

LEARNING UNIT 9

t Tests: One-Sample, Two-Independent-Sample, and Related-Samples Designs

Excel Toolbox

Mathematical operators

- +
- -
- ()
- *
- /
- ^2 [square]
- ^,5 [square root]

Functions

- AVERAGE
- COUNT
- STDEV.S
- SUM
- VAR.S
- T.TEST

(Continued)

(Continued)

Other tools

- format cells
- freeze panes
- fill down or paste
- inserting equations
- Analysis ToolPak

In this Learning Unit, we explore the nature of hypothesis testing when one group or two groups are observed; for two groups we explore situations in which the same or different participants are observed in each group. We further explore the informativeness of hypothesis testing for making decisions, and explore other ways of adding information about the nature of observed effects and how to appropriately interpret them. We do this with three different versions of a *t* test:

- one-sample *t* test,
- independent-sample *t* test, and
- related-samples *t* test.

Origins of the *t* Tests

An alternative to the *z* statistic was proposed by William Sealy Gosset (Student, 1908), a scientist working with the Guinness brewing company to improve brewing processes in the early 1900s. Because Guinness prohibited its employees from publishing “trade secrets,” Gosset obtained approval to publish his work only under the condition that he used a pseudonym (“Student”). He proposed substituting the sample variance for the population variance in the formula for standard error. When this substitution is made, the formula for error is called the **estimated standard error** (s_M):

$$\text{Estimated standard error: } s_M = \sqrt{\frac{s^2}{n}} = \frac{SD}{\sqrt{n}}$$

The substitution is possible because, as explained in learning units 2 and 7, the sample variance is an unbiased estimator of the population variance: On average, the sample variance equals the population variance. Using this substitution, an alternative test statistic can be introduced for one sample when the population variance is unknown. The formula, known as a ***t* statistic**, is as follows for one sample:

The **estimated standard error** is an estimate of the standard deviation of a sampling distribution of sample means selected from a population with an unknown variance. It is an estimate of the standard error, or standard distance that sample means can be expected to deviate from the value of the population mean stated in the null hypothesis.

The ***t* statistic**, known as ***t* observed** or ***t* obtained**, is an inferential statistic used to determine the number of standard deviations in a *t* distribution that a sample mean deviates from the mean value or mean difference stated in the null hypothesis.

$$t_{\text{obt}} = \frac{M - m}{s_M}, \text{ where } s_M = \frac{SD}{\sqrt{n}}$$

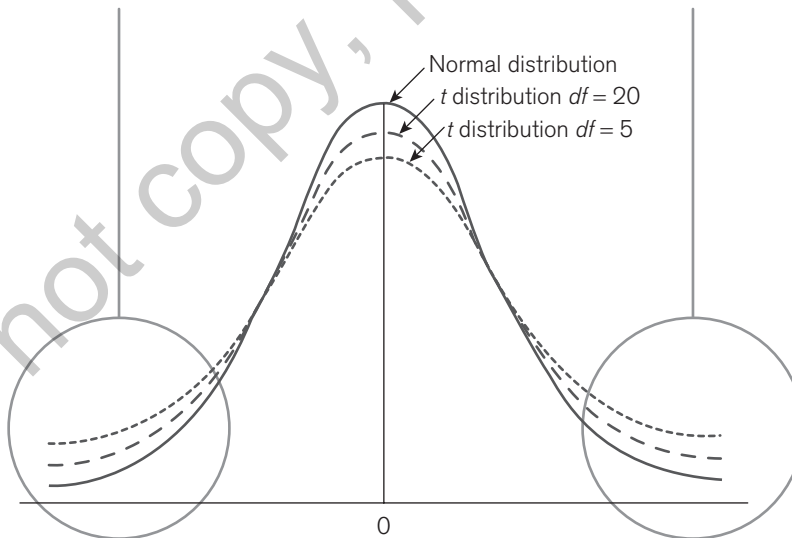
Gosset showed that substituting the sample variance for the population variance led to a new sampling distribution known as the ***t* distribution**, which is also known as **Student's *t***, referring to the pseudonym Gosset used when publishing his work. In Figure 9.1, you can see how similar the *t* distribution is to the normal distribution. The difference is that the *t* distribution has greater variability in the tails, because the sample variance is not always equal to the population variance. Sometimes the estimate for variance is too large; sometimes the estimate is too small. This leads to a larger probability of obtaining sample means farther from the population mean. Otherwise, the *t* distribution shares all the same characteristics of the normal distribution: It is symmetrical and asymptotic, and its mean, median, and mode are all located at the center of the distribution.

The Degrees of Freedom

The *t* distribution is associated with **degrees of freedom (*df*)**. In Learning Unit 2, we identified that the degrees of freedom for sample variance equal $n - 1$. Because the estimate of standard error for the *t* distribution is computed using the

FIGURE 9.1 • A normal distribution and two *t* distributions.

The tails of a *t* distribution are thicker, which reflects the greater variability in values resulting from not knowing the population variance.



The ***t* distribution**, or **Student's *t***, is a normal-like distribution with greater variability in the tails than a normal distribution, because the sample variance is substituted for the population variance to estimate the standard error in this distribution.

The **degrees of freedom (*df*)** for a *t* distribution are equal to the degrees of freedom for sample variance for a given sample: $n - 1$. Each *t* distribution is associated with specified degrees of freedom; as sample size increases, the degrees of freedom also increase.

Notice that the normal distribution has less variability in the tails; otherwise, these distributions share the same characteristics.

Source: www.unl.edu

sample variance, the degrees of freedom for the t distribution are also $n - 1$. The t distribution is a sampling distribution in which the estimated standard error is computed using the sample variance in the formula. As sample size increases, the sample variance more closely approximates the population variance. The result is that there is less

Appendix A12

See **Appendix A12**, p. 298, for more on degrees of freedom for parametric tests.

variability in the tails as sample size increases. So the shape of the t distribution changes (the tails approach the x -axis faster) as the sample size is increased. Each changing t distribution is thus associated with the same degrees of freedom as for sample variance: $df = n - 1$.

To locate probabilities and critical values in a t distribution, we use a t table, such as Table 9.1, which reproduces part of Table C.2 in Appendix C. In the t table, there are six columns of values listing alpha levels for one-tailed tests (top heading) and two-tailed tests (lower heading). The rows show the degrees of freedom (df) for a t distribution.

To use this table, you need to know the sample size (n), the alpha level (α), and the location of the rejection region (in one or both tails). For example, if we select a sample

TABLE 9.1 • A portion of the t table adapted from Table C.2 in Appendix C.

df	Proportion in One Tail					
	.25	.10	.05	.025	.01	.005
	Proportion in Two Tails Combined					
	.50	.20	.10	.05	.02	.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
7	0.711	1.415	1.895	2.365	2.998	3.499
8	0.706	1.397	1.860	2.306	2.896	3.355
9	0.703	1.383	1.833	2.282	2.821	3.250
10	0.700	1.372	1.812	2.228	2.764	3.169

Source: Table III in Fisher, R. A., & Yates, F. (1974). *Statistical tables for biological, agricultural and medical research* (6th ed). London, England: Longman Group Ltd. (previously published by Oliver and Boyd Ltd., Edinburgh). Adapted and reprinted with permission of Addison Wesley Longman.

of 11 students, then $n = 11$, and $df = 10$ ($n - 1 = 10$). To find the t distribution with 10 degrees of freedom, we look for 10 listed in the rows. The critical values for this distribution at a .05 level of significance appear in the column with that probability listed: For a one-tailed test, the critical value is 1.812 for an upper-tail critical test and -1.812 for a lower-tail critical test. For a two-tailed test, the critical values are ± 2.228 . Each critical value identifies the cutoff for the rejection region, beyond which the decision will be to reject the null hypothesis for a hypothesis test.

Keep in mind that a t distribution is an estimate of a normal distribution. The larger the sample size, the more closely a t distribution estimates a normal distribution. When the sample size is so large that it equals the population size, we describe the sample size as infinite. In this case, the t distribution is a normal distribution. You can see this in the t table in Appendix C. The critical values at a .05 level of significance are ± 1.96 for a two-tailed t test with infinite (∞) degrees of freedom and 1.645 (upper-tail critical) or -1.645 (lower-tail critical) for a one-tailed test. These are the same critical values listed in the unit normal table at a .05 level of significance. In terms of the null hypothesis, in a small sample, there is a greater probability of obtaining sample means that are farther from the value stated in the null hypothesis. As sample size increases, obtaining sample means that are farther from the value stated in the null hypothesis becomes less likely. The result is that critical values get smaller as sample size increases.

Computing the One-Sample t Test

In this section, we compute the **one-sample t test**, which is used to compare a mean value measured in a sample to a known value in the population. Specifically, this test is used to test hypotheses concerning a single group mean selected from a population with an unknown variance. To compute the one-sample t test, we make three assumptions:

1. *Normality.* We assume that data in the population being sampled are normally distributed. This assumption is particularly important for small samples. In larger samples ($n > 30$), the standard error is smaller, and this assumption becomes less critical as a result.
2. *Random sampling.* We assume that the data we measure were obtained from a sample that was selected using a random sampling procedure. It is considered inappropriate to conduct hypothesis tests with nonrandom samples.
3. *Independence.* We assume that each outcome or observation is independent, meaning that one outcome does not influence another. Specifically, outcomes are independent when the probability of one outcome has no effect on the probability of another outcome. Using random sampling usually satisfies this assumption.

Keep in mind that satisfying the assumptions for the t test is critically important. That said, for each example in this book, the data are intentionally constructed such that the assumptions for conducting the tests have been met. In Example 9.1 we

The **one-sample t test** is a statistical procedure used to compare a mean value measured in a sample to a known value in the population. It is specifically used to test hypotheses concerning the mean in a single population with an unknown variance.

follow the four steps to hypothesis testing introduced in Learning Unit 7 to compute a one-sample t test at a two-tailed .05 level of significance using an example adapted from published research.

Example 9.1. Learning is a common construct that behavioral sciences study. One common type of learning is the ability to recognize new objects, referred to as novelty recognition (Fisher-Thompson, 2017; Privitera, Mayeaux, Schey, & Lapp, 2013). An example with animals is the object recognition task. A mouse placed in an environment with two identical objects for five minutes is later returned to the same environment, but one of the objects has been replaced with a new or *novel* object. Because mice are naturally curious, we expect that the mouse will spend more time investigating the novel object, thus demonstrating object recognition. To operationalize, or make measurable, the percentage of time spent investigating the novel object relative to the familiar object, we make the following calculation:

$$\frac{\text{Time Spent Investigating Novel Object}}{\text{Time Spent Investigating Novel Object} + \text{Time Spent Investigating Familiar Object}} \times 100$$

Using this formula, if a mouse spends the same amount of time investigating each object (in other words, the mouse fails to show object recognition), then the result will be 50%. Thus, our standard we will compare against in the null hypothesis for this test will be 50%. A score below 50% indicates that subjects recognized the novel object but preferred the familiar object. Although unlikely, familiarity preference is a remote possibility.

Using a sample data set adapted from published research, we will use the four steps to hypothesis testing introduced in Learning Unit 7 to test whether the mean score in sample data significantly differs from the expected value of 50% at a .05 level of significance.

Step 1: State the hypotheses. The population mean is 50%, and we are testing whether or not the population mean differs from the sample mean:

$H_0 : \mu = 50\%$ For mice given the opportunity to investigate a novel and a familiar object, the mean percentage of time spent investigating the novel object is equal to 50%, as would be expected by chance.

$H_1 : \mu \neq 50\%$ For mice given the opportunity to investigate a novel and a familiar object, the mean percentage of time spent investigating the novel object is not equal to 50%.

Again, if a mouse spends the same amount of time investigating each object (in other words, the mouse fails to show object recognition), then the result will be 50%. Thus, our standard we will compare against in the null hypothesis for this test is 50%. The higher the percentage above 50%, the more time the mouse spent investigating the novel object, and thus the more likely we will be to reject the null hypothesis and conclude that object recognition occurred.

Step 2: Set the criteria for a decision. The level of significance for this test is .05. We are computing a two-tailed test with $n - 1$ degrees of freedom. We will use a

data set with 15 scores, each from a different mouse, a sample size that is appropriate for this behavioral task in research with nonhumans. With $n = 15$, the degrees of freedom for this test are $15 - 1 = 14$. To locate the critical values, we find 14 listed in the rows of Table C.2 in Appendix C and go across to the column for a .05 proportion in two tails combined. The critical values are ± 2.145 .

We will compare the value of the test statistic with these critical values. If the value of the test statistic is beyond a critical value (either greater than 2.145 or less than -2.145), then there is less than a 5% chance we would obtain that outcome if the null hypothesis were correct, so we reject the null hypothesis; otherwise, we retain the null hypothesis.

Step 3: Compute the test statistic. Download Novel_Objects.xlsx from the student study site: <http://study.sagepub.com/priviteraexcel1e>. As shown in Figure 9.2, Column A contains an ID number for each animal; Column B contains the percentage of total investigation time each animal devoted to the novel object. Column C, which we save for use later, contains the expected percentage of time each animal would have devoted to the novel object if it did not show a preference for either the novel or the familiar object.

As shown in Figure 9.2, we insert in column D some labels to keep track of our calculations in column E for the one-sample *t* test:

D4: Mean (M)

D5: Sample size (n)

D6: Standard deviation (SD)

D7: Degrees of freedom (df)

D8: Critical value of t (t_{crit})

We covered mean in Learning Unit 1 and standard deviation in Learning Unit 2.

To the right of the cells mentioned above, we type these functions and formulas into column E:

E4: =AVERAGE(B4:B18)

E5: =COUNT(B4:B18)

E6: =STDEV.S(B4:B18)

E7: =E5-1

E8: 2.145

At this point we have what we need to proceed with our calculation. From the values we have calculated already in rows 4 to 8, we prepare column D with three more labels:

Appendix A12

See **Appendix A12**, p. 298, for more on degrees of freedom for parametric tests.

Appendix B

See **Appendix B2**, p. 301, for formatting cells.

D9: Estimated standard error (s_M)

D10: Obtained value of t (t_{obt})

D11: p value

On our way to finding the t statistic, we compute the estimated standard error. To compute the estimated standard error, we divide the sample standard deviation by the square root of the sample size:

$$s_M = \frac{SD}{\sqrt{n}}$$

In column E,

E9: =E6/E5^0.5

which yields

$$s_M = \frac{9.4}{\sqrt{15}} = 2.42$$

in cell E9 in Figure 9.2b.

We will compare the sample mean to the population mean stated in the null hypothesis: $\mu = 50$. The estimated sample standard deviation is the denominator of the t statistic.

$$t_{obt} = \frac{M - \mu}{s_M}$$

Find the t statistic by substituting the values for the sample mean, $M = 59.1$; the population mean stated in the null hypothesis, $\mu = 50$; and the estimated standard error we just calculated, $s_M = 2.42$. In column E,

E10: =(E4-50)/E9

which yields

$$t_{obt} = \frac{59.1 - 50}{2.42} = 3.74 \text{ in cell E10 in Figure 9.2b.}$$

Note that although there is no function in Excel to calculate a t value, there is a function to calculate the p value associated with a t test. To calculate an exact p value for a one-sample t test, we use a second column of expected values equal to 50% for each of the 15 mice, shown in column C in Figure 9.2. We use the T.TEST function built into Excel.

In column E,

E11: =T.TEST(B4:B18,C4:C18,2,1).

This function requires two cell ranges of data: B4:B18 contains the observed percentage of time spent investigating the novel object, and C4:C18 contains the expected percentage of time spent sniffing the novel object: 50% in each cell. After those two

FIGURE 9.2 • One-sample *t* test. (a) Functions and formulas. (b) Resulting calculations from functions and formulas.

(a)

	A	B	C	D	E
1		Percentage of total			
2		time spent exploring	Expected		
3	Mouse #	novel object	percentage		
4	1	55	50	Mean (<i>M</i>) =AVERAGE(B4:B18)	
5	2	63	50	Sample size (<i>n</i>) =COUNT(B4:B18)	
6	3	56	50	Standard deviation (<i>SD</i>) =STDEV.S(B4:B18)	
7	4	73	50	Degrees of freedom (<i>df</i>) =E5-1	
8	5	54	50	Critical value of <i>t</i> (<i>t</i> _{crit}) =2.145	
9	6	74	50	Estimated standard error (<i>s_M</i>) =E6/E5^0.5	
10	7	57	50	Obtained value of <i>t</i> (<i>t</i> _{obt}) =(E4-50)/E9	
11	8	53	50	<i>p</i> value =T.TEST(B4:B18,C4:C18,2,1)	
12	9	63	50		
13	10	76	50	Estimated Cohen's <i>d</i> =(E4-50)/E6	
14	11	45	50	Eta-squared (η^2) =E10^2/(E10^2+E7)	
15	12	46	50		
16	13	61	50	<i>t</i> (<i>s_M</i>) =E8*E9	
17	14	55	50	95% CI upper limit =E4+E16	
18	15	55	50	95% CI lower limit =E4-E16	

(b)

	A	B	C	D	E
1		Percentage of total			
2		time spent exploring	Expected		
3	Mouse #	novel object	percentage		
4	1	55	50	Mean (<i>M</i>)	59.1
5	2	63	50	Sample size (<i>n</i>)	15
6	3	56	50	Standard deviation (<i>SD</i>)	9.4
7	4	73	50	Degrees of freedom (<i>df</i>)	14
8	5	54	50	Critical value of <i>t</i> (<i>t</i> _{crit})	2.145
9	6	74	50	Estimated standard error (<i>s_M</i>)	2.42
10	7	57	50	Obtained value of <i>t</i> (<i>t</i> _{obt})	3.74
11	8	53	50	<i>p</i> value	.002
12	9	63	50		
13	10	76	50	Estimated Cohen's <i>d</i>	0.97
14	11	45	50	Eta-squared (η^2)	0.50
15	12	46	50		
16	13	61	50	<i>t</i> (<i>s_M</i>)	5.20
17	14	55	50	95% CI upper limit	64.3
18	15	55	50	95% CI lower limit	53.9

Appendix B

See **Appendix B2**, p. 301, on formatting cells to add superscripts.

ranges of data, the next argument required in the function is the number of tails, for which we specify 2. The final argument is the type of t test, which we specify as related-samples, which Excel terms “Paired,” with a 1. As expected with such a large t_{obt} , the p value returned of .002 is small, shown in cell E11 in Figure 9.2b.

Step 4: Make a decision. To decide to reject or retain the null hypothesis, we compare the obtained value ($t_{\text{obt}} = 3.74$) to the critical values in the t table in Appendix C2. For $df = n - 1$, $15 - 1 = 14$, the critical value at $\alpha = .05$ is 2.145. Because t_{obt} of 3.74 exceeds the critical value, the decision is to reject the null hypothesis. This t_{obt} indicates that our observed value of 9.1 percentage points above the expected value of 50 percentage points is 3.74 times larger than the average deviation of 2.42 percentage points of a mean based on 15 samples. If this result were reported in a research journal, it would look something like this following APA format (APA, 2010):

The percentage of time mice explored the novel object ($M = 59.1$, $SD = 9.4$) was significantly higher than the percentage expected by chance, $t(14) = 3.74$, $p = .002$. Thus, the results support the conclusion that the mice demonstrated object recognition.

Effect Size for the One-Sample t Test

As described in Learning Unit 7, hypothesis testing identifies whether an effect exists in a population. When we decide to retain the null hypothesis, we conclude that an effect does not exist in the population. When we decide to reject the null hypothesis, we conclude that an effect does exist in the population. However, hypothesis testing does not tell us how large the effect is.

In Example 9.1, we concluded that mice investigated a novel object more than they investigated a familiar object. To determine the size of an effect, we compute effect size, which gives an estimate of the size of an effect in the population. Two measures of effect size for the one-sample t test are described in this section: estimated Cohen's d and proportion of variance (eta squared).

To label these calculations, in column D we enter

D13: Estimated Cohen's d

D14: Eta squared (η^2)

Estimated Cohen's d . The estimate of effect size that is most often used with a t test is the **estimated Cohen's d** . As described at the beginning of this learning unit on the t test, when the population standard deviation is unknown, we use the sample standard deviation, because it gives an unbiased estimate of the population standard deviation. Similarly, with the estimated Cohen's d formula, we use the sample standard deviation as follows:

$$d = \frac{M - \mu}{SD}$$

Estimated Cohen's d is a measure of effect size in terms of the number of standard deviations that mean scores shifted above or below the population mean stated by the null hypothesis. The larger the value of d , the larger the effect in the population.

In column E,

$$E13: =(E4-50)/E6$$

which yields

$$d = \frac{59.1 - 50.0}{9.4} = \frac{9.1}{9.4} = 0.97$$

in cell E13 in Figure 9.2b.

We conclude that novelty of an object will increase investigation by mice of that object by 0.97 standard deviations above the expectation of equal investigation of familiar and novel objects. The effect size conventions (Cohen, 1988) given in the middle column of Table 9.2 show that this is a large effect size. We could report this measure with the significant *t* test in Example 9.1 by stating,

The percentage of time mice explored the novel object ($M = 59.1$, $SD = 9.4$) was significantly higher than the percentage expected by chance, $t(14) = 3.74$, $p < .01$, $d = 0.97$. Thus, the results support the conclusion that the mice demonstrated object recognition.

Proportion of Variance: Eta squared (η^2). Another measure of effect size is to estimate the **proportion of variance** that can be accounted for by some **treatment**. A treatment, which is any unique characteristic of a sample or any unique way that a researcher treats a sample, can change the value of a dependent variable. A treatment is associated with variability in a study. Proportion of variance estimates how much of the variability in a dependent variable can be accounted for by the treatment. In the proportion of variance formula, the variability explained by a treatment is divided by the total variability observed:

$$\text{Proportion of variance} = \frac{\text{variance explained}}{\text{total variance}}$$

In Example 9.1, we found that mice investigated a novel object more than they investigated a familiar object. The unique characteristic of the sample in this study was that the mice encountered a novel object that attracted their attention, not just two familiar objects. The variable we measured (i.e., the dependent variable) was percentage of total time investigating that was devoted to the novel object. Measuring proportion of variance determines how much of the variability in the dependent variable (percentage of investigation time) can be explained by the treatment (the fact that one of the objects was novel). Here, we describe a measure of proportion of variance, eta squared (η^2).

Eta squared is a measure of proportion of variance that can be expressed in a single formula based on the result of a *t* test:

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

Proportion of variance

is a measure of effect size in terms of the proportion or percentage of variability in a dependent variable that can be explained or accounted for by a treatment.

In hypothesis testing, a **treatment** is any unique characteristic of a sample or any unique way that a researcher treats a sample.

In this formula, t is the value of the t statistic, and df is the degrees of freedom. In this example, $t = 3.74$, and $df = 14$. To find variance, we square the standard deviation. Thus, in the eta squared formula, we square the value of t to find the proportion of variance. In column E,

$$E14: =E10^2/(E10^2+E7)$$

which yields

$$\eta^2 = \frac{3.74^2}{3.74^2 + 14} = \frac{13.9876}{13.9876 + 14} = 0.50$$

in cell E14 in Figure 9.2b.

We conclude that 50% of the variability in the percentage of time spent investigating objects (the dependent variable) can be explained by the fact that one of the objects was novel (the treatment). We could report this measure with the significant t test in Example 9.1 by stating,

The percentage of time mice explored the novel object ($M = 59.1$, $SD = 9.4$) was significantly higher than the percentage expected by chance, $t(14) = 3.74$, $p < .01$ ($\eta^2 = .50$). Thus, the results support the conclusion that the mice demonstrated object recognition.

The third column in Table 9.2 displays guidelines for interpreting a trivial, small, medium, and large effect for a variety of measures for effect size, including η^2 . Using this table, we find that $\eta^2 = .4998$ is a large effect. Although eta squared is a popular measure of proportion of variance, it tends to overestimate the proportion of variance explained by a treatment. To correct for this bias, many researchers use a modified eta squared formula, called omega-squared. Coverage of omega-squared is beyond the scope of this book.

TABLE 9.2 • The size of an effect using estimated Cohen's d and proportion of variance (eta squared).

Description of Effect	d	η^2	ω^2
Trivial	—	$\eta^2 < .01$	$\omega^2 < .01$
Small	$d < 0.2$	$.01 < \eta^2 < .09$	$.01 < \omega^2 < .09$
Medium	$0.2 < d < 0.8$	$.10 < \eta^2 < .25$	$.10 < \omega^2 < .25$
Large	$d > 0.8$	$\eta^2 > .25$	$\omega^2 > .25$

Note that Cohen's d is interpreted the same with negative values. The sign (+, -) simply indicates the direction of the effect.

Confidence Intervals for the One-Sample *t* Test

In Example 9.1, we stated a null hypothesis regarding the value of the mean in a population. We can further describe the nature of the effect by determining where the effect is likely to be in the population by computing the confidence intervals.

As introduced in Learning Unit 7, there are two types of estimates: a point estimate and an interval estimate. When using one sample, a *point estimate* is the sample mean we measure. The interval estimate, reported as a *confidence interval*, is stated within a given *level of confidence*, which is the likelihood that an interval contains an unknown population mean.

To illustrate confidence intervals for the one-sample *t* test, we will revisit Example 9.1 to compute the confidence intervals at a 95% level of confidence for the data analyzed using the one-sample *t* test. To find the confidence intervals, we need to evaluate an estimation formula. We will use the estimation formula to identify the upper and lower confidence limits within which the unknown population mean is likely to be contained. The estimation formula for the one-sample *t* test is as follows:

$$M \pm t(s_M)$$

In all, we follow three steps to estimate the value of a population mean using a point estimate and an interval estimate:

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.

Step 1: Compute the sample mean and standard error. We have already computed the sample mean, which is the point estimate of the population mean, $M = 59.1$ in cell E4 of Figure 9.2b. We have also already computed the standard error of the mean, which is the sample standard deviation divided by the square root of the sample size, $S_M = 2.42$ in cell E9 of Figure 9.2b.

Step 2: Choose the level of confidence and find the critical values at that level of confidence. In this example, we chose the 95% confidence interval (CI). The critical value at this level of confidence will be the same as we found in Step 2 for Example 9.1 using hypothesis testing. As shown in Table 9.3, the 95% level of confidence corresponds to a two-tailed test at a .05 level of significance using hypothesis testing. Thus, the critical value for the interval estimate is 2.145, as shown in cell E8 of Figure 9.2b.

To explain further how this critical value was determined, remember that in a sampling distribution, 50% of sample means fall above the sample mean we selected, and 50% fall below it. We are looking for the 95% of sample means that surround the sample mean we selected, meaning the 47.5% of sample means above and the 47.5% of sample means below the sample mean we selected. This leaves only 2.5% of sample means remaining in the upper tail and 2.5% in the lower tail. Table 9.3 shows how

TABLE 9.3 • Levels of significance using hypothesis testing and the corresponding levels of confidence using estimation.

Level of Confidence	Level of Significance (α level, two-tailed)
99%	.01
95%	.05
90%	.10
80%	.20

different levels of confidence using estimation correspond to different two-tailed levels of significance (α) using hypothesis testing. Referring to Table 9.3, we find that 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 C.2 in Appendix C. The degrees of freedom are 14 ($df = n - 1$ for a one-sample t test). The critical value for the interval estimate is 2.145. Multiplying the observed standard error of the mean by the critical value of t tells us how far above the sample mean 47.5% of all sample means would fall and how far below the mean another 47.5% of all sample means would fall. This range above and below the sample mean encompasses 95% of sample means.

Step 3: Compute the estimation formula to find the confidence limits for a 95% confidence interval. In column D,

$$D16: t(s_M)$$

$$D17: 95\% \text{ CI upper limit}$$

$$D18: 95\% \text{ CI lower limit}$$

To compute the formula, multiply t by the estimated standard error. In column E,

$$E16: =E8 * E9$$

which yields

$$t(s_M) = 2.145(2.42) = 5.20$$

in cell E16 in Figure 9.2b.

Add 5.20 to the sample mean to find the upper confidence limit, and subtract 5.20 from the sample mean to find the lower confidence limit. In column E,

$$E17: =E4 + E16$$

$$E18: =E4 - E16$$

which yields

$$M + t(s_M) = 59.1 + 5.20 = 64.3$$

in cell E17 in Figure 9.2b, and

$$M - t(s_M) = 59.1 - 5.20 = 53.9$$

in cell E18 in Figure 9.2b.

As shown in Figure 9.3, the 95% confidence interval in this population is between a percentage of 53.9% and 64.3% of investigation time directed toward a novel object. We can estimate within a 95% level of confidence that the mean percentage of time investigating a novel object is between 53.9% and 64.3% in the population. We are 95% confident that the population mean falls within this range, because 95% of all sample means we could have selected from this population fall within the range of sample means we specified.

Computing the One-Sample *t* Test Using the Analysis Toolpak

We can also calculate this *t* test using the Analysis ToolPak available in Excel for easy and accurate calculation. We'll guide you through the steps to do the analysis we did for the one-sample *t* test.

Return to the workbook *Novel_Objects.xlsx*. Click on the Data tab, and then on the Data Analysis icon all the way to the right. Select "t-Test: Paired Two Sample for Means," as shown in Figure 9.4a. As we mentioned above, we can get the same result from a one-sample *t* test as we can from a related-samples *t* test (which is called a Paired *t*-Test in Excel) when we pair the value predicted by the null hypothesis, column C in Figure 9.2, with each score that was measured, column B in Figure 9.2.

Selecting "t-Test: Paired Two Sample for Means" yields the dialog box in Figure 9.4b. For Variable 1, we select the observed values of the percentage of time spent exploring the novel object in cells B3 through B18, which includes in B3 a label for the data.

FIGURE 9.3 • At a 95% CI, the trust population mean score falls between 53.9 and 64.3 in this population of curious mice.

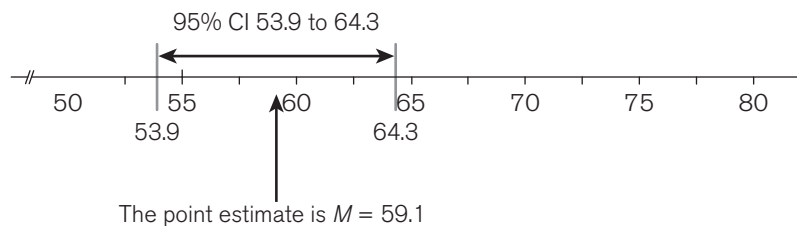
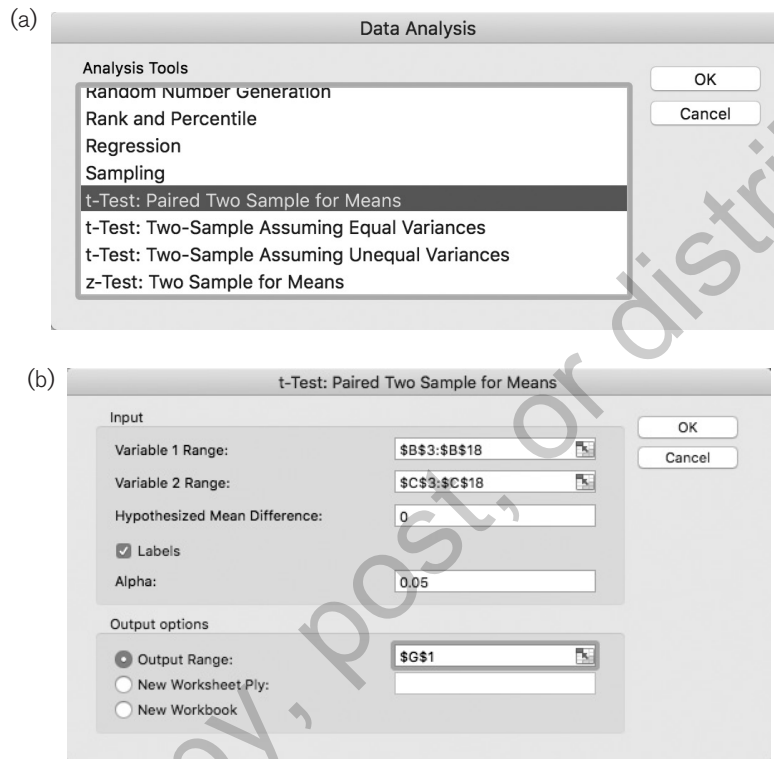


FIGURE 9.4 • Performing a one-sample t test with the Analysis ToolPak in Excel. (a) Selecting “t-Test: Paired Two Sample for Means” to perform a one-sample t test. (b) Specifying the location of the data and parameters for the t test.



For Variable 2, we select the expected percentages in C3 through C18, all of which are 50, and include in C3 a label for the data. The Hypothesized Mean Difference we expect to be 0. (Zero is the default if this box remains blank.) Check the Labels box so that the output contains the labels from B3 and C3. We keep our output on the same page by selecting Output Range and clicking in cell G1.

Clicking OK on the dialog box returns the output table in Figure 9.5. The labels that we included in the cell range for the analysis are in H3 and I3. We can change these labels as we desire. Notice that in Figure 9.5, we get the same mean of 59.1% in cell H4, same t_{obt} of 3.74 in cell H10, and same two-tailed p value of .002 in cell H13 as we did in Figure 9.2b. Although neither an estimate of effect size nor confidence intervals are generated automatically, the output table gives the mean from which we would subtract the expected value of 50%, the variance, and the degrees of freedom. With this information we can calculate effect size and confidence intervals as we did above.

FIGURE 9.5 • Results of one-sample *t* test using the Analysis ToolPak.

	G	H	I
1	t-Test: Paired Two Sample for Means		
2			
3		<i>novel object</i>	<i>percentage</i>
4	Mean	59.1	50.0
5	Variance	88.1	0
6	Observations	15	15
7	Pearson Correlation	#DIV/0!	
8	Hypothesized Mean Difference	0	
9	df	14	
10	t Stat	3.74	
11	P(T<=t) one-tail	.001	
12	t Critical one-tail	1.76	
13	P(T<=t) two-tail	.002	
14	t Critical two-tail	2.14	

Computing the Two-Independent-Sample *t* Test

In this section, we compute the **two-independent-sample *t* test**, which is used to compare the mean difference between two groups; specifically, to test hypotheses regarding the difference between two population means. In terms of the null hypothesis, we state the mean difference that we expect in the population and compare it to the difference we observe between the two sample means in our sample. Often, a visual inspection of data from two groups can be quite insightful in terms of determining whether groups differ. Appendix A8 provides an illustration for how to inspect grouped data visually. For a two-independent-sample *t* test concerning two population means, we make four assumptions:

1. **Normality.** We assume that data in each population being sampled are normally distributed. This assumption is particularly important for small samples, because the standard error is typically much larger. In larger sample sizes ($n > 30$), the standard error is smaller, and this assumption becomes less critical as a result.
2. **Random sampling.** We assume that the data we measure were obtained from samples that were selected using a random sampling procedure.
3. **Independence.** We assume that each measured outcome or observation is independent, meaning that one outcome does not influence another. Specifically, outcomes are independent when the probability of one outcome has no effect on the probability of another outcome. Using random sampling usually satisfies this assumption.

The **two-independent-sample *t* test** is a statistical procedure used to compare the mean difference between two independent groups. This test is specifically used to test hypotheses concerning the difference between two population means, where the variance in one or both populations is unknown.

4. *Equal variances.* We assume that the variances in each population are equal to each other. This assumption is usually satisfied when the larger sample variance is not greater than two times the smaller:

$$\frac{\text{larger } s^2}{\text{smaller } s^2} < 2$$

Appendix A8

See **Appendix A8**, p. 290, for how to visually inspect data to compare differences between two groups.

Keep in mind that satisfying the assumptions for the t test is critically important. That said, for each example in this book, the data are intentionally constructed such that the assumptions for conducting the tests have been met. In Example 9.2 we follow the four steps to hypothesis testing introduced in Learning Unit 7 to compute a two-independent-sample t test using an example adapted from published research.

Example 9.2. For an example, let us consider the impact of safety training in the workplace. Nonfatal workplace injuries can be expressed as a rate: the number of injuries per 200,000 hours worked by all employees. A nonfatal incidence rate of 5 means that 5 nonfatal injuries in 200,000 hours of work were accumulated by all employees at a company. Thus, the incident rate that we analyze has been adjusted for size of company.

Using a sample data set adapted from published research, we will use the four steps to hypothesis testing introduced in Learning Unit 7. We examine at a .05 level of significance whether safety training for 40 companies produces a difference in incidence rate as compared to 40 other companies without safety training.

Step 1: State the hypotheses. The null hypothesis states that there is no difference between the two groups, and we are testing whether or not there is a difference:

$H_0 : \mu_1 - \mu_2 = 0$ There is no difference; safety training has no effect on the incidence rate of nonfatal injuries.

$H_0 : \mu_1 - \mu_2 \neq 0$ Safety training does have an effect on the incidence rate of nonfatal injuries.

Step 2: Set the criteria for a decision. The level of significance for this test is .05. We are computing a two-tailed test, so we place the rejection region in both tails. For the t test, the degrees of freedom for each group or sample are $n - 1$. Table 9.4 compares degrees of freedom for one-sample and for two-independent-sample t tests. To find the degrees of freedom for two samples, then, we add the degrees of freedom in each sample. This can be found using one of three methods:

Method 1: df for two-independent-sample t test = $df_1 + df_2$

Method 2: df for two-independent-sample t test = $(n_1 - 1) + (n_2 - 1)$

Method 3: df for two-independent-sample t test = $N - 2$

As summarized in Table 9.4, we can add the degrees of freedom for each sample using the first two methods. In the third method, N is the total sample size for both groups combined, and we subtract 2 from this value. All three methods will produce the same result for degrees of freedom. The degrees of freedom for each sample here are $40 - 1 = 39$. Thus, the degrees of freedom for the two-independent-sample t test are the sum of these degrees of freedom:

$$df = 39 + 39 = 78$$

In Table C.2 in Appendix C, p. 318, the degrees of freedom in the leftmost column increase by 1 up to $df = 30$. After 30, they increase by 10. Because there is no entry for $df = 78$, we use the next smallest value, which is $df = 60$. Move across the columns to find the critical value for a .05 proportion in two tails combined. The critical values for this test are ± 2.000 .

We will compare the value of the test statistic with these critical values. If the value of the test statistic is beyond a critical value (either greater than $+2.000$ or less than -2.000), then there is less than a 5% chance we would obtain that outcome if the null hypothesis were correct, so we reject the null hypothesis; otherwise, we retain the null hypothesis.

Step 3: Compute the test statistic. Download Employee_Safety_Training.xlsx from the student study site: <http://study.sagepub.com/priviteraexcel1e>. Column A contains the rate of nonfatal injuries per 200,000 hours that employees worked at companies

Appendix A12

See **Appendix A12**, p. 298, for more on degrees of freedom for parametric tests.

TABLE 9.4 • Computing the degrees of freedom for a t test.

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	$(310 - 285) = 25$

with safety training; column B contains that same measure at companies without safety training. We can copy A2 to B3 and paste them to D2 to E3 as shown in Figure 9.6. These column headers label the two treatments of the independent variable.

Also as shown in Figure 9.6, we insert in column C labels to keep track of our calculations in columns D and E for the two-independent-sample t test:

C4: Mean (M)

C5: Sample size (n)

C6: Variance (s^2)

C7: Degrees of freedom (df)

To the right of the cells mentioned above, we type these functions and formulas into column D and E:

D4: =AVERAGE(A4:A43) E4: =AVERAGE(B4:B43)

D5: =COUNT(A4:A43) E5: =COUNT(B4:B43)

D6: =VAR.S(A4:A43) E6: =VAR.S(B4:B43)

D7: =D5-1 E7: =E5-1

As mentioned above, the critical value of t taken from Table C.2 in Appendix C is 2.000.

C8: Critical value of t , $df = 60$ in Table C.2 (t_{crit})

D8: 2.000

These values in D4 through E7 allow us to proceed with the calculation of the two-independent-sample t test and compare that result with the t_{crit} of 2.000. We prepare column C with six more labels:

C9: Sample mean difference ($M_1 - M_2$)

C10: Hypothesized mean difference ($\mu_1 - \mu_2$)

C11: Pooled sample variance (s_p^2)

C12: Standard error for difference ($s_{M_1 - M_2}$)

C13: Obtained value of t (t_{obt})

C14: p value

In the formula for a two-independent-sample t test, we subtract the mean difference between the sample means, cell D9, from the mean difference stated in the null, cell D10. To column D we add:

D9: =D4-E4

D10: 0

We divide this difference by the combined standard error in both samples, called the **estimated standard error for the difference**, which is computed as

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

Notice that the numerator in the estimated standard error for the difference formula is s_p^2 , which is called the **pooled sample variance**. The first step, then, to compute the estimated standard error for the difference is to compute the pooled sample variance. Because we have equal sample sizes in the two groups, we can average the two sample variances using the following formula:

$$s_p^2 = \frac{s_1^2 + s_2^2}{2}$$

Appendix A9 provides more detail regarding the calculation and interpretation of the pooled sample variance.

In column D11, we calculate pooled sample variance:

D11: =(D6+E6)/2

which yields

$$s_p^2 = \frac{3.61 + 4.92}{2} = 4.27$$

in cell D11 in Figure 9.6b.

Having the pooled sample variance allows us to then calculate the estimated standard error for the difference in column D (notice that 4.27 is now the numerator in the estimated standard error for the difference formula):

D12: =(D11/D5)+(D11/E5)^0.5

which yields

$$s_{M_1-M_2} = \sqrt{\frac{4.27}{39} + \frac{4.27}{39}} = 0.46$$

in cell D12 in Figure 9.6b.

Appendix B

See **Appendix B2**, p. 301, on formatting cells to add superscripts or subscripts.

See **Appendix B8**, p. 312, on inserting equations, especially to use both a superscript and subscript or add multiple subscripts.

Appendix A9

See **Appendix A9**, p. 292, for more detail regarding the calculation and interpretation of the pooled sample variance.

The **estimated standard error for the difference** is an estimate of the standard deviation of a sampling distribution of mean differences between two sample means. It is an estimate of the standard error or standard distance that mean differences can be expected to deviate from the mean difference stated in the null hypothesis.

The **pooled sample variance** is the mean sample variance of two samples. When the sample size is unequal, the variance in each group or sample is weighted by its respective degrees of freedom.

Now we have the three components needed to calculate the two-independent-sample t test:

$$t_{\text{obt}} = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{S_{M_1 - M_2}}$$

In column D we insert

D13: =(D9-D10)/D12

which yields

$$t_{\text{obt}} = \frac{-1.95 - 0}{0.46} = -4.222$$

in cell D13 in Figure 9.6b.

Finally, in column D we calculate a p value:

D14: =T.TEST(A4:A43,B4:B43,2,2)

which yields a p value of .000065 in cell D14 in Figure 9.6b.

Step 4: Make a decision. The t_{obt} value in cell D13 is -4.222. This value far exceeds our two-tailed critical value at $\alpha = .05$ of 2.000 for $df = 60$ from Table C2 in Appendix C. In fact, the exact p value that we can calculate with Excel indicates that the probability of such an outcome occurring, if the null hypothesis were true, is very unlikely: $p = .000065$ in cell D14 in Figure 9.5b. If this result were reported in a research journal, it would look something like this following APA format (American Psychological Association, 2010):

The mean nonfatal incidence rate at companies with employee safety training ($M = 8.23$, $SD = 1.90$) was significantly lower than was the rate at companies without employee safety training ($M = 10.18$, $SD = 2.22$), $t(78) = -4.222$, $p < .001$.

Effect Size for the Two-Independent-Sample t Test

Hypothesis testing is used to identify whether an effect exists in one or more populations of interest. When we reject the null hypothesis, we conclude that an effect does exist in the population. When we retain the null hypothesis, we conclude that an effect does not exist in the population. In Example 9.2, we concluded that an effect does exist. We will compute effect size for the test in Example 9.2 to determine the effect size of this result or mean difference. We can identify two measures of effect

FIGURE 9.6 • Two-independent-sample *t* test. (a) Functions and formulas. (b) Results of the calculations.

(a)

	A	B	C	D	E
1					
2	Training			Training	
3	With	Without		With	Without
4	8	13	Mean (<i>M</i>) =AVERAGE(A4:A43)		=AVERAGE(B4:B43)
5	4	6	Sample size (<i>n</i>) =COUNT(A4:A43)		=COUNT(B4:B43)
6	11	9	Variance (<i>s</i>²) =VAR.S(A4:A43)		=VAR.S(B4:B43)
7	6	9	Degrees of freedom (<i>df</i>) =D5-1		=E5-1
8	4	11	Critical value of <i>t</i>, <i>df</i>=60 in Table C.2 (<i>t</i>_{crit}) 2		
9	8	12	Sample mean difference (<i>M</i>₁-<i>M</i>₂) =D4-E4		
10	10	14	Hypothesized mean difference (<i>μ</i>₁-<i>μ</i>₂) 0		
11	10	5	Pooled sample variance (<i>s</i>_p²) =(D6+E6)/2		
12	8	13	Standard error for difference (<i>s</i>_{<i>M</i>₁-<i>M</i>₂}) =((D11/D5)+(D11/E5))^0.5		
13	5	9	Obtained value of <i>t</i> (<i>t</i>_{obt}) =(D9-D10)/D12		
14	11	10	<i>p</i> value =T.TEST(A4:A43,B4:B43,2,2)		
15	7	11			
16	6	10	Estimated Cohen's <i>d</i> =(D4-E4)/D11		
17	10	13	Eta-squared (<i>η</i>²) =D13^2/(D13^2+78)		
18	11	11			
19	7	9	<i>t</i>(<i>s</i>_{<i>M</i>₁-<i>M</i>₂}) =D8*D12		
20	9	9	95% CI upper limit =D9+D19		
21	8	10	95% CI lower limit =D9-D19		

(b)

	A	B	C	D	E
1					
2	Training			Training	
3	With	Without		With	Without
4	8	13	Mean (<i>M</i>)	8.23	10.18
5	4	6	Sample size (<i>n</i>)	40	40
6	11	9	Variance (<i>s</i>²)	3.61	4.92
7	6	9	Degrees of freedom (<i>df</i>)	39	39
8	4	11	Critical value of <i>t</i>, <i>df</i>=60 in Table C.2 (<i>t</i>_{crit})	2.000	
9	8	12	Sample mean difference (<i>M</i>₁-<i>M</i>₂)	-1.95	
10	10	14	Hypothesized mean difference (<i>μ</i>₁-<i>μ</i>₂)	0	
11	10	5	Pooled sample variance (<i>s</i>_p²)	4.27	
12	8	13	Standard error for difference (<i>s</i>_{<i>M</i>₁-<i>M</i>₂})	0.46	
13	5	9	Obtained value of <i>t</i> (<i>t</i>_{obt})	-4.222	
14	11	10	<i>p</i> value	.000065	
15	7	11			
16	6	10	Estimated Cohen's <i>d</i>	-0.46	
17	10	13	Eta-squared (<i>η</i>²)	.19	
18	11	11			
19	7	9	<i>t</i>(<i>s</i>_{<i>M</i>₁-<i>M</i>₂})	0.92	
20	9	9	95% CI upper limit	-1.03	
21	8	10	95% CI lower limit	-2.87	

size for the two-independent-sample t test: estimated Cohen's d and proportion of variance with eta squared.

To label these calculations, in column C we enter

C16: Estimated Cohen's d

C17: Eta squared (η^2)

Estimated Cohen's d . As stated in Example 9.1 above, estimated Cohen's d is most often used with the t test. When the estimated Cohen's d is used with the two-independent-sample t test, we place the difference between two sample means in the numerator and the **pooled sample standard deviation** (or square root of the pooled sample variance) in the denominator. The pooled sample standard deviation is an estimate for the pooled or mean standard deviation for the difference between two population means. The formula for an estimated Cohen's d for the two-independent-sample t test is

$$\frac{M_1 - M_2}{\sqrt{s_p^2}}$$

In column D, enter

D16: =(D4-E4)/D11

which yields

$$d = \frac{8.23 - 10.18}{4.27} = -0.46$$

in cell D 16 in Figure 9.6b.

We conclude safety training decreases nonfatal incident rate by 0.46 standard deviations below the mean as compared to no safety training. The effect size conventions given in the middle column of Table 9.2 show that this is a medium effect size. We could report this measure with the significant t test in Example 9.2 by stating,

The mean nonfatal incidence rate at companies with employee safety training ($M = 8.23$, $SD = 1.90$) was significantly lower than was the rate at companies without employee safety training ($M = 10.18$, $SD = 2.22$), $t(78) = -4.222$, $p < .01$, $d = -0.46$.

Proportion of Variance: Eta squared (η^2). Another measure of effect size is proportion of variance, which estimates the proportion of variance in a dependent variable that can be explained by some treatment. In Example 9.2, this measure can describe the proportion of variance in the nonfatal incident rate (the dependent variable) that can be explained by whether companies did or did not have safety training (the treatment). One measure of proportion of variance for the two-independent-sample t test is eta squared, η^2 .

Pooled sample standard deviation is the combined sample standard deviation of two samples. It is computed by taking the square root of the pooled sample variance. This measure estimates the standard deviation for the difference between two population means.

Eta squared can be expressed in a single formula based on the result of a *t* test:

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

In Example 9.2, $t = -4.22$, and $df = 78$. To find proportion of variance using the eta squared formula, we then square the value of *t* in the numerator and the denominator. In column D, insert:

$$D17: =D13^2/(D13^2+78)$$

which yields

$$\eta^2 = \frac{-4.222^2}{-4.222^2 + 78} = \frac{17.83}{17.83 + 78} = .19$$

in cell D17 in Figure 9.6b.

We conclude that only 19% of the variability in nonfatal incident rates can be explained by whether companies did or did not provide safety training. Based on the effect size conventions in Table 9.2, this result indicates a medium effect size. We can report this estimate with the significant *t* test in Example 9.2 by stating,

The mean nonfatal incidence rate at companies with employee safety training ($M = 8.23$, $SD = 1.90$) was significantly lower than was the rate at companies without employee safety training ($M = 10.18$, $SD = 2.22$), $t(78) = -4.222$, $p < .01$, $\eta^2 = .19$.

Confidence Intervals for the Two-Independent-Sample *t* Test

In Example 9.2, we stated a null hypothesis regarding the mean difference in a population. We can further describe the nature of the effect by determining where the effect is likely to be in the population by computing the confidence intervals.

As introduced in Learning Unit 7, there are two types of estimates: a point estimate and an interval estimate. When comparing two samples, a *point estimate* is the sample mean difference we measure. The interval estimate, often reported as a *confidence interval*, is stated within a given *level of confidence*, which is the likelihood that an interval contains an unknown population mean difference.

To illustrate confidence intervals for the two-independent-sample *t* test, we will revisit Example 9.2, 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 Example 9.1. For a two-independent-sample *t* test, the estimation formula is

$$M_1 - M_2 \pm t(s_{M_1 - M_2})$$

Step 1: Compute the sample mean and standard error. The difference between the two sample means is $M_1 - M_2 = -1.95$ nonfatal injuries per 200,000 hours that employees worked. Therefore, the mean difference or point estimate of the population mean difference is -1.95 . (We already computed this value for Example 9.2 in Step 3 of hypothesis testing.)

The estimated standard error for the difference, $s_{M_1 - M_2}$, is equal to 0.46. (We already computed this value as well for Example 9.2 in Step 3 of hypothesis testing.)

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 (CI), so we choose a 95% level of confidence. Remember, in a sampling distribution, 50% of the differences between two sample means fall above the mean difference we selected in our sample, and 50% fall below it. We are looking for the 95% of differences between two sample means that surround the mean difference we measured 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 C.2 in Appendix C. As explained in Step 2 in Example 9.2, we use $df = 60$. The critical value for the interval estimate is $t = 2.000$.

Step 3: Compute the estimation formula to find the confidence limits for a 95% confidence interval. Refer again to Figure 9.6. In column C, insert

C19: $t(s_{M_1 - M_2})$

C20: 95% CI upper limit

C21: 95% CI lower limit

Because we are estimating the difference between two sample means in the population with an unknown variance, we use the $M_1 - M_2 \pm t(s_{M_1 - M_2})$ estimation formula.

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

$$t(s_{M_1 - M_2})$$

In column D,

D19: =D8*D12

which yields

$$t(s_{M_1 - M_2}) = 2.000(0.46) = 0.92$$

in cell D19 in Figure 9.6b.

Add 0.92 to the sample mean difference to find the upper confidence limit, and subtract 0.92 from the sample mean difference to find the lower confidence limit. In column D,

D20: =D9+D19

D21: =D9-D19

which yields

$$M_1 - M_2 + t(s_{M_1 - M_2}) = -1.95 + 0.92 = -1.03$$

in cell D20 in Figure 9.6b, and

$$M_1 - M_2 - t(s_{M_1 - M_2}) = -1.95 - 0.92 = -2.87$$

in cell D21 in Figure 9.6b.

As shown in Figure 9.7, the 95% confidence interval in this population is between a mean difference in nonfatal injury incidence rate of -2.87 and -1.03 per 200,000 hours worked. We can estimate within a 95% level of confidence that the difference between groups in nonfatal injury incidence rate is between -2.87 and -1.03 per 200,000 hours worked. We are 95% confident that the mean difference in the population falls within this range, because 95% of all sample mean differences we could have selected from this population fall within the range of sample mean differences we specified.

Computing the Two-Independent-Sample *t* Test Using the Analysis Toolpak

We can also use the Analysis ToolPak available in Excel for easy and accurate calculation. We'll guide you through the steps to do the analysis for the two-independent-sample *t* test.

Click on the Data tab, and then on the Data Analysis icon all the way to the right. Select "t-Test: Two-Sample Assuming Equal Variances" (Figure 9.8a). According to the fourth assumption of the two-independent-sample *t* test described above, variances in the two samples must be equal. The rule of thumb we use is that the larger variance is no more than twice the smaller variance. That is the case with these data, as is shown in Figure 9.6b, cells D6 and E6. The variances for the two groups are 3.61 and 4.92.

FIGURE 9.7 • At a 95% CI, the mean difference nonfatal injury incidents falls between -2.87 and -1.03 .

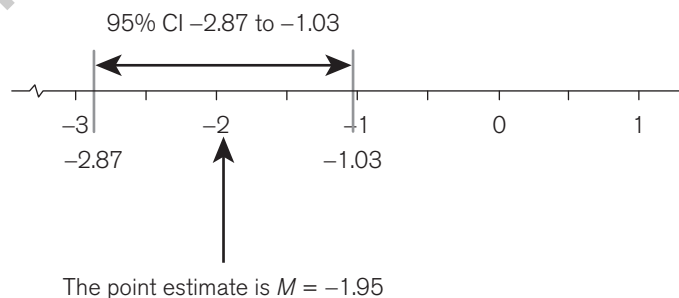
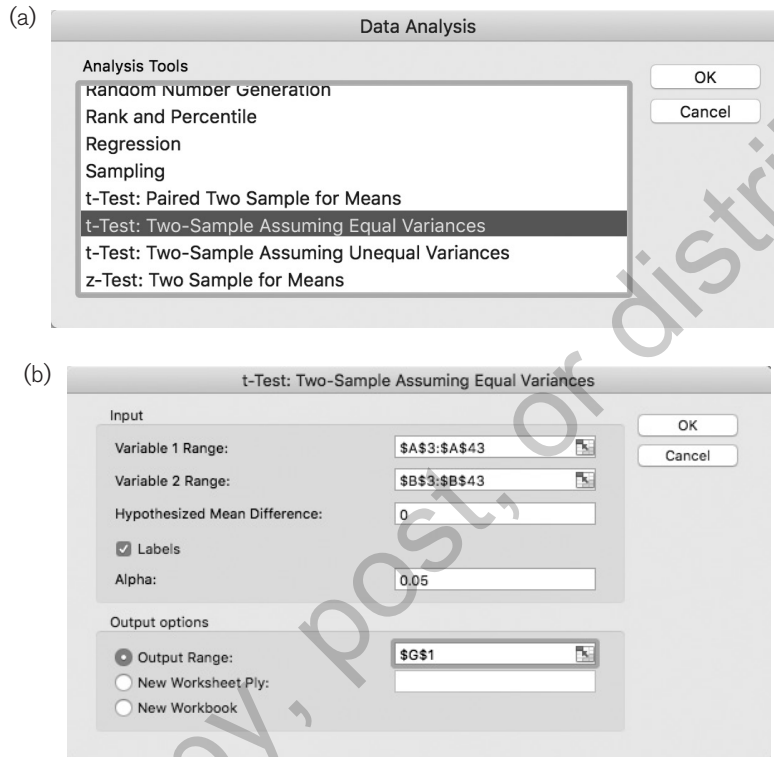


FIGURE 9.8 • Performing a two-independent-sample t test with the Analysis ToolPak in Excel. (a) Selecting “t-Test: Two-Sample Assuming Equal Variance” to perform a one-sample t test. (b) Specifying the location of the data and parameters for the t test.



Selecting “t-Test: Two-Sample Assuming Equal Variance” yields the dialog box in Figure 9.8b. For Variable 1, we select the nonfatal injury incident rate for companies with safety training in cells A3 through A43, which includes in A3 a label for the data. For Variable 2, we select the data for companies without safety training in cells B3 through B43, and include in B3 a label for the data. The Hypothesized Mean Difference is 0. Check the Labels box so that the output contains the labels from A3 and B3. We keep our output on the same page by selecting Output Range and clicking in cell F3.

Clicking “OK” on the dialog box returns the output table in Figure 9.9. Notice that in Figure 9.9, we get the same means of 8.23 in cell H4 and 10.18 in cell I4 as we obtained in Figure 9.6b. We also get the same t_{obt} of -4.222 in cell H10 and same p value of .000065 in cell H13 as we did in Figure 9.7b. The t_{crit} in cell H14 in Figure 9.9 is for $df = 78$ and is thus more precise than the one stated in Step 2 above and shown in Figure 9.6. Although neither an estimate of effect size nor confidence intervals are generated automatically, the output table gives the means of the two groups, the variance, and the degrees of freedom. With this information we can calculate effect size and confidence intervals as we did above.

FIGURE 9.9 • Results of two-independent-sample *t* test using the Analysis ToolPak.

	G	H	I
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		<i>With</i>	<i>Without</i>
4	Mean	8.23	10.18
5	Variance	3.61	4.92
6	Observations	40	40
7	Pooled Variance	4.27	
8	Hypothesized Mean Difference	0	
9	df	78	
10	t Stat	-4.222	
11	P(T<=t) one-tail	.000032	
12	t Critical one-tail	1.665	
13	P(T<=t) two-tail	.000065	
14	t Critical two-tail	1.991	

Computing the Related-Samples *t* Test

In this section, we compute the **related-samples *t* test**, which is used to compare the mean difference between pairs of scores. In terms of the null hypothesis, we start by stating the null hypothesis for the mean difference between pairs of scores in a population, and we 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. Appendix A10 provides an overview for the reason we compute difference scores. For a related-samples *t* test, we make two assumptions:

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.

Again, keep in mind that satisfying the assumptions for the *t* test is critically important. That said, for each example in this book, the data are intentionally constructed such that the assumptions for conducting the tests have been met. In Example 9.3, we follow the four steps to hypothesis testing introduced in Learning Unit 7 to compute a related-samples *t* test using an example adapted from published

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.

Appendix A10

See **Appendix A10**, p. 293, for more detail regarding why difference scores are calculated for a related-samples *t* test.

Appendix A11

See **Appendix A11**, p. 295, for more detail regarding the types of designs that are considered related-samples designs.

research. Note that there are many types of designs that fit into the category of *related-samples*. An overview of the types of designs that fit into this category is provided in Appendix A11.

Example 9.3. One area of focus in cognitive psychology is attention. Psychologists have examined what kinds of visual stimuli capture our attention most quickly. The course of human evolution may have predisposed us to notice animals more readily than we notice inanimate objects (Hagen & Laeng, 2016; New, Cosmides, & Tooby, 2007). In our evolutionary past, animals could have been predators that would harm us or food that would nourish us. Thus animal objects in the environment may have held more meaning than nonanimal objects such as plants or rocks. Changes to animate stimuli may capture our attention more quickly than changes in other stimuli. Suppose we conduct a study of whether people are faster to detect change in animate targets (e.g.,

people or animals) than in inanimate targets (e.g., plants, cars). We show participants several pairs of scenes that are virtually identical except for one change. That change could be to an animate object or to an inanimate object. For 35 participants, we record to the nearest 0.01 second the time taken for correct identification of a change. Using a sample data set adapted from published research, we will use the four steps to hypothesis testing introduced in Learning Unit 7 to test for a difference in their responses to each kind of change, animate versus inanimate, at a .05 level of significance.

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

$H_0 : \mu_1 - \mu_2 = 0$ No difference; changes in animate as compared to inanimate objects do not differ in time to detection.

$H_a : \mu_1 - \mu_2 \neq 0$ Changes in animate as compared to inanimate objects differ in time to detection.

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 $df = 35 - 1 = 34$.

Because $df = 34$ is not available in Table C.2 in Appendix C, we take the closest smaller value, which is $df = 30$. Move across the columns to find the critical value for a .05 proportion in two tails combined. The critical values for this test are ± 2.042 .

We will compare the value of the test statistic with these critical values. If the value of the test

Appendix A12

See **Appendix A12**, p. 298, for more on degrees of freedom for parametric tests.

statistic is beyond a critical value (either greater than +2.042 or less than -2.042), then there is less than a 5% chance we would obtain that outcome if the null hypothesis were correct, so we reject the null hypothesis; otherwise, we retain the null hypothesis.

Step 3: Compute the test statistic. Download Visual_Change.xlsx from the student study site: <http://study.sagepub.com/priviteraexcel1e>. As shown in Figure 9.10, participants are identified in column A, and their times to detect visual change in animate and inanimate objects are listed in the same row. Note that, in this spreadsheet, the information for a participant is all on one row, and that one row contains information from only a single participant.

To compute the test statistic, we (1) compute a difference score by subtracting for each participant one measure from the other measure; (2) compute the mean, variance, and standard deviation of difference scores; (3) compute the estimated standard error for difference scores; and then (4) compute the test statistic.

(1) Compute the difference scores. In cell D3, type “D” to signify that the column will contain difference scores, as in Figure 9.10. To calculate a difference score for the first participant, enter into cell D4 =B4-C4. Select cell D4. Fill down to cell D38, or copy D4 and paste from D5 to D38. Keep in mind that the sign (negative or positive) of difference scores matters when we compute the mean and standard deviation.

(2) Compute the mean, variance, and standard deviation of difference scores, and the **estimated standard error for the difference scores (s_{MD})**. We’ll reserve column E for calculating D^2 , which we need on our way to calculating the variance and standard deviation of the difference scores. Use column F, as in Figure 9.10, for labels to keep track of what we calculate:

F4: Mean difference score (M_D)

F5: Sample size (n)

F6: Variance of the difference scores (s_D^2)

F7: Standard deviation of difference scores (s_D)

F8: Standard error for difference scores (s_{MD})

To the right of the cells in column F mentioned above, type into Column G functions and formulas to calculate the values, as shown in Figure 9.10:

G4: =AVERAGE(D4:D38)

G5: =COUNT(D4:D38)

G6: =VAR.S(D4:D38)

Appendix B

See **Appendix B2**, p. 301, on formatting cells to add superscripts or subscripts.

See **Appendix B8**, p. 312, on inserting equations, especially to use both a superscript and subscript or add multiple subscripts.

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.

FIGURE 9.10 • Related-samples *t* test. (a) Functions and formulas. (b) Results of calculations.

(a)

	A	B	C	D	E	F	G
1							
2		Object Changed					
3	Participant	Animate	Inanimate	D			
4	1	1.82	2.54	=B4-C4		Mean difference score (M_D) =AVERAGE(D4:D38)	
5	2	0.64	0.27	=B5-C5		Sample size (n) =COUNT(D4:D38)	
6	3	2.89	4.98	=B6-C6		Variance of the difference scores (s_D²) =VAR.S(D4:D38)	
7	4	1.06	2.23	=B7-C7		Standard deviation of difference scores (s_D) =G6^0.5	
8	5	2.38	5.96	=B8-C8		Standard error for difference scores (s_{MD}) =G7/G5^0.5	
9	6	0.75	2.29	=B9-C9		Degrees of freedom (df) =G5-1	
10	7	3.74	3.38	=B10-C10		Critical value of t (t_{crit}) 2.042	
11	8	3.49	4.05	=B11-C11		Obtained value of t (t_{obt}) =(G4-0)/G8	
12	9	4.31	4.72	=B12-C12		p value =T.TEST(B4:B38,C4:C38,2,1)	
13	10	2.68	4.94	=B13-C13			
14	11	1.08	3.48	=B14-C14		Estimated Cohen's d =G4/G7	
15	12	0.85	4.16	=B15-C15		Eta-squared (η²) =G11^2/(G11^2+G9)	
16	13	4.51	3.28	=B16-C16			
17	14	2.86	5.75	=B17-C17		t(s_{MD}) =G10*G8	
18	15	5.46	8.24	=B18-C18		95% CI upper limit =G4+G17	
19	16	2.46	0.66	=B19-C19		95% CI lower limit =G4-G17	

(b)

	A	B	C	D	E	F	G
1							
2		Object Changed					
3	Participant	Animate	Inanimate	D			
4	1	1.82	2.54	-0.72		Mean difference score (M_D)	-0.674
5	2	0.64	0.27	0.37		Sample size (n)	35
6	3	2.89	4.98	-2.09		Variance of the difference scores (s_D²)	2.434
7	4	1.06	2.23	-1.17		Standard deviation of difference scores (s_D)	1.560
8	5	2.38	5.96	-3.58		Standard error for difference scores (s_{MD})	0.264
9	6	0.75	2.29	-1.54		Degrees of freedom (df)	34
10	7	3.74	3.38	0.36		Critical value of t (t_{crit})	2.042
11	8	3.49	4.05	-0.56		Obtained value of t (t_{obt})	-2.555
12	9	4.31	4.72	-0.41		p value	.015
13	10	2.68	4.94	-2.26			
14	11	1.08	3.48	-2.40		Estimated Cohen's d	-0.432
15	12	0.85	4.16	-3.31		Eta-squared (η²)	.161
16	13	4.51	3.28	1.23			
17	14	2.86	5.75	-2.89		t(s_{MD})	0.539
18	15	5.46	8.24	-2.78		95% CI upper limit	-0.135
19	16	2.46	0.66	1.80		95% CI lower limit	-1.212

G7: =G6^0.5

G8: =G7/G5^0.5

(3) Compute the test statistic. At this point we are ready to proceed with the calculation and evaluation of the related-samples *t* test. Use column F, as in Figure 9.10, for labels to keep track of what we calculate:

F9: Degrees of freedom (*df*)

F10: Critical value of *t* (t_{crit})

F11: Obtained value of *t* (t_{obt})

F12: *p* value

To the right of the cells in column F mentioned above, type into Column G functions and formulas to calculate the values, as shown in Figure 9.10:

G9: =G5-1

G10: 2.042

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. Similar to the other *t* tests, the mean difference is placed in the numerator, and the estimate of the standard error is placed in the denominator. 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:

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

In column G,

G11: =(G4-0)/G8

which yields

$$t_{\text{obt}} = \frac{-0.674 - 0}{.264} = -2.555$$

in cell G11 in Figure 9.10b.

Excel allows us to calculate an exact *p* value, as shown in cell G12 in Figure 9.10a:

G12: =T.TEST(B4:B38,C4:C38,2,1)

This function requires two cell ranges of data: B4:B38 contains the times to identify change in the animate object, C4:C38 contains the times to identify change in the inanimate object. After those two ranges of data, the next argument required in the function is the number of tails, for which we specify 2. The final argument is the type of *t* test, which we specify as “paired” 1. “Paired” is the term used in Excel to calculate a related-samples *t* test. As expected with the t_{obt} , the *p* value is .015, as shown in cell G12 of Figure 9.10b.

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 9.10 reveals that the obtained value ($t_{\text{obt}} = -2.555$) exceeds the lower 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:

Changes to animate objects were identified significantly more quickly than were changes to inanimate objects, $t(34) = -2.555, p = .015$.

Effect Size for the Related-Samples t Test

Hypothesis testing identifies whether or not an effect exists. In Example 9.3, we concluded that an effect does exist—people noticed changes to animate objects more quickly than they noticed changes to inanimate objects; we rejected the null hypothesis. The size of this effect is determined by measures of effect size. We will compute effect size for Example 9.3, because the decision was to reject the null hypothesis for that hypothesis test. There are two measures of effect size for the related-samples t test: estimated Cohen's d and proportion of variance with eta squared.

To label these calculations, in column C we enter

F14: Estimated Cohen's d

F15: Eta-squared (η^2)

Estimated Cohen's d . As stated in Example 9.1 above, estimated Cohen's d is most often used with the t test. When the estimated Cohen's d is used with the related-samples t test, it 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 column G,

G14: =G4/G7

which yields

$$d = \frac{-0.674}{1.560} = -0.432$$

in cell G14 of Figure 9.10b.

We conclude that time to recognize a change in animate objects is 0.432 standard deviations shorter than time to recognize changes in inanimate objects. The effect size conventions listed in Table 9.2 show that this is a medium effect size ($-0.8 < d < -0.2$). We could report this measure with the significant *t* test in Example 9.3 by stating,

Changes to animate objects were identified significantly more quickly than were changes to inanimate objects, $t(34) = -2.555$, $p < .05$ ($d = -0.432$).

Proportion of Variance: Eta squared (η^2). Another measure of effect size is proportion of variance, which estimates the proportion of variance in a dependent variable that can be explained by some treatment. In Example 9.3, this measure can describe the proportion of variance in the difference in recognition time (the dependent variable) that can be explained by whether the changed object was animate or inanimate (the treatment). One measure of proportion of variance for the two-independent-sample *t* test is eta squared, η^2 .

Eta squared can be expressed in a single formula based on the result of a *t* test:

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

In column G,

$$G15: =G11^2/(G11^2+G9)$$

which yields

$$\eta^2 = \frac{-2.555^2}{-2.555^2 + 34} = \frac{6.528}{6.528 + 34} = .161$$

in cell G15 in Figure 9.10b.

Typically, we report proportions to the hundredths place. So with rounding, we conclude that 16% of the variability in reaction time can be explained by whether the object that changed was animate or inanimate. Based on the effect size conventions in Table 9.2, this result indicates a medium effect size. We can report this estimate with the significant *t* test in Example 9.3 by stating,

Changes to animate objects were identified significantly more quickly than were changes to inanimate objects, $t(34) = -2.555$, $p < .05$ ($\eta^2 = .16$).

Confidence Intervals for the Related-Samples *t* Test

In Example 9.3, we stated a null hypothesis regarding the mean difference in a population. We can further describe the nature of the effect by determining where the effect is likely to be in the population by computing the confidence intervals.

As introduced in Learning Unit 7, there are two types of estimates: a point estimate and an interval estimate. When using two related samples, a *point estimate* is the sample mean difference score we measure. The interval estimate, often reported as a *confidence interval*, is stated within a given *level of confidence*, which is the likelihood that an interval contains an unknown population mean.

To illustrate confidence intervals for the related-samples *t* test, we will revisit Example 9.3, 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 Example 9.1. 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 $M_D = -0.674$. The estimated standard error for difference scores $s_{MD} = 0.264$.

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 C.2 in Appendix C. The degrees of freedom are 34 ($df = n_D - 1$) for two related samples. The critical value for the interval estimate is $t = 2.042$.

Step 3: Compute the estimation formula to find the confidence limits for a 95% confidence interval. In column G,

F17: $t(s_{MD})$

F18: 95% CI upper limit

F19: 95% CI lower limit

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. In column G,

G17: =G10*G8

which yields

$$t(s_{MD}) = 2.042(0.264) = 0.539$$

in cell G17 of Figure 9.10b.

Add 0.539 to the sample mean difference to find the upper confidence limit, and subtract 0.539 from the sample mean to find the lower confidence limit. In column G,

$$G18: =G4+G17$$

$$G19: =G4-G17$$

which yields

$$M_D + t(s_{MD}) = -0.674 + 0.539 = -0.135$$

in cell G18 of Figure 9.10b, and

$$M_D - t(s_{MD}) = -0.674 - 0.539 = -1.212$$

in cell G19 of Figure 9.10b.

As shown in Figure 9.11, the 95% confidence interval in this population is between -1.212 seconds and -0.135 seconds. We can estimate within a 95% level of confidence that people take more time to notice a change in an inanimate object than they take to notice a change in an animate object.

Computing the Related-Samples *t* Test Using the Analysis Toolpak

We can also use the Analysis ToolPak available in Excel for easy and accurate calculation. We will guide you through the steps to do the analysis for the related-samples *t* test.

Click on the Data tab, and then on the Data Analysis icon all the way to the right. Select “t-Test: Paired Two Sample for Means,” as shown in Figure 9.12a, which yields the dialog box in Figure 9.12b. For Variable 1, we select reaction times when the animate object changed, cells B3 through B38, which includes in B3 a label for the data. For Variable 2, we select reaction times when the inanimate object changed, cells C3 to C38, and include in C3 a label for the data. The Hypothesized Mean Difference is 0. Check the Labels box so that the output contains the labels from B3 and C3. We keep our output on the same page by selecting Output Range and clicking in cell I1.

Clicking “OK” on the dialog box returns the output table in Figure 9.13. Notice that we get the same t_{obt} of -2.555 as in cell G11 of Figure 9.10b, and the same p value of $.015$ as in cell G12 of Figure 9.10b. The t_{crit} for $df = 34$ in Figure 9.13 cell J14 is 2.032 .

FIGURE 9.11 • At a 95% CI, the mean difference in response time falls between -1.212 and -0.135 .

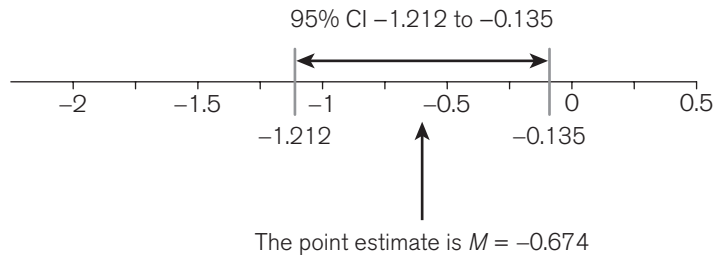


FIGURE 9.12 • Performing a one-sample *t* test with the Analysis ToolPak in Excel. (a) Selecting “t-Test: Paired Two Sample for Means” to perform a one-sample *t* test. (b) Specifying the location of the data and parameters for the *t* test.

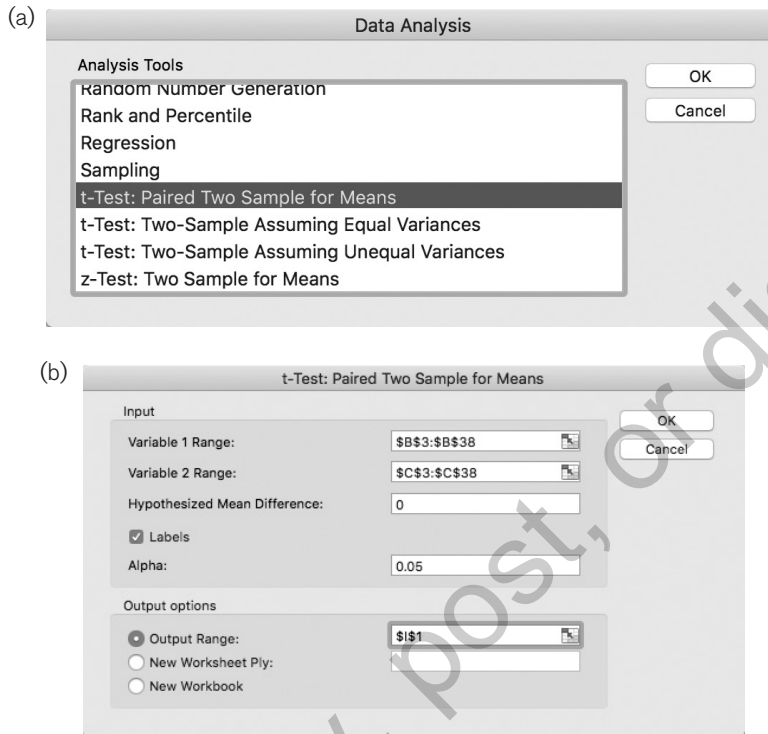


FIGURE 9.13 • Results of related-samples *t* test using the Analysis ToolPak.

	I	J	K
1	t-Test: Paired Two Sample for Means		
2			
3		<i>Animate</i>	<i>Inanimate</i>
4	Mean	2.987	3.661
5	Variance	2.291	3.761
6	Observations	35	35
7	Pearson Correlation	.616	
8	Hypothesized Mean Difference	0	
9	df	34	
10	t Stat	-2.555	
11	P(T<=t) one-tail	.008	
12	t Critical one-tail	1.691	
13	P(T<=t) two-tail	.015	
14	t Critical two-tail	2.032	