Statistics for the Behavioral Sciences

9e

Frederick J Gravetter
Larry B. Wallnau

C H A P T E R

10

The *t* Test for Two Independent Samples

Tools You Will Need

The following items are considered essential background material for this chapter. If you doubt your knowledge of any of these items, you should review the appropriate chapter or section before proceeding.

- Sample variance (Chapter 4)
- Standard error formulas (Chapter 7)
- The *t* statistic (Chapter 9)
  - Distribution of *t* values
  - *df* for the *t* statistic
  - Estimated standard error

Preview

In a classic study in the area of problem solving, Katona (1940) compared the effectiveness of two methods of instruction. One group of participants was shown the exact, step-by-step procedure for solving a problem, and then these participants were required to memorize the solution. This method was called *learning by memorization* (later called the *expository* method). Participants in a second group were encouraged to study the problem and find the solution on their own. Although these participants were given helpful hints and clues, the exact solution was never explained. This method was called *learning by understanding* (later called the *discovery* method).

Katona's experiment included the problem shown in Figure 10.1. This figure shows a pattern of five squares made of matchsticks. The problem is to change the pattern into exactly four squares by moving only three matches. (All matches must be used, none can be removed, and all the squares must be the same size.) Two groups of participants learned the solution to this problem. One group learned by understanding, and the other group learned by memorization. After 3 weeks, both groups returned to be tested again. The two groups did equally well on the matchstick problem they had learned earlier. But when they were given two new problems (similar to the matchstick problem), the *understanding* group performed much better than the *memorization* group.
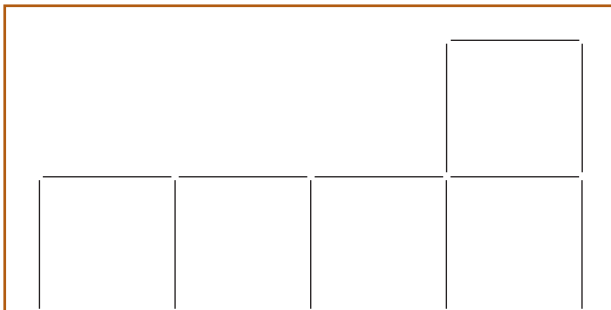
The Problem: Although the data a show a mean difference between the two groups in Katona's study, you cannot automatically conclude that the difference was caused by the method they used to solve the first problem. Specifically, the two groups consist of different people with different backgrounds, different skills, different IQs, and so on. Because the two different groups consist of different individuals, you should expect them to have different scores and different means. This issue was first presented in Chapter 1 when we introduced the concept of sampling error (see Figure 1.2 on p. 9). Thus, there are two possible explanations for the difference between the two groups.

1. It is possible that there really is a difference between the two treatment conditions so that the method of understanding produces better learning than the method of memorization.
2. It is possible that there is no difference between the two treatment conditions and the mean difference obtained in the experiment is simply the result of sampling error.

A hypothesis test is necessary to determine which of the two explanations is most plausible. However, the hypothesis tests we have examined thus far are intended to evaluate the data from only one sample. In this study there are two separate samples.

The Solution: In this chapter we introduce the *independent-measures* t *test,* which is a hypothesis test that uses two separate samples to evaluate the mean difference between two treatment conditions or between two different populations. Like the *t* test introduced in Chapter 9, the independent-measures *t* test uses the sample variance to compute an estimated standard error. This test, however, combines the variance from the two separate samples to evaluate the difference between two separate sample means.

Incidentally, if you still have not discovered the solution to the matchstick problem, keep trying. According to Katona's results, it would be a very poor teaching strategy for us to give you the answer to the matchstick problem. If you still have not discovered the solution, however, check Appendix C at the beginning of the Chapter 10 problem solutions; there we show you how it is done.

FIGURE 10.1

A pattern of five squares made of matchsticks. The problem is to change the pattern into exactly four squares by moving only three matchsticks.

10.1   INTRODUCTION TO THE INDEPENDENT-MEASURES DESIGN

Until this point, all the inferential statistics we have considered involve using one sample as the basis for drawing conclusions about one population. Although these *single-sample* techniques are used occasionally in real research, most research studies require the comparison of two (or more) sets of data. For example, a social psychologist may want to compare men and women in terms of their political attitudes, an educational psychologist may want to compare two methods for teaching mathematics, or a clinical psychologist may want to evaluate a therapy technique by comparing depression scores for patients before therapy with their scores after therapy. In each case, the research question concerns a mean difference between two sets of data.

There are two general research designs that can be used to obtain the two sets of data to be compared:

1. The two sets of data could come from two completely separate groups of participants. For example, the study could involve a sample of men compared with a sample of women. Or the study could compare grades for one group of freshmen who are given laptop computers with grades for a second group who are not given computers.

2. The two sets of data could come from the same group of participants. For example, the researcher could obtain one set of scores by measuring depression for a sample of patients before they begin therapy and then obtain a second set of data by measuring the same individuals after 6 weeks of therapy.

The first research strategy, using completely separate groups, is called an *independent-measures* research design or a *between-subjects* design. These terms emphasize the fact that the design involves separate and independent samples and makes a comparison between two groups of individuals. The structure of an independent-measures research design is shown in Figure 10.2. Notice that the research study uses two separate

FIGURE 10.2

Do the achievement scores for children taught by method A differ from the scores for children taught by method B? In statistical terms, are the two population means the same or different? Because neither of the two population means is known, it will be necessary to take two samples, one from each population. The first sample provides information about the mean for the first population, and the second sample provides information about the second population.

Population A
Taught by method A

Unknown
$\mu = ?$

Sample A

Population B
Taught by method B

Unknown
$\mu = ?$
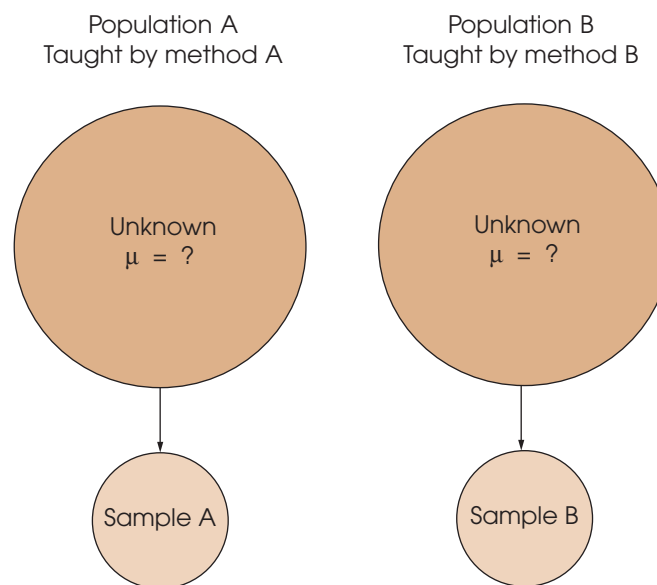
Sample B

samples to represent the two different populations (or two different treatments) being compared.

DEFINITION

> A research design that uses a separate group of participants for each treatment condition (or for each population) is called an **independent-measures research design** or a **between-subjects research design.**

In this chapter, we examine the statistical techniques used to evaluate the data from an independent-measures design. More precisely, we introduce the hypothesis test that allows researchers to use the data from two separate samples to evaluate the mean difference between two populations or between two treatment conditions.

The second research strategy, in which the two sets of data are obtained from the same group of participants, is called a *repeated-measures* research design or a *within-subjects* design. The statistics for evaluating the results from a repeated-measures design are introduced in Chapter 11. Also, at the end of Chapter 11, we discuss some of the advantages and disadvantages of independent-measures and repeated-measures designs.

10.2   THE *t* STATISTIC FOR AN INDEPENDENT-MEASURES RESEARCH DESIGN

Because an independent-measures study involves two separate samples, we need some special notation to help specify which data go with which sample. This notation involves the use of subscripts, which are small numbers written beside a sample statistic. For example, the number of scores in the first sample would be identified by n_1; for the second sample, the number of scores is n_2. The sample means would be identified by M_1 and M_2. The sums of squares would be SS_1 and SS_2.

THE HYPOTHESIS FOR AN INDEPENDENT-MEASURES TEST

The goal of an independent-measures research study is to evaluate the mean difference between two populations (or between two treatment conditions). Using subscripts to differentiate the two populations, the mean for the first population is μ_1, and the second population mean is μ_2. The difference between means is simply $\mu_1 - \mu_2$. As always, the null hypothesis states that there is no change, no effect, or, in this case, no difference. Thus, in symbols, the null hypothesis for the independent-measures test is

H_0: $\mu_1 - \mu_2 = 0$ (No difference between the population means)

You should notice that the null hypothesis could also be stated as $\mu_1 = \mu_2$. However, the first version of H_0 produces a specific numerical value (zero) that is used in the calculation of the *t* statistic. Therefore, we prefer to phrase the null hypothesis in terms of the difference between the two population means.

The alternative hypothesis states that there is a mean difference between the two populations,

H_1: $\mu_1 - \mu_2 \neq 0$ (There is a mean difference.)

Equivalently, the alternative hypothesis can simply state that the two population means are not equal: $\mu_1 \neq \mu_2$.

THE FORMULAS FOR AN INDEPENDENT-MEASURES HYPOTHESIS TEST

The independent-measures hypothesis test uses another version of the *t* statistic. The formula for this new *t* statistic has the same general structure as the *t* statistic formula that was introduced in Chapter 9. To help distinguish between the two *t* formulas, we refer to the original formula (Chapter 9) as the *single-sample* t *statistic* and we refer to the new formula as the *independent-measures* t *statistic*. Because the new independent-measures *t* includes data from two separate samples and hypotheses about two populations, the formulas may appear to be a bit overpowering. However, the new formulas are easier to understand if you view them in relation to the single-sample *t* formulas from Chapter 9. In particular, there are two points to remember:

1. The basic structure of the *t* statistic is the same for both the independent-measures and the single-sample hypothesis tests. In both cases,

$$t = \frac{\text{sample statistic} - \text{hypothesized population parameter}}{\text{estimated standard error}}$$

2. The independent-measures *t* is basically a *two-sample* t *that doubles all the elements of the single-sample* t *formulas*.

To demonstrate the second point, we examine the two *t* formulas piece by piece.

The overall *t* formula   The single-sample *t* uses one sample mean to test a hypothesis about one population mean. The sample mean and the population mean appear in the numerator of the *t* formula, which measures how much difference there is between the sample data and the population hypothesis.

$$t = \frac{\text{sample mean} - \text{population mean}}{\text{estimated standard error}} = \frac{M - \mu}{s_M}$$

The independent-measures *t* uses the difference between *two* sample means to evaluate a hypothesis about the difference between *two* population means. Thus, the independent-measures *t* formula is

$$t = \frac{\text{sample mean difference} - \text{population mean difference}}{\text{estimated standard error}} = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{(M_1 - M_2)}}$$

In this formula, the value of $M_1 - M_2$ is obtained from the sample data and the value for $\mu_1 - \mu_2$ comes from the null hypothesis.

The estimated standard error   In each of the *t*-score formulas, the standard error in the denominator measures how accurately the sample statistic represents the population parameter. In the single-sample *t* formula, the standard error measures the amount of error expected for a sample mean and is represented by the symbol s_M. For the independent-measures *t* formula, the standard error measures the amount of error that is expected when you use a sample mean difference $(M_1 - M_2)$ to represent a population mean difference $(\mu_1 - \mu_2)$. The standard error for the sample mean difference is represented by the symbol $s_{(M_1 - M_2)}$.

Caution: Do not let the notation for standard error confuse you. In general, standard error measures how accurately a statistic represents a parameter. The symbol for standard error takes the form $s_{\text{statistic}}$. When the statistic is a sample mean, M, the symbol for standard error is s_M. For the independent-measures test, the statistic is a sample

mean difference ($M_1 - M_2$), and the symbol for standard error is $s_{(M_1 - M_2)}$. In each case, the standard error tells how much discrepancy is reasonable to expect between the sample statistic and the corresponding population parameter.

Interpreting the estimated standard error   The *estimated standard error of* $M_1 - M_2$ that appears in the bottom of the independent-measures *t* statistic can be interpreted in two ways. First, the standard error is defined as a measure of the standard, or average, distance between a sample statistic ($M_1 - M_2$) and the corresponding population parameter ($\mu_1 - \mu_2$). As always, samples are not expected to be perfectly accurate and the standard error measures how much difference is reasonable to expect between a sample statistic and the population parameter.

$$
\begin{array}{ccc}
\text{Sample mean} & & \text{Population mean} \\
\text{difference} & \xleftarrow[\text{(average distance)}]{\text{estimated standard error}} & \text{difference} \\
(M_1 - M_2) & & (\mu_1 - \mu_2)
\end{array}
$$

When the null hypothesis is true, however, the population mean difference is zero.

$$
\begin{array}{cc}
\text{Sample mean} & \\
\text{difference} & \xleftarrow[\text{(average distance)}]{\text{estimated standard error}} \; 0\left(\text{If } H_0 \text{ is true}\right) \\
(M_1 - M_2) &
\end{array}
$$

The standard error is measuring how close the sample mean difference is to zero, which is equivalent to measuring how much difference there is between the two sample means.

$$
M_1 \xleftarrow[\text{(average distance)}]{\text{estimated standard error}} M_2
$$

This produces a second interpretation for the estimated standard error. Specifically, the standard error can be viewed as a measure of how much difference is reasonable to expect between two sample means if the null hypothesis is true.

The second interpretation of the estimated standard error produces a simplified version of the independent-measures *t* statistic.

$$
t = \frac{\text{sample mean difference}}{\text{estimated standard error}}
$$

$$
= \frac{\text{actual difference between } M_1 \text{ and } M_2}{\text{standard difference (If } H_0 \text{ is true) between } M_1 \text{ and } M_2}
$$

In this version, the numerator of the *t* statistic measures how much difference *actually exists* between the two sample means, including any difference that is caused by the different treatments. The denominator measures how much difference *should exist* between the two sample means if there is no treatment effect that causes them to be different. A large value for the *t* statistic is evidence for the existence of a treatment effect.

CALCULATING THE ESTIMATED STANDARD ERROR

To develop the formula for $s_{(M_1 - M_2)}$, we consider the following three points:

1. Each of the two sample means represents it own population mean, but in each case there is some error.

M_1 approximates μ_1 with some error.

M_2 approximates μ_2 with some error.

Thus, there are two sources of error.

2. The amount of error associated with each sample mean is measured by the estimated standard error of *M*. Using Equation 9.1 (p. 285), the estimated standard error for each sample mean is computed as follows:

$$\text{For } M_1 \ s_M = \sqrt{\frac{s_1^2}{n_1}} \qquad \text{For } M_2 \ s_M = \sqrt{\frac{s_2^2}{n_2}}$$

3. For the independent-measures *t* statistic, we want to know the total amount of error involved in using *two* sample means to approximate *two* population means. To do this, we find the error from each sample separately and then add the two errors together. The resulting formula for standard error is

$$s_{(M_1 - M_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{10.1}$$

Because the independent-measures *t* statistic uses two sample means, the formula for the estimated standard error simply combines the error for the first sample mean and the error for the second sample mean (Box 10.1).

POOLED VARIANCE   Although Equation 10.1 accurately presents the concept of standard error for the independent-measures *t* statistic, this formula is limited to situations in which the

---

BOX 10.1	**THE VARIABILITY OF DIFFERENCE SCORES**

It may seem odd that the independent-measures *t* statistic *adds* together the two sample errors when it *subtracts* to find the difference between the two sample means. The logic behind this apparently unusual procedure is demonstrated here.

We begin with two populations, I and II (Figure 10.3). The scores in population I range from a high of 70 to a low of 50. The scores in population II range from 30 to 20. We use the range as a measure of how spread out (variable) each population is:

For population I, the scores cover a range of 20 points.
For population II, the scores cover a range of 10 points.

If we randomly select one score from population I and one score from population II and compute the difference between these two scores ($X_1 - X_2$), what range of values is possible for these differences? To answer this question, we need to find the biggest

possible difference and the smallest possible difference. Look at Figure 10.3; the biggest difference occurs when $X_1 = 70$ and $X_2 = 20$. This is a difference of $X_1 - X_2 = 50$ points. The smallest difference occurs when $X_1 = 50$ and $X_2 = 30$. This is a difference of $X_1 - X_2 = 20$ points. Notice that the differences go from a high of 50 to a low of 20. This is a range of 30 points:
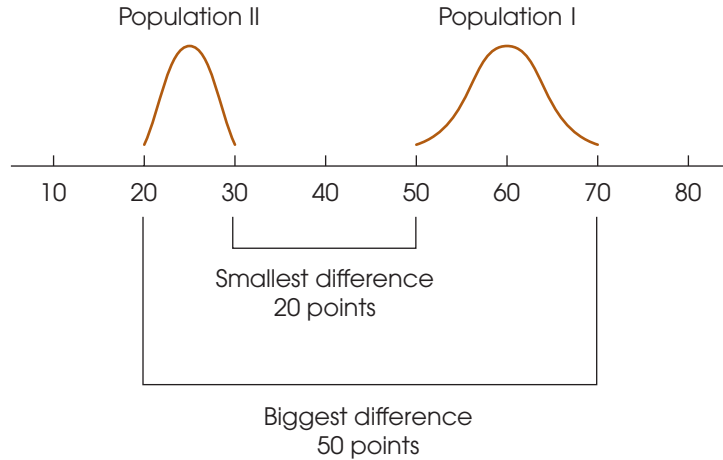
range for population I (X_1 scores) = 20 points
range for population II (X_2 scores) = 10 points
range for the differences ($X_1 - X_2$) = 30 points

The variability for the difference scores is found by *adding* together the variability for each of the two populations.

In the independent-measures *t* statistics, we compute the variability (standard error) for a sample mean difference. To compute this value, we add together the variability for each of the two sample means.

FIGURE 10.3

Two population distributions. The scores in population I vary from 50 to 70 (a 20-point spread), and the scores in population II range from 20 to 30 (a 10-point spread). If you select one score from each of these two populations, the closest two values are $X_1 = 50$ and $X_2 = 30$. The two values that are farthest apart are $X_1 = 70$ and $X_2 = 20$.

Population II   Population I

10   20   30   40   50   60   70   80

Smallest difference
20 points

Biggest difference
50 points

An alternative to computing pooled variance is presented in Box 10.2, p. 339.

two samples are exactly the same size (that is, $n_1 = n_2$). For situations in which the two sample sizes are different, the formula is *biased* and, therefore, inappropriate. The bias comes from the fact that Equation 10.1 treats the two sample variances equally. However, when the sample sizes are different, the two sample variances are not equally good and should not be treated equally. In Chapter 7, we introduced the law of large numbers, which states that statistics obtained from large samples tend to be better (more accurate) estimates of population parameters than statistics obtained from small samples. This same fact holds for sample variances: The variance obtained from a large sample is a more accurate estimate of σ^2 than the variance obtained from a small sample.

One method for correcting the bias in the standard error is to combine the two sample variances into a single value called the *pooled variance*. The pooled variance is obtained by averaging or "pooling" the two sample variances using a procedure that allows the bigger sample to carry more weight in determining the final value.

You should recall that when there is only one sample, the sample variance is computed as

$$s^2 = \frac{SS}{df}$$

For the independent-measures *t* statistic, there are two *SS* values and two *df* values (one from each sample). The values from the two samples are combined to compute what is called the *pooled variance*. The pooled variance is identified by the symbol s_p^2 and is computed as

$$\text{pooled variance} = s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} \tag{10.2}$$

With one sample, the variance is computed as *SS* divided by *df*. With two samples, the pooled variance is computed by combining the two *SS* values and then dividing by the combination of the two *df* values.

As we mentioned earlier, the pooled variance is actually an average of the two sample variances, but the average is computed so that the larger sample carries more weight in determining the final value. The following examples demonstrate this point.

Equal samples sizes   We begin with two samples that are exactly the same size. The first sample has $n = 6$ scores with $SS = 50$, and the second sample has $n = 6$ scores with $SS = 30$. Individually, the two sample variances are

$$\text{Variance for sample } 1 : s^2 = \frac{SS}{df} = \frac{50}{5} = 10$$

$$\text{Variance for sample } 2 : s^2 = \frac{SS}{df} = \frac{30}{5} = 6$$

The pooled variance for these two samples is

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{50 + 30}{5 + 5} = \frac{80}{10} = 8.00$$

Note that the pooled variance is exactly halfway between the two sample variances. Because the two samples are exactly the same size, the pooled variance is simply the average of the two sample variances.

Unequal samples sizes   Now consider what happens when the samples are not the same size. This time the first sample has $n = 3$ scores with $SS = 20$, and the second sample has n = 9 scores with SS = 48. Individually, the two sample variances are

$$\text{Variance for sample } 1 : s^2 = \frac{SS}{df} = \frac{20}{2} = 10$$

$$\text{Variance for sample } 2 : s^2 = \frac{SS}{df} = \frac{48}{8} = 6$$

The pooled variance for these two samples is

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{20 + 48}{2 + 8} = \frac{68}{10} = 6.80$$

This time the pooled variance is not located halfway between the two sample variances. Instead, the pooled value is closer to the variance for the larger sample ($n = 9$ and $s^2 = 6$) than to the variance for the smaller sample ($n = 3$ and $s^2 = 10$). The larger sample carries more weight when the pooled variance is computed.

When computing the pooled variance, the weight for each of the individual sample variances is determined by its degrees of freedom. Because the larger sample has a larger *df* value, it carries more weight when averaging the two variances. This produces an alternative formula for computing pooled variance.

$$\text{pooled variance } = s_p^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2} \tag{10.3}$$

For example, if the first sample has $df_1 = 3$ and the second sample has $df_2 = 7$, then the formula instructs you to take 3 of the first sample variance and 7 of the second sample variance for a total of 10 variances. You then divide by 10 to obtain the average. The alternative formula is especially useful if the sample data are summarized as means and variances. Finally, you should note that because the pooled variance is an average of the two sample variances, the value obtained for the pooled variance is always located between the two sample variances.

ESTIMATED STANDARD ERROR

Using the pooled variance in place of the individual sample variances, we can now obtain an unbiased measure of the standard error for a sample mean difference. The resulting formula for the independent-measures estimated standard error is

$$\text{estimated standard error of } M_1 - M_2 = s_{(M_1 - M_2)} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \qquad (10.4)$$

Conceptually, this standard error measures how accurately the difference between two sample means represents the difference between the two population means. In a hypothesis test, H_0 specifies that $\mu_1 - \mu_2 = 0$, and the standard error also measures how much difference is expected, on average, between the two sample means. In either case, the formula combines the error for the first sample mean with the error for the second sample mean. Also note that the pooled variance from the two samples is used to compute the standard error for the sample mean difference.

THE FINAL FORMULA AND DEGREES OF FREEDOM

The complete formula for the independent-measures *t* statistic is as follows:

$$t = \frac{\left(M_1 - M_2\right) - \left(\mu_1 - \mu_2\right)}{s_{(M_1 - M_2)}}$$

$$= \frac{\text{sample mean difference} - \text{population mean difference}}{\text{estimated standard error}} \qquad (10.5)$$

In the formula, the estimated standard error in the denominator is calculated using Equation 10.4, and requires calculation of the pooled variance using either Equation 10.2 or 10.3.

The degrees of freedom for the independent-measures *t* statistic are determined by the *df* values for the two separate samples:

$$df \text{ for the } t \text{ statistic} = df \text{ for the first sample} + df \text{ for the second sample}$$
$$= df_1 + df_2$$
$$= (n_1 - 1) + (n_2 - 1) \qquad (10.6)$$

Equivalently, the *df* value for the independent-measures *t* statistic can be expressed as

$$df = n_1 + n_2 - 2 \qquad (10.7)$$

Note that the *df* formula subtracts 2 points from the total number of scores; 1 point for the first sample and 1 for the second.

The independent-measures *t* statistic is used for hypothesis testing. Specifically, we use the difference between two sample means $(M_1 - M_2)$ as the basis for testing hypotheses about the difference between two population means $(\mu_1 - \mu_2)$. In this context, the overall structure of the *t* statistic can be reduced to the following:

$$t = \frac{\text{data} - \text{hypothesis}}{\text{error}}$$

This same structure is used for both the single-sample *t* from Chapter 9 and the new independent-measures *t* that was introduced in the preceding pages. Table 10.1 identifies each component of these two *t* statistics and should help reinforce the point that we made earlier in the chapter; that is, the independent-measures *t* statistic simply doubles each aspect of the single-sample *t* statistic.

TABLE 10.1

The basic elements of a *t* statistic for the single-sample *t* and the independent-measures *t*.

	Sample Data	Hypothesized Population Parameter	Estimated Standard Error	Sample Variance
Single-sample *t* statistic	M	μ	$\sqrt{\dfrac{s^2}{n}}$	$s^2 = \dfrac{SS}{df}$
Independent-measures *t* statistic	$(M_1 - M_2)$	$(\mu_1 - \mu_2)$	$\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}$	$s_p^2 = \dfrac{SS_1 + SS_2}{df_1 + df_2}$

LEARNING CHECK

1. What is the defining characteristic of an independent-measures research study?

2. Explain what is measured by the estimated standard error in the denominator of the independent-measures *t* statistic.

3. One sample from an independent-measures study has $n = 4$ with $SS = 100$. The other sample has $n = 8$ and $SS = 140$.
   a. Compute the pooled variance. (*Note:* Equation 10.2 works well with these data.)
   b. Compute the estimated standard error for the mean difference.

4. One sample from an independent-measures study has $n = 9$ with a variance of $s^2 = 35$. The other sample has $n = 3$ and $s^2 = 40$.
   a. Compute the pooled variance. (*Note:* Equation 10.3 works well with these data.)
   b. Compute the estimated standard error for the mean difference.

5. An independent-measures *t* statistic is used to evaluate the mean difference between two treatments with $n = 8$ in one treatment and $n = 12$ in the other. What is the *df* value for the *t* statistic?

ANSWERS

1. An independent-measures study uses a separate group of participants to represent each of the populations or treatment conditions being compared.

2. The estimated standard error measures how much difference is expected, on average, between a sample mean difference and the population mean difference. In a hypothesis test, $\mu_1 - \mu_2$ is set to zero and the standard error measures how much difference is expected between the two sample means.

3. a. The pooled variance is 240/10 = 24.
   b. The estimated standard error is 3.

4. a. The pooled variance is 36.
   b. The estimated standard error is 4.

5. $df = df_1 + df_2 = 7 + 11 = 18$.

10.3 HYPOTHESIS TESTS AND EFFECT SIZE WITH THE INDEPENDENT-MEASURES *t* STATISTIC

The independent-measures *t* statistic uses the data from two separate samples to help decide whether there is a significant mean difference between two populations or between two treatment conditions. A complete example of a hypothesis test with two independent samples follows.

EXAMPLE 10.1   Research results suggest a relationship between the TV viewing habits of 5-year-old children and their future performance in high school. For example, Anderson, Huston, Wright, and Collins (1998) report that high school students who had regularly watched Sesame Street as children had better grades in high school than their peers who had not watched Sesame Street. Suppose that a researcher intends to examine this phenomenon using a sample of 20 high school students.

The researcher first surveys the students' parents to obtain information on the family's TV-viewing habits during the time that the students were 5 years old. Based on the survey results, the researcher selects a sample of $n = 10$ students with a history of watching Sesame Street and a sample of $n = 10$ students who did not watch the program. The average high school grade is recorded for each student and the data are as follows:

Average High School Grade			
Watched Sesame Street		Did Not Watch Sesame Street	
86	99	90	79
87	97	89	83
91	94	82	86
97	89	83	81
98	92	85	92
$n = 10$		$n = 10$	
$M = 93$		$M = 85$	
$SS = 200$		$SS = 160$	

Note that this is an independent-measures study using two separate samples representing two distinct populations of high school students. The researcher would like to know whether there is a significant difference between the two types of high school student.

STEP 1   State the hypotheses and select the alpha level.

$H_0: \mu_1 - \mu_2 = 0$ (No difference.)

$H_1: \mu_1 - \mu_2 \neq 0$ (There is a difference.)

We set $\alpha = .01$.

Directional hypotheses could be used and would specify whether the students who watched Sesame Street should have higher or lower grades.

STEP 2   This is an independent-measures design. The *t* statistic for these data has degrees of freedom determined by

$$df = df_1 + df_2$$

$$= (n_1 - 1) + (n_2 - 1)$$

$$= 9 + 9$$

$$= 18$$

The *t* distribution for $df = 18$ is presented in Figure 10.4. For $\alpha = .01$, the critical region consists of the extreme 1% of the distribution and has boundaries of $t = +2.878$ and $t = -2.878$.

FIGURE 10.4

The critical region for the independent-measures hypothesis test in Example 10.1 with $df = 18$ and $\alpha = .01$.

t distribution
$df = 18$

Reject H_0 ←

→ Reject H_0

$t = -2.878$

$t = 0$

$t = +2.878$
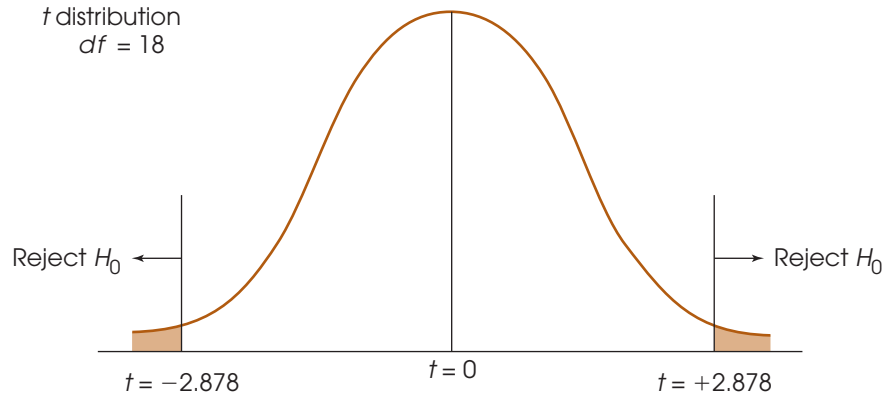
S T E P 3 Obtain the data and compute the test statistic. The data are given, so all that remains is to compute the *t* statistic. As with the single-sample *t* test in Chapter 9, we recommend that the calculations be divided into three parts.

First, find the pooled variance for the two samples:

Caution: The pooled variance combines the two samples to obtain a single estimate of variance. In the formula, the two samples are combined in a single fraction.

$$s^2_p = \frac{SS_1 + SS_2}{df_1 + df_2}$$
$$= \frac{200 + 160}{9 + 9}$$
$$= \frac{360}{18}$$
$$= 20$$

Second, use the pooled variance to compute the estimated standard error:

Caution: The standard error adds the errors from two separate samples. In the formula, these two errors are added as two separate fractions. In this case, the two errors are equal because the sample sizes are the same.

$$s_{(M_1 - M_2)} = \sqrt{\frac{s^2_p}{n_1} + \frac{s^2_p}{n_2}} = \sqrt{\frac{20}{10} + \frac{20}{10}}$$
$$= \sqrt{2 + 2}$$
$$= \sqrt{4}$$
$$= 2$$

Third, compute the *t* statistic:

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{(M_1 - M_2)}} = \frac{(93 - 85) - 0}{2}$$
$$= \frac{8}{2}$$
$$= 4$$

S T E P 4 Make a decision. The obtained value ($t = 4.00$) is in the critical region. In this example, the obtained sample mean difference is four times greater than would be expected if there were no difference between the two populations. In other words, this result is very unlikely if H_0 is true. Therefore, we reject H_0 and conclude that there is a significant difference between the high school grades for students who watched Sesame Street and those who did not. Specifically, the students who watched Sesame Street had significantly higher grades than those who did not watch the program.

Note that the Sesame Street study in Example 10.1 is an example of nonexperimental research (see Chapter 1, p. 17). Specifically, the researcher did not manipulate the TV programs watched by the children and did not control a variety of variables that could influence high school grades. As a result, we cannot conclude that watching Sesame Street *causes* higher high school grades. In particular, many other, uncontrolled factors, such as the parents' level of education or family economic status, might explain the difference between the two groups. Thus, we do not know exactly why there is a relationship between watching Sesame Street and high school grades, but we do know that a relationship exists.

MEASURING EFFECT SIZE FOR THE INDEPENDENT-MEASURES *t* TEST

As noted in Chapters 8 and 9, a hypothesis test is usually accompanied by a report of effect size to provide an indication of the absolute magnitude of the treatment effect. One technique for measuring effect size is Cohen's *d*, which produces a standardized measure of mean difference. In its general form, Cohen's *d* is defined as

$$d = \frac{\text{mean difference}}{\text{standard deviation}} = \frac{\mu_1 - \mu_2}{\sigma}$$

In the context of an independent-measures research study, the difference between the two sample means ($M_1 - M_2$) is used as the best estimate of the mean difference between the two populations, and the pooled standard deviation (the square root of the pooled variance) is used to estimate the population standard deviation. Thus, the formula for estimating Cohen's *d* becomes

$$\text{estimated } d = \frac{\text{estimated mean difference}}{\text{estimated standard deviation}} = \frac{M_1 - M_2}{\sqrt{s_p^2}} \tag{10.8}$$

For the data from Example 10.1, the two sample means are 93 and 85, and the pooled variance is 20. The estimated *d* for these data is

$$d = \frac{M_1 - M_2}{\sqrt{s_p^2}} = \frac{93 - 85}{\sqrt{20}} = \frac{8}{4.7} = 1.79$$

Using the criteria established to evaluate Cohen's *d* (see Table 8.2 on p. 264), this value indicates a very large treatment effect.

The independent-measures *t* test also allows for measuring effect size by computing the percentage of variance accounted for, r^2. As we saw in Chapter 9, r^2 measures how much of the variability in the scores can be explained by the treatment effects. For example, some of the variability in the high school grades from the Sesame Street study can be explained by knowing whether a particular student watched the program; students who watched Sesame Street tend to have higher grades and students who did not watch the show tend to have lower grades. By measuring exactly how much of the variability can be explained, we can obtain a measure of how big the treatment effect

actually is. The calculation of r^2 for the independent-measures *t* test is exactly the same as it was for the single-sample *t* test in Chapter 9.

$$r^2 = \frac{t^2}{t^2 + df} \tag{10.9}$$

For the data in Example 10.1, we obtained $t = 4.00$ with $df = 18$. These values produce an r^2 of

$$r^2 = \frac{4^2}{4^2 + 18} = \frac{16}{16 + 18} = \frac{16}{34} = 0.47$$

According to the standards used to evaluate r^2 (see Table 9.3 on p. 299), this value also indicates a very large treatment effect.

Although the value of r^2 is usually obtained by using Equation 10.9, it is possible to determine the percentage of variability directly by computing *SS* values for the set of scores. The following example demonstrates this process using the data from the Sesame Street study in Example 10.1.

E X A M P L E 1 0 . 2   The Sesame Street study described in Example 10.1 compared high school grades for two groups of students; one group who had watched Sesame Street when they were children and one group who had not watched the program. If we assume that the null hypothesis is true and that there is no difference between the two populations of students, then there should be no systematic difference between the two samples. In this case, the two samples can be combined to form a single set of $n = 20$ scores with an overall mean of $M = 89$. The two samples are shown as a single distribution in Figure 10.5(a).
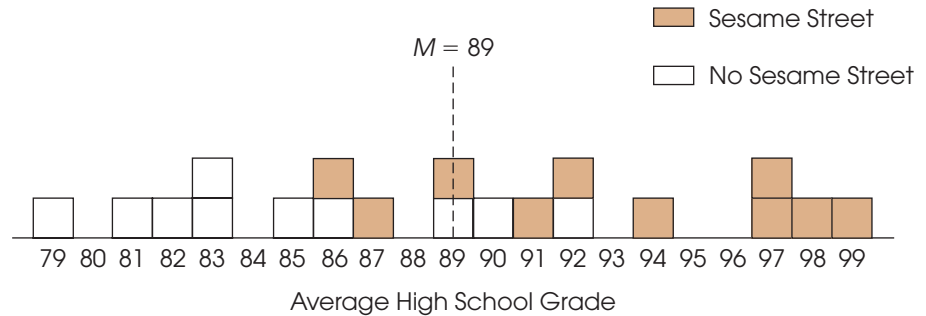
For this example, however, the conclusion from the hypothesis test is that there is a real difference between the two groups. The students who watched Sesame Street have a mean score of $M = 93$, which is 4 points above the overall average. Similarly, the students who did not watch the program had a mean score of $M = 85$, 4 points below the overall average. Thus, the Sesame Street effect causes one group of scores to move toward the right of the distribution, away from the middle, and causes the other group to move toward the left, away from the middle. The result is that the Sesame Street effect causes the scores to spread out and increases the variability.

To determine how much the treatment effect has increased the variability, we remove the treatment effect and examine the resulting scores. To remove the effect, we add 4 points to the score for each student who did not watch Sesame Street and we subtract 4 points from the score for each student who did watch. This adjustment causes both groups to have a mean of $M = 89$, so there is no longer any mean difference between the two groups. The adjusted scores are shown in Figure 10.5(b).

It should be clear that the adjusted scores in Figure 10.5(b) are less variable (more closely clustered) than the original scores in Figure 10.5(a). That is, removing the treatment effect has reduced the variability. To determine exactly how much the treatment influences variability, we have computed *SS*, the sum of squared deviations, for each set of scores. For the scores in Figure 10.5(a), including the treatment effect, we obtain $SS = 680$. When the treatment effect is removed, in Figure 10.5(b), the variability is reduced to $SS = 360$. The difference between these two values is 320 points. Thus, the treatment effect accounts for 320 points of the total variability in the original scores. When expressed as a proportion of the total variability, we obtain

$$\frac{\text{variability explained by the treatment}}{\text{total variability}} = \frac{320}{680} = 0.47 = 47\%$$
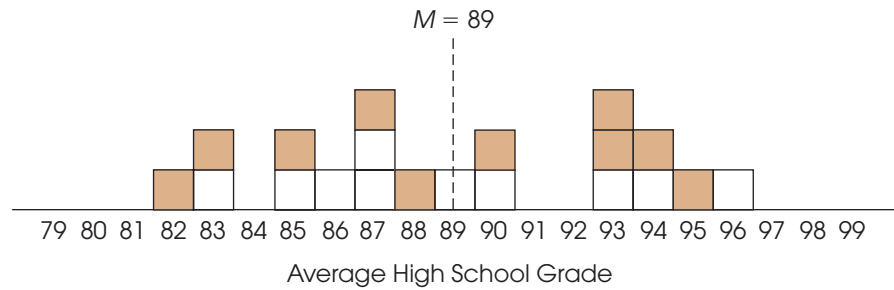
FIGURE 10.5

The two groups of scores from Example 10.1 combined into a single distribution. The original scores, including the treatment effect, are shown in part (a). Part (b) shows the adjusted scores, after the treatment effect has been removed.

You should recognize that this is exactly the same value we obtained for r^2 using Equation 10.9.

CONFIDENCE INTERVALS FOR ESTIMATING $\mu_1 - \mu_2$

As noted in Chapter 9, it is possible to compute a confidence interval as an alternative method for measuring and describing the size of the treatment effect. For the single-sample *t*, we used a single sample mean, *M*, to estimate a single population mean. For the independent-measures *t*, we use a sample mean difference, $M_1 - M_2$, to estimate the population mean difference, $\mu_1 - \mu_2$. In this case, the confidence interval literally estimates the size of the population mean difference between the two populations or treatment conditions.

As with the single-sample *t*, the first step is to solve the *t* equation for the unknown parameter. For the independent-measures *t* statistic, we obtain

$$\mu_1 - \mu_2 = M_1 - M_2 \pm t s_{(M_1 - M_2)} \tag{10.10}$$

In the equation, the values for $M_1 - M_2$ and for $s_{(M_1 - M_2)}$ are obtained from the sample data. Although the value for the *t* statistic is unknown, we can use the degrees of freedom for the *t* statistic and the *t* distribution table to estimate the *t* value. Using the

estimated *t* and the known values from the sample, we can then compute the value of $\mu_1 - \mu_2$. The following example demonstrates the process of constructing a confidence interval for a population mean difference.

E X A M P L E   1 0 . 3   Earlier we presented a research study comparing high school grades for students who had watched Sesame Street as children with the grades for students who had not watched the program (p. 326). The results of the hypothesis test indicated a significant mean difference between the two populations of students. Now, we construct a 95% confidence interval to estimate the size of the population mean difference.

The data from the study produced a mean grade of $M = 93$ for the Sesame Street group and a mean of $M = 85$ for the no-Sesame Street group, and the estimated standard error for the mean difference was $s_{(M_1 - M_2)} = 2$. With $n = 10$ scores in each sample, the independent-measures *t* statistic has $df = 18$. To have 95% confidence, we simply estimate that the *t* statistic for the sample mean difference is located somewhere in the middle 95% of all the possible *t* values. According to the *t* distribution table, with $df = 18$, 95% of the *t* values are located between t = +2.101 and $t = -2.101$. Using these values in the estimation equation, we obtain

$$\mu_1 - \mu_2 = M_1 - M_2 \pm ts_{(M_1 - M_2)}$$
$$= 93 - 85 \pm 2.101(2)$$
$$= 8 \pm 4.202$$

This produces an interval of values ranging from $8 - 4.202 = 3.798$ to $8 + 4.202 = 12.202$. Thus, our conclusion is that students who watched Sesame Street have higher grades that those who did not, and the mean difference between the two populations is somewhere between 3.798 points and 12.202 points. Furthermore, we are 95% confident that the true mean difference is in this interval because the only value estimated during the calculations was the *t* statistic, and we are 95% confident that the *t* value is located in the middle 95% of the distribution. Finally note that the confidence interval is constructed around the sample mean difference. As a result, the sample mean difference, $M_1 - M_2 = 93 - 83 = 8$ points, is located exactly in the center of the interval.

As with the confidence interval for the single-sample *t* (p. 302), the confidence interval for an independent-measures *t* is influenced by a variety of factors other than the actual size of the treatment effect. In particular, the width of the interval depends on the percentage of confidence used so that a larger percentage produces a wider interval. Also, the width of the interval depends on the sample size, so that a larger sample produces a narrower interval. Because the interval width is related to sample size, the confidence interval is not a pure measure of effect size like Cohen's *d* or r^2.

CONFIDENCE INTERVALS AND HYPOTHESIS TESTS   In addition to describing the size of a treatment effect, estimation can be used to get an indication of the *significance* of the effect. Example 10.3 presented an independent-measures research study examining the effect on high school grades of having watched Sesame Street as a child. Based on the results of the study, the 95% confidence interval estimated that the population mean difference for the two groups of students was between
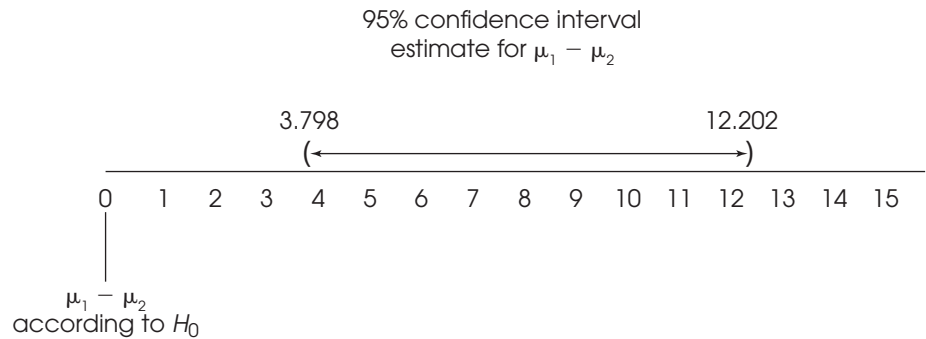
FIGURE 10.6

The 95% confidence interval for the population mean difference ($\mu_1 - \mu_2$) from Example 10.3. Note that $\mu_1 - \mu_2 = 0$ is excluded from the confidence interval, indicating that a zero difference is not an acceptable value (H_0 would be rejected in a hypothesis test with $\alpha = .05$).

3.798 and 12.202 points. The confidence interval estimate is shown in Figure 10.6. In addition to the confidence interval for $\mu_1 - \mu_2$, we have marked the spot where the mean difference is equal to zero. You should recognize that a mean difference of zero is exactly what would be predicted by the null hypothesis if we were doing a hypothesis test. You also should realize that a zero difference ($\mu_1 - \mu_2 = 0$) is *outside* of the 95% confidence interval. In other words, $\mu_1 - \mu_2 = 0$ is not an acceptable value if we want 95% confidence in our estimate. To conclude that a value of zero is *not acceptable* with 95% confidence is equivalent to concluding that a value of zero is *rejected* with 95% confidence. This conclusion is equivalent to rejecting H_0 with $\alpha = .05$. On the other hand, if a mean difference of zero were included within the 95% confidence interval, then we would have to conclude that $\mu_1 - \mu_2 = 0$ is an acceptable value, which is the same as failing to reject H_0.

The hypothesis test for these data was conducted in Example 10.1 (p. 326) and the decision was to reject H_0.

IN THE LITERATURE

REPORTING THE RESULTS OF AN INDEPENDENT-MEASURES *t* TEST

A research report typically presents the descriptive statistics followed by the results of the hypothesis test and measures of effect size (inferential statistics). In Chapter 4 (p. 123), we demonstrated how the mean and the standard deviation are reported in APA format. In Chapter 9 (p. 302), we illustrated the APA style for reporting the results of a *t* test. Now we use the APA format to report the results of Example 10.1, an independent-measures *t* test. A concise statement might read as follows:

> The students who watched Sesame Street as children had higher high school grades ($M = 93$, $SD = 4.71$) than the students who did not watch the program ($M = 85$, $SD = 4.22$). The mean difference was significant, $t(18) = 4.00$, $p < .01$, $d = 1.79$.

You should note that standard deviation is not a step in the computations for the independent-measures *t* test, yet it is useful when providing descriptive statistics for each treatment group. It is easily computed when doing the *t* test because you need *SS* and *df* for both groups to determine the pooled variance. Note that the format for reporting *t* is exactly the same as that described in Chapter 9 (p. 302) and that the measure of effect size is reported immediately after the results of the hypothesis test.

Also, as we noted in Chapter 9, if an exact probability is available from a computer analysis, it should be reported. For the data in Example 10.1, the computer analysis reports a probability value of $p = .001$ for $t = 4.00$ with $df = 18$. In the research report, this value would be included as follows:

The difference was significant, $t(18) = 4.00$, $p = .001$, $d = 1.79$.

Finally, if a confidence interval is reported to describe effect size, it appears immediately after the results from the hypothesis test. For the Sesame Street examples (Example 10.1 and Example 10.3), the report would be as follows:

The difference was significant, $t(18) = 4.00$, $p = .001$, 95% CI [3.798, 12.202].

LEARNING CHECK

1. An educational psychologist would like to determine whether access to computers has an effect on grades for high school students. One group of $n = 16$ students has home room each day in a computer classroom in which each student has a computer. A comparison group of $n = 16$ students has home room in a traditional classroom. At the end of the school year, the average grade is recorded for each student. The data are as follows:

Computer	Traditional
$M = 86$	$M = 82.5$
$SS = 1005$	$SS = 1155$

   a. Is there a significant difference between the two groups? Use a two-tailed test with $\alpha = .05$.
   b. Compute Cohen's d to measure the size of the difference.
   c. Write a sentence that demonstrates how the outcome of the hypothesis test and the measure of effect size would appear in a research report.
   d. Compute the 90% confidence interval for the population mean difference between a computer classroom and a regular classroom.

2. A researcher report states that there is a significant difference between treatments for an independent-measures design with $t(28) = 2.27$.
   a. How many individuals participated in the research study? (*Hint:* Start with the *df* value.)
   b. Should the report state that $p > .05$ or $p < .05$?

ANSWERS

1. a. The pooled variance is 72, the standard error is 3, and $t = 1.17$. With a critical value of $t = 2.042$, fail to reject the null hypothesis.
   b. Cohen's $d = 3.5/\sqrt{72} = 0.412$
   c. The results show no significant difference in grades for students with computers compared to students without computers, $t(30) = 1.17$, $p > .05$, $d = 0.412$.
   d. With $df = 30$ and 90% confidence, the t values for the confidence interval are ± 1.697. The interval is $\mu_1 - \mu_2 = 3.5 \pm 1.697(3)$. Thus, the population mean difference is estimated to be between -1.591 and 8.591. The fact that zero is an acceptable value (inside the interval) is consistent with the decision that there is no significant difference between the two population means.

2. a. The $df = 28$, so the total number of participants is 30.
   b. A significant result is indicated by $p < .05$.

DIRECTIONAL HYPOTHESES AND ONE-TAILED TESTS

When planning an independent-measures study, a researcher usually has some expectation or specific prediction for the outcome. For the Sesame Street study in Example 10.1, the researcher clearly expects the students who watched Sesame Street to have higher grades than the students who did not watch. This kind of directional prediction can be incorporated into the statement of the hypotheses, resulting in a directional, or one-tailed, test. Recall from Chapter 8 that one-tailed tests can lead to rejecting H_0 when the mean difference is relatively small compared to the magnitude required by a two-tailed test. As a result, one-tailed tests should be used when clearly justified by theory or previous findings. The following example demonstrates the procedure for stating hypotheses and locating the critical region for a one-tailed test using the independent-measures t statistic.

E X A M P L E   1 0 . 4

We use the same research situation that was described in Example 10.1. The researcher is using an independent-measures design to examine the relationship between watching educational TV as a child and academic performance as a high school student. The prediction is that high school students who watched Sesame Street regularly as 5-year-old children have higher grades.

STEP 1

State the hypotheses and select the alpha level. As always, the null hypothesis says that there is no effect, and the alternative hypothesis says that there is an effect. For this example, the predicted effect is that the students who watched Sesame Street have higher grades. Thus, the two hypotheses are as follows.

H_0: $\mu_{\text{Sesame Street}} \leq \mu_{\text{No Sesame Street}}$ (Grades are not higher with Sesame Street)

H_1: $\mu_{\text{Sesame Street}} > \mu_{\text{No Sesame Street}}$ (Grades are higher with Sesame Street)

Note that it is usually easier to state the hypotheses in words before you try to write them in symbols. Also, it usually is easier to begin with the alternative hypothesis (H_1), which states that the treatment works as predicted. Also note that the equal sign goes in the null hypothesis, indicating *no difference* between the two treatment conditions. The idea of zero difference is the essence of the null hypothesis, and the numerical value of zero is used for $(\mu_1 - \mu_2)$ during the calculation of the t statistic. For this test we use $\alpha = .01$.

STEP 2

Locate the critical region. For a directional test, the critical region is located entirely in one tail of the distribution. Rather than trying to determine which tail, positive or negative, is the correct location, we suggest that you identify the criteria for the critical region in a two-step process as follows. First, look at the data and determine whether the sample mean difference is in the direction that was predicted. If the answer is no, then the data obviously do not support the predicted treatment effect, and you can stop the analysis. On the other hand, if the difference is in the predicted direction, then the second step is to determine whether the difference is large enough to be significant. To test for significance, simply find the one-tailed critical value in the t distribution table. If the calculated t statistic is more extreme (either positive or negative) than the critical value, then the difference is significant.

For this example, the students who watched Sesame Street had higher grades, as predicted. With $df = 18$, the one-tailed critical value for $\alpha = .01$ is $t = 2.552$.

STEP 3

Collect the data and calculate the test statistic. The details of the calculations were shown in Example 10.1. The data produce a t statistic of $t = 4.00$.

STEP 4   Make a decision. The *t* statistic of $t = 4.00$ is well beyond the critical boundary of $t = 2.552$. Therefore, we reject the null hypothesis and conclude that grades for students who watched Sesame Street are significantly higher than grades for students who did not watch the program. In a research report, the one-tailed test would be clearly noted:

> Grades were significantly higher for students who watched Sesame Street, $t(18) = 4.00$, $p < .01$, one tailed.

THE ROLE OF SAMPLE VARIANCE AND SAMPLE SIZE IN THE INDEPENDENT-MEASURES *t* TEST

In Chapter 9 (p. 294), we identified several factors that can influence the outcome of a hypothesis test. Two factors that play important roles are the variability of the scores and the size of the samples. Both factors influence the magnitude of the estimated standard error in the denominator of the *t* statistic. However, the standard error is directly related to sample variance (larger variance leads to larger error) but it is inversely related to sample size (larger size leads to smaller error). As a result, larger variance produces a smaller value for the *t* statistic (closer to zero) and reduces the likelihood of finding a significant result. By contrast, a larger sample produces a larger value for the *t* statistic (farther from zero) and increases the likelihood of rejecting H_0.

Although variance and sample size both influence the hypothesis test, only variance has a large influence on measures of effect size such as Cohen's *d* and r^2; larger variance produces smaller measures of effect size. Sample size, on the other hand, has no effect on the value of Cohen's *d* and only a small influence on r^2.

The following example provides a visual demonstration of how large sample variance can obscure a mean difference between samples and lower the likelihood of rejecting H_0 for an independent-measures study.

EXAMPLE 10.5   We use the data in Figure 10.7 to demonstrate the influence of sample variance. The figure shows the results from a research study comparing two treatments. Notice that the study uses two separate samples, each with $n = 9$, and there is a 5-point mean difference
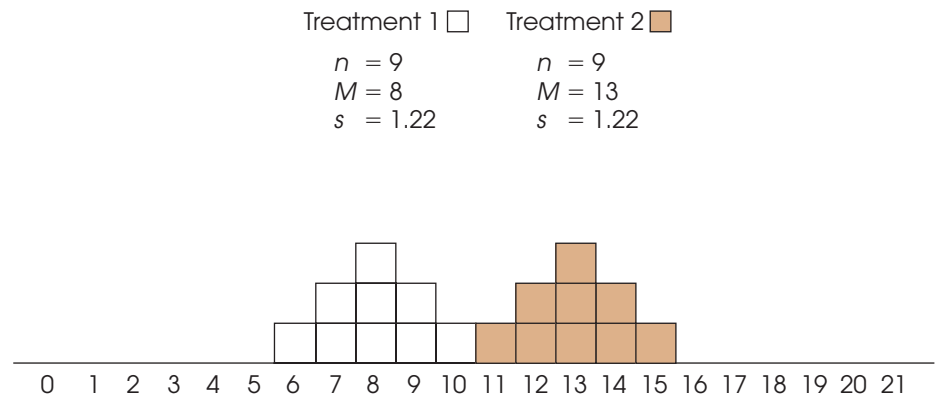

FIGURE 10.7

Two sample distributions representing two different treatments. These data show a significant difference between treatments, $t(16) = 8.62$, $p < .01$, and both measures of effect size indicate a very large treatment effect, $d = 4.10$ and $r^2 = 0.82$.

between the two samples: $M = 8$ for treatment 1 and $M = 13$ for treatment 2. Also notice that there is a clear difference between the two distributions; the scores for treatment 2 are clearly higher than the scores for treatment 1.

For the hypothesis test, the data produce a pooled variance of 1.50 and an estimated standard error of 0.58. The t statistic is

$$t = \frac{\text{mean difference}}{\text{estimated standard error}} = \frac{5}{0.58} = 8.62$$

With $df = 16$, this value is far into the critical region (for $\alpha = .05$ or $\alpha = .01$), so we reject the null hypothesis and conclude that there is a significant difference between the two treatments.

Now consider the effect of increasing sample variance. Figure 10.8 shows the results from a second research study comparing two treatments. Notice that there are still $n = 9$ scores in each sample, and the two sample means are still $M = 8$ and $M = 13$. However, the sample variances have been greatly increased: Each sample now has $s^2 = 44.25$ as compared with $s^2 = 1.5$ for the data in Figure 10.7. Notice that the increased variance means that the scores are now spread out over a wider range, with the result that the two samples are mixed together without any clear distinction between them.

The absence of a clear difference between the two samples is supported by the hypothesis test. The pooled variance is 44.25, the estimated standard error is 3.14, and the independent-measures t statistic is

$$t = \frac{\text{mean difference}}{\text{estimated standard error}} = \frac{5}{3.14} = 1.59$$

With $df = 16$ and $\alpha = .05$, this value is not in the critical region. Therefore, we fail to reject the null hypothesis and conclude that there is no significant difference between the two treatments. Although there is still a 5-point difference between sample means (as in Figure 10.7), the 5-point difference is not significant with the increased variance. In general, large sample variance can obscure any mean difference that exists in the data and reduces the likelihood of obtaining a significant difference in a hypothesis test.

Treatment 1 ☐   Treatment 2 ▧

$n = 9$	$n = 9$
$M = 8$	$M = 13$
$s = 6.65$	$s = 6.65$

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

FIGURE 10.8

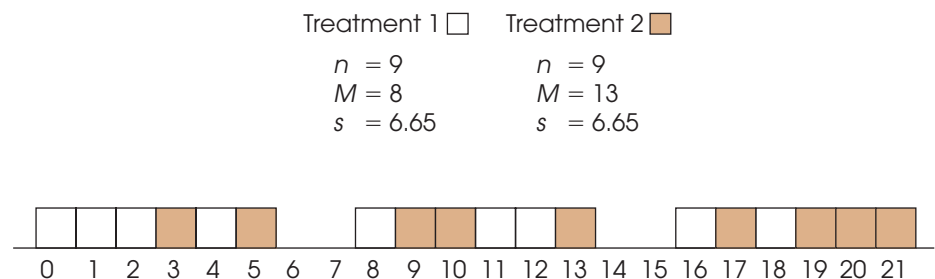Two sample distributions representing two different treatments. These data show exactly the same mean difference as the scores in Figure 10.7; however, the variance has been greatly increased. With the increased variance, there is no longer a significant difference between treatments, $t(16) = 1.59$, $p > .05$, and both measures of effect size are substantially reduced, $d = 0.75$ and $r^2 = 0.14$.

Finally, we should note that the problems associated with high variance often can be minimized by transforming the original scores to ranks and then conducting an alternative statistical analysis known as the *Mann-Whitney test,* which is designed specifically for ordinal data. The Mann-Whitney test is presented in Appendix E, which also discusses the general purpose and process of converting numerical scores into ranks. The Mann-Whitney test also can be used if the data violate one of the assumptions for the independent-measures *t* test outlined in the next section.

10.4 ASSUMPTIONS UNDERLYING THE INDEPENDENT-MEASURES *t* FORMULA

There are three assumptions that should be satisfied before you use the independent-measures *t* formula for hypothesis testing:

1. The observations within each sample must be independent (see p. 254).

2. The two populations from which the samples are selected must be normal.

3. The two populations from which the samples are selected must have equal variances.

The first two assumptions should be familiar from the single-sample *t* hypothesis test presented in Chapter 9. As before, the normality assumption is the less important of the two, especially with large samples. When there is reason to suspect that the populations are far from normal, you should compensate by ensuring that the samples are relatively large.

The third assumption is referred to as *homogeneity of variance* and states that the two populations being compared must have the same variance. You may recall a similar assumption for the *z*-score hypothesis test in Chapter 8. For that test, we assumed that the effect of the treatment was to add a constant amount to (or subtract a constant amount from) each individual score. As a result, the population standard deviation after treatment was the same as it had been before treatment. We now are making essentially the same assumption, but phrasing it in terms of variances.

Recall that the pooled variance in the *t*-statistic formula is obtained by averaging together the two sample variances. It makes sense to average these two values only if they both are estimating the same population variance—that is, if the homogeneity of variance assumption is satisfied. If the two sample variances are estimating different population variances, then the average is meaningless. (*Note*: If two people are asked to estimate the same thing—for example, what your IQ is—it is reasonable to average the two estimates. However, it is not meaningful to average estimates of two different things. If one person estimates your IQ and another estimates the number of grapes in a pound, it is meaningless to average the two numbers.)

Homogeneity of variance is most important when there is a large discrepancy between the sample sizes. With equal (or nearly equal) sample sizes, this assumption is less critical, but still important. Violating the homogeneity of variance assumption can negate any meaningful interpretation of the data from an independent-measures experiment. Specifically, when you compute the *t* statistic in a hypothesis test, all of the numbers in the formula come from the data except for the population mean difference, which you get from H_0. Thus, you are sure of all of the numbers in the formula except one. If you obtain an extreme result for the *t* statistic (a value in the critical region), then you conclude that the hypothesized value was wrong. But consider what happens when

Remember: Adding a constant to (or subtracting a constant from) each score does not change the standard deviation.

you violate the homogeneity of variance assumption. In this case, you have two questionable values in the formula (the hypothesized population value and the meaningless average of the two variances). Now if you obtain an extreme *t* statistic, you do not know which of these two values is responsible. Specifically, you cannot reject the hypothesis because it may have been the pooled variance that produced the extreme *t* statistic. Without satisfying the homogeneity of variance requirement, you cannot accurately interpret a *t* statistic, and the hypothesis test becomes meaningless.

HARTLEY'S *F*-MAX TEST How do you know whether the homogeneity of variance assumption is satisfied? One simple test involves just looking at the two sample variances. Logically, if the two population variances are equal, then the two sample variances should be very similar. When the two sample variances are reasonably close, you can be reasonably confident that the homogeneity assumption has been satisfied and proceed with the test. However, if one sample variance is more than three or four times larger than the other, then there is reason for concern. A more objective procedure involves a statistical test to evaluate the homogeneity assumption. Although there are many different statistical methods for determining whether the homogeneity of variance assumption has been satisfied, Hartley's *F*-max test is one of the simplest to compute and to understand. An additional advantage is that this test can also be used to check homogeneity of variance with more than two independent samples. Later, in Chapter 12, we examine statistical methods for comparing several different samples, and Hartley's test is useful again. The following example demonstrates the *F*-max test for two independent samples.

E X A M P L E   1 0 . 6 The *F*-max test is based on the principle that a sample variance provides an unbiased estimate of the population variance. The null hypothesis for this test states that the population variances are equal, therefore, the sample variances should be very similar. The procedure for using the *F*-max test is as follows:

1. Compute the sample variance, $s^2 = \dfrac{SS}{df}$, for each of the separate samples.

2. Select the largest and the smallest of these sample variances and compute

$$F\text{-max} = \frac{s^2(\text{largest})}{s^2(\text{smallest})}$$

A relatively large value for *F*-max indicates a large difference between the sample variances. In this case, the data suggest that the population variances are different and that the homogeneity assumption has been violated. On the other hand, a small value of *F*-max (near 1.00) indicates that the sample variances are similar and that the homogeneity assumption is reasonable.

3. The *F*-max value computed for the sample data is compared with the critical value found in Table B.3 (Appendix B). If the sample value is larger than the table value, then you conclude that the variances are different and that the homogeneity assumption is not valid.

    To locate the critical value in the table, you need to know:

    a. k = number of separate samples. (For the independent-measures *t* test, $k = 2$.)

    b. $df = n - 1$ for each sample variance. The Hartley test assumes that all samples are the same size.

    c. The alpha level. The table provides critical values for $\alpha = .05$ and $\alpha = .01$. Generally a test for homogeneity would use the larger alpha level.

Suppose, for example, that two independent samples each have $n = 10$ with sample variances of 12.34 and 9.15. For these data,

$$F\text{-max} = \frac{s^2\left(\text{largest}\right)}{s^2\left(\text{smallest}\right)} = \frac{12.34}{9.15} = 1.35$$

With $\alpha = .05$, $k = 2$, and $df = n - 1 = 9$, the critical value from the table is 4.03. Because the obtained F-max is smaller than this critical value, you conclude that the data do not provide evidence that the homogeneity of variance assumption has been violated.

The goal for most hypothesis tests is to reject the null hypothesis to demonstrate a significant difference or a significant treatment effect. However, when testing for homogeneity of variance, the preferred outcome is to fail to reject H_0. Failing to reject H_0 with the F-max test means that there is no significant difference between the two population variances and the homogeneity assumption is satisfied. In this case, you may proceed with the independent-measures t test using pooled variance.

If the F-max test rejects the hypothesis of equal variances, or if you simply suspect that the homogeneity of variance assumption is not justified, you should not compute an independent-measures t statistic using pooled variance. However, there is an alternative formula for the t statistic that does not pool the two sample variances and does not require the homogeneity assumption. The alternative formula is presented in Box 10.2.

BOX 10.2   AN ALTERNATIVE TO POOLED VARIANCE

Computing the independent-measures t statistic using pooled variance requires that the data satisfy the homogeneity of variance assumption. Specifically, the two distributions from which the samples are obtained must have equal variances. To avoid this assumption, many statisticians recommend an alternative formula for computing the independent-measures t statistic that does not require pooled variance or the homogeneity assumption. The alternative procedure consists of two steps:

1. The standard error is computed using the two separate sample variances as in Equation 10.1.

2. The value of degrees of freedom for the t statistic is adjusted using the following equation:

$$df = \frac{\left(V_1 + V_2\right)^2}{\dfrac{V_1^2}{n_1 - 1} + \dfrac{V_2^2}{n_2 - 1}} \quad \text{where } V_1 = \frac{s_1^2}{n_1} \text{ and } V_2 = \frac{s_2^2}{n_2}$$

Decimal values for df should be rounded down to the next lower integer.

The adjustment to degrees of freedom lowers the value of df, which pushes the boundaries for the critical region farther out. Thus, the adjustment makes the test more demanding and therefore corrects for the same bias problem that the pooled variance attempts to avoid.

Note: Many computer programs that perform statistical analysis (such as SPSS) report two versions of the independent-measures t statistic; one using pooled variance (with equal variances assumed) and one using the adjustment shown here (with equal variances not assumed).

LEARNING CHECK

1. A researcher is using an independent-measures design to evaluate the difference between two treatment conditions with $n = 8$ in each treatment. The first treatment produces $M = 63$ with a variance of $s^2 = 18$, and the second treatment has $M = 58$ with $s^2 = 14$.

   a. Use a one-tailed test with $\alpha = .05$ to determine whether the scores in the first treatment are significantly greater than the scores in the second. (*Note*: Because the two samples are the same size, the pooled variance is simply the average of the two sample variances.)

   b. Predict how the value for the *t* statistic would be affected if the two sample variances were increased to $s^2 = 68$ and $s^2 = 60$. Compute the new *t* to confirm your answer.

   c. Predict how the value for the *t* statistic for the original samples would be affected if each sample had $n = 32$ scores (instead of $n = 8$). Compute the new *t* to confirm your answer.

2. The homogeneity of variance assumption requires that the two sample variances be equal. (True or false?)

3. When you are using an *F*-max test to evaluate the homogeneity of variance assumption, you usually do not want to find a significant difference between the variances. (True or false?)

ANSWERS

1. a. The pooled variance is 16, the estimated standard error is 2, and $t(14) = 2.50$. With a one-tailed critical value of 1.761, reject the null hypothesis. Scores in the first treatment are significantly higher than scores in the second.

   b. Increasing the variance should lower the value of *t*. The new pooled variance is 64, the estimated standard error is 4, and $t(14) = 1.25$.

   c. Increasing the sample sizes should increase the value of *t*. The pooled variance is still 16, but the new standard error is 1, and $t(62) = 5.00$.

2. False. The assumption is that the two *population* variances are equal.

3. True. If there is a significant difference between the two variances, you cannot do the *t* test with pooled variance.

SUMMARY

1. The independent-measures *t* statistic uses the data from two separate samples to draw inferences about the mean difference between two populations or between two different treatment conditions.

2. The formula for the independent-measures *t* statistic has the same structure as the original *z*-score or the single-sample *t*:

$$t = \frac{\text{sample statistic} - \text{population parameter}}{\text{estimated standard error}}$$

For the independent-measures *t*, the sample statistic is the sample mean difference $(M_1 - M_2)$. The population

parameter is the population mean difference, $(\mu_1 - \mu_2)$. The estimated standard error for the sample mean difference is computed by combining the errors for the two sample means. The resulting formula is

$$t = \frac{\left(M_1 - M_2\right) - \left(\mu_1 - \mu_2\right)}{s_{\left(M_1 - M_2\right)}}$$

where the estimated standard error is

$$s_{\left(M_1 - M_2\right)} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

The pooled variance in the formula, s_p^2, is the weighted mean of the two sample variances:

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

This t statistic has degrees of freedom determined by the sum of the df values for the two samples:

$$df = df_1 + df_2$$
$$= (n_1 - 1) + (n_2 - 1)$$

3. For hypothesis testing, the null hypothesis states that there is no difference between the two population means:

$$H_0: \mu_1 = \mu_2 \quad \text{or} \quad \mu_1 - \mu_2 = 0$$

4. When a hypothesis test with an independent-measures t statistic indicates a significant difference, you should also compute a measure of the effect size. One measure of effect size is Cohen's d, which is a standardized measure of the mean difference. For the independent-measures t statistic, Cohen's d is estimated as follows:

$$\text{estimated } d = \frac{M_1 - M_2}{\sqrt{s_p^2}}$$

A second common measure of effect size is the percentage of variance accounted for by the treatment effect. This measure is identified by r^2 and is computed as

$$r^2 = \frac{t^2}{t^2 + df}$$

5. An alternative method for describing the size of the treatment effect is to construct a confidence interval for the population mean difference, $\mu_1 - \mu_2$. The confidence interval uses the independent-measures t equation, solved for the unknown mean difference:

$$\mu_1 - \mu_2 = M_1 - M_2 \pm ts_{(M_1 - M_2)}$$

First, select a level of confidence and then look up the corresponding t values. For example, for 95% confidence, use the range of t values that determine the middle 95% of the distribution. The t values are then used in the equation along with the values for the sample mean difference and the standard error, which are computed from the sample data.

6. Appropriate use and interpretation of the t statistic using pooled variance require that the data satisfy the homogeneity of variance assumption. This assumption stipulates that the two populations have equal variances. An informal test of the assumption can be made by verifying that the two sample variances are approximately equal. Hartley's F-max test provides a statistical technique for determining whether the data satisfy the homogeneity assumption. An alternative technique that avoids pooling variances and eliminates the need for the homogeneity assumption is presented in Box 10.2.

KEY TERMS

independent-measures research design (318)

between-subjects research design (318)

repeated-measures research design (318)

within-subjects research design (318)

independent-measures t statistic (319)

estimated standard error of $M_1 - M_2$ (320)

pooled variance (322)

Mann-Whitney test (337)

homogeneity of variance (337)

RESOURCES

Book Companion Website: www.cengage.com/psychology/gravetter
    You can find a tutorial quiz and other learning exercises for Chapter 10 on the book companion website. The website also provides access to a workshop entitled *Independent vs. Repeated t-tests,* which compares the t test presented in this chapter with the repeated-measures test presented in Chapter 11.

aplia

Improve your understanding of statistics with Aplia's auto-graded problem sets and immediate, detailed explanations for every question. To learn more, visit **www.aplia.com/statistics**.

CENGAGE**brain**.com

Psychology CourseMate brings course concepts to life with interactive learning, study, and exam preparation tools that support the printed textbook. A textbook-specific website, Psychology CourseMate includes an integrated interactive eBook and other interactive learning tools including quizzes, flashcards, and more.

Visit www.cengagebrain.com to access your account and purchase materials.

SPSS

General instructions for using SPSS are presented in Appendix D. Following are detailed instructions for using SPSS to perform **The Independent-Measures *t* Test** presented in this chapter.

Data Entry

1. The scores are entered in what is called *stacked format,* which means that all of the scores from *both samples* are entered in one column of the data editor (probably VAR00001). Enter the scores for sample #2 directly beneath the scores from sample #1 with no gaps or extra spaces.
2. Values are then entered into a second column (VAR00002) to identify the sample or treatment condition corresponding to each of the scores. For example, enter a 1 beside each score from sample #1 and enter a 2 beside each score from sample #2.

Data Analysis

1. Click **Analyze** on the tool bar, select **Compare Means,** and click on **Independent-Samples *t* Test.**
2. Highlight the column label for the set of scores (VAR0001) in the left box and click the arrow to move it into the **Test Variable(s)** box.
3. Highlight the label from the column containing the sample numbers (VAR0002) in the left box and click the arrow to move it into the **Group Variable** box.
4. Click on **Define Groups.**
5. Assuming that you used the numbers 1 and 2 to identify the two sets of scores, enter the values 1 and 2 into the appropriate group boxes.
6. Click **Continue.**
7. In addition to performing the hypothesis test, the program computes a confidence interval for the population mean difference. The confidence level is automatically set at 95% but you can select **Options** and change the percentage.
8. Click **OK.**

SPSS Output

We used the SPSS program to analyze the data from the Sesame Street study in Example 10.1 and the program output is shown in Figure 10.9. The output includes a table of sample statistics with the mean, standard deviation, and standard error of the mean for each group. A second table, which is split into two sections in Figure 10.9, begins with the results of Levene's test for homogeneity of variance. This test should *not* be significant (you do not want the two variances to be different), so you want the reported Sig. value to be greater than .05. Next, the results of the independent-measures *t* test are presented using two different assumptions. The top row shows the outcome assuming equal variances, using the pooled variance to compute *t*. The second row does not assume equal variances and computes the *t* statistic using the alternative method presented in Box 10.2. Each row reports the calculated *t* value, the degrees of freedom, the level of significance (the *p* value for the test), the size of the mean difference and the standard error for the mean difference (the denominator of the *t* statistic). Finally, the output includes a 95% confidence interval for the mean difference.

Group Statistics

	VAR00002	N	Mean	Std. Deviation	Std. Error Mean
VAR00001	1.00	10	93.0000	4.71405	1.49071
	2.00	10	85.0000	4.21637	1.33333

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
VAR00001	Equal variances assumed	.384	.543	4.000	18
	Equal variances not assumed			4.000	17.780

Independent Samples Test

		t-test for Equality of Means			95% Confidence Interval of the Difference	
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
VAR00001	Equal variances assumed	.001	8.00000	2.00000	3.79816	12.20184
	Equal variances not assumed	.001	8.00000	2.00000	3.79443	12.20557

FIGURE 10.9

The SPSS output for the independent-measures hypothesis test in Example 10.1.

FOCUS ON PROBLEM SOLVING

1. As you learn more about different statistical methods, one basic problem is deciding which method is appropriate for a particular set of data. Fortunately, it is easy to identify situations in which the independent-measures *t* statistic is used. First, the data always consist of two separate samples (two *n*s, two *M*s, two *SS*s, and so on). Second, this *t* statistic is always used to answer questions about a mean difference: On the average, is one group different (better, faster, smarter) than the other group? If you examine the data and identify the type of question that a researcher is asking, you should be able to decide whether an independent-measures *t* is appropriate.

2. When computing an independent-measures *t* statistic from sample data, we suggest that you routinely divide the formula into separate stages rather than trying to do all of the calculations at once. First, find the pooled variance. Second, compute the standard error. Third, compute the *t* statistic.

3. One of the most common errors for students involves confusing the formulas for pooled variance and standard error. When computing pooled variance, you are "pooling" the two samples together into a single variance. This variance is computed as a *single fraction,* with two *SS* values in the numerator and two *df* values in the denominator. When computing the standard error, you are adding the error from the first sample and the error from the second sample. These two separate errors are added as *two separate fractions* under the square root symbol.

DEMONSTRATION 10.1

THE INDEPENDENT-MEASURES *t* TEST

In a study of jury behavior, two samples of participants were provided details about a trial in which the defendant was obviously guilty. Although group 2 received the same details as group 1, the second group was also told that some evidence had been withheld from the jury by the judge. Later the participants were asked to recommend a jail sentence. The length of term suggested by each participant is presented here. Is there a significant difference between the two groups in their responses?

Group 1	Group 2
4	3
4	7
3	8
2	5
5	4
1	7
1	6
4	8

for Group 1: $M = 3$ and $SS = 16$

for Group 2: $M = 6$ and $SS = 24$

There are two separate samples in this study. Therefore, the analysis uses the independent-measures t test.

STEP 1 State the hypothesis, and select an alpha level.

H_0: $\mu_1 - \mu_2 = 0$ (For the population, knowing that evidence has been withheld has no effect on the suggested sentence.)

H_1: $\mu_1 \neq \mu_2 \neq 0$ (For the population, knowing that evidence has been withheld has an effect on the jury's response.)

We set the level of significance to $\alpha = .05$, two tails.

STEP 2 **Identify the critical region.** For the independent-measures t statistic, degrees of freedom are determined by

$$df = n_1 + n_2 - 2$$
$$= 8 + 8 - 2$$
$$= 14$$

The t distribution table is consulted, for a two-tailed test with $\alpha = .05$ and $df = 14$. The critical t values are $+2.145$ and -2.145.

STEP 3 **Compute the test statistic.** As usual, we recommend that the calculation of the t statistic be separated into three stages.

Pooled variance: For these data, the pooled variance equals

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{16 + 24}{7 + 7} = \frac{40}{14} = 2.86$$

Estimated standard error: Now we can calculate the estimated standard error for mean differences.

$$s_{(M_1 - M_2)} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{\frac{2.86}{8} + \frac{2.86}{8}} = \sqrt{0.358 + 0.358} = \sqrt{0.716} = 0.85$$

The t *statistic:* Finally, the t statistic can be computed.

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{(M_1 - M_2)}} = \frac{(3 - 6) - 0}{0.85} = \frac{-3}{0.85} = -3.53$$

STEP 4 **Make a decision about H_0, and state a conclusion.** The obtained t value of -3.53 falls in the critical region of the left tail (critical $t = \pm 2.145$). Therefore, the null hypothesis is rejected. The participants who were informed about the withheld evidence gave significantly longer sentences, $t(14) = -3.53$, $p < .05$, two tails.

DEMONSTRATION 10.2

EFFECT SIZE FOR THE INDEPENDENT-MEASURES *t*

We estimate Cohen's *d* and compute r^2 for the jury decision data in Demonstration 10.1. For these data, the two sample means are $M_1 = 3$ and $M_2 = 6$, and the pooled variance is 2.86. Therefore, our estimate of Cohen's *d* is

$$\text{estimated } d = \frac{M_1 - M_2}{\sqrt{s_p^2}} = \frac{3 - 6}{\sqrt{2.86}} = \frac{3}{1.69} = 1.78$$

With a *t* value of $t = 3.53$ and $df = 14$, the percentage of variance accounted for is

$$r^2 = \frac{t^2}{t^2 + df} = \frac{(3.53)^2}{(3.53)^2 + 14} = \frac{12.46}{26.46} = 0.47 \left(\text{or } 47\%\right)$$

PROBLEMS

1. Describe the basic characteristics of an independent-measures, or a between-subjects, research study.

2. Describe what is measured by the estimated standard error in the bottom of the independent-measures *t* statistic.

3. If other factors are held constant, explain how each of the following influences the value of the independent-measures *t* statistic and the likelihood of rejecting the null hypothesis:
   a. Increasing the number of scores in each sample.
   b. Increasing the variance for each sample.

4. Describe the homogeneity of variance assumption and explain why it is important for the independent-measures *t* test.

5. One sample has $SS = 48$ and a second sample has $SS = 32$.
   a. If $n = 5$ for both samples, find each of the sample variances and compute the pooled variance. Because the samples are the same size, you should find that the pooled variance is exactly halfway between the two sample variances.
   b. Now assume that $n = 5$ for the first sample and $n = 9$ for the second. Again, calculate the two sample variances and the pooled variance. You should find that the pooled variance is closer to the variance for the larger sample.

6. One sample has $SS = 70$ and a second sample has $SS = 42$.
   a. If $n = 8$ for both samples, find each of the sample variances, and calculate the pooled variance. Because the samples are the same size, you should

find that the pooled variance is exactly halfway between the two sample variances.
   b. Now assume that $n = 8$ for the first sample and $n = 4$ for the second. Again, calculate the two sample variances and the pooled variance. You should find that the pooled variance is closer to the variance for the larger sample.

7. As noted on page 320, when the two population means are equal, the estimated standard error for the independent-measures *t* test provides a measure of how much difference to expect between two sample means. For each of the following situations, assume that $\mu_1 = \mu_2$ and calculate how much difference should be expected between the two sample means.
   a. One sample has $n = 8$ scores with $SS = 45$ and the second sample has $n = 4$ scores with $SS = 15$.
   b. One sample has $n = 8$ scores with $SS = 150$ and the second sample has $n = 4$ scores with $SS = 90$.
   c. In part b, the samples have larger variability (bigger SS values) than in part a, but the sample sizes are unchanged. How does larger variability affect the size of the standard error for the sample mean difference?

8. Two separate samples, each with $n = 12$ individuals, receive two different treatments. After treatment, the first sample has $SS = 1740$ and the second has $SS = 1560$.
   a. Find the pooled variance for the two samples.
   b. Compute the estimated standard error for the sample mean difference.
   c. If the sample mean difference is 8 points, is this enough to reject the null hypothesis and conclude that there is a significant difference for a two-tailed test at the .05 level?

d. If the sample mean difference is 12 points, is this enough to indicate a significant difference for a two-tailed test at the .05 level?

e. Calculate the percentage of variance accounted for (r^2) to measure the effect size for an 8-point mean difference and for a 12-point mean difference.

9. Two separate samples receive two different treatments. The first sample has $n = 9$ with $SS = 710$, and the second has $n = 6$ with $SS = 460$.

a. Compute the pooled variance for the two samples.

b. Calculate the estimated standard error for the sample mean difference.

c. If the sample mean difference is 10 points, is this enough to reject the null hypothesis using a two-tailed test with $\alpha = .05$?

d. If the sample mean difference is 13 points, is this enough to reject the null hypothesis using a two tailed test with $\alpha = .05$?

10. For each of the following, assume that the two samples are selected from populations with equal means and calculate how much difference should be expected, on average, between the two sample means.

a. Each sample has $n = 5$ scores with $s^2 = 38$ for the first sample and $s^2 = 42$ for the second. (*Note:* Because the two samples are the same size, the pooled variance is equal to the average of the two sample variances.)

b. Each sample has $n = 20$ scores with $s^2 = 38$ for the first sample and $s^2 = 42$ for the second.

c. In part b, the two samples are bigger than in part a, but the variances are unchanged. How does sample size affect the size of the standard error for the sample mean difference?

11. For each of the following, calculate the pooled variance and the estimated standard error for the sample mean difference

a. The first sample has $n = 4$ scores and a variance of $s^2 = 55$, and the second sample has $n = 6$ scores and a variance of $s^2 = 63$.

b. Now the sample variances are increased so that the first sample has $n = 4$ scores and a variance of $s^2 = 220$, and the second sample has $n = 6$ scores and a variance of $s^2 = 252$.

c. Comparing your answers for parts a and b, how does increased variance influence the size of the estimated standard error?

12. A researcher conducts an independent-measures study comparing two treatments and reports the t statistic as $t(30) = 2.085$.

a. How many individuals participated in the entire study?

b. Using a two-tailed test with $\alpha = .05$, is there a significant difference between the two treatments?

c. Compute r^2 to measure the percentage of variance accounted for by the treatment effect.

13. Hallam, Price, and Katsarou (2002) investigated the influence of background noise on classroom performance for children aged 10 to 12. In one part of the study, calming music led to better performance on an arithmetic task compared to a no-music condition. Suppose that a researcher selects one class of $n = 18$ students who listen to calming music each day while working on arithmetic problems. A second class of $n = 18$ serves as a control group with no music. Accuracy scores are measured for each child and the average for students in the music condition is $M = 86.4$ with $SS = 1550$ compared to an average of $M = 78.8$ with $SS = 1204$ for students in the no-music condition.

a. Is there a significant difference between the two music conditions? Use a two-tailed test with $\alpha = .05$.

b. Compute the 90% confidence interval for the population mean difference.

c. Write a sentence demonstrating how the results from the hypothesis test and the confidence interval would appear in a research report.

14. Do you view a chocolate bar as delicious or as fattening? Your attitude may depend on your gender. In a study of American college students, Rozin, Bauer, and Catanese (2003) examined the importance of food as a source of pleasure versus concerns about food associated with weight gain and health. The following results are similar to those obtained in the study. The scores are a measure of concern about the negative aspects of eating.

Males	Females
$n = 9$	$n = 15$
$M = 33$	$M = 42$
$SS = 740$	$SS = 1240$

a. Based on these results, is there a significant difference between the attitudes for males and for females? Use a two-tailed test with $\alpha = .05$.

b. Compute r^2, the percentage of variance accounted for by the gender difference, to measure effect size for this study.

c. Write a sentence demonstrating how the result of the hypothesis test and the measure of effect size would appear in a research report.

15. In a study examining overweight and obese college football players, Mathews and Wagner (2008) found that on average both offensive and defensive linemen exceeded the at-risk criterion for body mass index (BMI). BMI is a ratio of body weight to height squared and is commonly used to classify people as overweight or obese. Any value greater than 30 kg/m^2 is considered to be at risk. In the study, a sample of $n = 17$ offensive linemen averaged $M = 34.4$ with a

standard deviation of $s = 4.0$. A sample of $n = 19$ defensive linemen averaged $M = 31.9$ with $s = 3.5$.

a. Use a single-sample *t* test to determine whether the offensive linemen are significantly above the at-risk criterion for BMI. Use a one-tailed test with $\alpha = .01$.

b. Use a single-sample *t* test to determine whether the defensive linemen are significantly above the at-risk criterion for BMI. Use a one-tailed test with $\alpha = .01$.

c. Use an independent-measures *t* test to determine whether there is a significant difference between the offensive linemen and the defensive linemen. Use a two-tailed test with $\alpha = .01$.

16. Functional foods are those containing nutritional supplements in addition to natural nutrients. Examples include orange juice with calcium and eggs with omega-3. Kolodinsky, et al. (2008) examined attitudes toward functional foods for college students. For American students, the results indicated that females had a more positive attitude toward functional foods and were more likely to purchase them compared to males. In a similar study, a researcher asked students to rate their general attitude toward functional foods on a 7-point scale (higher score is more positive). The results are as follows:

Females	Male
$n = 8$	$n = 12$
$M = 4.69$	$M = 4.43$
$SS = 1.60$	$SS = 2.72$

a. Do the data indicate a significant difference in attitude for males and females? Use a two-tailed test with $\alpha = .05$.

b. Compute r^2, the amount of variance accounted for by the gender difference, to measure effect size.

c. Write a sentence demonstrating how the results of the hypothesis test and the measure of effect size would appear in a research report.

17. In 1974, Loftus and Palmer conducted a classic study demonstrating how the language used to ask a question can influence eyewitness memory. In the study, college students watched a film of an automobile accident and then were asked questions about what they saw. One group was asked, "About how fast were the cars going when they smashed into each other?" Another group was asked the same question except the verb was changed to "hit" instead of "smashed into." The "smashed into" group reported significantly higher estimates of speed than the "hit" group. Suppose a researcher repeats this study with a sample of today's college students and obtains the following results.

Estimated Speed	
Smashed into	Hit
$n = 15$	$n = 15$
$M = 40.8$	$M = 34.0$
$SS = 510$	$SS = 414$

a. Do the results indicate a significantly higher estimated speed for the "smashed into" group? Use a one-tailed test with $\alpha = .01$.

b. Compute the estimated value for Cohen's *d* to measure the size of the effect.

c. Write a sentence demonstrating how the results of the hypothesis test and the measure of effect size would appear in a research report.

18. Numerous studies have found that males report higher self-esteem than females, especially for adolescents (Kling, Hyde, Showers, & Buswell, 1999). Typical results show a mean self-esteem score of $M = 39.0$ with $SS = 60.2$ for a sample of $n = 10$ male adolescents and a mean of $M = 35.4$ with $SS = 69.4$ for a sample of $n = 10$ female adolescents.

a. Do the results indicate that self-esteem is significantly higher for males? Use a one-tailed test with $\alpha = .01$.

b. Use the data to make a 95% confidence interval estimate of the mean difference in self-esteem between male and female adolescents.

c. Write a sentence demonstrating how the results from the hypothesis test and the confidence interval would appear in a research report.

19. A researcher is comparing the effectiveness of two sets of instructions for assembling a child's bike. A sample of eight fathers is obtained. Half of the fathers are given one set of instructions and the other half receives the second set. The researcher measures how much time is needed for each father to assemble the bike. The scores are the number of minutes needed by each participant.

Instruction Set I	Instruction Set II
8	14
4	10
8	6
4	10

a. Is there a significant difference in time for the two sets of instructions? Use a two-tailed test at the .05 level of significance.

b. Calculate the estimated Cohen's *d* and r^2 to measure effect size for this study.

20. When people learn a new task, their performance usually improves when they are tested the next day, but only if they get at least 6 hours of sleep (Stickgold, Whidbee, Schirmer, Patel, & Hobson, 2000). The following data demonstrate this phenomenon. The participants learned a visual discrimination task on one day, and then were tested on the task the following day. Half of the participants were allowed to have at least 6 hours of sleep and the other half were kept awake all night. Is there a significant difference between the two conditions? Use a two-tailed test with $\alpha = .05$.

Performance Scores	
6 Hours of Sleep	No Sleep
$n = 14$	$n = 14$
$M = 72$	$M = 65$
$SS = 932$	$SS = 706$

21. Steven Schmidt (1994) conducted a series of experiments examining the effects of humor on memory. In one study, participants were given a mix of humorous and nonhumorous sentences and significantly more humorous sentences were recalled. However, Schmidt argued that the humorous sentences were not necessarily easier to remember, they were simply preferred when participants had a choice between the two types of sentence. To test this argument, he switched to an independent-measures design in which one group got a set of exclusively humorous sentences and another group got a set of exclusively nonhumorous sentences. The following data are similar to the results from the independent-measures study.

Humorous Sentences				Nonhumorous Sentences			
4	5	2	4	6	3	5	3
6	7	6	6	3	4	2	6
2	5	4	3	4	3	4	4
3	3	5	3	5	2	6	4

Do the results indicate a significant difference in the recall of humorous versus nonhumorous sentences? Use a two-tailed test with $\alpha = .05$.

22. Downs and Abwender (2002) evaluated soccer players and swimmers to determine whether the routine blows to the head experienced by soccer players produced long-term neurological deficits. In the study, neurological tests were administered to mature soccer players and swimmers and the results indicated significant differences. In a similar study, a researcher obtained the following data.

Swimmers	Soccer players
10	7
8	4
7	9
9	3
13	7
7	
6	
12	

a. Are the neurological test scores significantly lower for the soccer players than for the swimmers in the control group? Use a one-tailed test with $\alpha = .05$.
b. Compute the value of r^2 (percentage of variance accounted for) for these data.

23. Research has shown that people are more likely to show dishonest and self-interested behaviors in darkness than in a well-lit environment (Zhong, Bohns, & Gino, 2010). In one experiment, participants were given a set of 20 puzzles and were paid $0.50 for each one solved in a 5-minute period. However, the participants reported their own performance and there was no obvious method for checking their honesty. Thus, the task provided a clear opportunity to cheat and receive undeserved money. One group of participants was tested in a room with dimmed lighting and a second group was tested in a well-lit room. The reported number of solved puzzles was recorded for each individual. The following data represent results similar to those obtained in the study.

Well-Lit Room	Dimly Lit Room
7	9
8	11
10	13
6	10
8	11
5	9
7	15
12	14
5	10

a. Is there a significant difference in reported performance between the two conditions? Use a two-tailed test with $\alpha = .01$.
b. Compute Cohen's d to estimate the size of the treatment effect.