



10

Testing Means Related-Samples t Test With Confidence Intervals

••• Learning Objectives

After reading this chapter, you should be able to:

1. Describe two types of research designs used when we select related samples.
2. Explain why difference scores are computed for the related-samples t test.
3. State three advantages for selecting related samples.
4. Calculate the degrees of freedom for a related-samples t test and locate critical values in the t table.
5. Identify the assumptions for the related-samples t test.
6. Compute a related-samples t test and interpret the results.
7. Compute and interpret effect size and proportion of variance for a related-samples t test.
8. Compute and interpret confidence intervals for a related-samples t test.
9. Summarize the results of a related-samples t test in APA format.
10. Compute a related-samples t test and identify confidence intervals using SPSS.

10.1 RELATED SAMPLES DESIGNS

In Chapter 9, we introduced a hypothesis test for situations in which we observe independent samples. An independent sample is one in which different participants are observed one time in each group. We can select two independent samples in one of two ways: We can select a sample from two different populations, or we can select a sample from a single population and randomly assign participants to two groups or conditions, as described in Chapter 9.

In this chapter we describe a hypothesis test for when we observe **related samples**—that is, samples in which the participants are observed in more than one treatment or matched on common characteristics. There are two types of research designs commonly used to select related samples: the repeated-measures design and the matched-pairs design. While these research designs are analyzed in the same way, the data are collected differently for each. Both research designs are briefly described in this section.

The Repeated-Measures Design

The most common related-samples design is the **repeated-measures design**, in which each participant is observed repeatedly. Table 10.1 shows a situation with two treatments. In a repeated-measures design, each participant (n_1, n_2 , etc.) is observed twice (once in each treatment). We can create repeated measures by using a pre-post design or a within-subjects design.

TABLE 10.1

In a Repeated-Measures Design With Two Treatments, n Participants Are Observed Two Times

Treatment 1		Treatment 2
n_1	–	n_1
n_2	–	n_2
n_3	–	n_3
n_4	–	n_4
n_5	–	n_5

In a **related sample**, also called a **dependent sample**, participants are related. Participants can be related in one of two ways: They are observed in more than one group (a repeated-measures design), or they are matched, experimentally or naturally, based on common characteristics or traits (a matched-pairs design).

The **repeated-measures design** is a research design in which the same participants are observed in each treatment. Two types of repeated-measures designs are the pre-post design and the within-subjects design.

Chapter Outline

- 10.1 Related and Independent Samples
- 10.2 Introduction to the Related-Samples t Test
- 10.3 Computing the Related-Samples t Test
- 10.4 Measuring Effect Size for the Related-Samples t Test
- 10.5 Confidence Intervals for the Related-Samples t Test
- 10.6 Inferring Significance and Effect Size From a Confidence Interval
- 10.7 SPSS in Focus: Related-Samples t Test and Confidence Intervals
- 10.8 APA in Focus: Reporting the t Statistic and Confidence Intervals

FVI

Two types of designs for selecting related samples are the repeated-measures and matched-pairs designs.

FVI

In a repeated-measures design, the same participants are observed in each treatment. Two types of repeated-measures designs are the pre-post and within-subjects designs.

FVI

In a matched-pairs design, participants are matched based on the characteristics that they share. Matched pairs can be obtained either experimentally or through natural occurrence.

Using the **pre-post design**, we measure a dependent variable for participants observed before (pre) and after (post) a treatment. For example, we can measure athletic performance (the dependent variable) in a sample of athletes before and after a training camp. We can compare the difference in athletic performance before and after the camp. This type of repeated-measures design is limited to observing participants two times (i.e., before and after a treatment).

Using the **within-subjects design**, we observe participants across many treatments but not necessarily before and after a treatment. For two samples, suppose we want to study the effects of exercise on memory. We can select a sample of participants and have them take a memory test after completing an anaerobic exercise, and again after completing an aerobic exercise. In this example, we observe participants twice, but not necessarily before and after a treatment. Anytime the same participants are observed in each group, either pre-post or within-subjects, we are using the repeated-measures design to observe related samples.

The Matched-Pairs Design

The matched-pairs design is also used to study related samples. In the **matched-pairs design**, participants are selected and then matched, experimentally or naturally, based on common characteristics or traits. The matched-pairs design is limited to observing two groups, where pairs of participants are matched. Using this design, Table 10.2 shows a situation with two treatments. In a matched-pairs design, different, yet matched, participants (n_1, n_2 , etc.) are observed in each treatment, and scores from each matched pair of participants are compared.

TABLE 10.2

In a Matched-Pairs Design, n Pairs Are Matched and Observed One Time

Treatment 1		Treatment 2
n_1	–	n_2
n_3	–	n_4
n_5	–	n_6
n_7	–	n_8
n_9	–	n_{10}

The **pre-post design** is a type of repeated-measures design in which researchers measure a dependent variable for participants before (pre) and after (post) a treatment.

The **within-subjects design** is a type of repeated-measures design in which researchers observe the same participants across many treatments but not necessarily before and after a treatment.

The **matched-pairs design**, also called the **matched-subjects design** or **matched-samples design**, is a research design in which participants are selected and then matched, experimentally or naturally, based on common characteristics or traits.

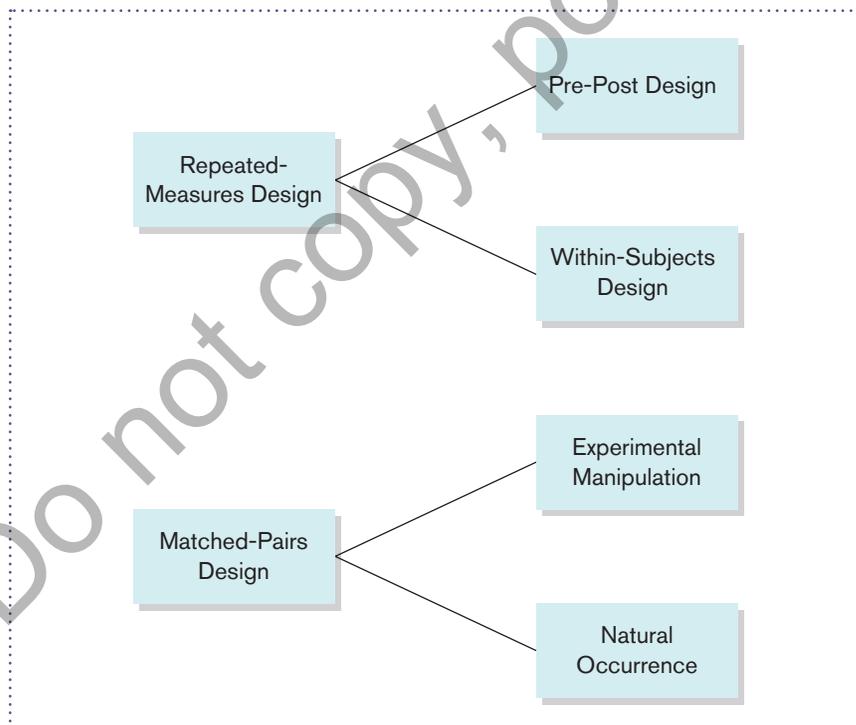
We can obtain matched pairs in one of two ways: experimental manipulation or natural occurrence. Matching through experimental manipulation is typical for research studies in which the researcher manipulates the traits or characteristics upon which participants are matched. We can match pairs of participants on any number of variables, including their level of intelligence, their personality type, their eating habits, their level of education, and their sleep patterns. We could measure these characteristics and then match

participants. For example, we could measure intelligence, then match the two participants scoring the highest, the two participants scoring the next highest, and so on. Matching through experimental manipulation requires that we measure some trait or characteristic before we match participants into pairs.

Matching through natural occurrence is typical for quasi-experiments in which participants are matched based on preexisting traits. The preexisting traits are typically biological or physiological in nature. For example, we could match participants based on genetics (e.g., biological twins) or family affiliation (e.g., brothers, sisters, or cousins). Each trait or characteristic is inherent to the participant. There is no need to measure this in order to pair participants. Instead, the participants are already matched naturally. Pairs of identical twins or brothers, for example, are paired together naturally. One member of each pair is assigned to a group or a treatment, and differences between pairs of scores are observed. Anytime participants are matched on common traits, either experimentally or through natural occurrence, we are using the matched-pairs design to select related samples. Figure 10.1 summarizes the designs used with related samples.

FIGURE 10.1

Two Designs Associated With Selecting Related Samples: the Repeated-Measures Design and the Matched-Pairs Design



Repeated measures can be selected by observing participants before and after a treatment (pre-post design) or across treatments (within-subjects design). Matched pairs can be selected through experimental manipulation or natural occurrence.

LEARNING CHECK 1

1. How do related samples differ from independent samples?
2. Name two types of repeated-measures designs.
3. Distinguish between the repeated-measures design and the matched-pairs design.
4. A researcher records the amount of computer use (in hours per day) in a sample of students, then matches the students based on their amount of computer use. Is this an example of matching through experimental manipulation or matching through natural occurrence?

Answers: 1. In related samples, participants are related because they are either observed more than once or matched on common traits. In independent samples, participants are not related because different participants are observed in each group. 2. Pre-post design and within-subjects design. 3. Each participant is observed at least two times in the repeated-measures design. In the matched-pairs design, each pair of participants is matched and observed one time. 4. Matching through experimental manipulation.

10.2 INTRODUCTION TO THE RELATED-SAMPLES t TEST

When we select two related samples, we can compare differences using the **related-samples t test**. To use this test, we start by stating the null hypothesis for the mean difference between pairs of scores in a population, and then compare this to the difference we observe between paired scores in a sample. The related-samples t test is different from the two-independent-sample t test in that first we subtract one score in each pair from the other to obtain the **difference score** for each participant; then we compute the test statistic.

There is a good reason for finding the difference between scores in each pair before computing the test statistic using a related-samples t test: It eliminates the source of error associated with observing different participants in each group or treatment. When we select related samples, we observe the same, or matched, participants in each group, not different participants. So we can eliminate this source of error.

Consider the hypothetical data shown in Table 10.3 for four participants (A, B, C, and D) observed in two groups (Q and Z). Table 10.3 identifies three places where differences can occur with two groups. The null hypothesis makes a statement about the mean difference between groups, which is the difference we are testing. Any other difference is called **error** because the differences cannot be attributed to having different groups.

As shown in Table 10.4, when we reduce pairs of scores to a single column of difference scores, we eliminate the between-persons error that was illustrated in Table 10.3. Between-persons error is associated with differences associated with observing different participants in each group or treatment. However, using the related-samples design, we observe the same (or matched) participants in each group, not different participants, so we can eliminate this source of error before computing the test statistic. Error in a study is measured by the estimate of standard error. Eliminating between-persons error makes the total value of error smaller, thereby reducing standard error. In Chapter 8, we showed that reducing standard error increases the power of

The **related-samples t test** is a statistical procedure used to test hypotheses concerning two related samples selected from populations in which the variance in one or both populations is unknown.

A **difference score** is a score or value obtained by subtracting one score from another. In a related-samples t test, difference scores are obtained prior to computing the test statistic.

For a t test, the term **error** refers to any unexplained difference that cannot be attributed to, or caused by, having different treatments. The standard error of the mean is used to measure the error or unexplained differences in a statistical design.

TABLE 10.3 A Hypothetical Set of Data of Four Participants

Participant	Groups	
	Q	Z
A	3	4
B	0	3
C	2	3
D	3	6
	$M = 2$	$M = 4$

Between-Persons Error. Differences occur between participants. For example, Person A increased 1 point from Group Q to Group Z, and Person B increased 3 points. This is not consistent change and is therefore a source of error—the inconsistent differences are not due to having different groups.

Within-Groups Error. Differences occur within each group, even though all participants experienced the same treatment in each group. This is a source of error—the differences are not due to having different groups.

Between-Groups Effect. We will use the test statistic to determine whether this difference is significant. This is the difference we are interested in testing.

There are three places where differences can occur. Mean differences (the between-groups effect) are the differences we are testing. The other two places where differences occur are regarded as errors in that they have nothing to do with having different groups.

TABLE 10.4 A Hypothetical Set of Data of the Same Four Participants in Table 10.3

Participant	Difference Scores (Q-Z)
A	-1
B	-3
C	-1
D	-3
	$M = -2$

Within-Groups Error. Differences occur within the column of difference scores, even though all participants experienced the same pair of groups. This is a source of error—these differences are not due to having different groups.

Between-Groups Effect. Mean difference score.

Notice that when we find the difference between paired scores, only one source of error remains (within-groups error). The between-persons error is eliminated when the same or matched participants are observed in each group.

detecting an effect. This is a key advantage of computing difference scores prior to computing the test statistic: It reduces standard error, thereby increasing the power to detect an effect.

FVI

Computing difference scores prior to computing the test statistic reduces error by eliminating one possible source of error, thereby increasing the power of a hypothesis test.

Advantages for Selecting Related Samples

Selecting related samples has many advantages and disadvantages. Because most of the disadvantages pertain specifically to a repeated-measures design and not a matched-pairs design, we introduce only the advantages that pertain to both designs. There are three key advantages for selecting related samples compared to selecting independent samples in behavioral research:

1. Selecting related samples can be more practical. It is more practical in that selecting related samples may provide a better way to test your hypotheses. This is especially true for research areas in learning and development in which researchers must observe changes in behavior or development over time or between matched pairs of participants. For example, it can be more practical to observe the behavior of the same participants before and after a treatment (repeated measures) or to compare how well participants of similar ability master a task (matched samples).

2. Selecting related samples reduces standard error. Computing difference scores prior to computing the test statistic eliminates the between-persons source of error, which reduces the estimate of standard error. Using the same data, this means that the value for the estimate of standard error for a related-samples t test will be smaller than that for a two-independent-sample t test.

3. Selecting related samples increases power. It follows from the second advantage that reducing the estimate for standard error will increase the value of the test statistic. Using the same data, a related-samples t test is more likely than a two-independent-sample t test to result in a decision to reject the null hypothesis. Hence, the related-samples t test is associated with greater power to detect an effect.

Thus, selecting related samples has distinct advantages that are practical and can increase power. The remainder of this section elucidates the test statistic and degrees of freedom for a related-samples t test, which is used to analyze data for related samples with two groups.

The Test Statistic

The test statistic for a related-samples t test is similar to the test statistics introduced in Chapters 8 and 9 for one group. The mean differences are placed in the numerator, and the estimate of standard error is placed in the denominator. For a related-samples t test, in the numerator, we subtract the mean difference between two related samples (M_D) from the mean difference stated in the null hypothesis (μ_D):

$$M_D - \mu_D.$$

The estimate of standard error is placed in the denominator of the test statistic for a related-samples t test. The standard error for a distribution of mean difference scores, called the **estimated standard error for difference scores** (s_{MD}), is computed using the following formula:

$$s_{MD} = \sqrt{\frac{s_D^2}{n_D}} = \frac{s_D}{\sqrt{n_D}}.$$

The **estimated standard error for difference scores** (s_{MD}) is an estimate of the standard deviation of a sampling distribution of mean difference scores. It is an estimate of the standard error or standard distance that the mean difference scores deviate from the mean difference score stated in the null hypothesis.

By placing the mean differences in the numerator and the estimated standard error for difference scores in the denominator, we obtain the formula for the test statistic for a related-samples t test:

$$t_{\text{obt}} = \frac{M_D - \mu_D}{s_{MD}}$$

The test statistic for a related-samples t test estimates the number of standard deviations in a t distribution that a sample mean difference falls from the population mean difference stated in the null hypothesis. The larger the value of the test statistic, the less likely a sample mean difference would occur if the null hypothesis were true, thereby making it more likely that we will decide to reject the null hypothesis.

Degrees of Freedom

To compute the test statistic, we first reduce each pair of scores to one column of difference scores. Hence, the degrees of freedom for the related-samples t test equal the number of difference scores minus 1: $df = (n_D - 1)$.

Assumptions

There are two assumptions we make to compute the related-samples t test:

1. *Normality.* We assume that data in the population of difference scores are normally distributed. Again, this assumption is most important for small sample sizes. With larger samples ($n > 30$), the standard error is smaller, and this assumption becomes less critical as a result.
2. *Independence within groups.* The samples are related or matched between groups. However, we must assume that difference scores were obtained from different individuals within each group or treatment.

FVI

The degrees of freedom for the related-samples t test equal the number of difference scores minus 1: $df = (n_D - 1)$.

FVI

Two assumptions for a related-samples t test are normality and independence within groups.

LEARNING CHECK 2

1. Why do we find the difference between pairs of scores (compute difference scores) before computing the test statistic for a related-samples t test?
2. What is the value for the degrees of freedom for each example listed below?
 - (a) A study comparing 10 matched pairs of scores
 - (b) A study involving 18 participants observed two times
3. What value is placed in the denominator of the test statistic for the related-samples t test?
4. What are the assumptions for a related-samples t test?

Answers: 1. Computing difference scores eliminates the between-persons error, thereby increasing the power of the test; 2. (a) $df = 9$, (b) $df = 17$; 3. Estimated standard error for difference scores; 4. Normality and independence within groups.

10.3 COMPUTING THE RELATED-SAMPLES t TEST

Two designs associated with selecting related samples are the repeated-measures design and the matched-pairs design. In Example 10.1, we compute the related-samples t test for a study using the repeated-measures design. However, the same procedures described here can also be applied to compute the related-samples t test for a study using the matched-pairs design.

Example 10.1



©iStockphoto.com/CEFutcher

One area of focus in many areas of psychology and in education is on understanding and promoting reading among children and adults (Kent, Wanzek, & Al Otaiba, 2012; White, Chen, & Forsyth, 2010). Suppose we conduct a study with this area of focus by testing if teacher supervision influences the time that elementary school children read. To test this, we stage two 6-minute reading sessions and record the time in seconds that children spend reading in each session. In one session, the children read with a teacher present in the room; in another session, the same group of children read without a teacher present. The difference in time spent reading in the presence versus absence of a teacher is recorded. Table 10.5 lists the results of this hypothetical study with difference scores given. Test whether or not reading times differ using a .05 level of significance.

FYI

The null and alternative hypotheses make statements about a population of mean difference scores.

TABLE 10.5

The Time (in seconds) Spent Reading in a Sample of Eight Students in the Presence and Absence of a Teacher

Participants	Teacher		Difference Scores
	Present	Absent	
1	220	210	$(220 - 210) = 10$
2	245	220	$(245 - 220) = 25$
3	215	195	$(215 - 195) = 20$
4	260	265	$(260 - 265) = -5$
5	300	275	$(300 - 275) = 25$
6	280	290	$(280 - 290) = -10$
7	250	220	$(250 - 220) = 30$
8	310	285	$(310 - 285) = 25$

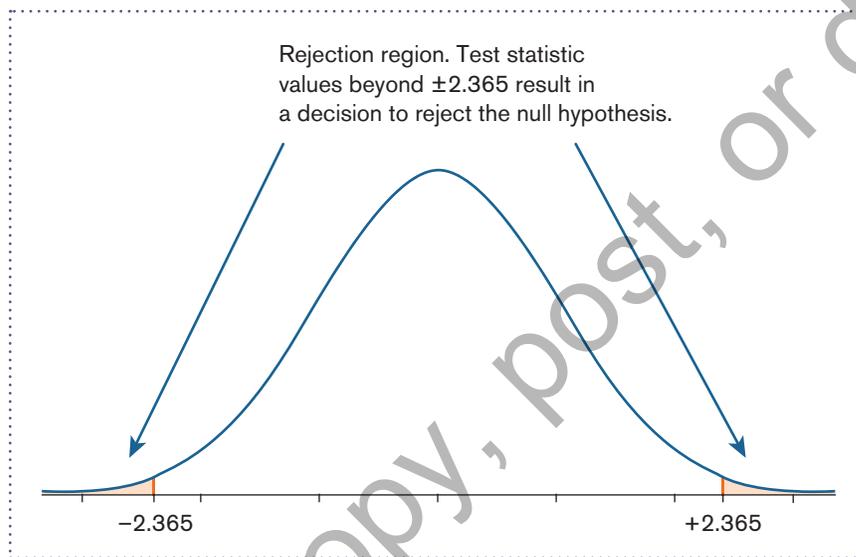
Difference scores are given in the last column.

Step 1: State the hypotheses. Because we are testing whether ($=$) or not (\neq) a difference exists, the null hypothesis states that there is no mean difference, and the alternative hypothesis states that there is a mean difference:

- $H_0: \mu_D = 0$ There is no mean difference in time spent reading in the presence versus absence of a teacher.
- $H_1: \mu_D \neq 0$ There is a mean difference in time spent reading in the presence versus absence of a teacher.

Step 2: Set the criteria for a decision. The level of significance for this test is .05. This is a two-tailed test for the mean difference between two related samples. The degrees of freedom for this test are $df = 8 - 1 = 7$. We locate 7 degrees of freedom in the far-left column of the t table in Table B.2 in Appendix B. Move across to the column to find the critical values for a .05 proportion in two tails. The critical values for this test are ± 2.365 . Figure 10.2 shows the t distribution and the rejection regions beyond these critical values.

FIGURE 10.2 Setting the Criteria for a Decision



The shaded areas show the rejection regions for a t distribution with $df = 7$. The critical value cutoffs are ± 2.365 . If the value of the test statistic falls in the shaded area, then we choose to reject the null hypothesis; otherwise, we retain the null hypothesis.

We will compare the value of the test statistic to these critical values. If the value of the test statistic falls beyond either critical value (± 2.365), then there is less than a 5% chance we would obtain that outcome if the null hypothesis were true, so we reject the null hypothesis; otherwise, we retain the null hypothesis.

Step 3: Compute the test statistic. To compute the test statistic, we (1) compute the mean, variance, and standard deviation of difference scores; (2) compute the estimated standard error for difference scores; then (3) compute the test statistic.

(1) Compute the mean, variance, and standard deviation of difference scores. Keep in mind that the sign (negative or positive) of difference scores matters when we compute the mean and standard deviation. To compute the mean, sum the difference scores ($\sum D$) and divide by the number of difference scores summed (n_D):

$$M_D = \frac{\sum D}{n_D} = \frac{120}{8} = 15.$$

To compute the variance, we will use the computational formula for sample variance (p. 105). Table 10.6 shows the calculations for computing D and D^2 using this formula:

$$s_D^2 = \frac{SS}{n_D - 1}, \text{ where } SS = \sum D^2 - \frac{(\sum D)^2}{n_D}$$

$$SS = 3,400 - \frac{(120)^2}{8} = 1,600.$$

$$s_D^2 = \frac{1600}{8-1} = 228.57.$$

TABLE 10.6 Calculations for the Variance of Difference Scores

D	D^2
10	100
25	625
20	400
-5	25
25	625
-10	100
30	900
25	625
$\sum D = 120$	$\sum D^2 = 3,400$

The first column lists the same difference scores as those in the last column of Table 10.5.

To compute the standard deviation, we take the square root of the variance:

$$s_D = \sqrt{s_D^2} = \sqrt{228.57} = 15.12.$$

(2) To compute the estimated standard error for difference scores (s_{MD}), we substitute 15.12 for s_D and 8 for n_D :

$$s_{MD} = \frac{s_D}{\sqrt{n_D}} = \frac{15.12}{\sqrt{8}} = 5.35.$$

(3) Compute the test statistic. We substitute 15 for M_D , 0 for μ_D (this is the value stated in the null hypothesis), and 5.35 for s_{MD} :

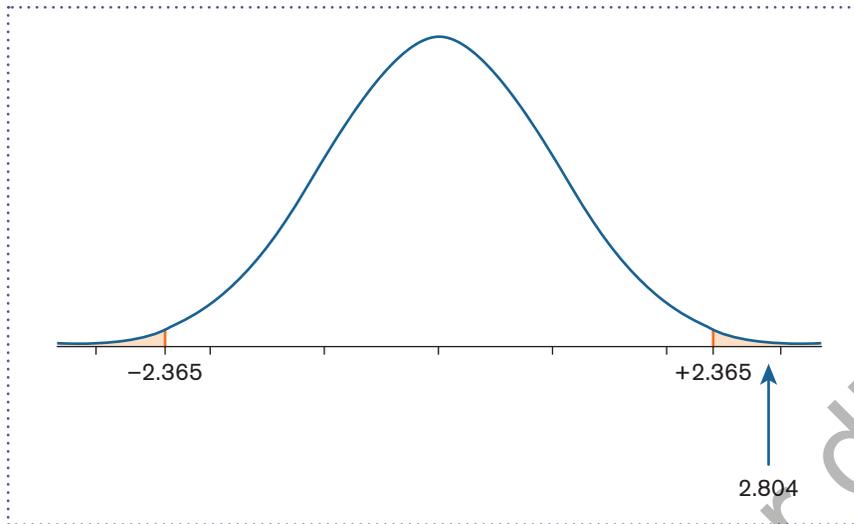
$$t_{\text{obt}} = \frac{M_D - \mu_D}{s_{MD}} = \frac{15 - 0}{5.35} = 2.804.$$

Step 4: Make a decision. To make a decision, we compare the obtained value to the critical value. We reject the null hypothesis if the obtained value exceeds the critical value. Figure 10.3 shows that the obtained value ($t_{\text{obt}} = 2.804$) exceeds the upper critical value; it falls in the rejection region. The decision is to reject the null hypothesis. If we were to report this result in a research journal, it would look something like this:

Elementary school children spent significantly more time reading in the presence of a teacher than when the teacher was absent, $t(7) = 2.804, p < .05$.

FVI

The mean difference and the estimated standard error for difference scores are entered in the formula for a related-samples t test.

FIGURE 10.3 Making a Decision for Example 10.1

The test statistic falls in the rejection region (it is beyond the critical value). Hence, we reject the null hypothesis.

MAKING SENSE INCREASING POWER BY REDUCING ERROR

As illustrated in Table 10.3, there are three places where scores can differ: Two are attributed to error (within-groups and between-persons), and one is tested by the null hypothesis (between-groups). When we compute difference scores, we eliminate one of the sources of error (i.e., between-persons error), as illustrated in Table 10.4, thereby increasing the power of a related-samples t test compared to the two-independent-sample t test. To see how this increases power, we can take another look at Example 10.1.

In Example 10.1, the null hypothesis was that the mean difference equals 0, and we computed the following test statistic:

$$t_{\text{obt}} = \frac{\text{mean difference}}{\text{estimate for error}} = \frac{15}{5.35} = 2.804, p < .05.$$

The test statistic reached significance; we decided to reject the null hypothesis. Now, let us suppose that different participants were assigned to each group and we obtained the same data shown in Table 10.5. In that case, we would compute the two-independent-sample t test, and we would not reduce the scores in each group to one column of difference scores.

When the two columns (teacher present, teacher absent) are not reduced to one column of difference scores, error is computed with both between-persons and within-groups as sources of error. When we compute the two-independent-sample t test, the mean difference is still 15, but our estimate for standard error will therefore be larger because both sources of error are included in its calculation. When we analyze these data using the test statistic for the two-independent-sample t test, we obtain

$$t_{\text{obt}} = \frac{\text{mean difference}}{\text{estimate for error}} = \frac{15}{18.10} = 0.829, p > .05.$$

Notice that the estimate for standard error in the denominator is indeed larger. In this case, our decision will change as a result. Using the two-independent-sample t test, we have not detected the effect or mean difference between groups. Instead, for the two-independent-sample t test, we decide to retain the null hypothesis. All other things being equal, the related-samples t test reduces our estimate of standard error. Thus, as a general rule, the related-samples t test will have greater power to detect an effect or mean difference than the two-independent-sample t test.

LEARNING CHECK 3

- What is the test statistic for a related-samples t test with a mean difference equal to 5 and an estimated standard error for difference scores equal to each of the following?
(a) 2.5 (b) 5.0 (c) 10.0
- The test statistic for a related-samples t test is +2.100. Using a .05 level of significance, would you retain or reject the null hypothesis under each of the following conditions?
 - $df = 10$, two-tailed test
 - $df = 30$, one-tailed test (upper-tail critical)
 - $df = 20$, two-tailed test
 - $df = 40$, one-tailed test (lower-tail critical)
- How does computing difference scores increase the power of a related-samples t test compared to a two-independent-sample t test?

Answers: 1. (a) $t_{\text{obt}} = \frac{5}{2.5} = 2.000$, (b) $t_{\text{obt}} = \frac{5}{5} = 1.000$, (c) $t_{\text{obt}} = \frac{5}{10} = 0.500$; 2. (a) Retain the null hypothesis, (b) Reject the null hypothesis, (c) Reject the null hypothesis, (d) Retain the null hypothesis; 3. Computing difference scores eliminates between-persons error, thereby reducing the estimate for standard error in the denominator.

10.4 MEASURING EFFECT SIZE FOR THE RELATED-SAMPLES t TEST

Hypothesis testing identifies whether or not an effect exists. In Example 10.1, we concluded that an effect does exist—elementary school children spent significantly more time reading in the presence of a teacher than when the teacher was absent; we rejected the null hypothesis. The size of this effect is determined by measures of effect size. We will compute effect size for Example 10.1 because the decision was to reject the null hypothesis for that hypothesis test. There are three measures of effect size for the related-samples t test: estimated Cohen's d , and two measures of proportion of variance (eta-squared and omega-squared).

FYI

The standard deviation of difference scores is used to estimate the population standard deviation in the formula for Cohen's d .

Estimated Cohen's d

An estimated Cohen's d is the most common measure of effect size used with the t test. For two related samples, d measures the number of standard deviations that mean difference scores shifted above or below the population mean difference stated in the null hypothesis. The larger the value of d , the larger the effect in the population. To compute estimated Cohen's d with two related samples, we place the mean difference between two samples in the numerator and the standard deviation of the difference scores to estimate the population standard deviation in the denominator:

$$d = \frac{M_D}{s_D}$$

In Example 10.1, the mean difference was $M_D = 15$, and the standard deviation of difference scores was $s_D = 15.12$. The estimated Cohen's d is

$$d = \frac{15}{15.12} = 0.99.$$

We conclude that time spent reading in the presence of a teacher is 0.99 standard deviations longer than when the teacher is absent for the population. The effect size conventions listed in Table 7.6 (Chapter 7, p. 197) and Table 8.4 (Chapter 8, p. 223) show that this is a large effect size ($d > 0.8$). We could report this measure with the significant t test in Example 10.1 by stating,

Elementary school children spent significantly more time reading in the presence of a teacher than when the teacher was absent, $t(7) = 2.804, p < .05 (d = 0.99)$.

Proportion of Variance

Another measure of effect size is proportion of variance. Specifically, this is an estimate of the proportion of variance in a dependent variable that can be explained by a treatment. In Example 10.1, we can measure the proportion of variance in time spent reading (the dependent variable) that can be explained by whether children read more in the presence or absence of a teacher (the treatment). Two measures of proportion of variance for the related-samples t test are eta-squared and omega-squared. The calculations are the same as those for the one-sample and two-independent-sample t tests.

In Example 10.1, we found that elementary school children spent significantly more time reading in the presence of a teacher than when the teacher was absent, $t(7) = 2.804, p < .05$. To compute eta-squared, we substitute $t = 2.804$ and $df = 7$:

$$\eta^2 = \frac{t^2}{t^2 + df} = \frac{(2.804)^2}{(2.804)^2 + 7} = .53.$$

Using eta-squared, we conclude that 53% of the variability in time spent reading can be explained by whether or not the teacher was present.

Again, eta-squared tends to overestimate the size of an effect. To correct for this, we can compute omega-squared. To compute omega-squared, we subtract 1 from t^2 in the numerator of the eta-squared formula, thereby reducing the value of the proportion of variance. To compute omega-squared, we again substitute $t = 2.804$ and $df = 7$:

$$\omega^2 = \frac{t^2 - 1}{t^2 + df} = \frac{(2.804)^2 - 1}{(2.804)^2 + 7} = .46.$$

Using omega-squared, we conclude that 46% of the variability in time spent reading can be explained by whether or not the teacher was present. The effect size conventions listed in Table 8.4 (Chapter 8, p. 223) show that both measures of proportion of variance estimate a large effect size in the population. We would report only one measure in a research journal. Using omega-squared, we could report this measure with the significant t test in Example 10.1 by stating,

Elementary school children spent significantly more time reading in the presence of a teacher than when the teacher was absent, $t(7) = 2.804, p < .05 (\omega^2 = .46)$.

FVI

Proportion of variance is computed in the same way for all t tests.

LEARNING CHECK 4

1. What is the denominator for the estimated Cohen's d when used with the related-samples t test?
2. The mean difference between two related samples is 4.00. What is the value for an estimated Cohen's d if a sample standard deviation for the difference scores in this example is equal to (a) 4 and (b) 40?
3. A psychologist reports that the value of the test statistic for two related samples is 2.400. If the degrees of freedom are 20, then what is the value for eta-squared? Is this a small, medium, or large effect size?
4. Using the same data as in Question 3, what is the proportion of variance explained using omega-squared? Is this a small, medium, or large effect size?

Answers: 1. The standard deviation for the difference scores; 2. (a) $d = \frac{4}{4} = 1.00$, (b) $d = \frac{40}{4} = 10.00$; 3. $\eta^2 = \frac{2.400^2}{2.400^2 + 20} = \frac{5.76}{5.76 + 20} = \frac{5.76}{25.76} = 0.22$ (medium effect size); 4. $\omega^2 = \frac{4}{(2.400)^2 + 20} = \frac{4}{5.76 + 20} = \frac{4}{25.76} = 0.155$ (medium effect size).

10.5 CONFIDENCE INTERVALS FOR THE RELATED-SAMPLES t TEST

In Example 10.1, we stated a null hypothesis regarding the mean difference in a population. We can also learn more about the mean difference in a population using a different procedure without ever deciding to retain or reject a null hypothesis. The alternative approach requires only that we set limits for the population parameter within which it is likely to be contained. The goal of this alternative approach, called *estimation*, is the same as that in hypothesis testing for Example 10.1—to learn more about the value of a mean in a population of interest.

To use estimation, we identify the sample mean (a point estimate) and give an interval within which a population mean is likely to be contained (an interval estimate). Same as we did for the t tests in Chapter 8 and Chapter 9, we find the interval estimate, often reported as a confidence interval, and state it within a given level of confidence, which is the likelihood that an interval contains an unknown population mean. To illustrate, we will revisit Example 10.1, and using the same data, we will compute the confidence intervals at a 95% level of confidence using the three steps to estimation first introduced in Chapter 8. For a related-samples t test, the estimation formula is:

$$M_D \pm t(s_{MD}).$$

Step 1: Compute the sample mean and standard error. The mean difference, which is the point estimate of the population mean difference, is equal to $M_D = 15$.

The estimated standard error for difference scores, s_{MD} , is equal to 5.35.

Step 2: Choose the level of confidence and find the critical values at that level of confidence. In this example, we want to find the 95% confidence interval, so we choose a 95% level of confidence. Remember, in a sampling distribution, 50% of the mean differences fall above the mean difference we selected in our sample, and 50% fall below it. We are looking for the 95% of mean differences that surround the mean difference we selected in our sample. A 95% CI corresponds to a two-tailed test at a .05 level of significance. To find the critical value at this level of confidence, we look in the t table in

Table B.2 in Appendix B. The degrees of freedom are 7 ($df = n_D - 1$ for two related samples). The critical value for the interval estimate is $t = 2.365$.

Step 3: Compute the estimation formula to find the confidence limits for a 95% confidence interval. Because we are estimating the mean difference between two related samples from a population with an unknown variance, we use the $M_D \pm t(S_{MD})$ estimation formula.

To compute the formula, multiply t by the estimated standard error for difference scores:

$$t(s_{MD}) = 2.365(5.35) = 12.65.$$

Add 12.65 to the sample mean difference to find the upper confidence limit:

$$M_D + t(s_{MD}) = 15 + 12.65 = 27.65.$$

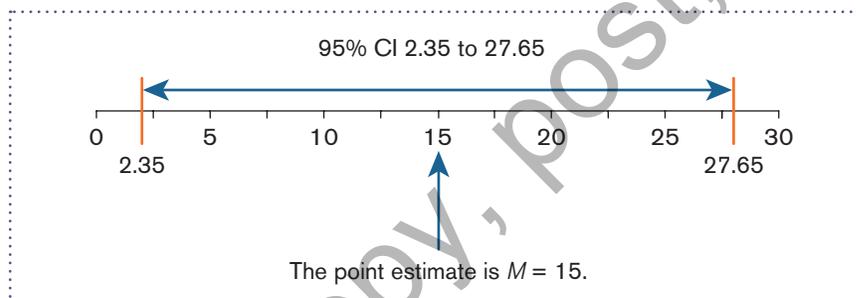
Subtract 12.65 from the sample mean difference to find the lower confidence limit:

$$M_D - t(s_{MD}) = 15 - 12.65 = 2.35.$$

As shown in Figure 10.4, the 95% confidence interval in this population is between 2.35 and 27.65 seconds. We can estimate within a 95% level of confidence that the children spent more time reading with the teacher present than with no teacher present during a 6-minute session.

FIGURE 10.4

In Example 10.1, the 95% Confidence Interval is 2.35 to 27.65 Seconds



10.6 INFERRING SIGNIFICANCE AND EFFECT SIZE FROM A CONFIDENCE INTERVAL

Notice that we did not make a decision using estimation, other than to state confidence limits for a 95% confidence interval. When we evaluated these same data with hypothesis testing in Example 10.1, we selected a sample to decide whether or not to reject the null hypothesis. While we do not “make a decision” per se using estimation, we can use the confidence limits to determine what the decision would have been using hypothesis testing. As first stated in Chapter 8, in terms of the decisions we make in hypothesis testing,

1. If the value stated by a null hypothesis is inside the confidence interval, the decision is to retain the null hypothesis (not significant).
2. If the value stated by a null hypothesis is outside the confidence interval, the decision is to reject the null hypothesis (significant).

Using these rules, we can thus determine the decision we would have made using hypothesis testing. In Example 10.1 we identified a null hypothesis that

FYI

When the value stated by a null hypothesis is outside the confidence interval, this indicates a significant effect in the population.

the mean difference was 0 between groups. Because a mean difference of 0 falls outside the 95% CI of 2.35 and 27.65 seconds (i.e., it is not a possible value for the mean difference in the population), we would have decided to reject the null hypothesis using hypothesis testing, which was the decision we made using hypothesis testing.

In terms of effect size, when the value stated by a null hypothesis is outside the confidence interval, we can interpret effect size using the confidence limits. Specifically, the effect size for a confidence interval is a range or interval, where the lower effect size estimate is the difference between the value stated in the null hypothesis and the lower confidence limit; the upper effect size estimate is the difference between the value stated in the null hypothesis and the upper confidence limit. Effect size can then be interpreted in terms of a shift in the population. For Example 10.1, we can estimate within a 95% level of confidence that the mean difference in time spent reading in the presence of a teacher is between 2.35 and 27.65 seconds longer than with no teacher present during a 6-minute session.

10.7 SPSS in Focus: Related-Samples *t* Test and Confidence Intervals

In Example 10.1, we tested whether teacher supervision influences the time that elementary school children read. We used a two-tailed test at a .05 level of significance and decided to reject the null hypothesis. Thus, we concluded that elementary school children spent significantly more time reading in the presence of a teacher than when the teacher was absent, $t(7) = 2.804$, $p < .05$. We can confirm this result using SPSS.

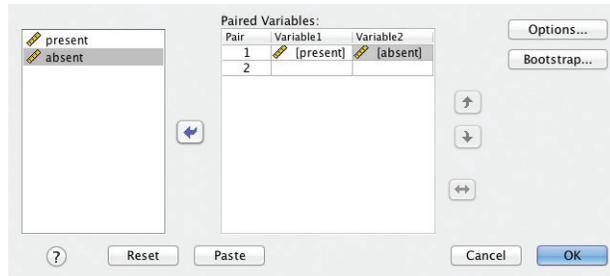
1. Click on the Variable View tab and enter *present* in the Name column. In the second row, enter *absent* in the Name column. We will enter whole numbers, so reduce the value in the Decimals column to 0.
2. Click on the Data View tab. Enter the data in each column as shown in Table 10.7.

TABLE 10.7 Data View in SPSS for Step 2

present	absent
220	210
245	220
215	195
260	265
300	275
280	290
250	220
310	285

- Go to the menu bar and click Analyze, then Compare Means and Paired-Samples T Test, to display the dialog box shown in Figure 10.5.

FIGURE 10.5 SPSS Dialog Box for Steps 3 to 5



- In the dialog box, select *present* and *absent* in the left box and move them to the right box using the arrow in the middle. The variables should be side by side in the box to the right.
- Select OK, or select Paste and click the Run command.

Notice that the calculations we made match the results displayed in the output table shown in Table 10.8. These same step-by-step directions for using SPSS can be used to compute the repeated-measures design (shown in this section) and the matched-pairs design. Finally, note that SPSS gives the confidence intervals for this test in the output table.

TABLE 10.8 The SPSS Output Table for the Related-Samples *t* Test

		Paired Differences							
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	present - absent	15.000	15.119	5.345	2.361	27.639	2.806	7	.026

The value of the test statistic.

Degrees of freedom are $n_b - 1 = 7$.

95% Confidence level

Mean difference or the numerator of the test statistic.

Estimate for standard error or the denominator of the test statistic.

The likelihood that something other than the manipulation (teacher present, absent) produced the differences observed is $p = .026$.

10-8

APA IN FOCUS: REPORTING THE t STATISTIC AND CONFIDENCE INTERVALS

To summarize a related-samples t test, we report the test statistic, degrees of freedom, and p value. In addition, we summarize the means and standard error or standard deviations measured in the study in a figure or table or in the main text. When reporting results, though, it is not necessary to identify the type of t test computed in a results section. The type of t test that was used is typically reported in a data analysis section that precedes the results section, where the statistics are reported.

CHAPTER SUMMARY ORGANIZED BY LEARNING OBJECTIVE

LO 1: Describe two types of research designs used when we select related samples.

- In a related sample, participants are related. Participants can be related in one of two ways: They are observed in more than one group (a repeated-measures design), or they are matched, experimentally or naturally, based on common characteristics or traits (a matched-pairs design).
- The repeated-measures design is a research design in which the same participants are observed in each treatment. Two types of repeated-measures designs are the pre-post design and the within-subjects design.
- The matched-pairs design is a research design in which participants are selected, then matched, experimentally or naturally, based on common characteristics or traits.

LO 2: Explain why difference scores are computed for the related-samples t test.

- To test the null hypothesis, we state the mean difference between paired scores in the population and compare this to the difference between paired scores in a sample. A related-samples t test is different from a two-independent-sample t test in that we first find the difference between the paired scores and then compute the test statistic. The difference between two scores in a pair is called a difference score.

- Computing difference scores eliminates between-persons error. This error is associated with differences associated with observing different participants in each group or treatment. Because we observe the same (or matched) participants in each treatment, not different participants, we can eliminate this source of error before computing the test statistic. Removing this error reduces the value of the estimate of standard error, which increases the power to detect an effect.

LO 3: State three advantages for selecting related samples.

- Three advantages for selecting related samples are that selecting related samples (1) can be more practical, (2) reduces standard error, and (3) increases power.

LO 4-6: Calculate the degrees of freedom for a related-samples t test and locate critical values in the t table; identify the assumptions for the related-samples t test; compute a related-samples t test and interpret the results.

- The related-samples t test is a statistical procedure used to test hypotheses concerning two related samples selected from related populations in which the variance in one or both populations is unknown.
- The degrees of freedom for a related-samples t test are the number of difference scores minus 1: $df = (n_D - 1)$.

- To compute a related-samples t test, we assume normality and independence within groups. The test statistic for a related-samples t test concerning the difference between two related samples is as follows:

$$t_{\text{obt}} = \frac{M_D - \mu_D}{s_{MD}}, \text{ where } s_{MD} = \frac{s_D}{\sqrt{n_D}}.$$

LO 7: Compute and interpret effect size and proportion of variance for a related-samples t test.

- Estimated Cohen's d is the most popular estimate of effect size used with the t test. It is a measure of effect size in terms of the number of standard deviations that mean difference scores shifted above or below the population mean difference stated in the null hypothesis. To compute estimated Cohen's d with two related samples, divide the mean difference (M_D) between two samples by the standard deviation of the difference scores (s_D):

$$d = \frac{M_D}{s_D}.$$

- Another measure of effect size is proportion of variance. Specifically, this is an estimate of the proportion of variance in the dependent variable that can be explained by a treatment. Two measures of proportion of variance for the related-samples t test are eta-squared and omega-squared. These measures are computed in the same way for all t tests:

$$\text{Using eta-squared: } \eta^2 = \frac{t^2}{t^2 + df}.$$

$$\text{Using omega-squared: } \omega^2 = \frac{t^2 - 1}{t^2 + df}.$$

LO 8: Compute and interpret confidence intervals for the related-samples t test.

- The three steps to estimation are as follows:
 - Step 1: Compute the sample mean and standard error.
 - Step 2: Choose the level of confidence and find the critical values at that level of confidence.
 - Step 3: Compute the estimation formula to find the confidence limits.
- The estimation formula for the related-samples t test is $M_D \pm t(s_{MD})$.

LO 9: Summarize the results of a related-samples t test in APA format.

- To report the results of a related-samples t test, state the test statistic, the degrees of freedom, the p value, and the effect size. In addition, summarize the means and the standard error or the standard deviations measured in the study in a figure or a table or in the text. Finally, note that the type of t test computed is reported in a data analysis section that precedes the results section, where the statistics are reported.

LO 10: Compute a related-samples t test and identify confidence intervals using SPSS.

- SPSS can be used to compute a related-samples t test using the Analyze, Compare Means, and Paired-Samples T Test options in the menu bar. These actions will display a dialog box that allows you to identify the groups and run the test (for more details, see Section 10.7).

KEY TERMS

dependent sample
difference scores
error

estimated standard error for
difference scores

matched-pairs design
matched-samples design
matched-subjects design
pre-post design
related sample

related-samples t test
repeated-measures design
within-subjects design

● ● ● END-OF-CHAPTER PROBLEMS

Factual Problems

1. Distinguish between related samples and independent samples.
2. Name two research designs in which related samples are selected.
3. Name and define two repeated-measures designs.
4. State two ways for matching pairs of participants using the matched-pairs research design.
5. State three advantages for using related samples in behavioral research.
6. Describe in words what the degrees of freedom are for a related-samples t test.
7. Define difference scores. How does using difference scores increase the power of a related-samples t test?
8. How does computing difference scores change the value of the estimate of standard error in the denominator of the test statistic for a related-samples t test?
9. What are the assumptions for a related-samples t test?
10. Is the related-samples t test computed differently for a repeated-measures design and a matched-pairs design? Explain.
11. Describe in words the formula for an estimated Cohen's d for two related samples.
12. What are the three steps to compute an estimation formula?

Concept and Application Problems

13. For each example, state whether the one-sample, two-independent-sample, or related-samples t test is most appropriate. If it is a related-samples t test, indicate whether the test is a repeated-measures design or a matched-pairs design.
 - (a) A professor tests whether students sitting in the front row score higher on an exam than students sitting in the back row.
 - (b) A researcher matches right-handed and left-handed siblings to test whether right-handed siblings express greater emotional intelligence than left-handed siblings.
 - (c) A graduate student selects a sample of 25 participants to test whether the average time students attend to a task is greater than 30 minutes.
 - (d) A principal at a local school wants to know how much students gain from being in an honors class. He gives students in an honors English class a test prior to the school year and again at the end of the school year to measure how much students learned during the year.
14. State the degrees of freedom and the type of test (repeated measures or matched pairs) for the following examples of measures for a related-samples t test:
 - (a) The difference in coping ability in a sample of 20 brothers and sisters paired based on their relatedness
 - (b) The difference in comprehension before and after an experimental training seminar in a sample of 30 students
 - (c) The difference in relationship satisfaction in a sample of 25 pairs of married couples
 - (d) The difference in athletic performance at the beginning and end of an athletic season for 12 athletes
15. What are the difference scores for the following list of scores for participants observed at two times?

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

16. Using the data listed in Question 15,
- Compute the mean difference (M_D), standard deviation (s_D), and standard error for the difference scores (s_{MD}).
 - Sketch a graph of the distribution of mean difference scores ($M_D \pm s_D$).
 - Sketch a graph of the sampling distribution of mean difference scores ($M_D \pm s_{MD}$).
17. Would each of the following increase, decrease, or have no effect on the value of the test statistic for a related-samples t test?
- The sample size is increased.
 - The level of significance is reduced from .05 to .01.
 - The estimated standard error for difference scores is doubled.
 - The mean difference is decreased.
18. What is the value of the test statistic for a related-samples t test given the following measurements?
- $n_D = 16$, $M_D = 4$, and $s_D = 8$
 - $M_D = 4$ and $s_{MD} = 8$
 - $n_D = 64$, $M_D = 8$, and $s_D = 16$
 - $M_D = 8$ and $s_{MD} = 16$
19. A statistics tutor wants to assess whether her remedial tutoring has been effective for her five students. Using a pre-post design, she records the following grades for a group of students prior to and after receiving her tutoring.

Tutoring	
Before	After
2.4	3.0
2.5	2.8
3.0	3.5
2.9	3.1
2.7	3.5

- Test whether or not her tutoring is effective at a .05 level of significance. State the value of the test statistic and the decision to retain or reject the null hypothesis.

- Compute effect size using estimated Cohen's d .
20. Published reports indicate that a brain region called the nucleus accumbens (NAC) is involved in interval timing, which is the perception of time in the seconds-to-minutes range. To test this, researchers investigated whether removing the NAC interferes with rats' ability to time the presentation of a liquid reward. Using a conditioning procedure, the researchers had rats press a lever for a reward that was delivered after 16 seconds. The time that eight rats responded the most (peak responding) was recorded before and after a surgery to remove the NAC. The peak responding times are given in the following table.

Peak Interval Timing	
Before NAC Surgery	After NAC Surgery
15	20
14	26
16	20
16	18
17	25
18	21
15	18
16	23

- Test whether or not the difference in peak responding changed at a .05 level of significance (two-tailed test). State the value of the test statistic and the decision to retain or reject the null hypothesis.
 - Compute effect size using estimated Cohen's d .
21. A psychologist wants to know whether wives and husbands who both serve in a foreign war have similar levels of satisfaction in their marriage. To test this, she uses a matched-pairs design by asking six married couples currently serving in a foreign war to rate how satisfied they are with their spouse on a 7-point scale ranging from 1 (*not satisfied at all*) to 7 (*very satisfied*). The following are the responses from husband-and-wife pairs.

Married Couples	
Wife	Husband
7	5
4	6
7	5
7	6
7	5
6	5

- (a) Test whether or not satisfaction ratings differ at a .05 level of significance. State the value of the test statistic and the decision to retain or reject the null hypothesis.
- (b) Compute effect size using eta-squared.
22. A health psychologist noticed that the siblings of his obese patients are often not overweight. He hypothesized that the normal-weight siblings consume fewer daily calories than the obese patients. To test this using a matched-pairs design, he compared the daily caloric intake of 20 obese patients to that of a respective normal-weight sibling. The following are the calories consumed for each sibling pair.

Normal-Weight Sibling	Overweight Sibling
1,600	2,000
1,800	2,400
2,100	2,000
1,800	3,000
2,400	2,400
2,800	1,900
1,900	2,600
2,300	2,450
2,000	2,000
2,050	1,950

- (a) Test whether or not obese patients consumed significantly more calories than their normal-weight siblings at a .05 level

of significance. State the value of the test statistic and the decision to retain or reject the null hypothesis.

- (b) Compute effect size using omega-squared.
- (c) Did the results support the researcher's hypothesis? Explain.
23. State whether each of the following related-samples t tests is significant for a two-tailed test at a .05 level of significance.
- (a) $t(30) = 3.220$
- (b) $t(18) = 2.034$
- (c) $t(12) = 2.346$
- (d) $t(60) = 1.985$
24. A researcher develops an advertisement aimed at increasing how much the public trusts a federal organization. She asks participants to rate their level of trust for the organization before and after viewing an advertisement. Higher ratings indicate greater trust. From the following findings reported in APA format, interpret these results by stating the research design used (repeated-measures or matched-pairs), the sample size, the decision, and the effect size.

Participants rated the federal organization as significantly more trustworthy ($M_D = +4$ points) after viewing the advertisement, $t(119) = 4.021, p < .05, d = 0.88$.

25. A researcher records the amount of time (in minutes) that parent-child pairs spend on social networking sites to test whether they show any generational differences. From the following findings reported in APA format, interpret these results by stating the research design used (repeated-measures or matched-pairs), the sample size, the decision, and the effect size.

Parents spent significantly less time on social networking sites compared to their children ($M_D = -42$ minutes), $t(29) = 4.021, p < .05, d = 0.49$.

26. Would each of the following increase, decrease, or have no effect on the value of estimated Cohen's d for the related-samples t test?
- (a) The sample standard deviation for difference scores is increased.

- (b) The estimated standard error for difference scores is increased.
- (c) The mean difference is increased.
27. An instructor believes that students do not retain as much information from a lecture on a Friday compared to a lecture on a Monday. To test this belief, the instructor teaches a small sample of college students preselected material from a single topic on statistics on a Friday and on a Monday. All students receive a test on the material. The differences in exam scores for material taught on Friday minus Monday are listed in the following table.

Difference Scores (Friday – Monday)	
	+3.4
	–1.6
	+4.4
	+6.3
	+1.0

Problems in Research

29. **The mental health of rescue workers deployed in the 2010 Haiti earthquake.** In a questionnaire-based study, van der Velden, van Loon, Benight, and Eckhardt (2012) looked at factors of mental health among rescue workers before (Time 1) and 3 months after (Time 2) deployment to Haiti following the January 2010 earthquake. The following table summarizes some of the results reported in this study. Based on the results given, which mental health problems were significantly worse following deployment to Haiti among the rescue workers?

Mental Health Problem	Deployment		Significance (p)
	Predeployment $M(SD)$	Postdeployment $M(SD)$	
Anxiety symptoms	10.08 (0.27)	10.40 (0.20)	.414
Depression symptoms	16.29 (0.73)	16.10 (0.30)	.048
Somatic problems	12.37 (0.82)	12.33 (0.68)	.771

- (a) Find the confidence limits at a 95% CI for these related samples.
- (b) Can we conclude that students retained more of the material taught in the Friday class?
28. A researcher hypothesizes that children will eat more of foods wrapped in familiar packaging than the same foods wrapped in plain packaging. To test this hypothesis, she records the number of bites that 24 children take of a food given to them wrapped in fast-food packaging versus plain packaging. If the mean difference (fast-food packaging minus plain packaging) is $M_D = 12$, and $s_{MD} = 2.4$, then:
- (a) Find the confidence limits at a 95% CI for these related samples.
- (b) Can we conclude that wrapping foods in familiar packaging increased the number of bites that children took compared to plain packaging?

Mental Health Problem	Deployment		Significance (p)
	Predeployment $M(SD)$	Postdeployment $M(SD)$	
Sleeping problems	3.37 (0.71)	3.27 (0.75)	.468
Hostility	6.13 (0.55)	6.10 (0.41)	.931
Interpersonal sensitivity	18.72 (1.54)	18.14 (0.35)	.002

30. **Liking for a vanilla milkshake.** Naleid and colleagues (2008) asked how the interaction of the ingredients corn oil and sucrose enhanced liking for vanilla milkshakes. To test this, rats responded on a lever to gain access to sucrose alone and to gain access to sucrose mixed with corn oil. The researchers hypothesized that rats would work harder (respond with a higher rate of lever pressing) for the mixture compared with sucrose alone. They found that adding 1.4% corn oil to a 12.5% sucrose solution enhanced liking (increased responding on the lever) compared to 12.5% sucrose alone, $p < .01$.

(a) Is this a repeated-measures design or a matched-pairs design?

(b) Did these authors find support for their hypothesis?

31. **Posttraumatic stress disorder (PTSD) following 9/11.** Levitt, Malta, Martin, Davis, and Cloitre (2007) evaluated the effectiveness of cognitive behavioral therapy (CBT) for treating PTSD and related symptoms for survivors of the 9/11 terrorist attacks on the World Trade Center (WTC). They used a pretest-posttest design to see if CBT was successful at reducing the symptoms of PTSD and related symptoms of depression. They used the Modified PTSD Symptom Scale Self-Report (MPSS-SR) questionnaire to measure symptoms of PTSD and the Beck Depression Inventory (BDI) self-report questionnaire to measure symptoms of depression. For both questionnaires, lower scores indicated fewer symptoms. The authors reported the following results:

Pre- to post treatment t tests for the WTC sample revealed significant decreases in scores on the MPSS-SR, $t(37) = 12.74$, $p < .01$; as well as on the BDI $t(34) = 7.36$, $p < .01$. (Levitt et al., 2007, p. 1427)

- (a) Was this a repeated-measures design or a matched-pairs design?
- (b) Which questionnaire (MPSS-SR or BDI) was completed by more participants?
- (c) Did the authors find support for their hypothesis? Explain.

32. **Teacher perceptions of children with cancer.** Liang, Chiang, Chien, and Yeh (2007) tested whether teachers for children with cancer (case) perceived these children similarly to a group of healthy children (control). They measured a variety of social factors, including social withdrawal, somatic complaints, and other social problems. The following table summarizes

many of their results using the related-samples t test.

Measures	Mean (<i>SD</i>)		t	p Value
	Case	Control		
Withdrawn	0.32 (2.58)	0.15 (1.45)	3.52	< .01
Somatic complaints	0.19 (2.22)	0.06 (1.24)	3.85	< .001
Social problems	0.26 (3.34)	0.08 (1.32)	4.31	< .001

- (a) Is this a repeated-measures design or a matched-pairs design?
- (b) Interpret the results displayed in the table. Do teachers perceive children with cancer as having more social problems than their peers?

33. **The impact of a mentoring program for highly aggressive children.** Faith, Fiala, Cavell, and Hughes (2011) examined the impact of a mentoring program for elementary school-aged children on attitudes that mentors and children have about mentoring. In their study, they reported that children's and mentor's ratings of the openness of the mentor significantly decreased from before to following the mentoring program, $t(101) = 12.51$.

- (a) What was the sample size for this test?
- (b) Was a repeated-measures or a matched-pairs design used in this study?
- (c) Compute eta-squared for the effect reported.

34. **Confidence intervals, significance, and effect size.** Zou (2007) noted in an article that confidence intervals "encompass significance tests and provide an estimate of the magnitude of the effect" (p. 399). What does "the magnitude of the effect" refer to in this citation?

Answers for even numbers are in Appendix C.



Sharpen your skills with SAGE edge at edge.sagepub.com/priviteraess

SAGE edge for Students provides a personalized approach to help you accomplish your coursework goals in an easy-to-use learning environment.