



9e

Statistics for the Behavioral Sciences

Frederick J Gravetter
Larry B. Wallnau

C H A P T E R

9

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 standard deviation (Chapter 4)
- Degrees of freedom (Chapter 4)
- Standard error (Chapter 7)
- Hypothesis testing (Chapter 8)

Introduction to the t Statistic

Preview

- 9.1 The t Statistic: An Alternative to z
- 9.2 Hypothesis Tests with the t Statistic
- 9.3 Measuring Effect Size for the t Statistic
- 9.4 Directional Hypotheses and One-Tailed Tests

Summary

Focus on Problem Solving

Demonstrations 9.1 and 9.2

Problems

Preview

Numerous accounts suggest that for many animals, including humans, direct stare from another animal is aversive (e.g., Cook, 1977). Try it out for yourself. Make direct eye contact with a stranger in a cafeteria. Chances are the person will display avoidance by averting his or her gaze or turning away from you. Some insects, such as moths, have even developed eye-spot patterns on the wings or body to ward off predators (mostly birds) who may have a natural fear of eyes (Blest, 1957). Suppose that a comparative psychologist is interested in determining whether the birds that feed on these insects show an avoidance of eye-spot patterns.

Using methods similar to those of Scaife (1976), the researcher performed the following experiment. A sample of $n = 16$ moth-eating birds is selected. The animals are tested in an apparatus that consists of a two-chambered box. The birds are free to roam from one side of the box to the other through a doorway in the partition that separates the two chambers. In one chamber, there are two eye-spot patterns painted on the wall. The other side of the box has plain walls. One at a time, the researcher tests each bird by placing it in the doorway between the chambers. Each subject is left in the apparatus for 60 minutes, and the amount of time spent in the plain chamber is recorded.

The null hypothesis for this study would state that eye-spot patterns have no effect on the behavior of moth-eating birds. If this is true, then birds placed in the apparatus should wander randomly from side to side during the 60-minute period, averaging half of the time on each side.

Therefore, for the general population of moth-eating birds, the null hypothesis states

$$H_0: \mu_{\text{plain side}} = 30 \text{ minutes}$$

The Problem: The researcher has most of the information needed to conduct a hypothesis test. In particular, the researcher has a hypothesis about the population ($\mu = 30$ minutes) and a sample of $n = 16$ scores that produces a sample mean (M). However, the researcher does not know the population standard deviation (σ). This value is needed to compute the standard error for the sample mean (σ_M) that appears in the denominator of the z -score equation. Recall that the standard error measures how much difference is reasonable to expect between a sample mean (M) and the population mean (μ). The value of the standard error is critical to deciding whether the sample data are consistent with the null hypothesis or refute the null hypothesis. Without the standard error, it is impossible to conduct a z -score hypothesis test.

The Solution: Because it is impossible to compute the standard error, a z -score cannot be used for the hypothesis test. However, it is possible to estimate the standard error using the sample data. The estimated standard error can then be used to compute a new statistic that is similar in structure to the z -score. The new statistic is called a t statistic and it can be used to conduct a new kind of hypothesis test.

9.1 THE t STATISTIC: AN ALTERNATIVE TO z

In the previous chapter, we presented the statistical procedures that permit researchers to use a sample mean to test hypotheses about an unknown population mean. These statistical procedures were based on a few basic concepts, which we summarize as follows:

1. A sample mean (M) is expected to approximate its population mean (μ). This permits us to use the sample mean to test a hypothesis about the population mean.
2. The standard error provides a measure of how well a sample mean approximates the population mean. Specifically, the standard error determines how much difference is reasonable to expect between a sample mean (M) and the population mean (μ).

Remember that the expected value of the distribution of sample means is μ , the population mean.

$$\sigma_M = \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \sigma_M = \sqrt{\frac{\sigma^2}{n}}$$

3. To quantify our inferences about the population, we compare the obtained sample mean (M) with the hypothesized population mean (μ) by computing a z -score test statistic.

$$z = \frac{M - \mu}{\sigma_M} = \frac{\text{obtained difference between data and hypothesis}}{\text{standard distance between } M \text{ and } \mu}$$

The goal of the hypothesis test is to determine whether the obtained difference between the data and the hypothesis is significantly greater than would be expected by chance. When the z -scores form a normal distribution, we are able to use the unit normal table (in Appendix B) to find the critical region for the hypothesis test.

THE PROBLEM WITH z -SCORES

The shortcoming of using a z -score for hypothesis testing is that the z -score formula requires more information than is usually available. Specifically, a z -score requires that we know the value of the population standard deviation (or variance), which is needed to compute the standard error. In most situations, however, the standard deviation for the population is not known. In fact, the whole reason for conducting a hypothesis test is to gain knowledge about an *unknown* population. This situation appears to create a paradox: You want to use a z -score to find out about an unknown population, but you must know about the population before you can compute a z -score. Fortunately, there is a relatively simple solution to this problem. When the variability for the population is not known, we use the sample variability in its place.

INTRODUCING THE t STATISTIC

In Chapter 4, the sample variance was developed specifically to provide an unbiased estimate of the corresponding population variance. Recall that the formulas for sample variance and sample standard deviation are as follows:

$$\text{sample variance} = s^2 = \frac{SS}{n - 1} = \frac{SS}{df}$$

$$\text{sample standard deviation} = s = \sqrt{\frac{SS}{n - 1}} = \sqrt{\frac{SS}{df}}$$

The concept of degrees of freedom, $df = n - 1$, was introduced in Chapter 4 (p. 117) and is discussed later in this chapter (p. 287).

Using the sample values, we can now *estimate* the standard error. Recall from Chapters 7 and 8 that the value of the standard error can be computed using either standard deviation or variance:

$$\text{standard error} = \sigma_M = \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \sigma_M = \sqrt{\frac{\sigma^2}{n}}$$

Now we estimate the standard error by simply substituting the sample variance or standard deviation in place of the unknown population value:

$$\text{estimated standard error} = s_M = \frac{s}{\sqrt{n}} \quad \text{or} \quad s_M = \sqrt{\frac{s^2}{n}} \quad (9.1)$$

Notice that the symbol for the *estimated standard error of M* is s_M instead of σ_M , indicating that the estimated value is computed from sample data rather than from the actual population parameter.

DEFINITION

The **estimated standard error** (s_M) is used as an estimate of the real standard error, σ_M , when the value of σ is unknown. It is computed from the sample variance or sample standard deviation and provides an estimate of the standard distance between a sample mean, M , and the population mean, μ .

Finally, you should recognize that we have shown formulas for standard error (actual or estimated) using both the standard deviation and the variance. In the past (Chapters 7 and 8), we concentrated on the formula using the standard deviation. At this point, however, we shift our focus to the formula based on variance. Thus, throughout the remainder of this chapter, and in following chapters, the estimated standard error of M typically is presented and computed using

$$s_M = \sqrt{\frac{s^2}{n}}$$

There are two reasons for making this shift from standard deviation to variance:

1. In Chapter 4 (p. 119) we saw that the sample variance is an *unbiased* statistic; on average, the sample variance (s^2) provides an accurate and unbiased estimate of the population variance (σ^2). Therefore, the most accurate way to estimate the standard error is to use the sample variance to estimate the population variance.
2. In future chapters we encounter other versions of the t statistic that require variance (instead of standard deviation) in the formulas for estimated standard error. To maximize the similarity from one version to another, we use variance in the formula for *all* of the different t statistics. Thus, whenever we present a t statistic, the estimated standard error is computed as

$$\text{estimated standard error} = \sqrt{\frac{\text{sample variance}}{\text{sample size}}}$$

Now we can substitute the estimated standard error in the denominator of the z -score formula. The result is a new test statistic called a t statistic:

$$t = \frac{M - \mu}{s_M} \quad (9.2)$$

DEFINITION

The **t statistic** is used to test hypotheses about an unknown population mean, μ , when the value of σ is unknown. The formula for the t statistic has the same structure as the z -score formula, except that the t statistic uses the estimated standard error in the denominator.

The only difference between the t formula and the z -score formula is that the z -score uses the actual population variance, σ^2 (or the standard deviation), and the

t formula uses the corresponding sample variance (or standard deviation) when the population value is not known.

$$z = \frac{M - \mu}{\sigma_M} = \frac{M - \mu}{\sqrt{\sigma^2/n}} \quad t = \frac{M - \mu}{s_M} = \frac{M - \mu}{\sqrt{s^2/n}}$$

DEGREES OF FREEDOM AND THE t STATISTIC

In this chapter, we have introduced the t statistic as a substitute for a z -score. The basic difference between these two is that the t statistic uses sample variance (s^2) and the z -score uses the population variance (σ^2). To determine how well a t statistic approximates a z -score, we must determine how well the sample variance approximates the population variance.

In Chapter 4, we introduced the concept of degrees of freedom (p. 117). Reviewing briefly, you must know the sample mean before you can compute sample variance. This places a restriction on sample variability such that only $n - 1$ scores in a sample are independent and free to vary. The value $n - 1$ is called the *degrees of freedom* (or df) for the sample variance.

$$\text{degrees of freedom} = df = n - 1 \quad (9.3)$$

DEFINITION

Degrees of freedom describe the number of scores in a sample that are independent and free to vary. Because the sample mean places a restriction on the value of one score in the sample, there are $n - 1$ degrees of freedom for a sample with n scores (see Chapter 4).

The greater the value of df for a sample, the better the sample variance, s^2 , represents the population variance, σ^2 , and the better the t statistic approximates the z -score. This should make sense because the larger the sample (n) is, the better the sample represents its population. Thus, the degrees of freedom associated with s^2 also describe how well t represents z .

THE t DISTRIBUTION

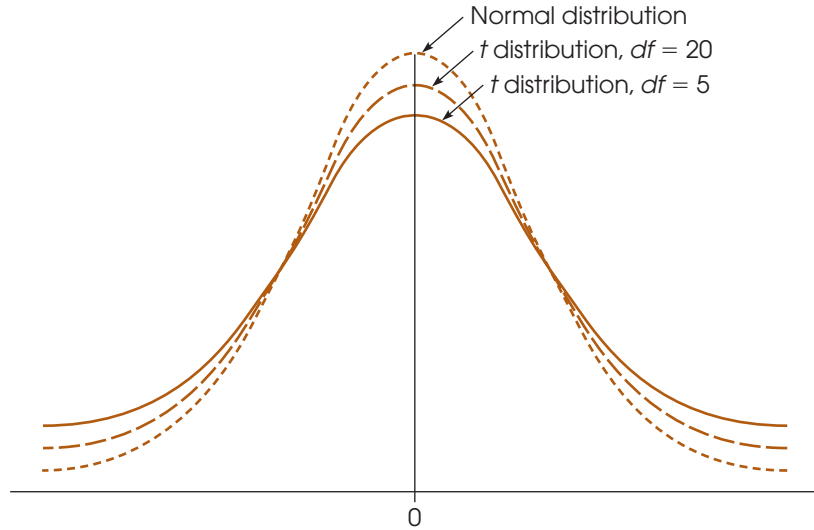
Every sample from a population can be used to compute a z -score or a t statistic. If you select all of the possible samples of a particular size (n), and compute the z -score for each sample mean, then the entire set of z -scores form a z -score distribution. In the same way, you can compute the t statistic for every sample and the entire set of t values form a t distribution. As we saw in Chapter 7, the distribution of z -scores for sample means tends to be a normal distribution. Specifically, if the sample size is large (around $n = 30$ or more) or if the sample is selected from a normal population, then the distribution of sample means is a nearly perfect normal distribution. In these same situations, the t distribution approximates a normal distribution, just as a t statistic approximates a z -score. How well a t distribution approximates a normal distributor is determined by degrees of freedom. In general, the greater the sample size (n) is, the larger the degrees of freedom ($n - 1$) are, and the better the t distribution approximates the normal distribution (Figure 9.1).

DEFINITION

A **t distribution** is the complete set of t values computed for every possible random sample for a specific sample size (n) or a specific degrees of freedom (df). The t distribution approximates the shape of a normal distribution.

FIGURE 9.1

Distributions of the t statistic for different values of degrees of freedom are compared to a normal z -score distribution. Like the normal distribution, t distributions are bell-shaped and symmetrical and have a mean of zero. However, t distributions have more variability, indicated by the flatter and more spread-out shape. The larger the value of df is, the more closely the t distribution approximates a normal distribution.



THE SHAPE OF THE t DISTRIBUTION

The exact shape of a t distribution changes with degrees of freedom. In fact, statisticians speak of a “family” of t distributions. That is, there is a different sampling distribution of t (a distribution of all possible sample t values) for each possible number of degrees of freedom. As df gets very large, the t distribution gets closer in shape to a normal z -score distribution. A quick glance at Figure 9.1 reveals that distributions of t are bell-shaped and symmetrical and have a mean of zero. However, the t distribution has more variability than a normal z distribution, especially when df values are small (see Figure 9.1). The t distribution tends to be flatter and more spread out, whereas the normal z distribution has more of a central peak.

The reason that the t distribution is flatter and more variable than the normal z -score distribution becomes clear if you look at the structure of the formulas for z and t . For both formulas, z and t , the top of the formula, $M - \mu$, can take on different values because the sample mean (M) varies from one sample to another. For z -scores, however, the bottom of the formula does not vary, provided that all of the samples are the same size and are selected from the same population. Specifically, all of the z -scores have the same standard error in the denominator, $\sigma_M = \sqrt{\sigma^2/n}$, because the population variance and the sample size are the same for every sample. For t statistics, on the other hand, the bottom of the formula varies from one sample to another. Specifically, the sample variance (s^2) changes from one sample to the next, so the estimated standard error also varies, $s_M = \sqrt{s^2/n}$. Thus, only the numerator of the z -score formula varies, but both the numerator and the denominator of the t statistic vary. As a result, t statistics are more variable than are z -scores, and the t distribution is flatter and more spread out. As sample size and df increase, however, the variability in the t distribution decreases, and it more closely resembles a normal distribution.

DETERMINING PROPORTIONS AND PROBABILITIES FOR t DISTRIBUTIONS

Just as we used the unit normal table to locate proportions associated with z -scores, we use a t distribution table to find proportions for t statistics. The complete t distribution table is presented in Appendix B, page 703, and a portion of this table is reproduced in

Table 9.1. The two rows at the top of the table show proportions of the t distribution contained in either one or two tails, depending on which row is used. The first column of the table lists degrees of freedom for the t statistic. Finally, the numbers in the body of the table are the t values that mark the boundary between the tails and the rest of the t distribution.

For example, with $df = 3$, exactly 5% of the t distribution is located in the tail beyond $t = 2.353$ (Figure 9.2). The process of finding this value is highlighted in Table 9.1. Begin by locating $df = 3$ in the first column of the table. Then locate a proportion of 0.05 (5%) in the one-tail proportion row. When you line up these two values in the table, you should find $t = 2.353$. Similarly, 5% of the t distribution is located in the tail beyond $t = -2.353$ (see Figure 9.2). Finally, notice that a total of 10% (or 0.10) is contained in the two tails beyond $t = \pm 2.353$ (check the proportion value in the “two-tails combined” row at the top of the table).

A close inspection of the t distribution table in Appendix B demonstrates a point we made earlier: As the value for df increases, the t distribution becomes more similar to a normal distribution. For example, examine the column containing t values for a 0.05 proportion in two tails. You will find that when $df = 1$, the t values that separate the extreme 5% (0.05) from the rest of the distribution are $t = \pm 12.706$. As

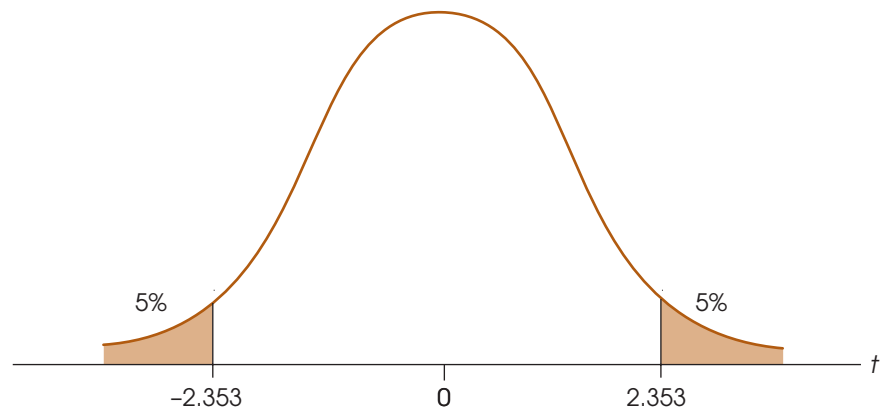
TABLE 9.1

A portion of the t -distribution table. The numbers in the table are the values of t that separate the tail from the main body of the distribution. Proportions for one or two tails are listed at the top of the table, and df values for t are listed in the first column.

	Proportion in One Tail					
	0.25	0.10	0.05	0.025	0.01	0.005
df	Proportion in Two Tails Combined					
	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707

FIGURE 9.2

The t distribution with $df = 3$. Note that 5% of the distribution is located in the tail beyond $t = 2.353$. Also, 5% is in the tail beyond $t = -2.353$. Thus, a total proportion of 10% (0.10) is in the two tails beyond $t = \pm 2.353$.



you read down the column, however, you will find that the critical t values become smaller and smaller, ultimately reaching ± 1.96 . You should recognize ± 1.96 as the z -score values that separate the extreme 5% in a normal distribution. Thus, as df increases, the proportions in a t distribution become more like the proportions in a normal distribution. When the sample size (and degrees of freedom) is sufficiently large, the difference between a t distribution and the normal distribution becomes negligible.

Caution: The t distribution table printed in this book has been abridged and does not include entries for every possible df value. For example, the table lists t values for $df = 40$ and for $df = 60$, but does not list any entries for df values between 40 and 60. Occasionally, you will encounter a situation in which your t statistic has a df value that is not listed in the table. In these situations, you should look up the critical t for both of the surrounding df values listed and then use the *larger* value for t . If, for example, you have $df = 53$ (not listed), look up the critical t value for both $df = 40$ and $df = 60$ and *then use the larger t value*. If your sample t statistic is greater than the larger value listed, then you can be certain that the data are in the critical region, and you can confidently reject the null hypothesis.

LEARNING CHECK

- Under what circumstances is a t statistic used instead of a z -score for a hypothesis test?
- A sample of $n = 9$ scores has $SS = 288$.
 - Compute the variance for the sample.
 - Compute the estimated standard error for the sample mean.
- In general, a distribution of t statistics is flatter and more spread out than the standard normal distribution. (True or false?)
- A researcher reports a t statistic with $df = 20$. How many individuals participated in the study?
- For $df = 15$, find the value(s) of t associated with each of the following:
 - The top 5% of the distribution.
 - The middle 95% of the distribution.
 - The middle 99% of the distribution.

ANSWERS

- A t statistic is used instead of a z -score when the population standard deviation and variance are not known.
- $s^2 = 36$
 - $s_M = 2$
- True.
- $n = 21$
- $t = +1.753$
 - $t = \pm 2.131$
 - $t = \pm 2.947$

9.2 HYPOTHESIS TESTS WITH THE t STATISTIC

In the hypothesis-testing situation, we begin with a population with an unknown mean and an unknown variance, often a population that has received some treatment (Figure 9.3). The goal is to use a sample from the treated population (a treated sample) as the basis for determining whether the treatment has any effect.

As always, the null hypothesis states that the treatment has no effect; specifically, H_0 states that the population mean is unchanged. Thus, the null hypothesis provides a specific value for the unknown population mean. The sample data provide a value for the sample mean. Finally, the variance and estimated standard error are computed from the sample data. When these values are used in the t formula, the result becomes

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

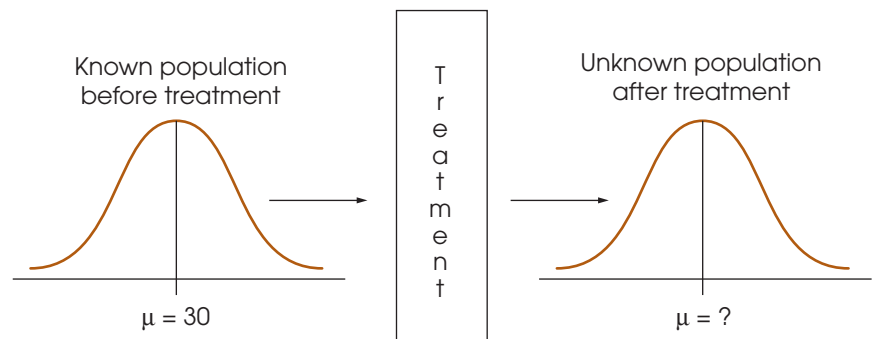
(from the data) (hypothesized from H_0)
(computed from the sample data)

As with the z -score formula, the t statistic forms a ratio. The numerator measures the actual difference between the sample data (M) and the population hypothesis (μ). The estimated standard error in the denominator measures how much difference is reasonable to expect between a sample mean and the population mean. When the obtained difference between the data and the hypothesis (numerator) is much greater than expected (denominator), we obtain a large value for t (either large positive or large negative). In this case, we conclude that the data are not consistent with the hypothesis, and our decision is to “reject H_0 .” On the other hand, when the difference between the data and the hypothesis is small relative to the standard error, we obtain a t statistic near zero, and our decision is “fail to reject H_0 .”

The Unknown Population As mentioned earlier, the hypothesis test often concerns a population that has received a treatment. This situation is shown in Figure 9.3. Note

FIGURE 9.3

The basic experimental situation for using the t statistic or the z -score is presented. It is assumed that the parameter μ is known for the population before treatment. The purpose of the experiment is to determine whether the treatment has an effect. Note that the population after treatment has unknown values for the mean and the variance. We will use a sample to test a hypothesis about the population mean.



that the value of the mean is known for the population before treatment. The question is whether the treatment influences the scores and causes the mean to change. In this case, the unknown population is the one that exists after the treatment is administered, and the null hypothesis simply states that the value of the mean is not changed by the treatment.

Although the t statistic can be used in the “before and after” type of research shown in Figure 9.3, it also permits hypothesis testing in situations for which you do not have a known population mean to serve as a standard. Specifically, the t test does not require any prior knowledge about the population mean or the population variance. All you need to compute a t statistic is a null hypothesis and a sample from the unknown population. Thus, a t test can be used in situations for which the null hypothesis is obtained from a theory, a logical prediction, or just wishful thinking. For example, many surveys contain rating-scale questions to determine how people feel about controversial issues. Participants are presented with a statement and asked to express their opinion on a scale from 1 to 7, with 1 indicating “strongly agree” and 7 indicating “strongly disagree.” A score of 4 indicates a neutral position, with no strong opinion one way or the other. In this situation, the null hypothesis would state that there is no preference in the population, $H_0: \mu = 4$. The data from a sample is then used to evaluate the hypothesis. Note that the researcher has no prior knowledge about the population mean and states a hypothesis that is based on logic.

HYPOTHESIS TESTING EXAMPLE

The following research situation demonstrates the procedures of hypothesis testing with the t statistic. Note that this is another example of a null hypothesis that is founded in logic rather than prior knowledge of a population mean.

EXAMPLE 9.1

Infants, even newborns, prefer to look at attractive faces compared to less attractive faces (Slater, et al., 1998). In the study, infants from 1 to 6 days old were shown two photographs of women’s faces. Previously, a group of adults had rated one of the faces as significantly more attractive than the other. The babies were positioned in front of a screen on which the photographs were presented. The pair of faces remained on the screen until the baby accumulated a total of 20 seconds of looking at one or the other. The number of seconds looking at the attractive face was recorded for each infant. Suppose that the study used a sample of $n = 9$ infants and the data produced an average of $M = 13$ seconds for the attractive face with $SS = 72$. Note that all of the available information comes from the sample. Specifically, we do not know the population mean or the population standard deviation.

STEP 1 State the hypotheses and select an alpha level. Although we have no information about the population of scores, it is possible to form a logical hypothesis about the value of μ . In this case, the null hypothesis states that the infants have no preference for either face. That is, they should average half of the 20 seconds looking at each of the two faces. In symbols, the null hypothesis states

$$H_0: \mu_{\text{attractive}} = 10 \text{ seconds}$$

The alternative hypothesis states that there is a preference and one of the faces is preferred over the other. A directional, one-tailed test would specify which of the two faces is preferred, but the nondirectional alternative hypothesis is expressed as follows:

$$H_1: \mu_{\text{attractive}} \neq 10 \text{ seconds}$$

We set the level of significance at $\alpha = .05$ for two tails.

- STEP 2** Locate the critical region. The test statistic is a t statistic because the population variance is not known. Therefore, the value for degrees of freedom must be determined before the critical region can be located. For this sample

$$df = n - 1 = 9 - 1 = 8$$

For a two-tailed test at the .05 level of significance and with 8 degrees of freedom, the critical region consists of t values greater than +2.306 or less than -2.306. Figure 9.4 depicts the critical region in this t distribution.

- STEP 3** Calculate the test statistic. The t statistic typically requires much more computation than is necessary for a z -score. Therefore, we recommend that you divide the calculations into a three-stage process as follows:

- a. First, calculate the sample variance. Remember that the population variance is unknown, and you must use the sample value in its place. (This is why we are using a t statistic instead of a z -score.)

$$s^2 = \frac{SS}{n-1} = \frac{SS}{df} = \frac{72}{8} = 9$$

- b. Next, use the sample variance (s^2) and the sample size (n) to compute the estimated standard error. This value is the denominator of the t statistic and measures how much difference is reasonable to expect by chance between a sample mean and the corresponding population mean.

$$s_M = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{9}{9}} = \sqrt{1} = 1$$

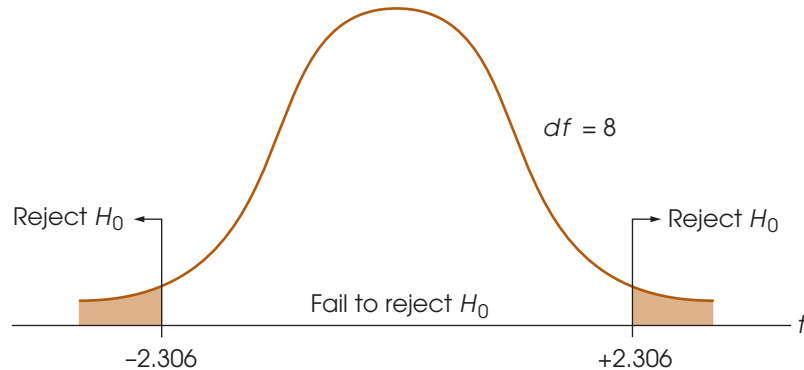
Finally, compute the t statistic for the sample data.

$$t = \frac{M - \mu}{s_M} = \frac{13 - 10}{1} = 3.00$$

- STEP 4** Make a decision regarding H_0 . The obtained t statistic of 3.00 falls into the critical region on the right-hand side of the t distribution (see Figure 9.4). Our statistical decision is to reject H_0 and conclude that babies do show a preference when given a

FIGURE 9.4

The critical region in the t distribution for $\alpha = .05$ and $df = 8$.



choice between an attractive and an unattractive face. Specifically, the average amount of time that the babies spent looking at the attractive face was significantly different from the 10 seconds that would be expected if there were no preference. As indicated by the sample mean, there is a tendency for the babies to spend more time looking at the attractive face.

ASSUMPTIONS OF THE t TEST

Two basic assumptions are necessary for hypothesis tests with the t statistic.

1. The values in the sample must consist of *independent* observations.

In everyday terms, two observations are independent if there is no consistent, predictable relationship between the first observation and the second. More precisely, two events (or observations) are independent if the occurrence of the first event has no effect on the probability of the second event. We examined specific examples of independence and non-independence in Box 8.1 (p. 254).

2. The population that is sampled must be normal.

This assumption is a necessary part of the mathematics underlying the development of the t statistic and the t distribution table. However, violating this assumption has little practical effect on the results obtained for a t statistic, especially when the sample size is relatively large. With very small samples, a normal population distribution is important. With larger samples, this assumption can be violated without affecting the validity of the hypothesis test. If you have reason to suspect that the population distribution is not normal, use a large sample to be safe.

THE INFLUENCE OF SAMPLE SIZE AND SAMPLE VARIANCE

As we noted in Chapter 8 (p. 252), a variety of factors can influence the outcome of a hypothesis test. In particular, the number of scores in the sample and the magnitude of the sample variance both have a large effect on the t statistic and thereby influence the statistical decision. The structure of the t formula makes these factors easier to understand.

$$t = \frac{M - \mu}{s_M} \quad \text{where } s_M = \sqrt{\frac{s^2}{n}}$$

Because the estimated standard error, s_M , appears in the denominator of the formula, a larger value for s_M produces a smaller value (closer to zero) for t . Thus, any factor that influences the standard error also affects the likelihood of rejecting the null hypothesis and finding a significant treatment effect. The two factors that determine the size of the standard error are the sample variance, s^2 , and the sample size, n .

The estimated standard error is directly related to the sample variance so that the larger the variance, the larger the error. Thus, large variance means that you are less likely to obtain a significant treatment effect. In general, large variance is bad for inferential statistics. Large variance means that the scores are widely scattered, which makes it difficult to see any consistent patterns or trends in the data. In general, high variance reduces the likelihood of rejecting the null hypothesis.

On the other hand, the estimated standard error is inversely related to the number of scores in the sample. The larger the sample is, the smaller the error is. If all other factors are held constant, large samples tend to produce bigger t statistics and therefore are more likely to produce significant results. For example, a 2-point mean difference with

a sample of $n = 4$ may not be convincing evidence of a treatment effect. However, the same 2-point difference with a sample of $n = 100$ is much more compelling.

LEARNING CHECK

1. A sample of $n = 4$ individuals is selected from a population with a mean of $\mu = 40$. A treatment is administered to the individuals in the sample and, after treatment, the sample has a mean of $M = 44$ and a variance of $s^2 = 16$.
 - a. Is this sample sufficient to conclude that the treatment has a significant effect? Use a two-tailed test with $\alpha = .05$.
 - b. If all other factors are held constant and the sample size is increased to $n = 16$, is the sample sufficient to conclude that the treatment has a significant effect? Again, use a two-tailed test with $\alpha = .05$.

- ANSWER**
1. a. $H_0: \mu = 40$ even after the treatment. With $n = 4$, the estimated standard error is 2, and $t = 4/2 = 2.00$. With $df = 3$, the critical boundaries are set at $t = \pm 3.182$. Fail to reject H_0 and conclude that the treatment does not have a significant effect.
 - b. With $n = 16$, the estimated standard error is 1 and $t = 4.00$. With $df = 15$, the critical boundary is ± 2.131 . The t value is beyond the critical boundary, so we reject H_0 and conclude that the treatment does have a significant effect.

9.3 MEASURING EFFECT SIZE FOR THE t STATISTIC

In Chapter 8, we noted that one criticism of a hypothesis test is that it does not really evaluate the size of the treatment effect. Instead, a hypothesis test simply determines whether the treatment effect is greater than chance, where “chance” is measured by the standard error. In particular, it is possible for a very small treatment effect to be “statistically significant,” especially when the sample size is very large. To correct for this problem, it is recommended that the results from a hypothesis test be accompanied by a report of effect size, such as Cohen’s d .

ESTIMATED COHEN’S d When Cohen’s d was originally introduced (p. 262), the formula was presented as

$$\text{Cohen's } d = \frac{\text{mean difference}}{\text{standard deviation}} = \frac{\mu_{\text{treatment}} - \mu_{\text{no treatment}}}{\sigma}$$

Cohen defined this measure of effect size in terms of the population mean difference and the population standard deviation. However, in most situations the population values are not known and you must substitute the corresponding sample values in their place. When this is done, many researchers prefer to identify the calculated value as an “*estimated d*” or name the value after one of the statisticians who first substituted sample statistics into Cohen’s formula (e.g., Glass’s g or Hedges’s g). For hypothesis tests using the t statistic, the population mean with no treatment is the value specified by the null hypothesis. However, the population mean with treatment and the standard deviation are both unknown. Therefore, we use the mean for the treated sample and the standard deviation for the sample after treatment as

estimates of the unknown parameters. With these substitutions, the formula for estimating Cohen's d becomes

$$\text{estimated } d = \frac{\text{mean difference}}{\text{sample standard deviation}} = \frac{M - \mu}{s} \quad (9.4)$$

The numerator measures that magnitude of the treatment effect by finding the difference between the mean for the treated sample and the mean for the untreated population (μ from H_0). The sample standard deviation in the denominator standardizes the mean difference into standard deviation units. Thus, an estimated d of 1.00 indicates that the size of the treatment effect is equivalent to one standard deviation. The following example demonstrates how the estimated d is used to measure effect size for a hypothesis test using a t statistic.

EXAMPLE 9.2

For the infant face-preference study in Example 9.1, the babies averaged $M = 13$ out of 20 seconds looking at the attractive face. If there were no preference (as stated by the null hypothesis), the population mean would be $\mu = 10$ seconds. Thus, the results show a 3-second difference between the mean with a preference ($M = 13$) and the mean with no preference ($\mu = 10$). Also, for this study the sample standard deviation is

$$s = \sqrt{\frac{SS}{df}} = \sqrt{\frac{72}{8}} = \sqrt{9} = 3$$

Thus, Cohen's d for this example is estimated to be

$$\text{Cohen's } d = \frac{M - \mu}{s} = \frac{13 - 10}{3} = 1.00$$

According to the standards suggested by Cohen (Table 8.2, p. 264), this is a large treatment effect.

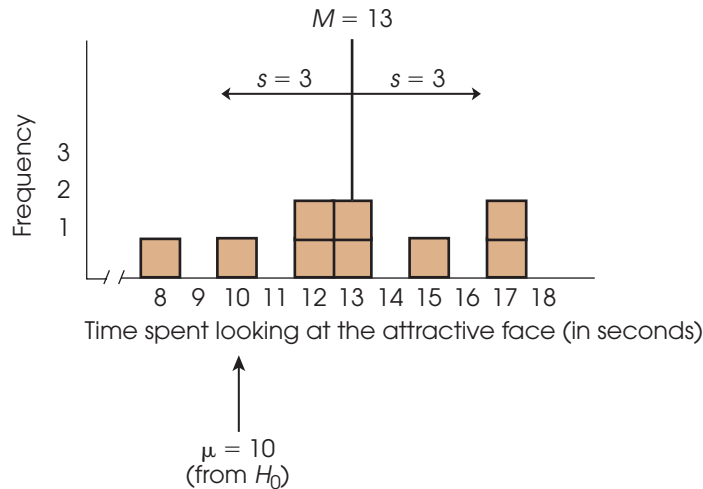
To help you visualize what is measured by Cohen's d , we have constructed a set of $n = 9$ scores with a mean of $M = 13$ and a standard deviation of $s = 3$ (the same values as in Examples 9.1 and 9.2). The set of scores is shown in Figure 9.5. Notice that the figure also includes an arrow that locates $\mu = 10$. Recall that $\mu = 10$ is the value specified by the null hypothesis and identifies what the mean ought to be if the treatment has no effect. Clearly, our sample is not centered around $\mu = 10$. Instead, the scores have been shifted to the right so that the sample mean is $M = 13$. This shift, from 10 to 13, is the 3-point mean difference that was caused by the treatment effect. Also notice that the 3-point mean difference is exactly equal to the standard deviation. Thus, the size of the treatment effect is equal to 1 standard deviation. In other words, Cohen's $d = 1.00$.

MEASURING THE PERCENTAGE OF VARIANCE EXPLAINED, r^2

An alternative method for measuring effect size is to determine how much of the variability in the scores is explained by the treatment effect. The concept behind this measure is that the treatment causes the scores to increase (or decrease), which means that the treatment is causing the scores to vary. If we can measure how much of the variability is explained by the treatment, we can obtain a measure of the size of the treatment effect.

FIGURE 9.5

The sample distribution for the scores that were used in Examples 9.1 and 9.2. The population mean, $\mu = 10$ seconds, is the value that would be expected if attractiveness has no effect on the infants' behavior. Note that the sample mean is displaced away from $\mu = 10$ by a distance equal to one standard deviation.



To demonstrate this concept we use the data from the hypothesis test in Example 9.1. Recall that the null hypothesis stated that the treatment (the attractiveness of the faces) has no effect on the infants' behavior. According to the null hypothesis, the infants should show no preference between the two photographs, and therefore should spend an average of $\mu = 10$ out of 20 seconds looking at the attractive face.

However, if you look at the data in Figure 9.5, the scores are not centered around $\mu = 10$. Instead, the scores are shifted to the right so that they are centered around the sample mean, $M = 13$. This shift is the treatment effect. To measure the size of the treatment effect, we calculate deviations from the mean and the sum of squared deviations, SS , in two different ways.

Figure 9.6(a) shows the original set of scores. For each score, the deviation from $\mu = 10$ is shown as a colored line. Recall that $\mu = 10$ comes from the null hypothesis and represents the population mean if the treatment has no effect. Note that almost all of the scores are located on the right-hand side of $\mu = 10$. This shift to the right is the treatment effect. Specifically, the preference for the attractive face has caused the infants to spend more time looking at the attractive photograph, which means that their scores are generally greater than 10. Thus, the treatment has pushed the scores away from $\mu = 10$ and has increased the size of the deviations.

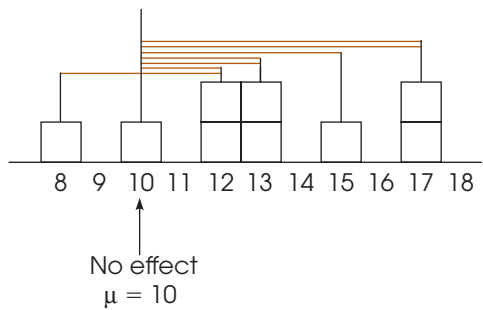
Next, we see what happens if the treatment effect is removed. In this example, the treatment has a 3-point effect (the average increases from $\mu = 10$ to $M = 13$). To remove the treatment effect, we simply subtract 3 points from each score. The adjusted scores are shown in Figure 9.6(b) and, once again, the deviations from $\mu = 10$ are shown as colored lines. First, notice that the adjusted scores are centered at $\mu = 10$, indicating that there is no treatment effect. Also notice that the deviations, the colored lines, are noticeably smaller when the treatment effect is removed.

To measure how much the variability is reduced when the treatment effect is removed, we compute the sum of squared deviations, SS , for each set of scores. The left-hand columns of Table 9.2 show the calculations for the original scores [Figure 9.6(a)], and the right-hand columns show the calculations for the adjusted scores [Figure 9.6(b)]. Note that the total variability, including the treatment effect, is $SS = 153$. However, when the treatment effect is removed, the variability is reduced to $SS = 72$. The difference

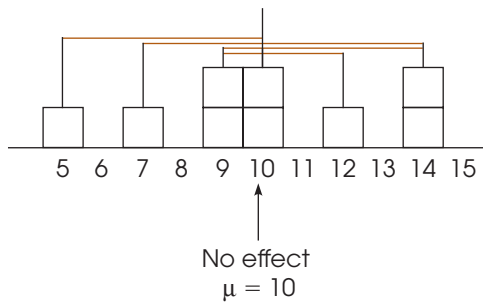
FIGURE 9.6

Deviations from $\mu = 10$ (no treatment effect) for the scores in Example 9.1. The colored lines in part (a) show the deviations for the original scores, including the treatment effect. In part (b) the colored lines show the deviations for the adjusted scores after the treatment effect has been removed.

(a) Original scores, including the treatment effect



(b) Adjusted scores with the treatment effect removed



between these two values, $153 - 72 = 81$ points, is the amount of variability that is accounted for by the treatment effect. This value is usually reported as a proportion or percentage of the total variability:

$$\frac{\text{variability accounted for}}{\text{total variability}} = \frac{81}{153} = 0.5294 \quad (52.94\%)$$

TABLE 9.2

Calculation of SS , the sum of squared deviations, for the data in Figure 9.6. The first three columns show the calculations for the original scores, including the treatment effect. The last three columns show the calculations for the adjusted scores after the treatment effect has been removed.

Calculation of SS including the treatment effect			Calculation of SS after the treatment effect is removed		
Score	Deviation from $\mu = 10$	Squared Deviation	Adjusted Score	Deviation from $\mu = 10$	Squared Deviation
8	-2	4	$8 - 3 = 5$	-5	25
10	0	0	$10 - 3 = 7$	-3	9
12	2	4	$12 - 3 = 9$	-1	1
12	2	4	$12 - 3 = 9$	-1	1
13	3	9	$13 - 3 = 10$	0	0
13	3	9	$13 - 3 = 10$	0	0
15	5	25	$15 - 3 = 12$	2	4
17	7	49	$17 - 3 = 14$	4	16
17	7	49	$17 - 3 = 14$	4	16
$SS = 153$			$SS = 72$		

Thus, removing the treatment effect reduces the variability by 52.94%. This value is called the *percentage of variance accounted for by the treatment* and is identified as r^2 .

Rather than computing r^2 directly by comparing two different calculations for SS , the value can be found from a single equation based on the outcome of the t test.

$$r^2 = \frac{t^2}{t^2 + df} \quad (9.5)$$

The letter r is the traditional symbol used for a correlation, and the concept of r^2 is discussed again when we consider correlations in Chapter 15. Also, in the context of t statistics, the percentage of variance that we are calling r^2 is often identified by the Greek letter omega squared (ω^2).

For the hypothesis test in Example 9.1, we obtained $t = 3.00$ with $df = 8$. These values produce

$$r^2 = \frac{3^2}{3^2 + 8} = \frac{9}{17} = 0.5294 \quad (52.94\%)$$

Note that this is exactly the same value we obtained with the direct calculation of the percentage of variability accounted for by the treatment.

Interpreting r^2 In addition to developing the Cohen's d measure of effect size, Cohen (1988) also proposed criteria for evaluating the size of a treatment effect that is measured by r^2 . The criteria were actually suggested for evaluating the size of a correlation, r , but are easily extended to apply to r^2 . Cohen's standards for interpreting r^2 are shown in Table 9.3.

According to these standards, the data we constructed for Examples 9.1 and 9.2 show a very large effect size with $r^2 = .5294$.

As a final note, we should remind you that, although sample size affects the hypothesis test, this factor has little or no effect on measures of effect size. In particular, estimates of Cohen's d are not influenced at all by sample size, and measures of r^2 are only slightly affected by changes in the size of the sample. The sample variance, on the other hand, influences hypothesis tests and measures of effect size. Specifically, high variance reduces both the likelihood of rejecting the null hypothesis and measures of effect size.

CONFIDENCE INTERVALS FOR ESTIMATING μ

An alternative technique for describing the size of a treatment effect is to compute an estimate of the population mean after treatment. For example, if the mean before treatment is known to be $\mu = 80$ and the mean after treatment is estimated to be $\mu = 86$, then we can conclude that the size of the treatment effect is around 6 points.

Estimating an unknown population mean involves constructing a *confidence interval*. A confidence interval is based on the observation that a sample mean tends to provide a reasonably accurate estimate of the population mean. The fact that a sample

TABLE 9.3

Criteria for interpreting the value of r^2 as proposed by Cohen (1988).

Percentage of Variance Explained, r^2	
$r^2 = 0.01$	Small effect
$r^2 = 0.09$	Medium effect
$r^2 = 0.25$	Large effect

mean tends to be near to the population mean implies that the population mean should be near to the sample mean. For example, if we obtain a sample mean of $M = 86$, we can be reasonably confident that the population mean is around 86. Thus, a confidence interval consists of an interval of values around a sample mean, and we can be reasonably confident that the unknown population mean is located somewhere in the interval.

DEFINITION

A **confidence interval** is an interval, or range of values, centered around a sample statistic. The logic behind a confidence interval is that a sample statistic, such as a sample mean, should be relatively near to the corresponding population parameter. Therefore, we can confidently estimate that the value of the parameter should be located in the interval.

CONSTRUCTING A CONFIDENCE INTERVAL

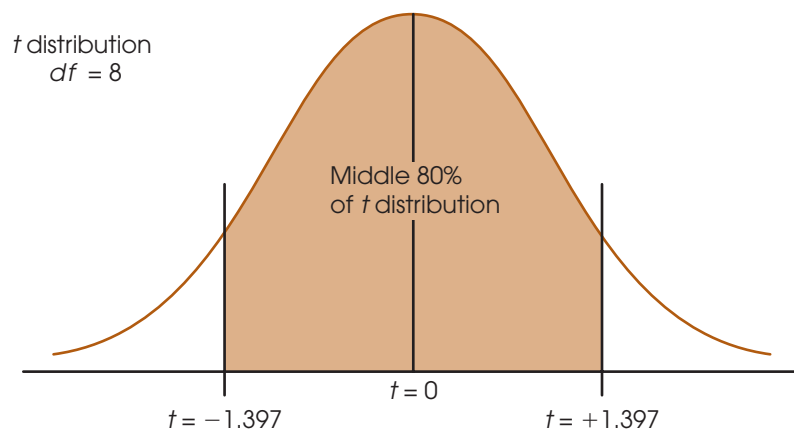
The construction of a confidence interval begins with the observation that every sample mean has a corresponding t value defined by the equation

$$t = \frac{M - \mu}{s_M}$$

Although the values for M and s_M are available from the sample data, we cannot use the equation to calculate t because the value for μ is unknown. Instead of calculating the t value, for a confidence interval we *estimate* the t value. For example, if the sample has $n = 9$ scores, then the t statistic has $df = 8$, and the distribution of all possible t values can be pictured as in Figure 9.7. Notice that the t values pile up around $t = 0$, so we can estimate that the t value for our sample should have a value around 0. Furthermore, the t distribution table lists a variety of different t values that correspond to specific proportions of the t distribution. With $df = 8$, for example, 80% of the t values are located between $t = +1.397$ and $t = -1.397$. To obtain these values, simply look up a two-tailed proportion of 0.20 (20%) for $df = 8$. Because 80% of all of the possible t values are located between ± 1.397 , we can be 80% confident that our sample mean corresponds to a t value in this interval. Similarly, we can be 95% confident that the mean for a sample of $n = 9$ scores corresponds to a t value between $+2.306$ and -2.306 . Notice that we are able to estimate the value of t with a specific level of

FIGURE 9.7

The distribution of t statistics with $df = 8$. The t values pile up around $t = 0$ and 80% of all of the possible values are located between $t = -1.397$ and $t = +1.397$.



confidence. To construct a confidence interval for μ , we plug the estimated t value into the t equation, and then we can calculate the value of μ .

Before we demonstrate the process of constructing a confidence interval for an unknown population mean, we simplify the calculations by regrouping the terms in the t equation. Because the goal is to compute the value of μ , we use simple algebra to solve the equation for μ . The result is

$$\mu = M \pm ts_M \quad (9.6)$$

The process of using this equation to construct a confidence interval is demonstrated in the following example.

EXAMPLE 9.3

Example 9.1 describes a study in which infants displayed a preference for the more attractive face by looking at it, instead of the less attractive face, for the majority of a 20-second viewing period. Specifically, a sample of $n = 9$ infants spent an average of $M = 13$ seconds out of a 20-second period looking at the more attractive face. The data produced an estimated standard error of $s_M = 1$. We use this sample to construct a confidence interval to estimate the mean amount of time that the population of infants spends looking at the more attractive face. That is, we construct an interval of values that is likely to contain the unknown population mean.

Again, the estimation formula is

$$\mu = M \pm t(s_M)$$

In the equation, the value of $M = 13$ and $s_M = 1$ are obtained from the sample data. The next step is to select a level of confidence that determines the value of t in the equation. The most commonly used confidence level is probably 95%, but values of 80%, 90%, and 99% are also common. For this example, we use a confidence level of 80%, which means that we construct the confidence interval so that we are 80% confident that the population mean is actually contained in the interval. Because we are using a confidence level of 80%, the resulting interval is called the *80% confidence interval for μ* .

To obtain the value for t in the equation, we simply estimate that the t statistic for our sample is located somewhere in the middle 80% of the t distribution. With $df = n - 1 = 8$, the middle 80% of the distribution is bounded by t values of $+1.397$ and -1.397 (see Figure 9.7). Using the sample data and the estimated range of t values, we obtain

$$\mu = M \pm t(s_M) = 13 \pm 1.397(1.00) = 13 \pm 1.397$$

At one end of the interval, we obtain $\mu = 13 + 1.397 = 14.397$, and at the other end we obtain $\mu = 13 - 1.397 = 11.603$. Our conclusion is that the average time looking at the more attractive fact for the population of infants is between $\mu = 11.603$ seconds and $\mu = 14.397$ seconds, and we are 80% confident that the true population mean is located within this interval. The confidence comes from the fact that the calculation was based on only one assumption. Specifically, we assumed that the t statistic was located between $+1.397$ and -1.397 , and we are 80% confident that this assumption is correct because 80% of all of the possible t values are located in this interval. Finally, note that the confidence interval is constructed around the sample mean. As a result, the sample mean, $M = 13$, is located exactly in the center of the interval.

To have 80% in the middle, there must be 20% (or .20) in the tails. To find the t values, look under two tails, .20 in the t table.

FACTORS AFFECTING THE WIDTH OF A CONFIDENCE INTERVAL

Two characteristics of the confidence interval should be noted. First, notice what happens to the width of the interval when you change the level of confidence (the percent confidence). To gain more confidence in your estimate, you must increase the width of the interval. Conversely, to have a smaller, more precise interval, you must give up confidence. In the estimation formula, the percentage of confidence influences the value of t . A larger level of confidence (the percentage), produces a larger t value and a wider interval. This relationship can be seen in Figure 9.7. In the figure, we identified the middle 80% of the t distribution to find an 80% confidence interval. It should be obvious that if we were to increase the confidence level to 95%, it would be necessary to increase the range of t values, and thereby increase the width of the interval.

Second, note what happens to the width of the interval if you change the sample size. This time the basic rule is as follows: The bigger the sample (n), the smaller the interval. This relationship is straightforward if you consider the sample size as a measure of the amount of information. A bigger sample gives you more information about the population and allows you to make a more precise estimate (a narrower interval). The sample size controls the magnitude of the standard error in the estimation formula. As the sample size increases, the standard error decreases, and the interval gets smaller. Because confidence intervals are influenced by sample size, they do not provide an unqualified measure of absolute effect size and are not an adequate substitute for Cohen's d or r^2 . Nonetheless, they can be used in a research report to provide a description of the size of the treatment effect.

LEARNING CHECK

1. If all other factors are held constant, an 80% confidence interval is wider than a 90% confidence interval. (True or false?)
2. If all other factors are held constant, a confidence interval computed from a sample of $n = 25$ is wider than a confidence interval computed from a sample of $n = 100$. (True or false?)

ANSWERS

1. False. Greater confidence requires a wider interval.
2. True. The smaller sample produces a wider interval.



IN THE LITERATURE REPORTING THE RESULTS OF A t TEST

In Chapter 8, we noted the conventional style for reporting the results of a hypothesis test, according to APA format. First, recall that a scientific report typically uses the term *significant* to indicate that the null hypothesis has been rejected and the term *not significant* to indicate failure to reject H_0 . Additionally, there is a prescribed format for reporting the calculated value of the test statistic, degrees of freedom, and alpha level for a t test. This format parallels the style introduced in Chapter 8 (p. 251).

In Example 9.1 we calculated a t statistic of 3.00 with $df = 8$, and we decided to reject H_0 with alpha set at .05. Using the same data from Example 9.1, we obtained $r^2 = 0.5294$ (52.94%) for the percentage of variance explained by the treatment effect. In a scientific report, this information is conveyed in a concise statement, as follows:

The infants spent an average of $M = 13$ out of 20 seconds looking at the attractive face, with $SD = 3.00$. Statistical analysis indicates that the time spent looking at the attractive face was significantly greater than would be expected if there were no preference, $t(8) = 3.00, p < .05, r^2 = 0.5294$.

The first statement reports the descriptive statistics, the mean ($M = 13$) and the standard deviation ($SD = 3$), as previously described (Chapter 4, p. 123). The next statement provides the results of the inferential statistical analysis. Note that the degrees of freedom are reported in parentheses immediately after the symbol t . The value for the obtained t statistic follows (3.00), and next is the probability of committing a Type I error (less than 5%). Finally, the effect size is reported, $r^2 = 52.94\%$. If the 80% confidence interval from Example 9.3 were included in the report as a description of effect size, it would be added after the results of the hypothesis test as follows:

$$t(8) = 3.00, p < .05, 80\% \text{ CI } [11.603, 14.397].$$

Often, researchers use a computer to perform a hypothesis test like the one in Example 9.1. In addition to calculating the mean, standard deviation, and the t statistic for the data, the computer usually calculates and reports the *exact probability* (or α level) associated with the t value. In Example 9.1 we determined that any t value beyond ± 2.306 has a probability of less than .05 (see Figure 9.4). Thus, the obtained t value, $t = 3.00$, is reported as being very unlikely, $p < .05$. A computer printout, however, would have included an exact probability for our specific t value.

Whenever a specific probability value is available, you are encouraged to use it in a research report. For example, the computer analysis of these data reports an exact p value of $p = .017$, and the research report would state “ $t(8) = 3.00$, $p = .017$ ” instead of using the less specific “ $p < .05$.” As one final caution, we note that occasionally a t value is so extreme that the computer reports $p = 0.000$. The zero value does not mean that the probability is literally zero; instead, it means that the computer has rounded off the probability value to three decimal places and obtained a result of 0.000. In this situation, you do not know the exact probability value, but you can report $p < .001$.

The statement $p < .05$ was explained in Chapter 8, page 251.

LEARNING CHECK

1. A sample of $n = 16$ individuals is selected from a population with a mean of $\mu = 80$. A treatment is administered to the sample and, after treatment, the sample mean is found to be $M = 86$ with a standard deviation of $s = 8$.
 - a. Does the sample provide sufficient evidence to conclude that the treatment has a significant effect? Test with $\alpha = .05$.
 - b. Compute Cohen's d and r^2 to measure the effect size.
 - c. Find the 95% confidence interval for the population mean after treatment.
2. How does sample size influence the outcome of a hypothesis test and measures of effect size? How does the standard deviation influence the outcome of a hypothesis test and measures of effect size?

ANSWERS

1. a. The estimated standard error is 2 points and the data produce $t = 6/2 = 3.00$. With $df = 15$, the critical values are $t = \pm 2.131$, so the decision is to reject H_0 and conclude that there is a significant treatment effect.
 - b. For these data, $d = 6/8 = 0.75$ and $r^2 = 9/24 = 0.375$ or 37.5%.
 - c. For 95% confidence and $df = 15$, use $t = \pm 2.131$. The confidence interval is $\mu = 86 \pm 2.131(2)$ and extends from 81.738 to 90.262.
2. Increasing sample size increases the likelihood of rejecting the null hypothesis but has little or no effect on measures of effect size. Increasing the sample variance reduces the likelihood of rejecting the null hypothesis and reduces measures of effect size.

9.4 DIRECTIONAL HYPOTHESES AND ONE-TAILED TESTS

As noted in Chapter 8, the nondirectional (two-tailed) test is more commonly used than the directional (one-tailed) alternative. On the other hand, a directional test may be used in some research situations, such as exploratory investigations or pilot studies or when there is *a priori* justification (for example, a theory or previous findings). The following example demonstrates a directional hypothesis test with a t statistic, using the same experimental situation presented in Example 9.1.

EXAMPLE 9.4 The research question is whether attractiveness affects the behavior of infants looking at photographs of women's faces. The researcher is expecting the infants to prefer the more attractive face. Therefore, the researcher predicts that the infants will spend more than half of the 20-second period looking at the attractive face. For this example, we use the same sample data that were used in the original hypothesis test in Example 9.1. Specifically, the researcher tested a sample of $n = 9$ infants and obtained a mean of $M = 13$ seconds looking at the attractive face with $SS = 72$.

STEP 1 State the hypotheses, and select an alpha level. With most directional tests, it is usually easier to state the hypothesis in words, including the directional prediction, and then convert the words into symbols. For this example, the researcher is predicting that attractiveness will cause the infants to increase the amount of time they spend looking at the attractive face; that is, more than half of the 20 seconds should be spent looking at the attractive face. In general, the null hypothesis states that the predicted effect will not happen. For this study, the null hypothesis states that the infants will not spend more than half of the 20 seconds looking at the attractive face. In symbols,

$$H_0: \mu_{\text{attractive}} \leq 10 \text{ seconds} \quad (\text{Not more than half of the 20 seconds looking at the attractive face})$$

Similarly, the alternative states that the treatment will work. In this case, H_1 states that the infants will spend more than half of the time looking at the attractive face. In symbols,

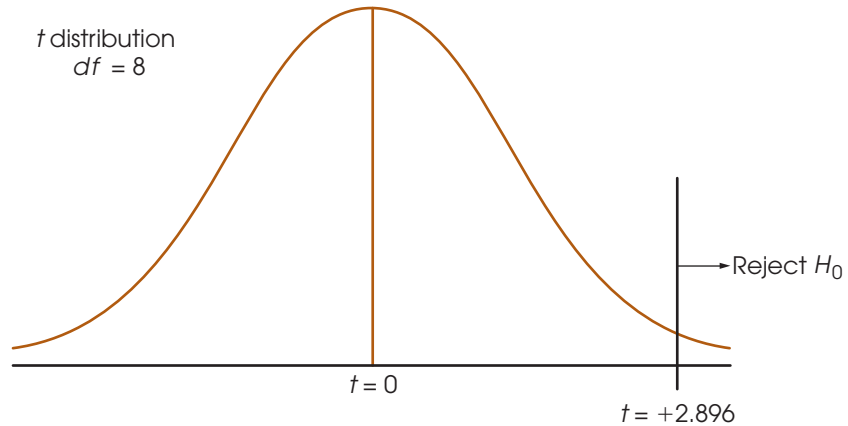
$$H_1: \mu_{\text{attractive}} > 10 \text{ seconds} \quad (\text{More than half of the 20 seconds looking at the attractive face})$$

This time, we set the level of significance at $\alpha = .01$.

STEP 2 Locate the critical region. In this example, the researcher is predicting that the sample mean (M) will be greater than 10 seconds. Thus, if the infants average more than 10 seconds looking at the attractive face, the data will provide support for the researcher's prediction and will tend to refute the null hypothesis. Also note that a sample mean greater than 10 will produce a positive value for the t statistic. Thus, the critical region for the one-tailed test will consist of positive t values located in the right-hand tail of the distribution. To find the critical value, you must look in the t distribution table using the one-tail proportions. With a sample of $n = 9$, the t statistic has $df = 8$; using $\alpha = .01$, you should find a critical value of $t = 2.896$. Therefore, if we obtain a t statistic greater than 2.896, we will reject the null hypothesis and conclude that the infants show a significant preference for the attractive face. Figure 9.8 shows the one-tailed critical region for this test.

FIGURE 9.8

The one-tailed critical region for the hypothesis test in Example 9.4 with $df = 8$ and $\alpha = .01$.



- STEP 3** Calculate the test statistic. The computation of the t statistic is the same for either a one-tailed or a two-tailed test. Earlier (in Example 9.1), we found that the data for this experiment produce a test statistic of $t = 3.00$.
- STEP 4** Make a decision. The test statistic is in the critical region, so we reject H_0 . In terms of the experimental variables, we have decided that the infants show a preference and spend significantly more time looking at the attractive face than they do looking at the unattractive face. In a research report, the results would be presented as follows:

The time spent looking at the attractive face was significantly greater than would be expected if there were no preference, $t(8) = 3.00$, $p < .01$, one tailed.

Note that the report clearly acknowledges that a one-tailed test was used.

THE CRITICAL REGION FOR A ONE-TAILED TEST

In step 2 of Example 9.4, we determined that the critical region is in the right-hand tail of the distribution. However, it is possible to divide this step into two stages that eliminate the need to determine which tail (right or left) should contain the critical region. The first stage in this process is simply to determine whether the sample mean is in the direction predicted by the original research question. For this example, the researcher predicted that the infants would prefer the attractive face and spend more time looking at it. Specifically, the researcher expects the infants to spend more than 10 out of 20 seconds focused on the attractive face. The obtained sample mean, $M = 13$ seconds, is in the correct direction. This first stage eliminates the need to determine whether the critical region is in the left- or right-hand tail. Because we already have determined that the effect is in the correct direction, the sign of the t statistic (+ or -) no longer matters. The second stage of the process is to determine whether the effect is large enough to be significant. For this example, the requirement is that the sample produces a t statistic greater than 2.896. If the magnitude of the t statistic, independent of its sign, is greater than 2.896, then the result is significant and H_0 is rejected.

LEARNING CHECK

1. A new over-the-counter cold medication includes a warning label stating that it “may cause drowsiness.” A researcher would like to evaluate this effect. It is known that under regular circumstances the distribution of reaction times is normal with $\mu = 200$. A sample of $n = 9$ participants is obtained. Each person is given the new cold medication, and, 1 hour later, reaction time is measured for each individual. The average reaction time for this sample is $M = 206$ with $SS = 648$. The researcher would like to use a hypothesis test with $\alpha = .05$ to evaluate the effect of the medication.
 - a. Use a two-tailed test with $\alpha = .05$ to determine whether the medication has a significant effect on reaction time.
 - b. Write a sentences that demonstrates how the outcome of the hypothesis test would appear in a research report.
 - c. Use a one-tailed test with $\alpha = .05$ to determine whether the medication produces a significant increase in reaction time.
 - d. Write a sentence that demonstrates how the outcome of the one-tailed hypothesis test would appear in a research report.

- ANSWER**
1. a. For the two-tailed test, $H_0: \mu = 200$. The sample variance is 81, the estimated standard error is 3, and $t = 6/3 = 2.00$. With $df = 8$, the critical boundaries are ± 2.306 . Fail to reject the null hypothesis.
 - b. The result indicates that the medication does not have a significant effect on reaction time, $t(8) = 2.00, p > .05$.
 - c. For a one-tailed test, $H_0: \mu \leq 200$ (no increase). The data product $t = 6/3 = 2.00$. With $df = 8$, the critical boundary is 1.860. Reject the null hypothesis.
 - d. The results indicate that the medication produces a significant increase in reaction time, $t(8) = 2.00, p < .05$, one tailed.

SUMMARY

1. The t statistic is used instead of a z -score for hypothesis testing when the population standard deviation (or variance) is unknown.
2. To compute the t statistic, you must first calculate the sample variance (or standard deviation) as a substitute for the unknown population value.

$$\text{sample variance} = s^2 = \frac{SS}{df}$$

Next, the standard error is *estimated* by substituting s^2 in the formula for standard error. The estimated standard error is calculated in the following manner:

$$\text{estimated standard error} = s_M = \sqrt{\frac{s^2}{n}}$$

Finally, a t statistic is computed using the estimated standard error. The t statistic is used as a substitute for a z -score that cannot be computed when the population variance or standard deviation is unknown.

$$t = \frac{M - \mu}{s_M}$$

The structure of the t formula is similar to that of the z -score in that

$$z \text{ or } t = \frac{\text{sample mean} - \text{population mean}}{(\text{estimated}) \text{ standard error}}$$

For a hypothesis test, you hypothesize a value for the unknown population mean and plug the hypothesized value into the equation along with the sample mean and

the estimated standard error, which are computed from the sample data. If the hypothesized mean produces an extreme value for t , then you conclude that the hypothesis was wrong.

4. The t distribution is an approximation of the normal z distribution. To evaluate a t statistic that is obtained for a sample mean, the critical region must be located in a t distribution. There is a family of t distributions, with the exact shape of a particular distribution of t values depending on degrees of freedom ($n - 1$). Therefore, the critical t values depend on the value for df associated with the t test. As df increases, the shape of the t distribution approaches a normal distribution.
5. When a t statistic is used for a hypothesis test, Cohen's d can be computed to measure effect size. In this situation, the sample standard deviation is used in the formula to obtain an estimated value for d :

$$\text{estimated } d = \frac{\text{mean difference}}{\text{standard deviation}} = \frac{M - \mu}{s}$$

6. A second measure of effect size is r^2 , which measures the percentage of the variability that is accounted for by the treatment effect. This value is computed as follows:

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

7. An alternative method for describing the size of a treatment effect is to use a confidence interval for μ . A confidence interval is a range of values that estimates the unknown population mean. The confidence interval uses the t equation, solved for the unknown mean:

$$\mu = M \pm t(s_M)$$

First, select a level of confidence and then look up the corresponding t values to use in the equation. For example, for 95% confidence, use the range of t values that determine the middle 95% of the distribution.

KEY TERMS

estimated standard error (286)

t statistic (286)

degrees of freedom (287)

t distribution (287)

estimated d (295)

percentage of variance accounted for by the treatment (r^2) (299)

confidence interval (300)

RESOURCES

Book Companion Website: www.cengage.com/psychology/gravetter

You can find a tutorial quiz and other learning exercises for Chapter 9 on the book companion website. The website also provides access to a workshop entitled *Single-Sample t Test*, which reviews the concepts and logic of hypothesis testing with the t statistic.



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.

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 t Test presented in this chapter.

Data Entry

Enter all of the scores from the sample in one column of the data editor, probably VAR00001.

Data Analysis

1. Click **Analyze** on the tool bar, select **Compare Means**, and click on **One-Sample 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. In the **Test Value** box at the bottom of the One-Sample t Test window, enter the hypothesized value for the population mean from the null hypothesis. *Note:* The value is automatically set at zero until you type in a new value.
4. 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.
5. Click **OK**.

SPSS Output

We used the SPSS program to analyze the data from the infants-and-attractive-faces study in Example 9.1 and the program output is shown in Figure 9.9. The output includes a table of sample statistics with the mean, standard deviation, and standard error for the sample mean. A second table shows the results of the hypothesis test, including the values for t , df , and the level of significance (the p value for the test), as well as the mean difference from the hypothesized value of $\mu = 10$ and a 95% confidence interval for the mean difference. To obtain a 95% confidence interval for the mean, simply add $\mu = 10$ points to the values in the table.

FOCUS ON PROBLEM SOLVING

1. The first problem we confront in analyzing data is determining the appropriate statistical test. Remember that you can use a z -score for the test statistic only when the value for σ is known. If the value for σ is not provided, then you must use the t statistic.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
VAR00001	9	13.0000	3.00000	1.00000

One-Sample Test

	Test Value = 10					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
VAR00001	3.000	8	.017	3.00000	.6940	5.3060

FIGURE 9.9

The SPSS output for the hypothesis test presented in Example 9.1.

- For the t test, the sample variance is used to find the value for the estimated standard error. Remember to use $n - 1$ in the denominator when computing the sample variance (see Chapter 4). When computing estimated standard error, use n in the denominator.

DEMONSTRATION 9.1**A HYPOTHESIS TEST WITH THE t STATISTIC**

A psychologist has prepared an “Optimism Test” that is administered yearly to graduating college seniors. The test measures how each graduating class feels about its future—the higher the score, the more optimistic the class. Last year’s class had a mean score of $\mu = 15$. A sample of $n = 9$ seniors from this year’s class was selected and tested. The scores for these seniors are 7, 12, 11, 15, 7, 8, 15, 9, and 6, which produce a sample mean of $M = 10$ with $SS = 94$.

On the basis of this sample, can the psychologist conclude that this year’s class has a different level of optimism than last year’s class?

Note that this hypothesis test uses a t statistic because the population variance (σ^2) is not known.

- STEP 1 State the hypotheses, and select an alpha level.** The statements for the null hypothesis and the alternative hypothesis follow the same form for the t statistic and the z -score test.

$$H_0: \mu = 15 \quad (\text{There is no change.})$$

$$H_1: \mu \neq 15 \quad (\text{This year’s mean is different.})$$

For this demonstration, we use $\alpha = .05$, two tails.

STEP 2 Locate the critical region. With a sample of $n = 9$ students, the t statistic has $df = n - 1 = 8$. For a two-tailed test with $\alpha = .05$ and $df = 8$, the critical t values are $t = \pm 2.306$. These critical t values define the boundaries of the critical region. The obtained t value must be more extreme than either of these critical values to reject H_0 .

STEP 3 Compute the test statistic. As we have noted, it is easier to separate the calculation of the t statistic into three stages.

Sample variance.

$$s^2 = \frac{SS}{n-1} = \frac{94}{8} = 11.75$$

Estimated standard error. The estimated standard error for these data is

$$s_M = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{11.75}{9}} = 1.14$$

The t statistic. Now that we have the estimated standard error and the sample mean, we can compute the t statistic. For this demonstration,

$$t = \frac{M - \mu}{s_M} = \frac{10 - 15}{1.14} = \frac{-5}{1.14} = -4.39$$

STEP 4 Make a decision about H_0 , and state a conclusion. The t statistic we obtained ($t = -4.39$) is in the critical region. Thus, our sample data are unusual enough to reject the null hypothesis at the .05 level of significance. We can conclude that there is a significant difference in level of optimism between this year's and last year's graduating classes, $t(8) = -4.39, p < .05$, two-tailed.

DEMONSTRATION 9.2

EFFECT SIZE: ESTIMATING COHEN'S d AND COMPUTING r^2

We will estimate Cohen's d for the same data used for the hypothesis test in Demonstration 9.1. The mean optimism score for the sample from this year's class was 5 points lower than the mean from last year ($M = 10$ versus $\mu = 15$). In Demonstration 9.1, we computed a sample variance of $s^2 = 11.75$, so the standard deviation is $\sqrt{11.75} = 3.43$. With these values,

$$\text{estimated } d = \frac{\text{mean difference}}{\text{standard deviation}} = \frac{5}{3.43} = 1.46$$

To calculate the percentage of variance explained by the treatment effect, r^2 , we need the value of t and the df value from the hypothesis test. In Demonstration 9.1, we obtained $t = -4.39$ with $df = 8$. Using these values in Equation 9.5, we obtain

$$r^2 = \frac{t^2}{t^2 + df} = \frac{(-4.39)^2}{(-4.39)^2 + 8} = \frac{19.27}{27.27} = 0.71$$

PROBLEMS

1. Under what circumstances is a t statistic used instead of a z -score for a hypothesis test?
2. A sample of $n = 25$ scores has a mean of $M = 83$ and a standard deviation of $s = 15$.
 - a. Explain what is measured by the sample standard deviation.
 - b. Compute the estimated standard error for the sample mean and explain what is measured by the standard error.
3. Find the estimated standard error for the sample mean for each of the following samples.
 - a. $n = 4$ with $SS = 48$
 - b. $n = 6$ with $SS = 270$
 - c. $n = 12$ with $SS = 132$
4. Explain why t distributions tend to be flatter and more spread out than the normal distribution.
5. Find the t values that form the boundaries of the critical region for a two-tailed test with $\alpha = .05$ for each of the following sample sizes:
 - a. $n = 6$
 - b. $n = 12$
 - c. $n = 24$
6. The following sample of $n = 6$ scores was obtained from a population with unknown parameters.
Scores: 7, 1, 6, 3, 6, 7
 - a. Compute the sample mean and standard deviation. (Note that these are descriptive values that summarize the sample data.)
 - b. Compute the estimated standard error for M . (Note that this is an inferential value that describes how accurately the sample mean represents the unknown population mean.)
7. The following sample was obtained from a population with unknown parameters.
Scores: 6, 12, 0, 3, 4
 - a. Compute the sample mean and standard deviation. (Note that these are descriptive values that summarize the sample data.)
 - b. Compute the estimated standard error for M . (Note that this is an inferential value that describes how accurately the sample mean represents the unknown population mean.)
8. To evaluate the effect of a treatment, a sample is obtained from a population with a mean of $\mu = 75$, and the treatment is administered to the individuals in the sample. After treatment, the sample mean is found to be $M = 79.6$ with a standard deviation of $s = 12$.
 - a. If the sample consists of $n = 16$ individuals, are the data sufficient to conclude that the treatment has a significant effect using a two-tailed test with $\alpha = .05$?
 - b. If the sample consists of $n = 36$ individuals, are the data sufficient to conclude that the treatment has a significant effect using a two-tailed test with $\alpha = .05$?
 - c. Comparing your answer for parts a and b, how does the size of the sample influence the outcome of a hypothesis test?
9. To evaluate the effect of a treatment, a sample of $n = 9$ is obtained from a population with a mean of $\mu = 40$, and the treatment is administered to the individuals in the sample. After treatment, the sample mean is found to be $M = 33$.
 - a. If the sample has a standard deviation of $s = 9$, are the data sufficient to conclude that the treatment has a significant effect using a two-tailed test with $\alpha = .05$?
 - b. If the sample standard deviation is $s = 15$, are the data sufficient to conclude that the treatment has a significant effect using a two-tailed test with $\alpha = .05$?
 - c. Comparing your answer for parts a and b, how does the variability of the scores in the sample influence the outcome of a hypothesis test?
10. A random sample of $n = 16$ individuals is selected from a population with $\mu = 70$, and a treatment is administered to each individual in the sample. After treatment, the sample mean is found to be $M = 76$ with $SS = 960$.
 - a. How much difference is there between the mean for the treated sample and the mean for the original population? (Note: In a hypothesis test, this value forms the numerator of the t statistic.)
 - b. How much difference is expected just by chance between the sample mean and its population mean? That is, find the standard error for M . (Note: In a hypothesis test, this value is the denominator of the t statistic.)
 - c. Based on the sample data, does the treatment have a significant effect? Use a two-tailed test with $\alpha = .05$.
11. The spotlight effect refers to overestimating the extent to which others notice your appearance or behavior, especially when you commit a social faux pas. Effectively, you feel as if you are suddenly standing in a spotlight with everyone looking. In one demonstration of this phenomenon, Gilovich, Medvec, and Savitsky (2000) asked college students to put on a Barry Manilow T-shirt that fellow students had previously judged to be embarrassing. The participants were

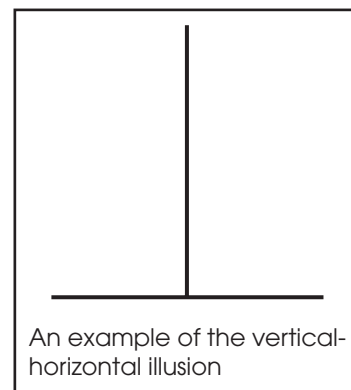
then led into a room in which other students were already participating in an experiment. After a few minutes, the participant was led back out of the room and was allowed to remove the shirt. Later, each participant was asked to estimate how many people in the room had noticed the shirt. The individuals who were in the room were also asked whether they noticed the shirt. In the study, the participants significantly overestimated the actual number of people who had noticed.

- a. In a similar study using a sample of $n = 9$ participants, the individuals who wore the shirt produced an average estimate of $M = 6.4$ with $SS = 162$. The average number who said they noticed was 3.1. Is the estimate from the participants significantly different from the actual number? Test the null hypothesis that the true mean is $\mu = 3.1$ using a two-tailed test with $\alpha = .05$.
 - b. Is the estimate from the participants significantly higher than the actual number ($\mu = 3.1$)? Use a one-tailed test with $\alpha = .05$.
12. Many animals, including humans, tend to avoid direct eye contact and even patterns that look like eyes. Some insects, including moths, have evolved eye-spot patterns on their wings to help ward off predators. Scaife (1976) reports a study examining how eye-spot patterns affect the behavior of birds. In the study, the birds were tested in a box with two chambers and were free to move from one chamber to another. In one chamber, two large eye-spots were painted on one wall. The other chamber had plain walls. The researcher recorded the amount of time each bird spent in the plain chamber during a 60-minute session. Suppose the study produced a mean of $M = 37$ minutes in the plain chamber with $SS = 288$ for a sample of $n = 9$ birds. (Note: If the eye-spots have no effect, then the birds should spend an average of $\mu = 30$ minutes in each chamber.)
 - a. Is this sample sufficient to conclude that the eye-spots have a significant influence on the birds' behavior? Use a two-tailed test with $\alpha = .05$.
 - b. Compute the estimated Cohen's d to measure the size of the treatment effect.
 - c. Construct the 95% confidence interval to estimate the mean amount of time spent on the plain side for the population of birds.
 13. Standardized measures seem to indicate that the average level of anxiety has increased gradually over the past 50 years (Twenge, 2000). In the 1950s, the average score on the Child Manifest Anxiety Scale was $\mu = 15.1$. A sample of $n = 16$ of today's children produces a mean score of $M = 23.3$ with $SS = 240$.
 - a. Based on the sample, has there been a significant change in the average level of anxiety since the 1950s? Use a two-tailed test with $\alpha = .01$.
 - b. Make a 90% confidence interval estimate of today's population mean level of anxiety.
 - c. Write a sentence that demonstrates how the outcome of the hypothesis test and the confidence interval would appear in a research report.
 14. The librarian at the local elementary school claims that, on average, the books in the library are more than 20 years old. To test this claim, a student takes a sample of $n = 30$ books and records the publication date for each. The sample produces an average age of $M = 23.8$ years with a variance of $s^2 = 67.5$. Use this sample to conduct a one-tailed test with $\alpha = .01$ to determine whether the average age of the library books is significantly greater than 20 years ($\mu > 20$).
 15. For several years researchers have noticed that there appears to be a regular, year-by-year increase in the average IQ for the general population. This phenomenon is called the Flynn effect after the researcher who first reported it (Flynn, 1984, 1999), and it means that psychologists must continuously update IQ tests to keep the population mean at $\mu = 100$. To evaluate the size of the effect, a researcher obtained a 10-year-old IQ test that was standardized to produce a mean IQ of $\mu = 100$ for the population 10 years ago. The test was then given to a sample of $n = 64$ of today's 20-year-old adults. The average score for the sample was $M = 107$ with a standard deviation of $s = 12$.
 - a. Based on the sample, is the average IQ for today's population significantly different from the average 10 years ago, when the test would have produced a mean of $\mu = 100$? Use a two-tailed test with $\alpha = .01$.
 - b. Make an 80% confidence interval estimate of today's population mean IQ for the 10-year-old test.
 16. In a classic study of infant attachment, Harlow (1959) placed infant monkeys in cages with two artificial surrogate mothers. One "mother" was made from bare wire mesh and contained a baby bottle from which the infants could feed. The other mother was made from soft terry cloth and did not provide any access to food. Harlow observed the infant monkeys and recorded how much time per day was spent with each mother. In a typical day, the infants spent a total of 18 hours clinging to one of the two mothers. If there were no preference between the two, you would expect the time to be divided evenly, with an average of $\mu = 9$ hours for each of the mothers. However, the typical monkey spent around 15 hours per day with the terry-cloth mother, indicating a strong preference for the soft, cuddly mother. Suppose a sample of $n = 9$ infant monkeys averaged $M = 15.3$ hours per day with $SS = 216$ with the terry-cloth mother. Is this result sufficient to conclude that the monkeys spent significantly more time

with the softer mother than would be expected if there were no preference? Use a two-tailed test with $\alpha = .05$.

17. Belsky, Weinraub, Owen, and Kelly (2001) reported on the effects of preschool childcare on the development of young children. One result suggests that children who spend more time away from their mothers are more likely to show behavioral problems in kindergarten. Using a standardized scale, the average rating of behavioral problems for kindergarten children is $\mu = 35$. A sample of $n = 16$ kindergarten children who had spent at least 20 hours per week in childcare during the previous year produced a mean score of $M = 42.7$ with a standard deviation of $s = 6$.
 - a. Are the data sufficient to conclude that children with a history of childcare show significantly more behavioral problems than the average kindergarten child? Use a one-tailed test with $\alpha = .01$.
 - b. Compute r^2 , the percentage of variance accounted for, to measure the size of the preschool effect.
 - c. Write a sentence showing how the outcome of the hypothesis test and the measure of effect size would appear in a research report.
18. Other research examining the effects of preschool childcare has found that children who spent time in day care, especially high-quality day care, perform better on math and language tests than children who stay home with their mothers (Broberg, Wessels, Lamb, & Hwang, 1997). Typical results, for example, show that a sample of $n = 25$ children who attended day care before starting school had an average score of $M = 87$ with $SS = 1536$ on a standardized math test for which the population mean is $\mu = 81$.
 - a. Is this sample sufficient to conclude that the children with a history of preschool day care are significantly different from the general population? Use a two-tailed test with $\alpha = .01$.
 - b. Compute Cohen's d to measure the size of the preschool effect.
 - c. Write a sentence showing how the outcome of the hypothesis test and the measure of effect size would appear in a research report.
19. A random sample of $n = 25$ scores is obtained from a population with a mean of $\mu = 45$. A treatment is administered to the individuals in the sample and, after treatment, the sample mean is $M = 48$.
 - a. Assuming that the sample standard deviation is $s = 6$ compute r^2 and the estimated Cohen's d to measure the size of the treatment effect.
 - b. Assuming that the sample standard deviation is $s = 15$, compute r^2 and the estimated Cohen's d to measure the size of the treatment effect.
 - c. Comparing your answers from parts a and b, how does the variability of the scores in the sample influence the measures of effect size?

20. A random sample is obtained from a population with a mean of $\mu = 70$. A treatment is administered to the individuals in the sample and, after treatment, the sample mean is $M = 78$ with a standard deviation of $s = 20$.
 - a. Assuming that the sample consists of $n = 25$ scores, compute r^2 and the estimated Cohen's d to measure the size of treatment effect.
 - b. Assuming that the sample consists of $n = 16$ scores, compute r^2 and the estimated Cohen's d to measure the size of treatment effect.
 - c. Comparing your answers from parts a and b, how does the number of scores in the sample influence the measures of effect size?
21. An example of the vertical-horizontal illusion is shown in the figure below. Although the two lines are exactly the same length, the vertical line appears to be much longer. To examine the strength of this illusion, a researcher prepared an example in which both lines were exactly 10 inches long. The example was shown to individual participants who were told that the horizontal line was 10 inches long and then were asked to estimate the length of the vertical line. For a sample of $n = 25$ participants, the average estimate was $M = 12.2$ inches with a standard deviation of $s = 1.00$.



- a. Use a one-tailed hypothesis test with $\alpha = .01$ to demonstrate that the individuals in the sample significantly overestimate the true length of the line. (Note: Accurate estimation would produce a mean of $\mu = 10$ inches.)
 - b. Calculate the estimated d and r^2 , the percentage of variance accounted for, to measure the size of this effect.
 - c. Construct a 95% confidence interval for the population mean estimated length of the vertical line.
22. In studies examining the effect of humor on interpersonal attractions, McGee and Shevlin (2009) found that an individual's sense of humor had a

significant effect on how the individual was perceived by others. In one part of the study, female college students were given brief descriptions of a potential romantic partner. The fictitious male was described positively as being single, ambitious, and having good job prospects. For one group of participants, the description also said that he had a great sense of humor. For another group, it said that he had no sense of humor. After reading the description, each participant was asked to rate the attractiveness of the man on a seven-point scale from 1 (very attractive) to 7 (very unattractive). A score of 4 indicates a neutral rating.

- a. The females who read the “great sense of humor” description gave the potential partner an average attractiveness score of $M = 4.53$ with a standard deviation of $s = 1.04$. If the sample consisted of $n = 16$ participants, is the average rating significantly higher than neutral ($\mu = 4$)? Use a one-tailed test with $\alpha = .05$.
- b. The females who read the description saying “no sense of humor” gave the potential partner an

average attractiveness score of $M = 3.30$ with a standard deviation of $s = 1.18$. If the sample consisted of $n = 16$ participants, is the average rating significantly lower than neutral ($\mu = 4$)? Use a one-tailed test with $\alpha = .05$.

23. A psychologist would like to determine whether there is a relationship between depression and aging. It is known that the general population averages $\mu = 40$ on a standardized depression test. The psychologist obtains a sample of $n = 9$ individuals who are all more than 70 years old. The depression scores for this sample are as follows: 37, 50, 43, 41, 39, 45, 49, 44, 48.
 - a. On the basis of this sample, is depression for elderly people significantly different from depression in the general population? Use a two-tailed test with $\alpha = .05$.
 - b. Compute the estimated Cohen’s d to measure the size of the difference.
 - c. Write a sentence showing how the outcome of the hypothesis test and the measure of effect size would appear in a research report.



Improve your statistical skills with ample practice exercises and detailed explanations on every question. Purchase www.aplia.com/statistics