeneity of variance?
  - c. What does it mean for a statistic to be robust?
  - d. What can researchers do when they believe assumptions have been violated?
2. How much do you think the average adult woman weighs? Let's imagine that a researcher, in response to available evidence from both the media and society, entertains the assumption that men believe that women weigh less than they actually do. Consequently, the researcher hypothesizes that men will give lower estimates of the average woman's weight than will women. To test this hypothesis, the researcher asks samples of men and women, "How much (in pounds) do you believe the average adult woman weighs?" The researcher reports the following descriptive statistics regarding the mean estimate of weight: Men:  $N = 13$ ,  $M = 135.62$ ,  $s = 19.78$ , Women:  $N = 13$ ,  $M = 137.69$ ,  $s = 12.68$ .
  - a. State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
  - b. Make a decision about the null hypothesis.
    - (1) Calculate the degrees of freedom ( $df$ ).
    - (2) For  $\alpha = .05$  (two-tailed), identify the critical values and state a decision rule.
    - (3) Calculate the  $t$ -test for independent means.
    - (4) Make a decision whether to reject the null hypothesis.
    - (5) Determine the level of significance.
  - c. Draw a conclusion from the analysis.
  - d. Relate the result of the analysis to the research hypothesis.

## 9.4 INFERENCE STATISTICS: TESTING THE DIFFERENCE BETWEEN TWO SAMPLE MEANS (UNEQUAL SAMPLE SIZES)

The example in this section tests the difference between the means of two samples with different sample sizes. With unequal  $N$ s, the calculation of the  $t$ -test is slightly more complicated because the formula for the standard error of the difference ( $s_{\bar{x}_1 - \bar{x}_2}$ ) must be modified.

### An Example From the Research: Kids' Motor Skills and Fitness

Childhood obesity is a serious concern in this country as it has been linked to psychological problems, such as lowered self-esteem and depression, and physical problems such as diabetes and high blood pressure. A variety of interventions have been developed to combat childhood obesity, some of which have the goal of changing children's lifestyles and eating habits. However, two researchers at the University of Northern Iowa, Oksana Matvienko and Iradge Ahrabi-Fard, wondered whether a brief intervention focusing on "the development of motor skills applied in popular sports and games is an effective approach for increasing physical activity particularly among young and preadolescent children" (Matvienko & Ahrabi-Fard, 2010, p. 299).

The researchers developed a short, 4-week intervention aimed at developing the motor skills of kindergarten and first-grade students. As part of an after-school program, students received classroom lessons on topics such as human anatomy and nutrition, played exercises and games designed to increase their physical strength and endurance, and learned different motor skills such as kicking balls and rope jumping.

The researchers obtained the permission of four elementary schools to participate in the study; students at two of the schools received the intervention, and students at the other two schools did not. This study is an example of quasi-experimental research, which involves comparing preexisting groups rather than randomly assigning participants to conditions. We will refer to the independent variable in this study as Group, which consists of two levels: Intervention and Control.

Students in both the Intervention and Control groups were measured on different physical activities. In this example, we'll focus on the number of times the student was able to successfully jump rope in 30 seconds; the dependent variable in this example will be called Jumps. To investigate how long the effect of the intervention may last, the number of jumps for each student was measured 4 months after the intervention.

Because the number of students attending each of the schools differed, the number of students in the Intervention and Control groups was not equal to each other. For this example, the number of students in the Intervention and Control groups will be 16 and 11, respectively. Data reflecting the results of this study for the two groups are presented in Table 9.3.

The descriptive statistics for the number of jumps variable are calculated in Table 9.5; the means of the two groups are displayed in Figure 9.5. Looking at this table and figure, we see that students in the Intervention group averaged a higher number of jumps in 30 seconds ( $M = 27.31$ ) than did students in the Control group ( $M = 11.91$ ).

### Inferential Statistics: Testing the Difference Between Two Sample Means (Unequal Sample Sizes)

As in the parking lot study, the next step in analyzing the motor skills study is to calculate an inferential statistic to test the study's research hypothesis. Except for one difference, the steps followed in testing the difference between two sample means are the same whether or not the two sample sizes are equal. The difference is that with unequal sample sizes, calculating the standard error of the difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ) is slightly more complicated.

#### State the Null and Alternative Hypotheses ( $H_0$ and $H_1$ )

In this example, the null hypothesis is that the average number of jumps for students who receive the intervention is the same as for students who do not receive the intervention; this represents the conclusion that the intervention does not affect students' physical fitness. The null hypothesis for the motor skills study is stated as follows:

$$H_0: \mu_{\text{Intervention}} = \mu_{\text{Control}}$$

**TABLE 9.3** NUMBER OF JUMPS FOR STUDENTS IN THE INTERVENTION AND CONTROL GROUPS

(a) Raw Data

Intervention				Control			
Student	Jumps	Student	Jumps	Student	Jumps	Student	Jumps
1	37	9	30	1	16	9	8
2	16	10	12	2	6	10	3
3	26	11	44	3	33	11	6
4	20	12	17	4	2		
5	39	13	35	5	13		
6	11	14	27	6	7		
7	22	15	38	7	11		
8	34	16	29	8	26		

(b) Grouped Frequency Distribution Tables

Intervention		
Jumps	<i>f</i>	%
> 40	1	6%
36–40	3	19%
31–35	2	13%
26–30	4	25%
21–25	1	6%
16–20	3	19%
11–15	2	13%
6–10	0	0%
≤ 5	0	0%
Total	16	100%

Control		
Jumps	<i>f</i>	%
> 40	0	0%
36–40	0	0%
31–35	1	9%
26–30	1	9%
21–25	0	0%
16–30	1	9%
11–15	2	18%
6–10	4	36%
≤ 5	2	18%
Total	11	100%

TABLE 9.4 ■ DESCRIPTIVE STATISTICS OF JUMPS FOR STUDENTS IN THE INTERVENTION AND CONTROL GROUPS

(a) Mean ( $\bar{X}_i$ )

Intervention
$\bar{X}_1 = \frac{\sum X}{N}$ $= \frac{37 + 16 + \dots + 38 + 29}{16} = \frac{437}{16}$ $= 27.31$

Control
$\bar{X}_2 = \frac{\sum X}{N}$ $= \frac{16 + 6 + \dots + 3 + 6}{11} = \frac{131}{11}$ $= 11.91$

(b) Standard Deviation ( $s_i$ )

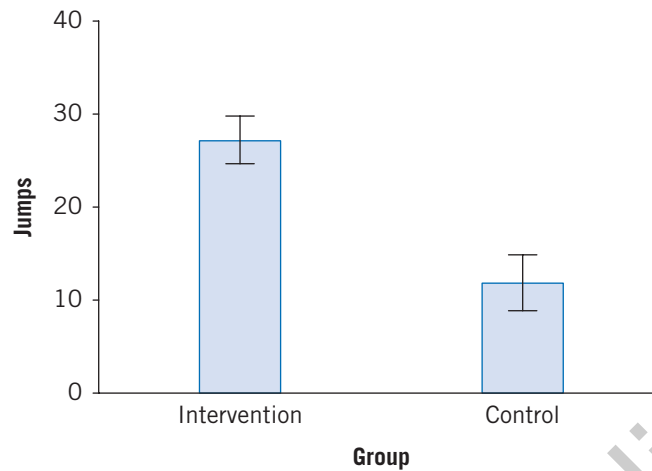
$s_1 = \sqrt{\frac{\sum(X - \bar{X})^2}{N - 1}}$ $= \sqrt{\frac{(37 - 27.31)^2 + \dots + (29 - 27.31)^2}{16 - 1}}$ $= \sqrt{\frac{93.85 + \dots + 2.85}{15}} = \sqrt{103.70}$ $= 10.18$
---

$s_2 = \sqrt{\frac{\sum(X - \bar{X})^2}{N - 1}}$ $= \sqrt{\frac{(16 - 11.91)^2 + \dots + (6 - 11.91)^2}{11 - 1}}$ $= \sqrt{\frac{16.74 + \dots + 34.92}{10}} = \sqrt{94.89}$ $= 9.74$
---

In this example, a nondirectional (two-tailed) alternative hypothesis will be used, stating that the mean number of jumps for the two groups of students in the population is *not* equal:

$$H_1: \mu_{\text{Intervention}} \neq \mu_{\text{Control}}$$

**FIGURE 9.5** BAR GRAPH OF JUMPS FOR STUDENTS IN THE INTERVENTION AND CONTROL GROUPS



### Make a Decision About the Null Hypothesis

The first step in making the decision whether to reject the null hypothesis is to *calculate the degrees of freedom (df)*. Inserting the sample sizes of  $N_1 = 16$  and  $N_2 = 11$  for the Intervention and Control groups into Formula 9-1, the degrees of freedom for the motor skills study are

$$\begin{aligned} df &= (N_1 - 1) + (N_2 - 1) \\ &= (16 - 1) + (11 - 1) = 15 + 10 \\ &= 25 \end{aligned}$$

The next step is to set alpha ( $\alpha$ ), identify the critical values, and state a decision rule. In this example, alpha will once again be defined as “ $\alpha = .05$  (two-tailed).” Next, the critical values are identified by moving down the *df* column in Table 3 until we reach the *df* = 25 row. For  $\alpha = .05$  (two-tailed), we find a critical value of 2.060. Consequently,

For  $\alpha = .05$  (two-tailed) and *df* = 25, critical value =  $\pm 2.060$ .

Based on our values of alpha and the critical values, we can state a decision rule used to make the decision about the null hypothesis. For the data in the motor skills study:

If  $t < -2.060$  or  $> 2.060$ , reject  $H_0$ ; otherwise, do not reject  $H_0$ .

The next step in making a decision about the null hypothesis is to *calculate a statistic*—in this case, the *t*-test for independent means. Looking back at Formula 9-2, the first part of calculating the *t*-statistic is to calculate the standard error of the difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ). This is the one aspect of the analysis that differs depending on whether the sample sizes are equal. Formula 9-4 presents the formula for the standard error of the difference for unequal sample sizes:

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right)} \quad (9-4)$$

where  $N_1$  and  $N_2$  are the sample sizes of the two groups, and  $s_1$  and  $s_2$  are the standard deviations of the two groups. Notice that this formula is more complicated algebraically than Formula 9-3 (the standard error of the difference with equal sample sizes) in that the sample size for each group ( $N_1$  and  $N_2$ ) must be represented multiple times. The standard error of the difference for the motor skills example is calculated as follows:

$$\begin{aligned} s_{\bar{X}_1 - \bar{X}_2} &= \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right)} \\ &= \sqrt{\frac{(16 - 1)(10.18)^2 + (11 - 1)(9.74)^2}{16 + 11 - 2} \left( \frac{1}{16} + \frac{1}{11} \right)} \\ &= \sqrt{\frac{(15)(103.63) + (10)(94.87)}{25} (.06 + .09)} \\ &= \sqrt{\frac{2,503.15}{25} (.15)} = \sqrt{15.02} \\ &= 3.88 \end{aligned}$$

Once the standard error of the difference has been calculated, we can calculate a value of the  $t$ -statistic using Formula 9-2:

$$\begin{aligned} t &= \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} \\ &= \frac{27.31 - 11.91}{3.88} = \frac{15.40}{3.88} \\ &= 3.97 \end{aligned}$$

Next, we make a decision whether to reject the null hypothesis. In this example:

$$t = 3.97 > 2.060 \therefore \text{reject } H_0 (p < .05)$$

Here, because the calculated  $t$ -statistic of 3.97 is greater than the critical value 2.060, the null hypothesis is rejected because 3.97 lies in the region of rejection at the right end of the  $t$ -distribution. This decision implies that the mean number of jumps for the two groups (27.31 for the Intervention group and 11.91 for the Control group) are significantly different from each other.

Because the decision has been made to reject the null hypothesis, it is appropriate to *determine the level of significance*. Returning to the table of critical values in Table 3, for  $df = 25$  and a .01 (two-tailed) probability we find a critical value of 2.787. Comparing the calculated  $t$ -statistic with this critical value, we draw the following conclusion:

$$t = 3.97 > 2.787 \therefore p < .01$$

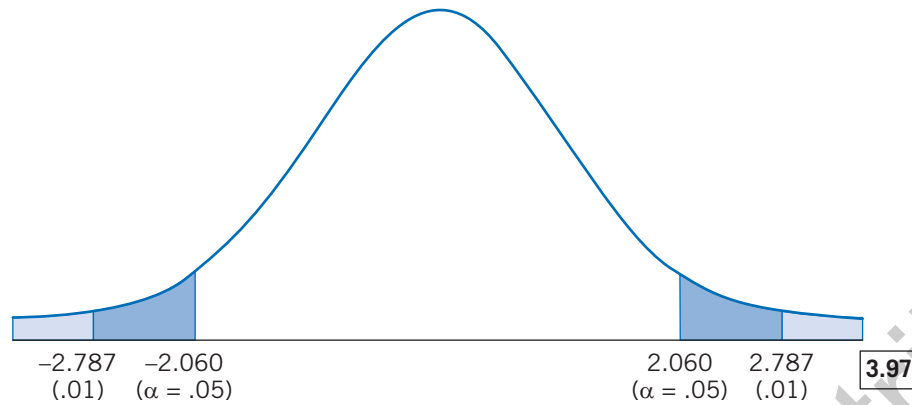
Because the  $t$ -statistic of 3.97 is greater than the .01 critical value of 2.787, its probability is not only less than .05 but also less than .01 (see Figure 9.6). Therefore, the probability of obtaining the observed difference in the mean number of jumps for the Intervention and Control groups is less than .01.

### Draw a Conclusion From the Analysis

What conclusions could we make on the basis of this analysis? Following the format of the earlier examples, the following statement could be made:

The average number of rope jumps in 30 seconds is significantly greater for the 16 students who received the intervention ( $M = 27.31$ ) than for the 11 students in the control group who did not receive the intervention ( $M = 11.91$ ),  $t(25) = 3.97$ ,  $p < .01$ .

FIGURE 9.6 DETERMINING THE LEVEL OF SIGNIFICANCE FOR THE MOTOR SKILLS STUDY



### Relate the Result of the Analysis to the Research Hypothesis

The purpose of the motor skills study was to evaluate the effectiveness of a brief intervention aimed at improving the physical activity levels of kindergarten and first-grade students. What are the implications of our statistical analysis for evaluating the effectiveness of this intervention? Here is what the authors of the study said:

This finding suggests that programs emphasizing the enhancement of basic motor skills that children apply in a variety of games and sports may be an effective approach to increasing overall activity and fitness levels of young children. (Matvienko & Ahrabi-Fard, 2010, p. 303)

The process of testing the difference between two sample means with unequal samples is summarized in Box 9.2.

### Assumptions of the $t$ -Test for Independent Means (Unequal Sample Sizes)

Earlier in this chapter, we introduced two assumptions related to the  $t$ -test for independent means: the assumption of normality, which is the assumption that the distribution of scores in the two populations is approximately normal, and the assumption of homogeneity of variance, in which the variance of scores in the two populations is the same. The data collected by researchers should meet these assumptions to draw appropriate conclusions from the results of statistical analyses. This is of particular concern when the sample sizes of the two groups are not equal to each other, as unequal sample sizes exacerbate the effects of differences in the shape and variance of the distributions of the two samples.

Statistics such as the  $t$ -statistic have been found to be robust, meaning that even if the data from a sample violate an assumption, the decision made regarding the null hypothesis is the same that would have been made if the assumption had not been violated. However, researchers have found that the robustness of statistics such as the  $t$ -statistic is lessened when the sample sizes of the groups are not equal to each other, especially when the variances are also unequal. The combination of unequal sample sizes and unequal variances may result in the need to consider the alternatives to the  $t$ -test for independent means discussed earlier in this chapter.

### Summary

The examples discussed thus far in this chapter have compared participants in different conditions or groups. For example, the parking lot study compared the departure times of drivers who either had or did not have an intruder, and the motor skills study measured the jumping rope ability of students who either received or did

not receive the after-school intervention. As such, these studies are examples of **between-subjects research designs**, defined as research designs in which each research participant appears in only one level or category of the independent variable. The word *between* indicates these designs involve testing differences between different groups of participants. The next section introduces a different type of research design, one in which each participant appears in all levels or categories of the independent variable, and as a result the researcher is testing differences *within* each participant.

### SUMMARY BOX 9.2 TESTING THE DIFFERENCE BETWEEN TWO SAMPLE MEANS (UNEQUAL SAMPLE SIZES) (MOTOR SKILLS EXAMPLE)

#### State the null and alternative hypotheses ( $H_0$ and $H_1$ )

$$H_0: \mu_{\text{Intervention}} = \mu_{\text{Control}} \quad H_1: \mu_{\text{Intervention}} \neq \mu_{\text{Control}}$$

#### Make a decision about the null hypothesis

Calculate the degrees of freedom ( $df$ )

$$df = (N_1 - 1) + (N_2 - 1) = (16 - 1) + (11 - 1) = 15 + 10 = 25$$

Set alpha ( $\alpha$ ), identify the critical values, and state a decision rule

$$\text{If } t < -2.060 \text{ or } > 2.060, \text{ reject } H_0; \text{ otherwise, do not reject } H_0$$

Calculate a statistic:  $t$ -test for independent means

Calculate the standard error of the difference (unequal sample sizes) ( $s_{\bar{X}_1 - \bar{X}_2}$ ).

$$\begin{aligned} s_{\bar{X}_1 - \bar{X}_2} &= \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2}\right)} \\ &= \sqrt{\frac{(16 - 1)(10.18)^2 + (11 - 1)(9.74)^2}{16 + 11 - 2} \left(\frac{1}{16} + \frac{1}{11}\right)} \\ &= \sqrt{\frac{2,503.15}{25} (.15)} = \sqrt{100.13(.15)} = \sqrt{15.02} = 3.88 \end{aligned}$$

Calculate the  $t$ -statistic ( $t$ )

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{27.31 - 11.91}{3.88} = \frac{15.40}{3.88} = 3.97$$

Make a decision whether to reject the null hypothesis

$$t = 3.97 > 2.060 \therefore \text{reject } H_0 (p < .05)$$

Determine the level of significance

$$t = 3.97 > 2.787 \therefore p < .01$$

#### Draw a conclusion from the analysis

The average number of rope jumps in 30 seconds is significantly greater for the 16 students who received the intervention ( $M = 27.31$ ) than for the 11 students in the control group who did not receive the intervention ( $M = 11.91$ ),  $t(25) = 3.97$ ,  $p < .01$ .

#### Relate the result of the analysis to the research hypothesis

"This finding suggests that programs emphasizing the enhancement of basic motor skills that children apply in a variety of games and sports may be an effective approach to increasing overall activity and fitness levels of young children" (Matvienko & Ahrabi-Fard, 2010, p. 303).

**LEARNING CHECK 4**

## Reviewing What You've Learned So Far

1. Review questions
  - a. How does the calculation of the  $t$ -test for the difference between two sample means change when the sample sizes of the two groups are unequal?
2. For each of the following situations, calculate the standard error of the difference ( $s_{\bar{x}_1 - \bar{x}_2}$ ) using the formula for unequal sample sizes.
  - a.  $N_1 = 4, s_1 = 2.00, N_2 = 5, s_2 = 3.00$
  - b.  $N_1 = 20, s_1 = 6.00, N_2 = 10, s_2 = 4.00$
  - c.  $N_1 = 11, s_1 = 8.64, N_2 = 17, s_2 = 11.75$
3. To assist parents whose children have been diagnosed as having the eating disorder anorexia nervosa, one study evaluated the impact of a program designed to provide parents access to information and support groups (Carlton & Pyle, 2007). At the end of the program, parents who participated in the program ( $N = 29$ ) and a group of parents of children with anorexia who did not participate ( $N = 53$ ) completed a survey asking them to what extent they know what to feed their child at home (the higher the score, the higher the level of knowledge). The researchers reported the following descriptive statistics: participated ( $M = 4.68, s = 1.57$ ) and did not participate ( $M = 2.77, s = 2.02$ ).
  - a. State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
  - b. Make a decision about the null hypothesis.
    - (1) Calculate the degrees of freedom ( $df$ ).
    - (2) For  $\alpha = .05$  (two-tailed), identify the critical values and state a decision rule.
    - (3) Calculate a value for the  $t$ -test for independent means (when calculating the standard error of the difference ( $s_{\bar{x}_1 - \bar{x}_2}$ ), be sure to use the formula for unequal sample sizes).
    - (4) Make a decision whether to reject the null hypothesis.
    - (5) Determine the level of significance.
  - c. Draw a conclusion from the analysis.
  - d. What conclusions might the researchers draw regarding the impact of the program on parents' knowledge of anorexia nervosa?

## 9.5 INFERENCE STATISTICS: TESTING THE DIFFERENCE BETWEEN PAIRED MEANS

This section will again use a research study to introduce a statistical procedure that tests the difference between two means. However, this study differs from the earlier examples in that each research participant appears in both levels of the independent variable rather than just one. It is an example of a **within-subjects research design**, which is a research design in which each participant appears in all levels or categories of the independent variable. Rather than testing the difference *between* different groups of participants, within-subjects research designs test differences *within* the same participant.

### Research Situations Appropriate for Within-Subjects Research Designs

There are several types of research situations where within-subjects research designs are used. First, these designs are used to examine differences within a person regarding different situations or stimuli. An example of this situation would be to have people taste two types of ice cream and rate both types in terms of their flavor. One research study that used a within-subjects design for this purpose looked at police officers' beliefs regarding eyewitnesses' ability to provide accurate information about a crime (Kebbell & Milne, 1998). A sample of police



officers were asked how often they believed eyewitnesses provided accurate information regarding four different aspects of a crime: the *person* who committed the crime, the *action* taken by the perpetrator, the *object* or target of the crime, and the *surroundings* in which the crime took place. Comparing these officers' ratings of these four aspects, the researchers found that officers believed eyewitnesses were more likely to provide useful information regarding the action involved in the crime than about the person committing the crime, the object of the crime, or the surroundings. That is, they believed witnesses are better able to describe the actions of a criminal than the actual criminal.

A second use of within-subjects designs involves collecting data on the same variable across repeated administrations. **Longitudinal research** is a research design in which the same information is collected from a research participant over two or more administrations to assess changes occurring within the person over time. An example of a longitudinal study, conducted by the author of this textbook (Tokunaga, 1985), examined the relationship between experiencing the death of a close friend or relative and changes in one's own fear of death during the initial bereavement period. People who had experienced this type of loss were contacted and asked to complete a survey measuring death-related fears every 4 months over a 12-month period.

Another type of longitudinal research, the **pretest-posttest research design**, involves collecting data from participants before and after the introduction of an intervention or experimental manipulation to determine whether the intervention or manipulation is associated with changes in the dependent variable. The research study described below is an example of a pretest-posttest design.

### An Example From the Research: Web-Based Family Interventions

Given the large number of families in which both parents work, concern has been expressed regarding parents' ability to be aware of their children's well-being. A team of researchers led by Diane Deitz set out to develop and evaluate a program designed to increase parents' knowledge of problems such as childhood anxiety and depression (Deitz, Cook, Billings, & Hendrickson, 2009). As the researchers wrote, "The prevalence of mental disorders in youth is substantial; however, these disorders often go unrecognized by parents and those closest to them . . . [therefore] the purpose of the project was to test a web-based program providing working parents with the knowledge and skills necessary for prevention and early intervention of mental health problems in youth" (p. 488).

In the study, the researchers developed a web-based program in which parents could work through a series of online modules covering such things as information about different mental disorders and treatment options for these disorders. As part of their study, the researchers hypothesized that "parents receiving the web-based program would exhibit significant gains in . . . knowledge of mental health issues in youth" (Deitz et al., 2009, p. 489).

The participants in the study, referred to as the web-based intervention study, were working parents with at least one child living at home. Before starting the program, each parent answered a test of 32 true/false items assessing their knowledge of depression, anxiety, treatment options, and parenting. Approximately 3 weeks later, after completing the program, parents completed the test a second time. In the study, the dependent variable, Knowledge, is the number of items each parent answered correctly. The independent variable, Time, consists of two levels: the test score before starting the program (Pretest) and after completing the program (Posttest). The purpose of the study was to see whether parents' knowledge of mental health issues in youth were higher after completing the program than at the start of the program.

The findings of the study will be illustrated using a sample of 20 parents. The parents' scores on the Knowledge variable both before (Pretest) and after (Posttest) the program are presented in Table 9.7(a). Notice that each of the research participants appears in both levels of the independent variable. For example, before the program was introduced, the first parent answered 16 items correctly; after the program, this same parent received a score of 24.

Table 9.8 calculates the mean and standard deviation for the Knowledge variable for the Pretest and Posttest time periods. From examining the descriptive statistics and the bar graph in Figure 9.7, we see that the mean Knowledge scores among these parents increased from the Pretest ( $M = 15.55$ ) to Posttest ( $M = 21.15$ ).

## Inferential Statistics: Testing the Difference Between Paired Means

To test the study's research hypothesis, the next step is to calculate an inferential statistic to determine whether the difference between the two means is statistically significant. As in the earlier examples, the inferential statistic will be a  $t$ -test. However, having each research participant appear in both levels of the independent variable alters the steps used to test the difference between the two means.

### Calculate the Difference Between the Paired Scores

In a between-subjects design, scores in the different groups are independent of each other. For example, in the parking lot study, the departure time for the first driver in the Intruder group is completely unrelated to the departure time of the first driver in the No Intruder group. However, in a pretest-posttest research design, the scores in the two groups are related to each other. For example, in the web-based intervention study, the first pretest and posttest knowledge scores of 16 and 24 in Table 9.5 were produced by the same person.

**TABLE 9.5** KNOWLEDGE OF ANXIETY AND DEPRESSION FOR PARENTS AT PRETEST AND POSTTEST

(a) Raw Data

Parent	Time	
	Pretest	Posttest
1	16	24
2	23	23
3	15	19
4	12	17
5	21	27
6	17	17
7	21	31
8	20	17
9	13	20
10	19	27

Parent	Time	
	Pretest	Posttest
11	12	22
12	14	19
13	17	28
14	8	9
15	12	20
16	16	22
17	10	15
18	14	27
19	12	21
20	19	18

(b) Grouped Frequency Distribution Tables

Pretest		
Knowledge	$f$	%
> 30	0	0%
26–30	0	0%
21–25	3	15%
16–20	7	35%
11–15	8	40%
6–10	2	10%
≤ 5	0	0%
Total	20	100%

Posttest		
Knowledge	$f$	%
> 30	1	5%
26–30	4	20%
21–25	5	25%
16–20	8	40%
11–15	1	5%
6–10	1	5%
≤ 5	0	0%
Total	20	100%

**TABLE 9.6** ■ DESCRIPTIVE STATISTICS OF KNOWLEDGE FOR PARENTS AT THE PRETEST AND POSTTEST TIME PERIODS

(a) Mean ( $\bar{X}_i$ )

Pretest	Posttest
$\bar{X}_1 = \frac{\sum X}{N}$ $= \frac{16 + 23 + \dots + 12 + 19}{20} = \frac{311}{20}$ $= 15.55$	$\bar{X}_2 = \frac{\sum X}{N}$ $= \frac{24 + 23 + \dots + 21 + 18}{20} = \frac{423}{20}$ $= 21.15$

(b) Standard Deviation ( $s_i$ )

$s_1 = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$ $= \sqrt{\frac{(16 - 15.55)^2 + \dots + (19 - 15.55)^2}{20 - 1}}$ $= \sqrt{\frac{.20 + \dots + 11.90}{19}} = \sqrt{16.47}$ $= 4.06$	$s_2 = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$ $= \sqrt{\frac{(24 - 21.15)^2 + \dots + (18 - 11.91)^2}{20 - 1}}$ $= \sqrt{\frac{8.12 + \dots + 9.92}{19}} = \sqrt{27.29}$ $= 5.22$
---	---

Having the same research participant appear in all levels of an independent variable has important consequences for how the data are analyzed. When we discussed sampling error in Chapters 6 and 7, we noted that samples may vary from each other as the result of random, chance factors. One of the factors that causes variability across samples is having different people in the different samples. Using a pretest-posttest design eliminates this source of random variability. Therefore, analyzing data collected using this type of design requires explicit recognition that the scores in the two groups are related to each other and may be paired together. To indicate this pairing, we calculate the difference between each participant's two scores.

For the web-based intervention study, the difference between the knowledge scores at the posttest and pretest for each of the 20 parents, represented by the symbol  $D$ , is calculated in the last column of Table 9.7(a). Notice that some of these differences are negative numbers. The presence of negative numbers simply indicates that a participant's posttest score is greater than his or her pretest score. For example, the first participant in Table 9.7 had a pretest score of 16 and a posttest score of 24, resulting in a difference of  $-8$ . Because they will be needed later in the analysis, Table 9.7(b) calculates the mean ( $\bar{X}_D$ ) and standard deviation ( $s_D$ ) of the difference scores.

The consequence of pairing the two scores is that rather than having two scores for each participant, we now have just one score—the difference score ( $D$ ). As a result, the steps in hypothesis testing for paired means are essentially identical to those presented in Chapter 7, in which we compared the mean of one sample against a hypothesized population mean. These steps are illustrated below.

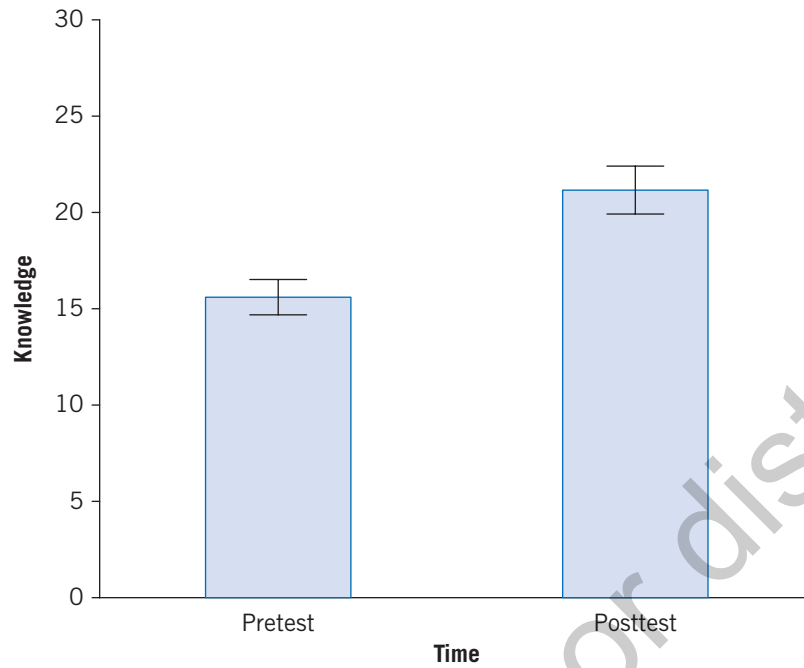
### State the Null and Alternative Hypotheses ( $H_0$ and $H_1$ )

The null hypothesis in a pretest-posttest research design reflects the belief that the two means in the population are equal to each other. In the web-based intervention study, the null hypothesis is that there is no difference between pretest and posttest knowledge scores. This absence of difference is reflected in the following null hypothesis:

$$H_0: \mu_D = 0$$

where the subscript  $D$  stands for “difference.” Stating that the mean difference between the two populations is equal to zero is simply another way of stating that the two population means are equal to each other ( $\mu_{\text{Pretest}} = \mu_{\text{Posttest}}$ ).

**FIGURE 9.7** BAR GRAPH OF KNOWLEDGE FOR PARENTS AT THE PRETEST AND POSTTEST TIMES



**TABLE 9.7** DIFFERENCE (*D*) SCORES OF KNOWLEDGE FOR PARENTS AT PRETEST AND POSTTEST

(a) Difference (*D*) Scores

Parent	Pretest	Posttest	Difference ( <i>D</i> )
1	16	24	-8
2	23	23	0
3	15	19	-4
4	12	17	-5
5	21	27	-6
6	17	17	0
7	21	31	-10
8	20	17	3
9	13	20	-7
10	19	27	-8

Parent	Pretest	Posttest	Difference ( <i>D</i> )
11	12	22	-10
12	14	19	-5
13	17	28	-11
14	8	9	-1
15	12	20	-8
16	16	22	-6
17	10	15	-5
18	14	27	-13
19	12	21	-9
20	19	18	1

(b) Descriptive Statistics

$$\begin{aligned} \bar{X}_D &= \frac{\sum D}{N_D} \\ &= \frac{-8+0+(-4)+\dots+(-13)+(-9)+1}{20} \\ &= \frac{-112}{20} \\ &= -5.60 \end{aligned}$$

$$\begin{aligned} s_D &= \sqrt{\frac{\sum (D - \bar{X}_D)^2}{N_D - 1}} \\ &= \sqrt{\frac{(-8 - (-5.60))^2 + \dots + (1 - (-5.60))^2}{20 - 1}} \\ &= \sqrt{\frac{5.76 + \dots + 43.56}{19}} = \sqrt{18.88} \\ &= 4.35 \end{aligned}$$

As in the previous examples, the alternative hypothesis may be directional or nondirectional. In the web-based intervention study, the following nondirectional alternative hypothesis will be used:

$$H_1: \mu_D \neq 0$$

Therefore, in this example, the null hypothesis will be rejected if the posttest knowledge scores are either significantly less *or* significantly greater than pretest knowledge scores.

### Make a Decision About the Null Hypothesis

The first step in making the decision to reject the null hypothesis is to calculate the degrees of freedom (*df*). Formula 9-5 presents the formula for the degrees of freedom for paired means:

$$df = N_D - 1 \quad (9-5)$$

where  $N_D$  is the number of difference scores. This degrees of freedom differs from the degrees of freedom for the difference between two sample means because we're now working with one set of scores (the difference [ $D$ ] scores) rather than two. For the web-based intervention study, because there are 20 parents, the degrees of freedom are

$$\begin{aligned} df &= N_D - 1 \\ &= 20 - 1 \\ &= 19 \end{aligned}$$

The next step is to set alpha ( $\alpha$ ), identify the critical values, and state a decision rule. Assuming alpha is set to .05, the nondirectional alternative hypothesis in this example ( $H_1: \mu_D \neq 0$ ) leads us to state that “ $\alpha = .05$  (two-tailed).” To determine the critical values, we move down the *df* column of Table 3 until we reach  $df = 19$  and then to the right until we are under the .05 heading within the “Level of significance for two-tailed test” columns. For the web-based intervention study:

$$\text{For } \alpha = .05 \text{ (two-tailed) and } df = 19, \text{ critical value} = \pm 2.093$$

We can now state a decision rule regarding the conditions under which the null hypothesis will be rejected. Given the critical values we've just identified:

$$\text{If } t < -2.093 \text{ or } > 2.093, \text{ reject } H_0; \text{ otherwise, do not reject } H_0.$$

The next step is to calculate an inferential statistic. As in this chapter's other examples, we'll calculate a *t*-test. However, in this situation we'll calculate the ***t*-test for dependent means (*t*)** to test the difference between two means based on the same participant or paired participants. The word *dependent* is used to indicate that the two means are from one sample drawn from one population, as opposed to the two samples and two populations that are the basis of the *t*-test for independent means discussed earlier in this chapter.

The formula for the *t*-test for dependent means is presented in Formula 9-6:

$$t = \frac{\bar{X}_D - \mu_D}{s_{\bar{D}}} \quad (9-6)$$

where  $\bar{X}_D$  is the mean of the difference scores,  $\mu_D$  is the hypothesized population mean of the difference scores, and  $s_{\bar{D}}$  is the standard error of the difference scores. Except for the  $D$  in the subscripts, Formula 9-6 is virtually identical to the formula for the *t*-test of one mean (Formula 7-4) presented in Chapter 7.

The three pieces of information needed to calculate the  $t$ -test for dependent means are located in different places. First, the value for  $\bar{X}_D$  is found in the descriptive statistics (Table 9.9(b)). For the web-based intervention study,  $\bar{X}_D = -5.60$ . Second, the population mean  $\mu_D$  is located in the null hypothesis. Because of how the null hypothesis in this example has been stated ( $H_0: \mu_D = 0$ ),  $\mu_D$  is equal to 0. The third piece of information, the standard error of the difference scores ( $s_{\bar{D}}$ ), is calculated using Formula 9-7:

$$s_{\bar{D}} = \frac{s_D}{\sqrt{N}} \quad (9-7)$$

where  $s_D$  is the standard deviation of the difference scores and  $N$  is the sample size. Obtaining the standard deviation of the difference scores ( $s_D$ ) for the 20 parents in the web-based intervention study from Table 9.9(b), the standard error of the difference scores ( $s_{\bar{D}}$ ) is as follows:

$$\begin{aligned} s_{\bar{D}} &= \frac{s_D}{\sqrt{N}} \\ &= \frac{4.35}{\sqrt{20}} = \frac{4.35}{4.47} \\ &= .97 \end{aligned}$$

Placing these three pieces of information into Formula 9-7, we may now calculate a value of the  $t$ -statistic:

$$\begin{aligned} t &= \frac{\bar{X}_D - \mu_D}{s_{\bar{D}}} \\ &= \frac{-5.60 - 0}{.97} = \frac{-5.60}{.97} \\ &= -5.77 \end{aligned}$$

The value of the  $t$ -statistic for the web-based intervention study is a negative number because the first mean ( $M_{\text{Pretest}} = 15.55$ ) is less than the second mean ( $M_{\text{Posttest}} = 21.15$ ).

Now that we've transformed the mean of the difference scores into a  $t$ -statistic, we're ready to make a decision whether to reject the null hypothesis. Comparing the  $t$ -statistic of  $-5.77$  calculated from this example with the critical values, we make the following decision:

$$t = -5.77 < -2.093 \therefore \text{reject } H_0 (p < .05)$$

Rejecting the null hypothesis in this study implies that the knowledge scores at the pretest and posttest are significantly different from each other.

Because the null hypothesis has been rejected, it's appropriate to determine the level of significance by determining whether the probability of our  $t$ -statistic is less than .01. In Table 3, for  $\alpha = .01$  (two-tailed) and  $df = 19$ , we find the critical value  $-2.861$ . Comparing the calculated  $t$ -statistic for this example with the .01 critical value:

$$t = -5.77 < -2.861 \therefore p < .01$$

Because the  $t$ -statistic of  $-5.77$  is less than the .01 critical value of  $-2.861$ , the level of significance for this set of data is  $p < .01$  (see Figure 9.8).

### Draw a Conclusion From the Analysis

For the web-based intervention study, we've found that the 20 parents' mean knowledge scores at the posttest are significantly greater than at the pretest. To present the results of this analysis in a more informative way, we could say the following:

The average knowledge scores for the 20 parents were significantly higher after completing the web-based intervention program ( $M = 21.15$ ) than before beginning the program ( $M = 15.55$ ),  $t(19) = -5.77$ ,  $p < .01$ .

### Relate the Result of the Analysis to the Research Hypothesis

One purpose of the web-based intervention study was to determine whether the program could provide working parents knowledge that may help them and their children with mental health–related problems. The researchers discuss the implication of finding a significant increase in knowledge from the pretest to the posttest below:

These findings indicate that the program can be an effective intervention for improving parents' knowledge of children's mental health problems and boost their confidence in handling such issues. . . . The study findings lend support to the growing literature on the utility of offering web-based programs to improve the health of the general population. (Deitz et al., 2009, p. 492)

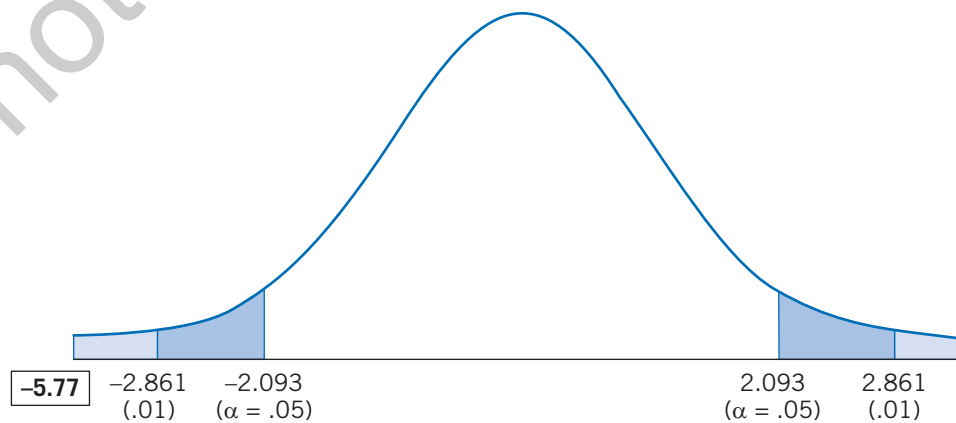
### Assumptions of the *t*-Test for Dependent Means

As with the *t*-test for independent means, the appropriate use of the *t*-test for dependent means requires that certain mathematical assumptions be met. One of the fundamental assumptions of statistics such as the *t*-test is the assumption of normality: that the data being analyzed are normally distributed. Given that the *t*-test for dependent means involves analyzing difference (*D*) scores, it's not surprising that this statistic is based on the assumption that difference scores in the population are normally distributed. The larger the sample size, the greater the likelihood of meeting this assumption.

### Summary

The steps in testing the difference between paired means are summarized in Box 9.3. Looking at this table, we see a great amount of similarity between these steps and those used to test a single mean in Chapter 7. Research studies may differ in their purpose, hypotheses, variables of interest, or method of data collection; however, the analysis of data follows a common sequencing and logic.

**FIGURE 9.8** ■ DETERMINING THE LEVEL OF SIGNIFICANCE FOR THE WEB-BASED INTERVENTION STUDY



### SUMMARY BOX 9.3 TESTING THE DIFFERENCE BETWEEN PAIRED MEANS (WEB-BASED INTERVENTION EXAMPLE)

#### State the null and alternative hypotheses ( $H_0$ and $H_1$ )

$$H_0: \mu_D = 0 \quad H_1: \mu_D \neq 0$$

#### Make a decision about the null hypothesis

Calculate the degrees of freedom ( $df$ )

$$df = N_D - 1 = 20 - 1 = 19$$

Set alpha ( $\alpha$ ), identify the critical values, and state a decision rule

$$\text{If } t < -2.093 \text{ or } > 2.093, \text{ reject } H_0; \text{ otherwise, do not reject } H_0$$

Calculate a statistic:  $t$ -test for dependent means

Calculate the standard error of the difference scores ( $s_{\bar{D}}$ )

$$s_{\bar{D}} = \frac{s_D}{\sqrt{N}} = \frac{4.35}{\sqrt{20}} = \frac{4.35}{4.47} = .97$$

Calculate the  $t$ -statistic ( $t$ )

$$t = \frac{\bar{X}_D - \mu_D}{s_{\bar{D}}} = \frac{-5.60 - 0}{.97} = \frac{-5.60}{.97} = -5.77$$

Make a decision whether to reject the null hypothesis

$$t = -5.77 < -2.093 \therefore \text{reject } H_0 (p < .05)$$

Determine the level of significance

$$t = -5.77 < -2.861 \therefore p < .01$$

#### Draw a conclusion from the analysis

The average knowledge scores for the 20 parents were significantly higher after completing the web-based intervention program ( $M = 21.15$ ) than before beginning the program ( $M = 15.55$ ),  $t(19) = -5.77$ ,  $p < .01$ .

#### Relate the result of the analysis to the research hypothesis

"These findings indicate that the program can be an effective intervention for improving parents' knowledge of children's mental health problems and boost their confidence in handling such issues. . . . The study findings lend support to the growing literature on the utility of offering web-based programs to improve the health of the general population" (Deitz et al., 2009, p. 492).



### LEARNING CHECK 5

#### Reviewing What You've Learned So Far

1. Review questions
  - a. What is the main difference between between-subjects and within-subjects research designs?
  - b. For what types of research situations might you use a within-subjects design?
  - c. In analyzing a pretest-posttest research design, why do we calculate the difference between the scores in the two groups?
  - d. When would you calculate the  $t$ -test for dependent means rather than the  $t$ -test for independent means?



2. You conduct a class project in which you divide classmates into seven pairs; each pair consists of a shy student and an outgoing student. You have each pair of students work on a puzzle but do not allow them to talk to each other. After completing the puzzle, you ask each student to indicate how much he or she enjoyed working on the puzzle:

Pair	Shy	Outgoing
1	8	4
2	4	6
3	8	2
4	6	6
5	3	7
6	6	3
7	5	4

- Calculate the difference score ( $D$ ) for each pair of students.
  - Calculate the mean ( $\bar{X}_D$ ), standard deviation ( $s_D$ ), and standard error ( $s_{\bar{D}}$ ), of the difference scores.
3. For each of the following situations, calculate the standard error of the difference scores ( $s_{\bar{D}}$ ) and the  $t$ -test for dependent means.
- $N_D = 9$ ,  $\bar{X}_D = 2.00$ ,  $s_{\bar{D}} = 1.75$
  - $N_D = 12$ ,  $\bar{X}_D = 1.50$ ,  $s_{\bar{D}} = 3.50$
  - $N_D = 21$ ,  $\bar{X}_D = 12.63$ ,  $s_{\bar{D}} = 24.82$
4. As adults get older, their lives may become increasingly sedentary. Given concerns about physical and psychological problems associated with a lack of physical activity, one study investigated whether some office tasks such as typing, working on a computer, and reading may be performed at a satisfactory level while walking on a treadmill rather than sitting (John, Bassett, Thompson, Fairbrother, & Baldwin, 2009). One part of the study involved having 20 adults take a typing test two times: once while sitting at a desk and once while walking on a treadmill. The average number of words per minute (WPM) typed correctly for the two conditions was as follows: sitting ( $M = 40.20$ ,  $s = 10.20$ ) and treadmill ( $M = 36.90$ ,  $s = 11.10$ ). Furthermore, the researchers reported a mean difference score ( $\bar{X}_D$ ) of 3.30 and a standard deviation of the difference scores ( $s_D$ ) of 4.70. Test the difference in typing speed for the two conditions.
- State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
  - Make a decision about the null hypothesis.
    - Calculate the degrees of freedom ( $df$ ).
    - For  $\alpha = .05$  (two-tailed), identify the critical values and state a decision rule.
    - Calculate a value for the  $t$ -test for dependent means.
    - Make a decision whether to reject the null hypothesis.
    - Determine the level of significance.
  - Draw a conclusion from the analysis.
  - What conclusions might the researchers draw regarding whether the task of typing may be performed at a satisfactory level while walking on a treadmill rather than sitting?

## 9.6 LOOKING AHEAD

The main purpose of the present chapter was to further your understanding of the process of hypothesis testing using situations slightly more complicated than those presented in earlier chapters. As you move further along in this book, you'll encounter research situations of growing complexity. Although the statistical procedures may become more challenging, the steps in hypothesis testing will remain essentially the same. The main

purpose of hypothesis testing is to make one of two decisions about the null hypothesis: reject or not reject. This decision is based on the probability of obtaining a calculated value of an inferential statistic when the null hypothesis is true. Relying on probability creates the possibility that the decision made about the null hypothesis may be in error. The next chapter discusses these errors in greater detail, as well as what researchers can do to minimize the possibility and impact of making these errors.

## 9.7 Summary

To test the difference between two sample means, we create the *sampling distribution of the difference*, which is the distribution of all possible values of the difference between two sample means when an infinite number of pairs of samples of size  $N$  are randomly selected from two populations. Three characteristics of the sampling distribution of the difference are that its mean is equal to zero (0), the distribution is approximately normal in shape, and the variability of this distribution is measured by the *standard error of the difference* ( $s_{\bar{X}_1 - \bar{X}_2}$ ), defined as the standard deviation of the sampling distribution of the difference. These characteristics of the sampling distribution of the difference enable researchers to determine the probability of obtaining any particular difference between the means of two samples.

The *t-test for independent means*, an inferential statistic that tests the difference between the means of two samples drawn from two populations, is used to test a hypothesis about the difference between two population means. The *t*-test may be calculated when the sample sizes for the two groups are either equal or unequal; with unequal sample sizes, the calculation of the *t*-test is slightly more complicated because the formula for the standard error of the difference must be modified.

The *t*-test for independent means is used in *between-subjects research designs*, in which each research participant appears in only one level or category of the independent variable, and these designs involve testing differences between different groups of participants. The *t-test for dependent means* is used in *within-subjects designs*, in which each participant appears in all levels or categories of the independent variable involve collecting information from each participant more than once. An example of within-subjects designs is *longitudinal research*, which involves collecting the same information from a research participant over two or more administrations to assess changes occurring within the person over time. One example of longitudinal research is the *pretest-posttest research design*, which involves collecting data from participants twice—before and after the introduction of an intervention or experimental manipulation—to determine whether the intervention or manipulation is associated with changes in the dependent variable.

## 9.8 Important Terms

bar graph (p. 267)	homogeneity of variance (p. 279)	pretest-posttest research design (p. 289)
sampling distribution of the difference (p. 270)	between-subjects research designs (p. 287)	<i>t</i> -test for dependent means ( <i>t</i> ) (p. 293)
standard error of the difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ) (p. 270)	within-subjects research designs (p. 288)	
<i>t</i> -test for independent means (p. 274)	longitudinal research (p. 289)	

## 9.9 Formulas Introduced in This Chapter

### Degrees of Freedom (*df*), Difference Between Two Sample Means

$$df = (N_1 - 1) + (N_2 - 1) \quad (9-1)$$

### *t*-Test for Independent Means

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}} \quad (9-2)$$

**Standard Error of the Difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ),**

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}} \quad (9-3)$$

**Standard Error of the Difference (Unequal Sample Sizes)**

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right)} \quad (9-4)$$

**Degrees of Freedom ( $df$ ), Difference Between Paired Means**

$$df = N_D - 1 \quad (9-5)$$

**t-Test for Dependent Means**

$$t = \frac{\bar{X}_D - \mu_D}{s_{\bar{D}}} \quad (9-6)$$

**Standard Error of Difference Scores ( $s_{\bar{D}}$ )**

$$s_{\bar{D}} = \frac{s_D}{\sqrt{N}} \quad (9-7)$$

**9.10 Using SPSS****Testing the Difference Between Two Sample Means: The Parking Lot Study (9.1)**

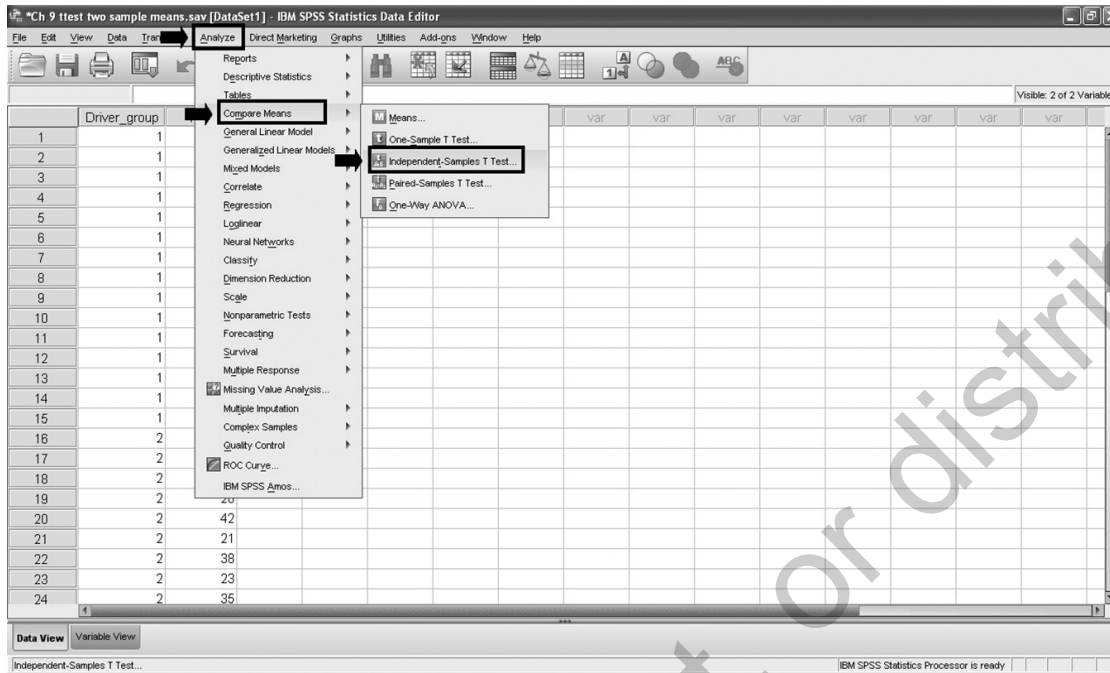
1. Define independent and dependent variables (name, # decimals, labels for the variables, labels for values of the independent variable) and enter data for the variables.

NOTE: Numerically code values of the independent variable (i.e., 1 = Intruder, 2 = No Intruder) and provide labels for these values in the **Values** box within **Variable View**.

Driver_group	Time
1	23
2	62
3	28
4	55
5	31
6	51
7	33
8	48
9	35
10	45
11	37
12	43
13	39
14	41
15	40
16	54
17	19
18	48
19	20
20	42
21	21
22	38
23	23
24	35

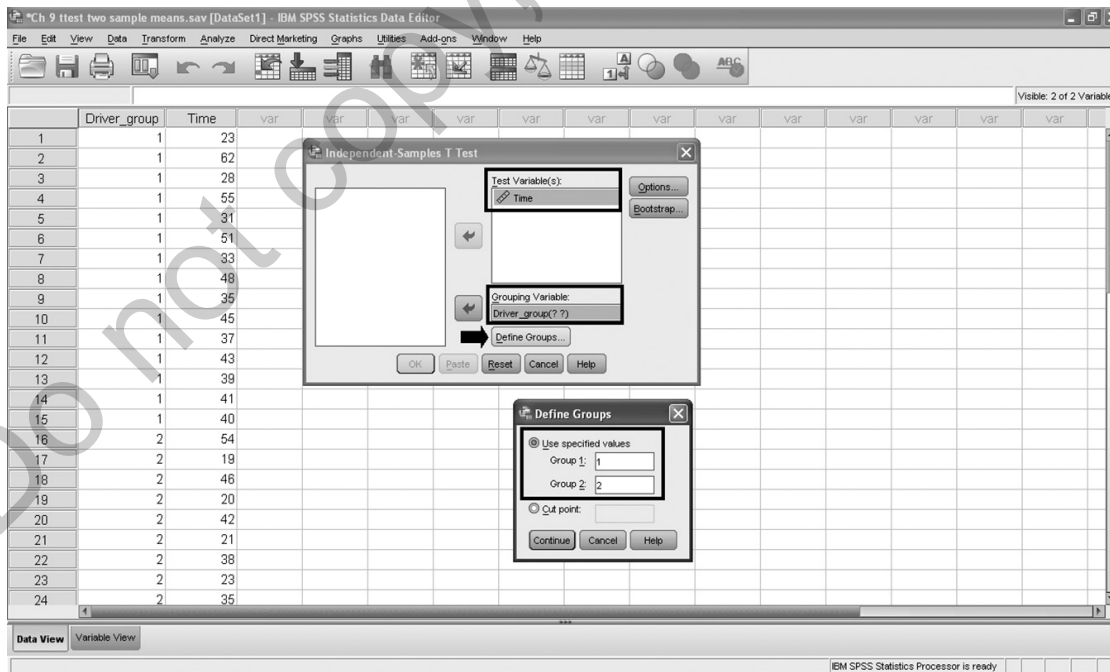
2. Select the  $t$ -test for independent means procedure within SPSS.

How? (1) Click **Analyze** menu, (2) click **Compare Means**, and (3) click **Independent-Samples T Test**.

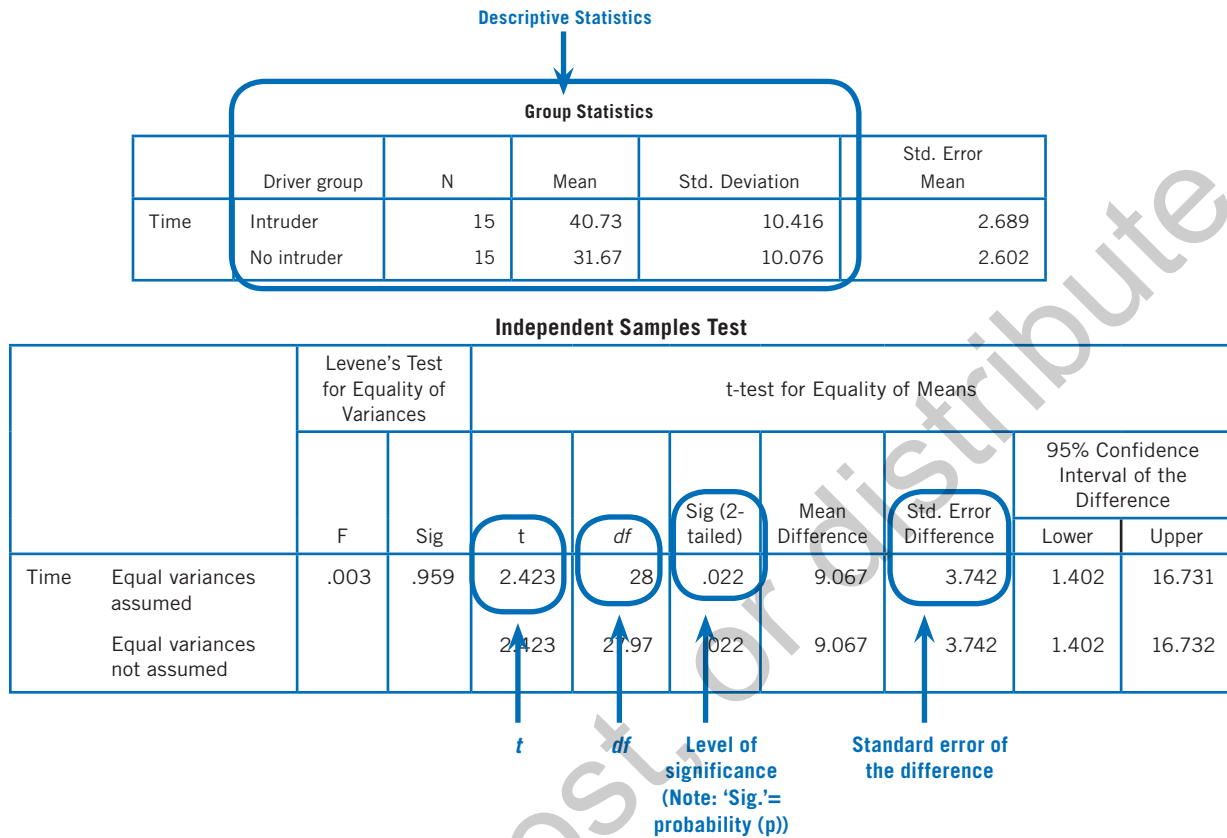


3. Identify the dependent variable, the independent variable, and the values of the independent variable.

How? (1) Click dependent variable and → **Test Variable**, (2) click independent variable and → **Grouping Variable**, (3) click **Define Groups** and type the values for the independent variable, (4) click **Continue**, and (5) click **OK**.



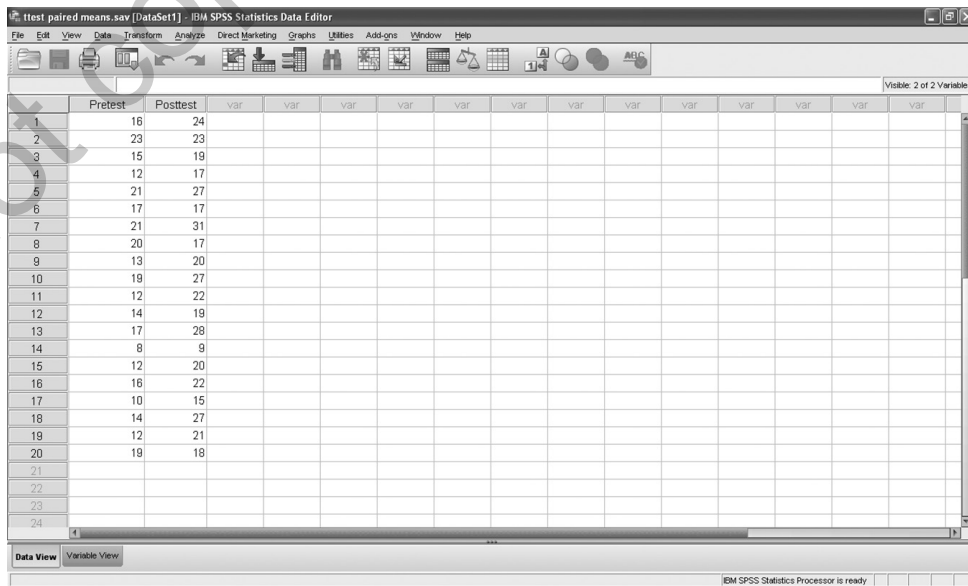
4. Examine output.



**Testing the Difference Between Paired Means: The Web-Based Intervention Study (9.5)**

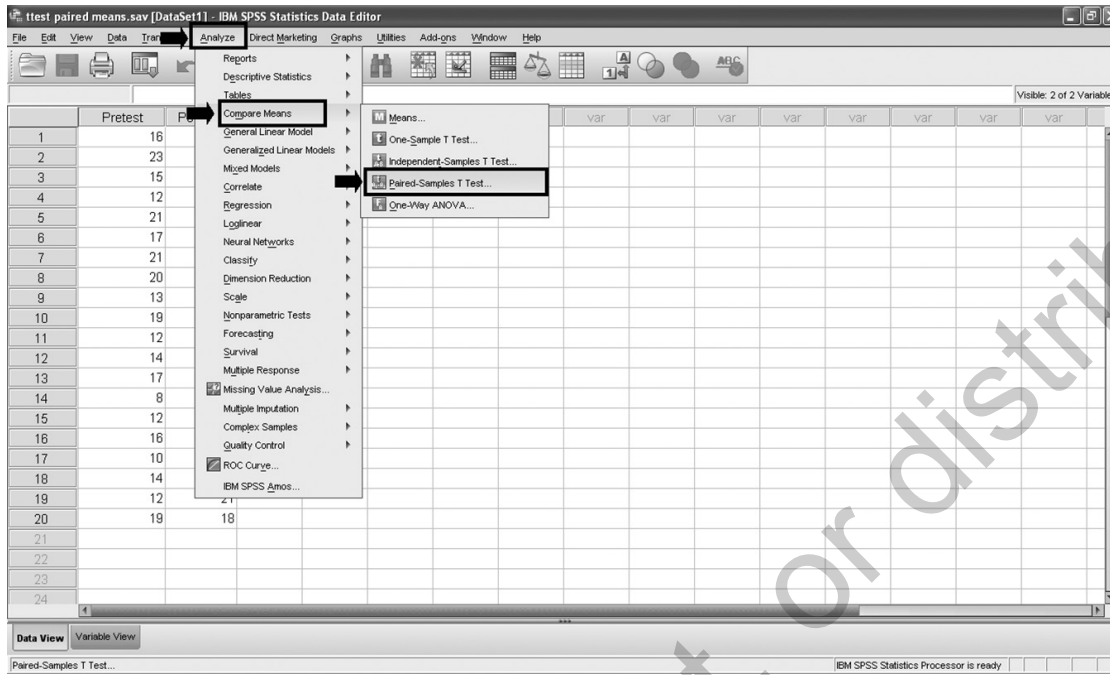
1. Define the two levels of the independent variable (name, # decimals, labels for the variables) and enter data for each level.

NOTE: Each participant has scores on two variables—each variable represents one of the two levels of the independent variable (i.e., pretest, posttest).



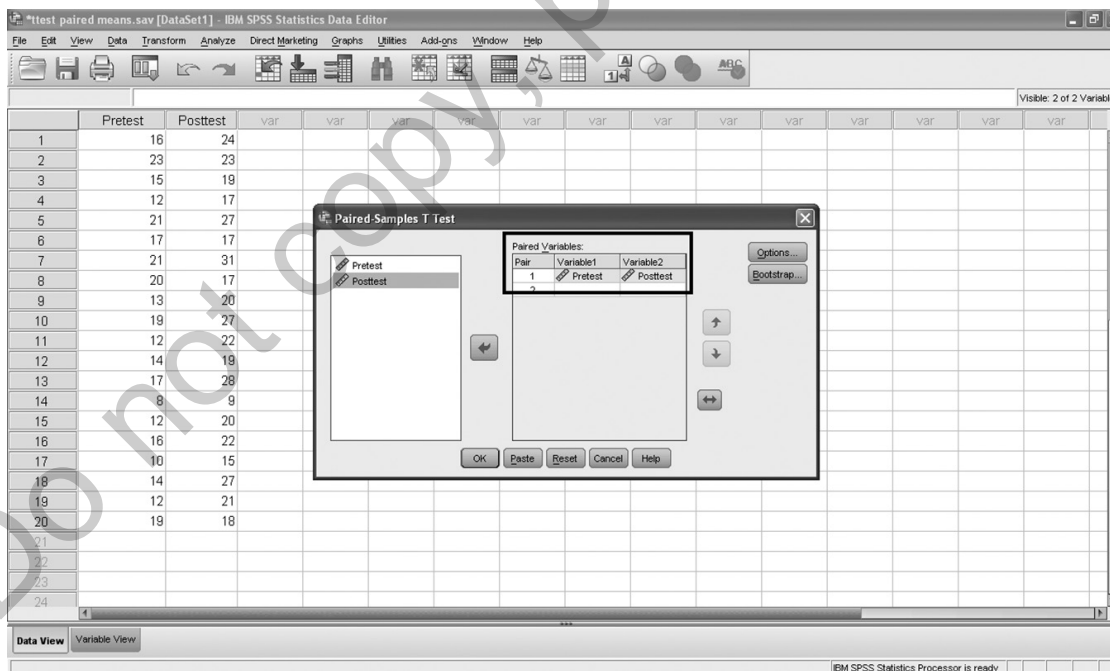
2. Begin the  $t$ -test for dependent means procedure within SPSS.

How? (1) Click **Analyze** menu, (2) click **Compare Means**, and (3) click **Paired-Samples T Test**.

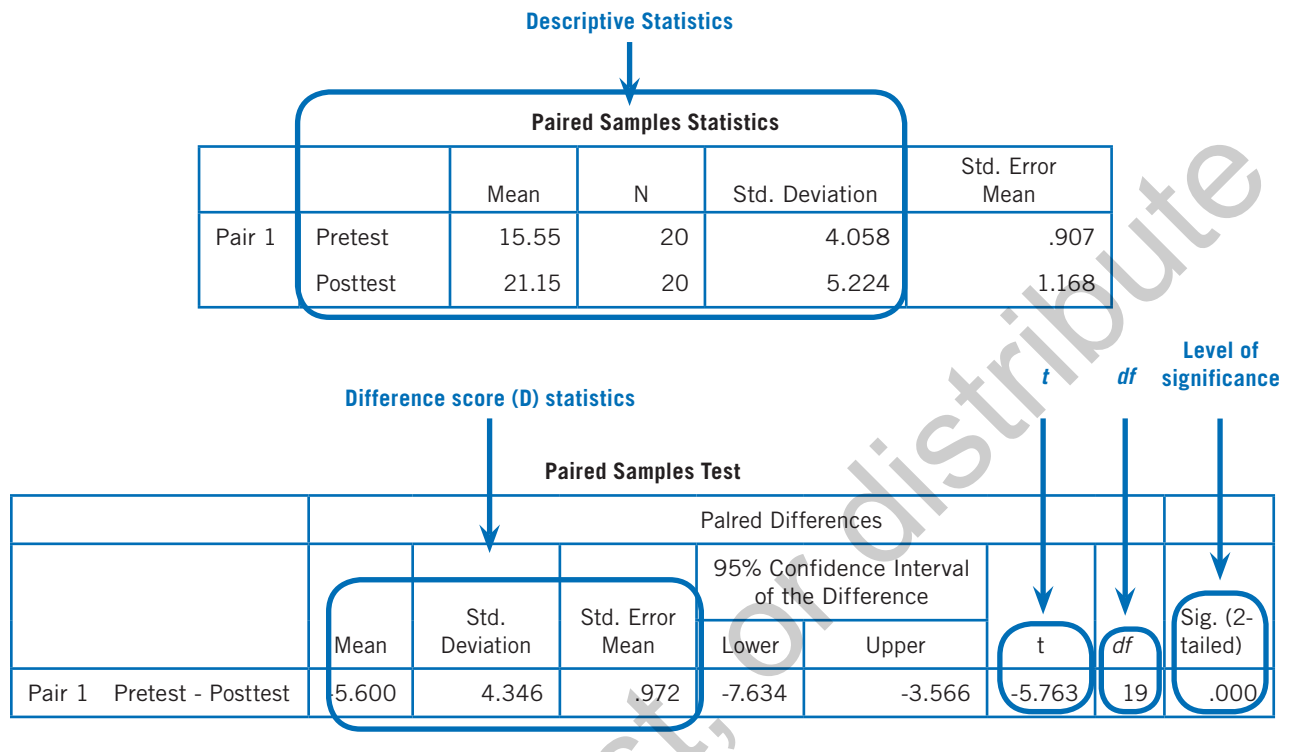


3. Identify the paired variables.

How? (1) Click the two variables and → **Paired Variables** and (2) click **OK**.



4. Examine output.



### 9.11 Exercises

1. Construct a bar graph for each of the following (assume the independent variable is group and the dependent variable is time):
  - a. Group A ( $N = 5, M = 4.00, s = 1.58$ ); Group B ( $N = 5, M = 6.00, s = 2.12$ )
  - b. Group A ( $N = 8, M = 26.00, s = 2.56$ ); Group B ( $N = 8, M = 23.00, s = 2.33$ )
  - c. Group A ( $N = 11, M = 1.77, s = .29$ ); Group B ( $N = 11, M = 1.53, s = .22$ )
  - d. Group A ( $N = 18, M = 63.59, s = 23.70$ ); Group B ( $N = 18, M = 71.42, s = 20.91$ )
  
2. Construct a bar graph for each of the following (assume the independent variable is group and the dependent variable is time):
  - a. Group A ( $N = 21, M = 14.05, s = 3.63$ ); Group B ( $N = 21, M = 12.33, s = 3.26$ )
  - b. Group A ( $N = 28, M = 6.79, s = 3.11$ ); Group B ( $N = 28, M = 7.93, s = 2.36$ )
  - c. Group A ( $N = 16, M = 52.56, s = 23.77$ ); Group B ( $N = 16, M = 60.38, s = 21.91$ )
  - d. Group A ( $N = 25, M = 5.76, s = 2.14$ ); Group B ( $N = 25, M = 4.43, s = 2.27$ )
  
3. For each of the following, (a) calculate the mean, standard deviation, and standard error of the mean for each group and (b) construct a bar graph (assume the independent variable is condition and the dependent variable is test score).
  - a. Experimental: 3, 7, 4, 1, 10, 4, 6  
Control: 10, 7, 12, 4, 8, 9

- b. Experimental: 15, 12, 10, 14, 17, 15, 18, 19  
Control: 22, 12, 17, 19, 20, 21, 16, 17
- c. Experimental: 3.89, 3.04, 3.95, 2.91, 2.72, 3.70, 3.16, 3.21, 2.86  
Control: 2.96, 3.38, 2.82, 2.07, 3.22, 2.56, 2.44, 3.11, 2.68
- d. Experimental: 2, 4, 3, 5, 1, 4, 3, 5, 4, 1, 4, 5, 3  
Control: 5, 2, 5, 1, 4, 3, 5, 2, 1, 4, 5, 1, 4
4. State the null and alternative hypotheses ( $H_0$  and  $H_1$ ) for each of the following research questions:
- Are the average starting salaries for clinical psychologists in private practice the same as or different from that of psychological researchers in business or the government?
  - In Chapter 3, we looked at psychology majors' and non-psychology majors' belief in the myth that we only use 10% of our brains. Do the two groups differ in this belief?
  - Are men paid more than women for doing the same job?
  - Do Republicans and Democrats similarly support a national health insurance program or does one group favor this more than the other?
5. For each of the following, calculate the degrees of freedom ( $df$ ) and determine the critical values of  $t$  (assume  $\alpha = .05$ ).
- $N_1 = 21, N_2 = 21, H_1: \mu_1 \neq \mu_2$
  - $N_1 = 14, N_2 = 14, H_1: \mu_1 \neq \mu_2$
  - $N_1 = 4, N_2 = 4, H_1: \mu_1 > \mu_2$
  - $N_1 = 32, N_2 = 32, H_1: \mu_1 < \mu_2$
6. For each of the following, calculate the degrees of freedom ( $df$ ) and determine the critical values of  $t$  (assume  $\alpha = .05$ ).
- $N_1 = 5, N_2 = 5, H_1: \mu_1 \neq \mu_2$
  - $N_1 = 12, N_2 = 12, H_1: \mu_1 < \mu_2$
  - $N_1 = 9, N_2 = 9, H_1: \mu_1 \neq \mu_2$
  - $N_1 = 16, N_2 = 16, H_1: \mu_1 > \mu_2$
7. For each of the following, calculate the standard error of the difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ).
- $N_1 = 35, s_1 = 1.50, N_2 = 35, s_2 = 3.25$
  - $N_1 = 4, s_1 = 4.30, N_2 = 4, s_2 = 2.10$
  - $N_1 = 23, s_1 = 8.10, N_2 = 23, s_2 = 7.50$
  - $N_1 = 20, s_1 = 1.20, N_2 = 20, s_2 = 1.75$
8. For each of the following, calculate the standard error of the difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ).
- $N_1 = 10, s_1 = 2.00, N_2 = 10, s_2 = 3.00$
  - $N_1 = 19, s_1 = 1.73, N_2 = 19, s_2 = 1.48$
  - $N_1 = 27, s_1 = 24.91, N_2 = 27, s_2 = 27.02$
  - $N_1 = 50, s_1 = 12.29, N_2 = 50, s_2 = 10.63$
9. For each of the following, calculate the  $t$ -test for independent means.
- $\bar{X}_1 = 3.49, \bar{X}_2 = 3.14, s_{\bar{X}_1 - \bar{X}_2} = .31$
  - $\bar{X}_1 = 13.27, \bar{X}_2 = 16.45, s_{\bar{X}_1 - \bar{X}_2} = 1.52$
  - $\bar{X}_1 = .76, \bar{X}_2 = .91, s_{\bar{X}_1 - \bar{X}_2} = .09$
  - $\bar{X}_1 = 1.52, \bar{X}_2 = 1.36, s_{\bar{X}_1 - \bar{X}_2} = .05$
10. For each of the following, calculate the  $t$ -test for independent means.
- $\bar{X}_1 = 7.00, \bar{X}_2 = 11.00, s_{\bar{X}_1 - \bar{X}_2} = 1.17$



- b.  $\bar{X}_1 = 65.56, \bar{X}_2 = 60.92, s_{\bar{X}_1 - \bar{X}_2} = 2.88$   
 c.  $\bar{X}_1 = 137.73, \bar{X}_2 = 114.09, s_{\bar{X}_1 - \bar{X}_2} = 10.71$   
 d.  $\bar{X}_1 = 73.24, \bar{X}_2 = 81.53, s_{\bar{X}_1 - \bar{X}_2} = 4.39$
11. For each of the following, calculate the standard error of the difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ) and the  $t$ -test for independent means.
- a.  $N_1 = 6, \bar{X}_1 = 18.50, s_1 = 2.00, N_2 = 6, \bar{X}_2 = 19.00, s_2 = 2.50$   
 b.  $N_1 = 13, \bar{X}_1 = 36.23, s_1 = 4.17, N_2 = 13, \bar{X}_2 = 29.59, s_2 = 6.01$   
 c.  $N_1 = 29, \bar{X}_1 = 7.80, s_1 = 1.25, N_2 = 29, \bar{X}_2 = 7.17, s_2 = 1.63$   
 d.  $N_1 = 38, \bar{X}_1 = 21.14, s_1 = 4.38, N_2 = 38, \bar{X}_2 = 23.29, s_2 = 3.91$
12. For each of the following, calculate the standard error of the difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ) and the  $t$ -test for independent means.
- a.  $N_1 = 11, \bar{X}_1 = 12.00, s_1 = 2.00, N_2 = 11, \bar{X}_2 = 10.00, s_2 = 3.00$   
 b.  $N_1 = 21, \bar{X}_1 = 39.85, s_1 = 5.23, N_2 = 21, \bar{X}_2 = 44.16, s_2 = 4.60$   
 c.  $N_1 = 33, \bar{X}_1 = 4.37, s_1 = 1.07, N_2 = 33, \bar{X}_2 = 4.92, s_2 = .94$   
 d.  $N_1 = 57, \bar{X}_1 = 53.98, s_1 = 6.96, N_2 = 57, \bar{X}_2 = 50.74, s_2 = 6.03$
13. When you're interviewing for a job, is your behavior influenced by beliefs the interviewer has about you? Researchers Ridge and Reber (2002) observed the interactions of 54 men, each of whom was interviewing a woman for a job. Right before conducting their interviews, half of the men were told the woman whom they were interviewing was attracted to them. The researchers hypothesized that "these same men would elicit relatively more flirtatious behavior from women than would men holding no such belief" (p. 2). The interviews were videotaped, and the number of flirtatious behaviors performed by each woman was counted. The researchers reported the following descriptive statistics regarding the mean number of flirtatious behaviors: attraction condition ( $N = 27, M = 37.44, s = 5.21$ ) and no attraction condition ( $N = 27, M = 34.59, s = 4.54$ ).
- a. State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
- b. Make a decision about the null hypothesis.
- (1) Calculate the degrees of freedom ( $df$ ).
  - (2) Set alpha ( $\alpha$ ), identify the critical values, and state a decision rule.
  - (3) Calculate a value for the  $t$ -test for independent means.
  - (4) Make a decision whether to reject the null hypothesis.
  - (5) Determine the level of significance.
- c. Draw a conclusion from the analysis.
- d. Relate the result of the analysis to the research hypothesis.
14. An advertising agency is interested in learning how to fit its commercials to the interests and needs of the viewing audience. It asked samples of 41 men and 41 women to report the average amount of television watched daily. The men reported a mean television time of 1.70 hours per day with a standard deviation of .70. The women reported a mean of 2.05 hours per day with a standard deviation of .80. Use these data to test the manager's claim that there is a significant gender difference in television viewing.
- a. State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
- b. Make a decision about the null hypothesis.
- (1) Calculate the degrees of freedom ( $df$ ).
  - (2) Set alpha ( $\alpha$ ), identify the critical values (draw the distribution), and state a decision rule.
  - (3) Calculate a value for the  $t$ -test for independent means.
  - (4) Make a decision whether to reject the null hypothesis.
  - (5) Determine the level of significance.

- c. Draw a conclusion from the analysis.
- d. What are the implications of this analysis for the advertising agency?
15. A pair of researchers was interested in studying men's preferences for a potential mate and how these preferences may change as men get older (Alterovitz & Mendelsohn, 2009). The researchers believed that men prefer women who are younger than themselves; furthermore, they hypothesized that the preferred difference in age is greater for older men than for younger men. As part of a larger study, they examined personal advertisements placed by men who were either 20 to 34 years old or 40 to 54 years old; in looking at these ads, the researchers calculated the difference between the man's age and the man's preferred age of a potential partner. The descriptive statistics for the difference in age variable for the two age groups are as follows:

20 to 34 years old:  $N_1 = 25$ ,  $\bar{X}_1 = 1.04$ ,  $s_1 = 2.72$

40 to 54 years old:  $N_1 = 25$ ,  $\bar{X}_1 = 4.98$ ,  $s_1 = 3.80$

- a. State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
- b. Make a decision about the null hypothesis.
- (1) Calculate the degrees of freedom ( $df$ ).
  - (2) Set alpha ( $\alpha$ ), identify the critical values (draw the distribution), and state a decision rule.
  - (3) Calculate a value for the  $t$ -test for independent means.
  - (4) Make a decision whether to reject the null hypothesis.
  - (5) Determine the level of significance.
- c. Draw a conclusion from the analysis.
- d. Relate the result of the analysis to the research hypothesis.
16. A third-grade teacher is interested in comparing the effectiveness of two styles of instruction in language comprehension: imagery, in which the students are asked to picture a situation involving the word, and repetition, in which the students repeat the definition of the word multiple times. After 6 weeks of instruction, she gives the students a language comprehension test. The following scores are the number of correct answers for each student. Determine whether the two styles of instruction differ in their effectiveness.
- Imagery: 12, 13, 11, 11, 13, 13, 15, 12, 9, 12
- Repetition: 6, 11, 10, 12, 9, 10, 11, 12, 10, 8
- a. For each group, calculate the sample size ( $N$ ), mean ( $\bar{X}$ ), and standard deviation ( $s$ ).
- b. State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
- c. Make a decision about the null hypothesis.
- (1) Calculate the degrees of freedom ( $df$ ).
  - (2) Set alpha ( $\alpha$ ), identify the critical values (draw the distribution), and state a decision rule.
  - (3) Calculate a value for the  $t$ -test for independent means.
  - (4) Make a decision whether to reject the null hypothesis.
  - (5) Determine the level of significance.
- d. Draw a conclusion from the analysis.
- e. What are the implications of this analysis for the teacher?
17. Imagine you believe that men and women have different beliefs regarding the average cost of a wedding. More specifically, because they are traditionally more involved in the planning of weddings and therefore have a more complete understanding of their costs, you hypothesize that young women will provide a higher estimate of the cost of the average wedding than will men. To test your hypothesis, you ask a sample of college students (13 women and 13 men), "How much (in thousands) do you believe the average wedding costs?" The estimates (in thousands of dollars) of these students are below:

Women: 20, 9, 14, 11, 25, 18, 12, 33, 24, 10, 30, 14, 22

Men: 15, 6, 19, 16, 33, 4, 21, 13, 10, 7, 24, 5, 20

- a. For each group, calculate the sample size ( $N$ ), mean ( $\bar{X}$ ), and standard deviation ( $s$ ).
  - b. State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
  - c. Make a decision about the null hypothesis.
    - (1) Calculate the degrees of freedom ( $df$ ).
    - (2) Set alpha ( $\alpha$ ), identify the critical values (draw the distribution), and state a decision rule.
    - (3) Calculate a value for the  $t$ -test for independent means.
    - (4) Make a decision whether to reject the null hypothesis.
    - (5) Determine the level of significance.
  - d. Draw a conclusion from the analysis.
  - e. Relate the result of the analysis to the research hypothesis.
18. For each of the following, calculate the standard error of the difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ) (be sure to use the formula for unequal sample sizes).
- a.  $N_1 = 8, s_1 = 2.30, N_2 = 12, s_2 = 2.00$
  - b.  $N_1 = 25, s_1 = 5.15, N_2 = 22, s_2 = 7.80$
  - c.  $N_1 = 13, s_1 = 4.50, N_2 = 19, s_2 = 5.89$
  - d.  $N_1 = 50, s_1 = 21.00, N_2 = 45, s_2 = 17.50$
19. For each of the following, calculate the standard error of the difference ( $s_{\bar{X}_1 - \bar{X}_2}$ ) (be sure to use the formula for unequal sample sizes).
- a.  $N_1 = 9, s_1 = 4.00, N_2 = 7, s_2 = 2.00$
  - b.  $N_1 = 10, s_1 = 8.50, N_2 = 20, s_2 = 7.80$
  - c.  $N_1 = 33, s_1 = 1.87, N_2 = 41, s_2 = 5.89$
  - d.  $N_1 = 21, s_1 = 16.21, N_2 = 19, s_2 = 3.56$
20. A team of researchers stated that effective methods of therapy help people access and process their emotions (Watson & Bedard, 2006). They looked at the level of emotional processing brought about by two different types of therapy for clients diagnosed with depression. The first group ( $N = 17$ ) took part in cognitive-behavioral therapy (CBT). The second group ( $N = 21$ ) took part in process-experiential therapy (PET). The Experiencing Scale measured their level of emotional processing. The CBT group scored a mean of 2.73 on the scale, with a standard deviation of .46. The PET group scored a mean of 3.04, with a standard deviation of .42. Use the steps of hypothesis testing to determine whether one therapeutic technique brings about more emotional processing than the other.
- a. State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
  - b. Make a decision about the null hypothesis.
    - (1) Calculate the degrees of freedom ( $df$ ).
    - (2) Set alpha ( $\alpha$ ), identify the critical values (draw the distribution), and state a decision rule.
    - (3) Calculate a value for the  $t$ -test for independent means (when calculating the standard error of the difference [ $s_{\bar{X}_1 - \bar{X}_2}$ ], be sure to use the formula for unequal sample sizes).
    - (4) Make a decision whether to reject the null hypothesis.
    - (5) Determine the level of significance.
  - c. Draw a conclusion from the analysis.
  - d. Does one of the therapy methods appear to be more effective than the other?

21. A team of audiologists was interested in examining whether their patients' satisfaction with their hearing aids was related to how long they had used hearing aids (Williams, Johnson, & Danhauer, 2009). They divided their patients into two categories, new users ( $N = 30$ ) and experienced users ( $N = 34$ ), and asked them to indicate how satisfied they were with their hearing aids; the higher the score, the greater the satisfaction. The new users reported a mean satisfaction of 26.90 on the scale (standard deviation = 3.96), and the experienced users reported a mean satisfaction of 28.03 (standard deviation = 5.04). Use the steps of hypothesis testing to test the difference in satisfaction between new and experienced users.
- State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
  - Make a decision about the null hypothesis.
    - Calculate the degrees of freedom ( $df$ ).
    - Set alpha ( $\alpha$ ), identify the critical values (draw the distribution), and state a decision rule.
    - Calculate a value for the  $t$ -test for independent means (when calculating the standard error of the difference [ $s_{\bar{x}_1 - \bar{x}_2}$ ], be sure to use the formula for unequal sample sizes).
    - Make a decision whether to reject the null hypothesis.
    - Determine the level of significance.
  - Draw a conclusion from the analysis.
  - Is patients' satisfaction with hearing aids related to how long they have used hearing aids?
22. Imagine a researcher asks a sample of five people to drive two types of cars and rate each of them on a 1 to 20 scale. Listed below are the data she collected:

Person	Type A	Type B
1	20	10
2	10	11
3	11	4
4	23	12
5	7	0

- Calculate the difference score ( $D$ ) for each of the five participants.
  - Calculate the mean ( $\bar{X}_D$ ), standard deviation ( $s_D$ ), and standard error ( $s_{\bar{D}}$ ) of the difference scores.
23. For each of the following, calculate the standard error of the difference scores ( $s_{\bar{D}}$ ) and the  $t$ -test for dependent means.
- $N_D = 25$ ,  $\bar{X}_D = 10.00$ ,  $s_D = 8.00$
  - $N_D = 5$ ,  $\bar{X}_D = 6.80$ ,  $s_D = 4.71$
  - $N_D = 16$ ,  $\bar{X}_D = 2.12$ ,  $s_D = 2.17$
  - $N_D = 20$ ,  $\bar{X}_D = .80$ ,  $s_D = 3.10$
24. For each of the following, calculate the standard error of the difference scores ( $s_{\bar{D}}$ ) and the  $t$ -test for dependent means.
- $N_D = 10$ ,  $\bar{X}_D = 12.00$ ,  $s_D = 9.00$
  - $N_D = 19$ ,  $\bar{X}_D = 3.52$ ,  $s_D = 6.14$
  - $N_D = 8$ ,  $\bar{X}_D = 2.89$ ,  $s_D = 4.86$
  - $N_D = 30$ ,  $\bar{X}_D = 1.44$ ,  $s_D = 8.32$
25. As people in our society live longer, there is a growing need for older adults to maintain a healthy lifestyle, part of which is their physical fitness. One study evaluated a program designed to increase older adults' level of fitness using

weight and strength training (Doll, 2009). In the study, eight adults (mean age = 75.60 years) received training on a weight machine over an 8-week period, doing exercises such as bench presses and abdominal crunches. Before and after beginning the training, each adult was measured on a number of functional tasks, one of which was the number of times they could lift a 5-pound weight over their heads. The descriptive statistics for the number of overhead lifts were the following: pretest ( $M = 45.60, s = 15.00$ ) and posttest ( $M = 53.90, s = 18.60$ ). Furthermore, the mean difference score ( $\bar{X}_D$ ) was  $-8.30$ ; the standard deviation of the difference scores ( $s_D$ ) was  $9.84$ . Test the difference in overhead lifts between the pretest and posttest.

- a. State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
  - b. Make a decision about the null hypothesis.
    - (1) Calculate the degrees of freedom ( $df$ ).
    - (2) Set alpha ( $\alpha$ ), identify the critical values (draw the distribution), and state a decision rule.
    - (3) Calculate a value for the  $t$ -test for dependent means.
    - (4) Make a decision whether to reject the null hypothesis.
    - (5) Determine the level of significance.
  - c. Draw a conclusion from the analysis.
  - d. What conclusions might the researchers draw regarding the effectiveness of a weight and strength training program designed for older adults?
26. Another part of the study discussed in Exercise 25 had a second sample of nine older adults take part in an 8-week calisthenics program in which they did squats, hamstring curls, and lunges. As the program was designed to help these adults perform everyday activities more easily, before and after the program, each adult was asked how difficult it was for them to get in and out of a bathtub (the higher the number, the greater the difficulty). The descriptive statistics for this bathtub-related difficulty variable were the following: pretest ( $M = 4.30, s = 2.00$ ) and posttest ( $M = 2.90, s = .90$ ). Furthermore, the mean difference score ( $\bar{X}_D$ ) was  $1.40$ ; the standard deviation of the difference scores ( $s_D$ ) was  $1.86$ . Test the difference in bathtub-related difficulty between the pretest and posttest.
- a. State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
  - b. Make a decision about the null hypothesis.
    - (1) Calculate the degrees of freedom ( $df$ ).
    - (2) Set alpha ( $\alpha$ ), identify the critical values (draw the distribution), and state a decision rule.
    - (3) Calculate a value for the  $t$ -test for dependent means.
    - (4) Make a decision whether to reject the null hypothesis.
    - (5) Determine the level of significance.
  - c. Draw a conclusion from the analysis.
  - d. What conclusions might the researchers draw regarding the effectiveness of a calisthenics program designed for older adults?
27. A group of friends decided to work together to decrease their cigarette smoking. They hypothesize that they can decrease their level of smoking by doing such things as giving one another encouragement, using carrots and candy to replace the physical action of smoking, and using deep breathing and counting through cravings. Each friend keeps track of how many cigarettes he or she smokes each day. The first set of data is each person's baseline (before beginning the plan), and the second set of data is after 4 weeks of using the tools. Complete the steps below to test their hypothesis.

Person	Before	After
1	8	0
2	25	13

(Continued)

(Continued)

Person	Before	After
3	17	0
4	5	1
5	12	14
6	20	10

- For each of the two time periods, calculate the sample size ( $N_i$ ), the mean ( $\bar{X}_i$ ), and the standard deviation ( $s_i$ ).
  - Calculate the mean ( $\bar{X}_D$ ) and the standard deviation of the difference scores ( $s_D$ ).
  - State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
  - Make a decision about the null hypothesis.
    - Calculate the degrees of freedom ( $df$ ).
    - Set alpha ( $\alpha$ ), identify the critical values (draw the distribution), and state a decision rule.
    - Calculate a value for the  $t$ -test for dependent means.
    - Make a decision whether to reject the null hypothesis.
    - Determine the level of significance.
  - Draw a conclusion from the analysis.
  - Relate the result of the analysis to the research hypothesis.
28. A researcher hypothesizes that an afternoon dose of caffeine lessens the amount of time a person sleeps at night. The participants are instructed to report their average hours of sleep per night for a week without any caffeine in the afternoons and their average amount of sleep for a week in which they drink two caffeinated beverages between 2 and 3 p.m. Using their reported data, run through the steps of hypothesis testing to test the research hypothesis.

Person	Without Caffeine	With Caffeine
1	9.00	6.00
2	6.00	5.50
3	8.00	8.00
4	7.00	6.25
5	7.50	5.00
6	7.00	5.00
7	6.00	7.00
8	9.00	8.00
9	8.00	7.00
10	7.00	6.00

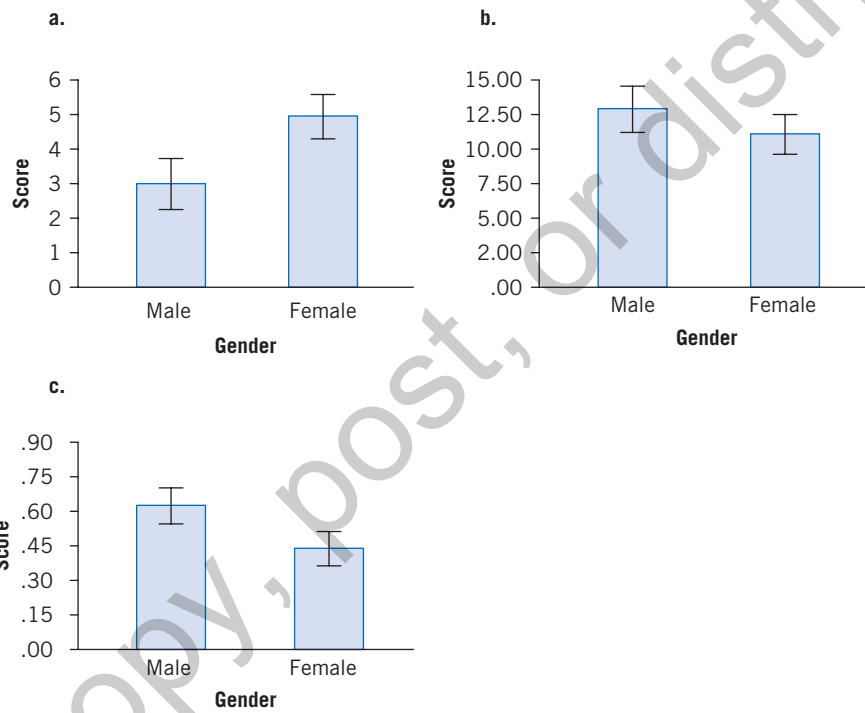
- For each of the two caffeine conditions, calculate the sample size ( $N_i$ ), the mean ( $\bar{X}_i$ ), and the standard deviation ( $s_i$ ).
- Calculate the mean ( $\bar{X}_D$ ) and the standard deviation of the difference scores ( $s_D$ ).
- State the null and alternative hypotheses ( $H_0$  and  $H_1$ ).
- Make a decision about the null hypothesis.
  - Calculate the degrees of freedom ( $df$ ).
  - Set alpha ( $\alpha$ ), identify the critical values (draw the distribution), and state a decision rule.

- (3) Calculate a value for the  $t$ -test for dependent means.
- (4) Make a decision whether to reject the null hypothesis.
- (5) Determine the level of significance.
- e. Draw a conclusion from the analysis.
- f. Relate the result of the analysis to the research hypothesis.

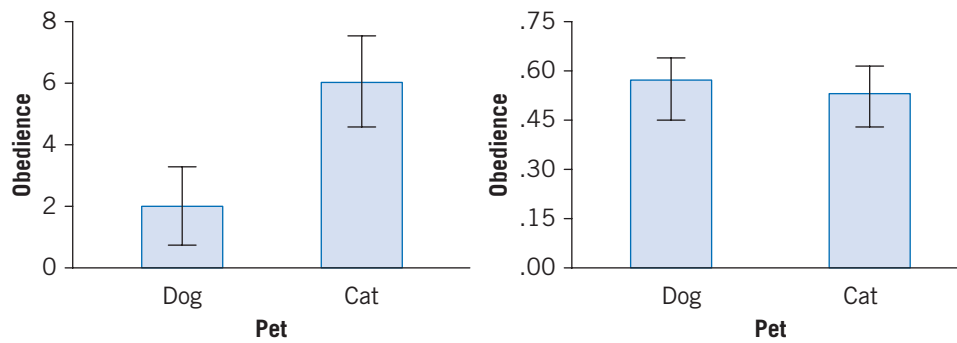
## Answers to Learning Checks

### Learning Check 1

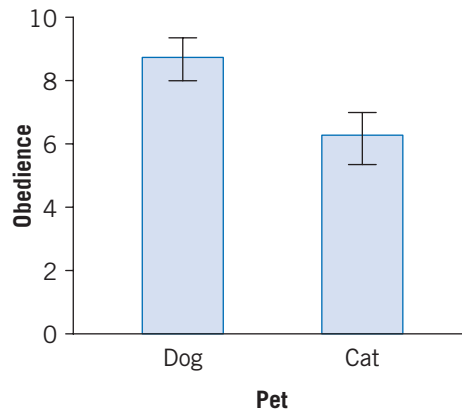
2.



3. a. Dog:  $\bar{X} = 2.00, s = 2.83, s_{\bar{X}} = 1.27$ ; Cat:  $\bar{X} = 6.00; s = 3.39, s_{\bar{X}} = 1.52$



- b. Dog:  $\bar{X} = .57, s = .20, s_{\bar{X}} = .06$ ; Cat:  $\bar{X} = .53, s = .27, s_{\bar{X}} = .09$   
 c. Dog:  $\bar{X} = 8.71, s = 2.30, s_{\bar{X}} = .62$ ; Cat:  $\bar{X} = 6.29, s = 2.02, s_{\bar{X}} = .54$



### Learning Check 2

- $H_0: \mu_{\text{Single parent}} = \mu_{\text{Intact}}; H_1: \mu_{\text{Single parent}} \neq \mu_{\text{Intact}}$
  - $H_0: \mu_{\text{Tap water}} = \mu_{\text{Bottled water}}; H_1: \mu_{\text{Tap water}} \neq \mu_{\text{Bottled water}}$
  - $H_0: \mu_{\text{Alone}} = \mu_{\text{With others}}; H_1: \mu_{\text{Alone}} \neq \mu_{\text{With others}}$
- $df = 12$ ; critical value =  $\pm 2.179$
  - $df = 20$ ; critical value =  $\pm 2.086$
  - $df = 14$ ; critical value = 1.761
- $s_{\bar{X}_1 - \bar{X}_2} = 2.24$
  - $s_{\bar{X}_1 - \bar{X}_2} = 3.61$
  - $s_{\bar{X}_1 - \bar{X}_2} = 1.03$
- $s_{\bar{X}_1 - \bar{X}_2} = 1.47; t = 2.04$
  - $s_{\bar{X}_1 - \bar{X}_2} = 4.65; t = .97$
  - $s_{\bar{X}_1 - \bar{X}_2} = .74; t = -3.88$

### Learning Check 3

- $H_0: \mu_{\text{Men}} = \mu_{\text{Women}}; H_1: \mu_{\text{Men}} \neq \mu_{\text{Women}}$
  - $df = 24$
    - If  $t < -2.064$  or  $> 2.064$ , reject  $H_0$ ; otherwise, do not reject  $H_0$
    - $s_{\bar{X}_1 - \bar{X}_2} = 6.52; t = -.32$
    - $t = -.32$  is not  $< -2.064$  or  $> 2.064 \therefore$  do not reject  $H_0$  ( $p > .05$ )
    - Not applicable ( $H_0$  not rejected)
  - The average estimate of a woman's weight in the sample of 13 men ( $M = 135.62$ ) and 13 women ( $M = 137.69$ ) was not significantly different,  $t(24) = -.32, p > .05$ .
  - The results of this analysis do not support the research hypothesis that men will give lower estimates of the average woman's weight than will women.

### Learning Check 4

- $s_{\bar{X}_1 - \bar{X}_2} = 1.76$
  - $s_{\bar{X}_1 - \bar{X}_2} = 2.11$



- c.  $s_{\bar{X}_1 - \bar{X}_2} = 4.13$
3. a.  $H_0: \mu_{\text{participate}} = \mu_{\text{not participate}}; H_1: \mu_{\text{participate}} \neq \mu_{\text{not participate}}$
- b. (1)  $df = 80$   
 (2) If  $t < -2.000$  or  $> 2.000$ , reject  $H_0$ ; otherwise, do not reject  $H_0$   
 (3)  $s_{\bar{X}_1 - \bar{X}_2} = .42; t = 4.55$   
 (4)  $t = 4.55 > 2.000 \therefore$  reject  $H_0$  ( $p < .05$ )  
 (5)  $t = 4.55 > 2.660 \therefore p < .01$
- c. The average level of knowledge regarding what to feed their children at home was significantly higher for the sample of 29 parents who participated in the program ( $M = 4.68$ ) than the sample of 53 parents who did not participate ( $M = 2.77$ ),  $t(80) = 4.55, p < .01$ .
- d. The results of this analysis suggest that the program may provide information for parents of children with anorexia nervosa.

**Learning Check 5**

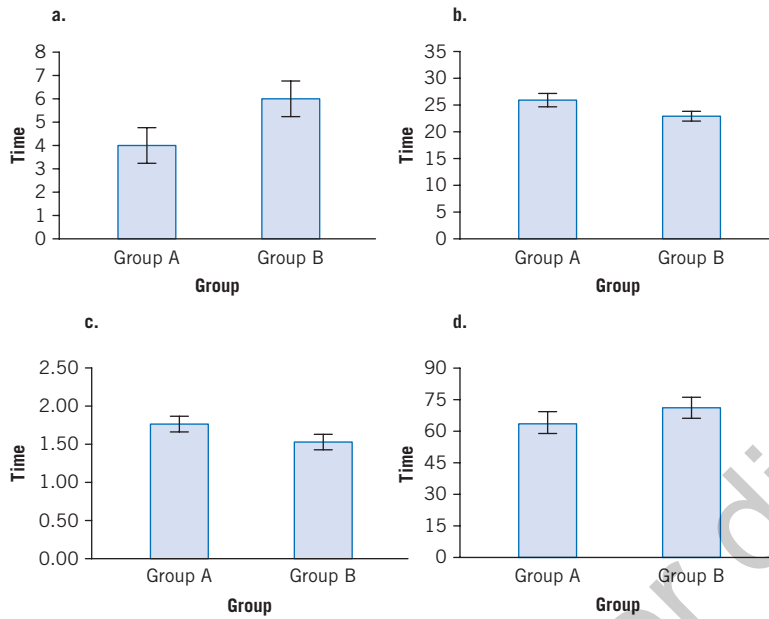
2. a.

Pair	Shy	Outgoing	Difference ( <i>D</i> )
1	8	4	4
2	4	6	-2
3	8	2	6
4	6	6	0
5	3	7	-4
6	6	3	3
7	5	4	1

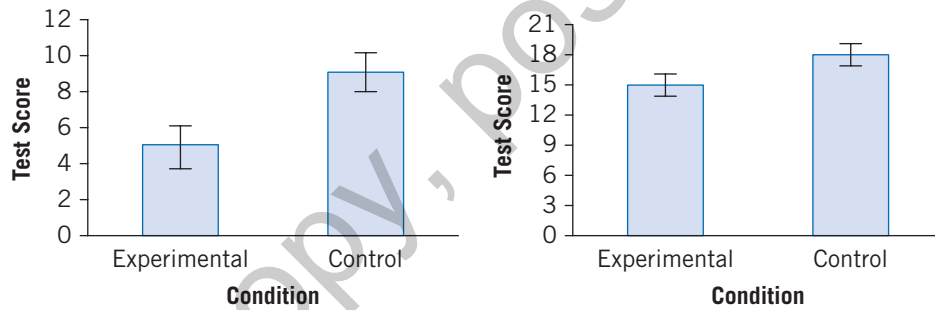
- b.  $\bar{X}_D, s_D = 1.14; s_D = 3.48; = 1.32$
3. a.  $s_D = .50; t = 6.00$   
 b.  $s_D = .38, t = 3.95$   
 c.  $s_D = 5.42, t = 2.33$
4. a.  $H_0: \mu_D = 0; H_1: \mu_D \neq 0$
- b. (1)  $df = 19$   
 (2) If  $t < -2.093$  or  $> 2.093$ , reject  $H_0$ ; otherwise, do not reject  $H_0$   
 (3)  $s_D = 1.05; t = 3.14$   
 (4)  $t = 3.14 > 2.093 \therefore$  reject  $H_0$  ( $p < .05$ )  
 (5)  $t = 3.14 > 2.861 \therefore p < .01$
- c. The average number of words per minute (WPM) typed correctly for a sample of 20 adults was significantly greater when tested while sitting ( $M = 40.20$ ) than while they were walking on a treadmill ( $M = 36.90$ ),  $t(19) = 3.14, p < .01$ .
- d. The results of this analysis suggest that typing is not a task that may be performed at a satisfactory level while walking on a treadmill rather than sitting.

## Answers to Odd-Numbered Exercises

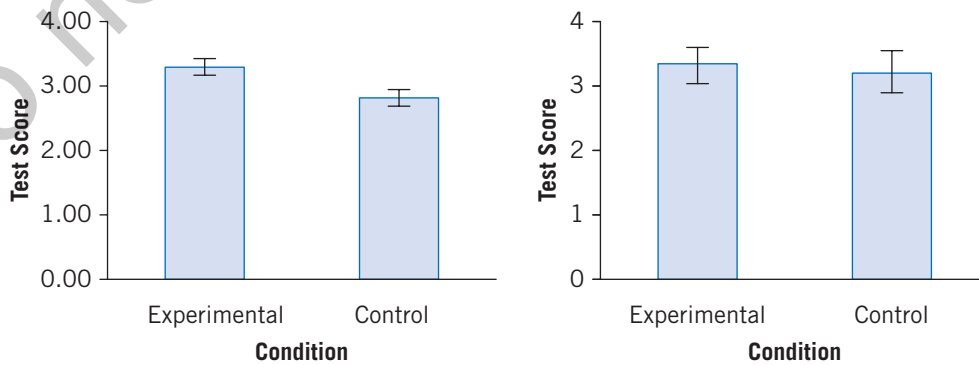
1.



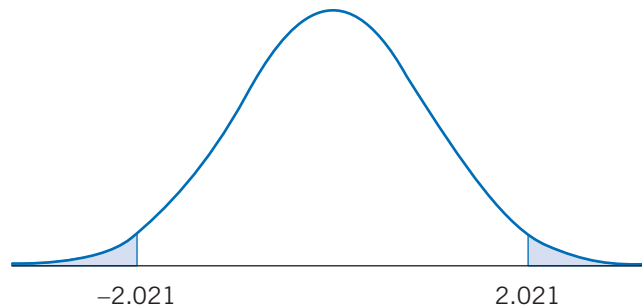
3. a. Experimental:  $\bar{X} = 5.00, s = 2.94, s_{\bar{X}} = 1.11$ ; Control:  $\bar{X} = 9.00, s = 3.06, s_{\bar{X}} = 1.16$   
 b. Experimental:  $\bar{X} = 15.00, s = 3.02, s_{\bar{X}} = 1.07$ ; Control:  $\bar{X} = 18.00, s = 3.21, s_{\bar{X}} = 1.13$



- c. Experimental:  $\bar{X} = 3.27, s = .46, s_{\bar{X}} = .15$ ; Control:  $\bar{X} = 2.80, s = .41, s_{\bar{X}} = .14$   
 d. Experimental:  $\bar{X} = 3.38, s = 1.39, s_{\bar{X}} = .39$ ; Control:  $\bar{X} = 3.23, s = 1.64, s_{\bar{X}} = .46$



5. a.  $df = 40$ ; critical value =  $\pm 2.021$   
 b.  $df = 26$ ; critical value =  $\pm 2.056$   
 c.  $df = 6$ ; critical value = 1.943  
 d.  $df = 62$ ; critical value =  $-1.671$
7. a.  $s_{\bar{X}_1 - \bar{X}_2} = .60$   
 b.  $s_{\bar{X}_1 - \bar{X}_2} = 2.39$   
 c.  $s_{\bar{X}_1 - \bar{X}_2} = 2.30$   
 d.  $s_{\bar{X}_1 - \bar{X}_2} = .47$
9. a.  $t = 1.13$   
 b.  $t = -2.09$   
 c.  $t = -1.67$   
 d.  $t = 3.20$
11. a.  $s_{\bar{X}_1 - \bar{X}_2} = 1.31$ ;  $t = -.38$   
 b.  $s_{\bar{X}_1 - \bar{X}_2} = 2.03$ ;  $t = 3.27$   
 c.  $s_{\bar{X}_1 - \bar{X}_2} = .37$ ;  $t = 1.70$   
 d.  $s_{\bar{X}_1 - \bar{X}_2} = .95$ ;  $t = -2.26$
13. a.  $H_0: \mu_{\text{Attraction}} = \mu_{\text{No attraction}}$ ;  $H_1: \mu_{\text{Attraction}} \neq \mu_{\text{No attraction}}$   
 b. (1)  $df = 52$   
 (2) If  $t < -2.009$  or  $> 2.009$ , reject  $H_0$ ; otherwise, do not reject  $H_0$   
 (3)  $s_{\bar{X}_1 - \bar{X}_2} = 1.33$ ;  $t = 2.14$   
 (4)  $t = 2.14 > 2.009 \therefore$  reject  $H_0$  ( $p < .05$ )  
 (5)  $t = 2.14 < 2.678 \therefore p < .05$  (but not  $< .01$ )  
 c. The average number of women's flirtatious behaviors was significantly higher in the sample of 13 interviewers in the attraction group ( $M = 37.44$ ) than the 13 interviewers in the no attraction group ( $M = 34.59$ ),  $t(24) = 2.14$ ,  $p < .05$ .  
 d. The results of this analysis support the research hypothesis that interviewers who believe that the women they interview are attracted to them elicit more flirtatious behavior from these women than interviewers who do not hold this belief.
15. a.  $H_0: \mu_{20-34} = \mu_{40-54}$ ;  $H_1: \mu_{20-34} \neq \mu_{40-54}$   
 b. (1)  $df = 48$   
 (2) If  $t < -2.021$  or  $> 2.021$ , reject  $H_0$ ; otherwise, do not reject  $H_0$



- (3)  $s_{\bar{X}_1 - \bar{X}_2} = .94$ ;  $t = -4.22$   
 (4)  $t = -4.19 < -2.021 \therefore$  reject  $H_0$  ( $p < .05$ )  
 (5)  $t = -4.19 < -2.704 \therefore p < .01$

- c. The average difference in age between themselves and their preferred female partners was significantly higher for the sample of 25 men aged 40 to 54 years ( $M = 4.98$ ) than the sample of 25 men aged 20 to 34 years ( $M = 1.04$ ),  $t(48) = -4.19$ ,  $p < .01$ .
- d. The results of this analysis support the research hypothesis that the preferred difference in age between themselves and their preferred partners is greater for older men than for younger men.

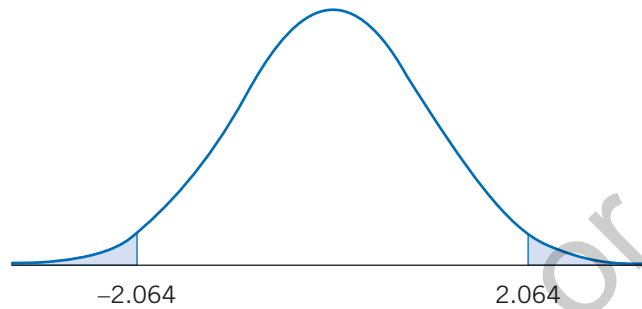
17. a. Women:  $N = 13$ ,  $\bar{X} = 18.62$ ,  $s = 7.81$

Men:  $N = 13$ ,  $\bar{X} = 14.85$ ,  $s = 8.55$

b.  $H_0: \mu_{\text{Women}} = \mu_{\text{Men}}$ ;  $H_1: \mu_{\text{Women}} \neq \mu_{\text{Men}}$

c. (1)  $df = 24$

(2) If  $t < -2.064$  or  $> 2.064$ , reject  $H_0$ ; otherwise, do not reject  $H_0$



(3)  $s_{\bar{X}_1 - \bar{X}_2} = 3.21$ ;  $t = 1.17$

(4)  $t = 1.17$  is not  $< -2.064$  or  $> 2.064$   $\therefore$  do not reject  $H_0$  ( $p > .05$ )

(5) Not applicable ( $H_0$  not rejected)

- d. The average estimates of the cost of a wedding (in thousands of dollars) for a sample of 13 women ( $M = 18.62$ ) and a sample of 13 men ( $M = 14.85$ ) were not significantly different,  $t(24) = 1.17$ ,  $p > .05$ .
- e. The result of this analysis does not support the research hypothesis that young women will provide a higher estimate of the cost of the average wedding than will men.

19. a.  $s_{\bar{X}_1 - \bar{X}_2} = 1.65$

b.  $s_{\bar{X}_1 - \bar{X}_2} = 3.11$

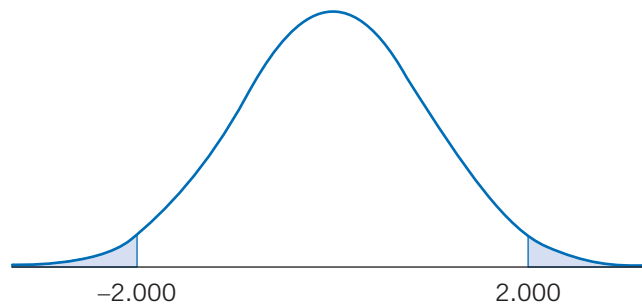
c.  $s_{\bar{X}_1 - \bar{X}_2} = 1.02$

d.  $s_{\bar{X}_1 - \bar{X}_2} = 3.08$

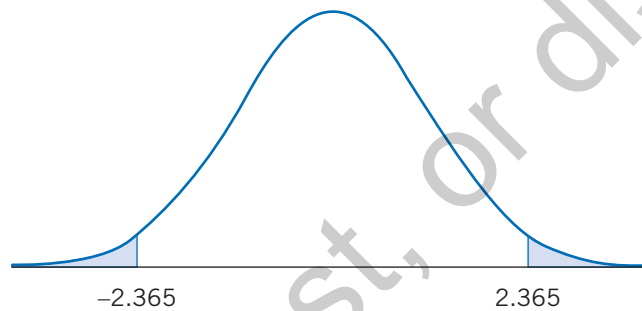
21. a.  $H_0: \mu_{\text{New user}} = \mu_{\text{Experienced user}}$ ;  $H_1: \mu_{\text{New user}} \neq \mu_{\text{Experienced user}}$

b. (1)  $df = 62$

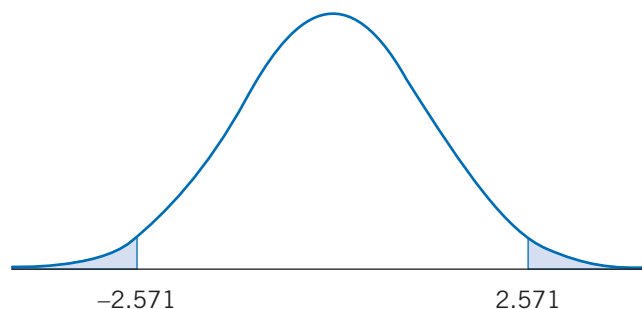
(2) If  $t < -2.000$  or  $> 2.000$ , reject  $H_0$ ; otherwise, do not reject  $H_0$



- (3)  $s_{\bar{X}_1 - \bar{X}_2} = 1.12; t = -1.01$   
 (4)  $t = -1.01$  is not  $< -2.000$  or  $> 2.000 \therefore$  do not reject  $H_0$  ( $p > .05$ )  
 (5) Not applicable ( $H_0$  not rejected)
- c. The average level of satisfaction with their hearing aids of the sample of 30 new users ( $M = 26.90$ ) and a sample of 34 experienced users ( $M = 28.03$ ) was not significantly different,  $t(62) = -1.01, p > .05$ .
- d. The result of this analysis indicates that these patients' satisfaction with their hearing aids is not related to how long they've used their hearing aids.
23. a.  $s_{\bar{D}} = 1.60; t = 6.25$   
 b.  $s_{\bar{D}} = 2.11; t = 3.22$   
 c.  $s_{\bar{D}} = .54; t = 3.93$   
 d.  $s_{\bar{D}} = .69; t = 1.16$
25. a.  $H_0: \mu_D = 0; H_1: \mu_D \neq 0$   
 b. (1)  $df = 7$



- (2) If  $t < -2.365$  or  $> 2.365$ , reject  $H_0$ ; otherwise, do not reject  $H_0$   
 (3)  $s_{\bar{D}} = 3.48; t = -2.39$   
 (4)  $t = -2.39 < -2.365 \therefore$  reject  $H_0$  ( $p < .05$ )  
 (5)  $t = -2.39 > -3.499 \therefore p < .05$  (but not  $< .01$ )
- e. The average number of overhead lifts for a sample of eight older adults in the weight and strength training program significantly increased over the 8-week period from the pretest ( $M = 45.60$ ) to the posttest ( $M = 53.90$ ),  $t(7) = -2.39, p < .05$ .
- f. The results of this analysis suggest that weight and strength training may increase older adults' levels of physical fitness.
27. a. Before:  $N_1 = 6, \bar{X}_1 = 14.50, s_1 = 7.56$   
 After:  $N_2 = 6, \bar{X}_2 = 6.33, s_2 = 6.71$   
 b.  $\bar{X}_D = 8.17; s_D = 6.59$   
 c.  $H_0: \mu_D = 0; H_1: \mu_D \neq 0$   
 d. (1)  $df = 5$   
 (2) If  $t < -2.571$  or  $> 2.571$ , reject  $H_0$ ; otherwise, do not reject  $H_0$



- (3)  $s_{\bar{D}} = 2.69; t = 3.04$   
(4)  $t = 3.04 > 2.571 \therefore$  reject  $H_0$  ( $p < .05$ )  
(5)  $t = 3.04 < 4.032 \therefore p < .05$  (but not  $< .01$ )
- e. The average number of cigarettes smoked per day for a sample of six smokers significantly decreased from before ( $M = 14.50$ ) to 4 weeks after ( $M = 6.33$ ) using the tools,  $t(5) < 3.04, p < .05$ .
- f. Using tools such as group support, substituting food for cigarettes, and deep breathing may be effective means of reducing cigarette smoking.



Sharpen your skills with **SAGE edge** at [edge.sagepub.com/tokunaga2e](https://edge.sagepub.com/tokunaga2e)

**SAGE edge for students** provides a personalized approach to help you accomplish your coursework goals in an easy-to-use learning environment. Log on to access:

- eFlashcards
- Web Quizzes
- SPSS Data Files
- Video and Audio Resources