CHAPTER 7

Comparing Two Group Means: The Independent Samples *t* Test

After reading this chapter, you will be able to

- Differentiate between the one-sample *t* test and the independent samples *t* test
- Summarize the relationship among an independent variable, a dependent variable, and random assignment
- Interpret the conceptual ingredients of the independent samples *t* test
- Interpret an APA style presentation of an independent samples *t* test
- Hand-calculate all ingredients for an independent samples *t* test
- Conduct and interpret an independent samples *t* test using SPSS

In the previous chapter, we discussed the basic principles of statistically testing a null hypothesis. We highlighted these principles by introducing two parametric inferential statistical tools, the *z* test and the one-sample *t* test. Recall that we use the *z* test when we want to compare a sample to the population and we know the population parameters, specifically the population mean and standard deviation. We use the one-sample *t* test when we do not have access to the population standard deviation. In this chapter, we will add another inferential statistical tool to our toolbox. Specifically, we will learn how to compare the difference between means from two groups drawn from the same population to learn whether that mean difference might exist in the population.

CONCEPTUAL UNDERSTANDING OF THE STATISTICAL TOOL

The Study

Photo 7.1

As a little kid, I was afraid of the dark. Of course, this is not an uncommon fear for children to have. I wasn't sure what I was afraid of as the dark contained the same things in my bedroom as the light contained. Of course, we know from developmental psychology research that young children are not yet capable of using this kind of logic. And let's face it: There is something quite functional about fearing the dark, even for adults. My bet is that most of us prefer to be physically alive than the alternative, and the dark is a reminder that we are vulnerable to the world around us. Such is the basic logic of terror management theory (TMT). In a nutshell, this theory contends that as humans, we are cognizant of our eventual deaths. When our ultimate demise is made salient to us, we espouse "worldviews" that allow us to feel our lives have meaning and that we fit into the culture in which we live.

TMT guided an experiment that Tim Kasser and Kennon Sheldon (2000) conducted. In discussing their research, we will walk through what is perhaps the most basic statistical tool needed to analyze data from an experimental study: the independent samples *t* test.

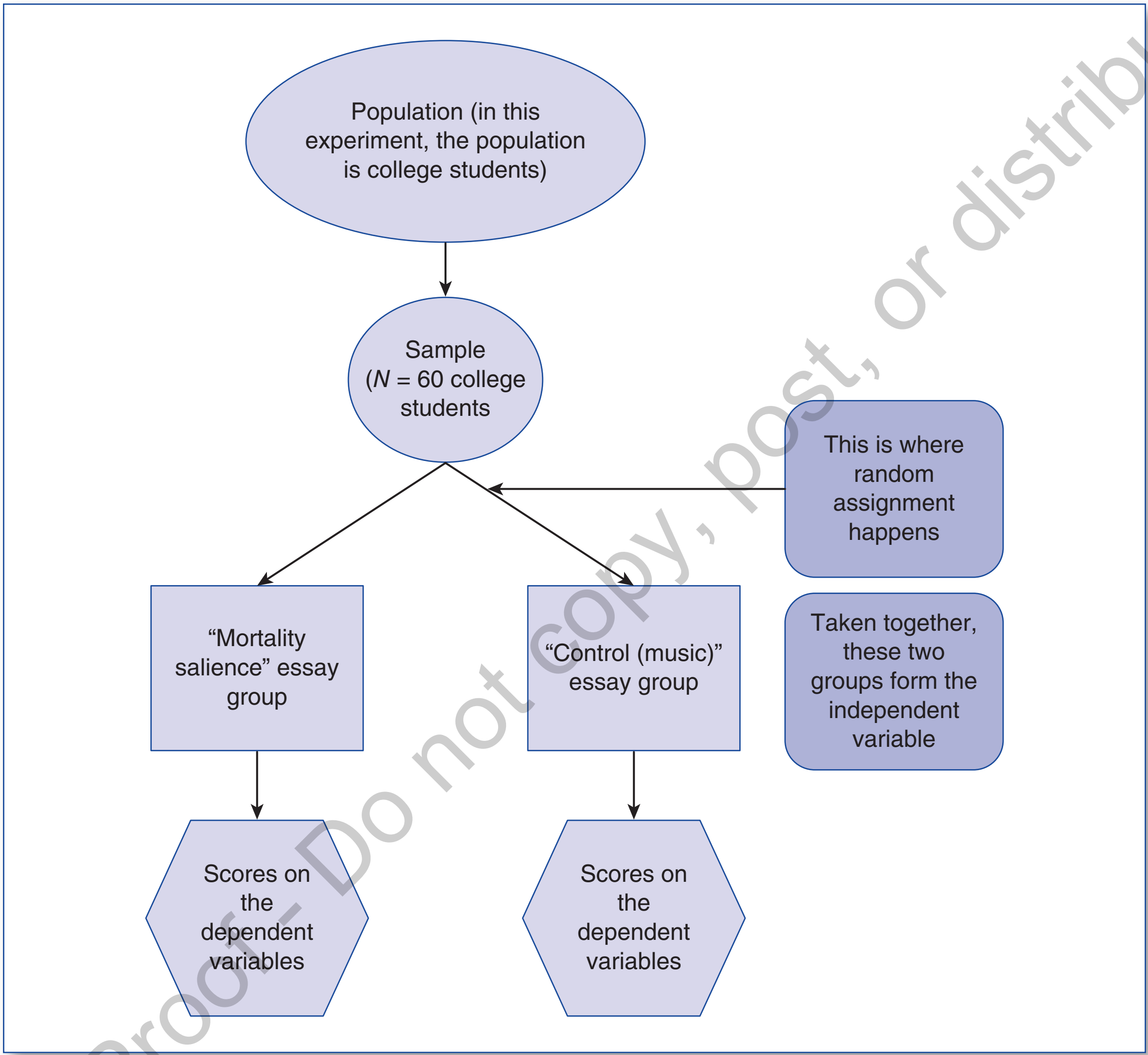
In Kasser and Sheldon's (2000) experiment, a sample of 60 college students was randomly assigned to one of two experimental groups (conditions). Thirty students wrote short essays about their attitudes toward listening to music, whereas the other 30 students wrote short essays about their attitudes toward their deaths. In this experiment, the *essay topic* about which students wrote is called the independent variable.

The independent variable is "variable" because the researchers used random assignment of participants to one of the two essay topics. We introduced the concept of random assignment in Chapter 1 when talking about explanatory research. People can differ from one another in many ways (e.g., gender, religious attitudes, and socioeconomic status). Not that such differences are trivial, but they are not of interest to the researchers in this particular experiment. Therefore, we want to control for their influence on how people in the sample behave, so that we can isolate the effect of the independent variable. Through the process of random assignment, we can minimize the influences of variables other than the independent variable. In doing so, any effects we find (from an independent samples *t* test to be discussed in this chapter) can be linked to the independent variable.

To assess the effects of the essay topic, after writing their essay, students imagined themselves 15 years in the future. In doing so, they made estimates about how much they would expect to spend on clothing, entertainment, and leisure activities (called "pleasure spending"). This estimate should depend on what experimental condition (group) people are randomly assigned to. That is, there should be differences between the two groups on responses about pleasure spending. In this example, pleasure spending is called a dependent variable because it "depends on" the independent variable (i.e., essay topic).According to TMT, people should respond to threats to their existence (i.e., writing essays about their deaths) by enhancing their worldviews. In this experiment, that would mean spending more on clothing, entertainment, and leisure activities. Doing so would suggest they are more successful, at least financially, in their culture.

Here is a visual depiction of this experiment (Figure 7.1):

Figure 7.1 Elements of Kasser and Sheldon's (2000) Experiment



LEARNING CHECK

1. What is the difference between an independent variable and a dependent variable?

   A: The independent variable is manipulated/controlled by the researchers to create two or more groups in an experiment. The independent variable is expected to affect behavior or mental processes. The dependent variable is the behavior or mental process that is influenced by the independent variable. The dependent variable is the outcome of the independent variable.

2. Why is random assignment a critical component of an experiment?

   A: Random assignment allows researchers to isolate the effects of the independent variable on the dependent variable. It does so by arbitrarily placing members of the sample into one of the groups created by the independent variable. Therefore, the differences between people in the sample are minimized, allowing researchers to connect the effects of the independent variable to the dependent variable.

3. Why must an experiment contain at least two groups?

   A: If an experiment contained only one group, there would be no way to compare scores on a dependent variable. To take a simple example, I am 6′ 2″ tall. Am I a tall person? We cannot answer this question without something with which to compare my height. Compared with an average man (who is about 5′ 9″ tall), yes, I am a tall person. Compared with most professional basketball players, no, I am not a tall person.

We will now focus on comparing scores on the dependent variable. Kasser and Sheldon (2000) used "standardized scores" to quantify responses on these dependent variables. "Standardized scores" may sound scary, but you know exactly what that means from reading Chapter 5 and learning about *z* scores. If you are not confident in your *z* score expertise, now is a great chance to go back and review that material in Chapter 5.

The Tool

Remember that the one-sample *t* test is used when we have one sample of data and want to compare its mean with the population mean. In Kasser and Sheldon's (2000) experiment, we have two groups of data (i.e., scores in the dependent variable) created by manipulating the independent variable. In this experiment, we will need to use the **independent samples *t* test**. It is called the "independent samples" *t* test because each member of the sample is randomly assigned to one and only one experimental group. This type of experiment is called a **between-subjects experiment**. Just to avoid confusion, the fact that this statistical tool is called the "independent" samples *t* test has nothing to do with the notion of an independent variable. Rather, *the word "independent" signifies that each member of the sample was randomly assigned to one experimental group*.

Independent samples *t* test: statistical tool used to compare means of two mutually exclusive groups of people.

Between-subjects design: experimental design in which participants are randomly assigned to one and only one experimental group (condition).

Ingredients

What is the logic of the independent samples *t* test? *It is a comparison of whether mean differences in the sample are generalizable to the population from which that sample was drawn.* We will now focus on the conceptual ingredients needed for an independent samples *t* test (which are not difficult to put into practice now that you know the one-sample *t* test).

First, *we must know the mean for each group on a dependent variable*. For instance, did the "death essay" group score higher or lower than the "music essay" group on estimates of future pleasure spending? We would need the mean estimate of pleasure spending for each of these two groups.

Second, not every member of the two groups will have the same score on a dependent variable. There will be variability among individual scores around the mean score for each group. To account for this variability, *we must consider the* ***standard error of the difference between the means.*** Recall from Chapter 4 the notion of a standard deviation. A *standard deviation* is a measure of how much a group of scores tends to stray from the group's mean. The standard error of the difference between the means serves the same purpose as a standard deviation, within the context of an independent samples *t* test. A standard deviation applies to one group of data; the standard error of the difference *between the means* applies to two groups of data. Like the standard deviation, the standard error of the difference between the means relies in part on the number of people in the groups (i.e., sample size). Thus, *the larger the sample size and the lower the variability of scores around the group means, the lower the standard error of the difference will be.* Therefore, as we said in the previous chapter, larger sample sizes are preferred, statistically speaking, because they reduce the standard error of the difference between the means. They are inherently more representative of the populations from which they were drawn.

Standard error of the difference between the means: standard deviation of a difference between two group means.

In short, to use an independent samples *t* test, we need to know (a) the mean of each group and (b) the standard error of the difference between those means. These two pieces of information give us a "*t* test statistic" that we can use to see whether there is a statistically significant difference between the means of these two groups. Here is the conceptual formula for the independent samples *t* test statistic:

$$t = \frac{\text{Mean difference between the two groups}}{\text{Standard error of the difference between the means}}$$

In thinking through this formula, there are three ways to increase the power of this statistical tool:

1. Larger mean differences (i.e., increase the numerator)
2. Larger sample sizes (i.e., decrease the denominator)
3. Less variability of scores within each group (i.e., decrease the denominator)

Hypothesis from Kasser and Sheldon (2000)

As a refresher from the previous chapter, remember that we are testing the hypothesis that there is no difference in the population between mean scores of the two groups. That is, we are testing the null hypothesis with our statistical tool. To be able to suggest that there are differences in the population based on these two mean scores, we must reject the notion that there is no difference between the mortality salience essay group and the music essay group. In other words, the mean difference between the two groups must be large enough to conclude that there is likely an effect that exists in the population. In plain English, the statistic tested the notion that there will be no difference between the mortality-salience condition and the control condition on scores on the dependent variable. Symbolically,

$$H_o: \mu_{\text{mortality-salience group}} = \mu_{\text{control group}}$$

Published research rarely if ever states a null hypothesis even though, as you know, it is always the null hypothesis that statistical tools are testing. Rather, published research tends to state the research hypothesis. For instance, in Kasser and Sheldon's (2000) article, they predicted that students who wrote essays about death would become more focused on the accumulation of wealth and possessions than would students who wrote about a neutral topic. (p. 348)

Symbolically, we have

$$H_r: \mu_{\text{mortality-salience group}} > \mu_{\text{control group}}$$

LEARNING CHECK

Let's drive home the conceptual logic of the independent samples *t* test with a few examples:

For the first two questions, calculate the *t* test statistic, plugging in the following data from the conceptual formula for the independent samples *t* test:

1. Mean of Group A = 50

   Mean of Group B = 35

   Standard error of the difference between the means = 7.50

   A:
   $$t = \frac{50-35}{7.50}$$
   $$t = \frac{15}{7.5}$$
   $$t = 2.00$$

2. Mean of Group C = 20

   Mean of Group D = 15

   Standard error of the difference between the means = 4

   A:
   $$t = \frac{20-15}{4}$$
   $$t = \frac{5}{4}$$
   $$t = 1.25$$

3. As the sample size increases, all else being equal, will the value of the *t* test statistic increase or decrease? Why?

   A: The *t* test statistic will increase in value because as the sample becomes larger, it will better represent the population from which it was drawn. Therefore, our results have a better chance of being generalized from the sample to the population. Mathematically speaking, a larger sample size decreases the denominator of the *t* test statistic formula (i.e., it decreases the standard error of the difference between the means), making the *t* test statistic bigger.

4. What is the difference between a standard deviation and a standard error of the difference between the means?

   A: The standard deviation is a measure of variability for one group of data. The standard error of the difference between the means is a pooled measure of variability for two groups of data.

Interpreting the Tool

Now that we have laid out all of the ingredients for the independent samples *t* test, let's take a look at Kasser and Sheldon's (2000) results. It might be a good time to take a look back at the visual depiction of this experiment (Figure 7.1). Then, consider the results of their experiment, as they are presented here:

> An independent samples *t* test suggested that participants primed to think about their deaths estimated spending more on pleasure items ($M = 0.22$, $SD = 0.96$) than participants primed to think about music ($M = -0.27$, $SD = 0.61$), $t(52.8) = 2.30$, $p = .02$, $d = 0.61$, 95% confidence interval (CI) [0.06, 0.92].

What does this text tell us? Let's break it down into its smallest, most digestible bites and indulge in them one at a time. We have the mean (*M*) and standard deviation (*SD*) for each group. Thinking about the ingredients for the *t* test, we of course must have the mean for each group. In addition, we need the standard error of the difference between the means. *To get the standard error of the difference between the means, we need to know the standard deviation for each group, as well as the sample size for each group*. Here, the standard deviation is reported. Although the sample size is not provided in the results, such information should always appear in the Method section of a journal article. In this experiment, the sample size was 60 people.

Recall that the dependent variable in this experiment was called "pleasure spending." It has a *t* test statistic of 2.30. How did that 2.30 magically appear? Let's review the conceptual formula for the *t* test:

$$t = \frac{\text{Mean difference between the two groups}}{\text{Standard error of the difference between the means}}$$

Now, let's fill in this formula with numbers from the results:

$$2.30 = \frac{0.22 - (-0.27)}{\text{Standard error of the difference between the means}}$$

The difference between the two group means was 0.49. Remember, we are dealing with *z* scores here, so when we see a difference score of 0.49, it means that there is almost one half of 1 standard deviation difference between the mean scores. It does not matter which mean you subtract from the other.

Typically in reported research, the standard error of the difference between the means is not given. However, with some basic algebra, we can figure it out (if we care to do so; for the purposes of making sense of these results, this step is not necessary, but let's be true to the process). From your algebra days, take 0.49 (the difference between the means of the two groups and the numerator of the *t* statistic) and divide it by 2.30. Voila, 0.213 is your standard error of the difference between the means.

Now that we know where the *t* test statistic comes from, there are some additional pieces of information we need to consider when using this tool. We need to consider *degrees of freedom (df)*, a concept introduced in the previous chapter. Degrees of freedom are the numbers of scores that are free to vary and still have the same mean. *Degrees of freedom are always dependent on the sample size.* The larger the sample, the more degrees of freedom a researcher has. The larger the sample, all else being equal, the more likely it is we can generalize a statistical result from the sample to the population. For an independent samples *t* test, the calculation of degrees of freedom is simple:

$$df = \text{Sample size} - 2$$

Why "minus 2"? Remember, that 2 is *the number of group means being compared* in the independent samples *t* test. Therefore, we are free to vary all but one score in each group (2 scores in total).

If you look at Kasser and Sheldon's (2000) results closely, you will notice something strange about the degrees of freedom that they reported. Remember that the sample size in this experiment was 60; therefore, shouldn't the degrees of freedom be 58 (i.e., 60 – 2)? What is going on here? Now is a good time to examine assumptions that are made when using an independent samples *t* test. Here are those assumptions:

Assumptions of the tool

1. The variability of scores in one group must be equal to the variability of scores in the other group. This assumption is called **homogeneity of variances**. That is, one group should not have a higher standard deviation than the other group (remember from Chapter 4 that the variance of scores in a group is obtained simply by squaring the standard deviation). Take a look back at the results from Kasser and Sheldon's (2000) experiment, and consider the standard deviation for each group. When conducting an independent samples *t* test, these standard deviations should be equal.

Homogeneity of variances: assumption of the independent samples *t* test that the variability of each group is approximately equal.

Indeed, this first assumption of the independent samples *t* test is a problem in this experiment. We will deal with how to tell whether this assumption is violated later in this chapter. For now, we can say that when this assumption is violated, researchers incur a fine when conducting the *t* test, much like a motorist incurs a fine when pulled over for speeding. Specifically, they lose degrees of freedom for this statistical tool. Remember that the higher the number of degrees of freedom a researcher has, the more likely he or she can generalize results from a sample to the population. Therefore, violating this assumption and, thus, losing degrees of freedom makes it less likely that a researcher can generalize results to the population.

In addition to homogeneity of variances, there are three other assumptions of the independent samples *t* test:

2. Perhaps most obvious, each observation (i.e., participant) must be in only one of the two groups.
3. The data were scale (interval or ratio) data; see Chapter 2 for a review of scales of data measurement.
4. The distributions of scores that comprise each group mean are normally distributed (see Chapter 5 for a review of normal data distributions).

Testing the null hypothesis

So, at this point, we have a *t* test statistic and our degrees of freedom. We can use these two pieces of information to tell whether we can generalize our result to the population. To do so, first recall from Chapter 6 that when testing a hypothesis, we need to know how likely some score or outcome was to occur by chance. The assumption that the distributions of scores are normally distributed becomes critical here. As you know, when a distribution of scores is normally distributed, we can specify how likely a specific score is to occur. In this example, we know what our specific score is; ours is the *t* test statistic. The *t* test statistic for pleasure spending was 2.30. As researchers, we want to know, "Can we generalize this outcome from our sample data to the larger population?" Armed with degrees of freedom, we can determine whether this difference between group means generalizes to our population. We must compare our *t* statistic of 2.30 to a critical value to learn whether the *t* statistic is large enough to allow us to draw conclusions about our population from our sample data.

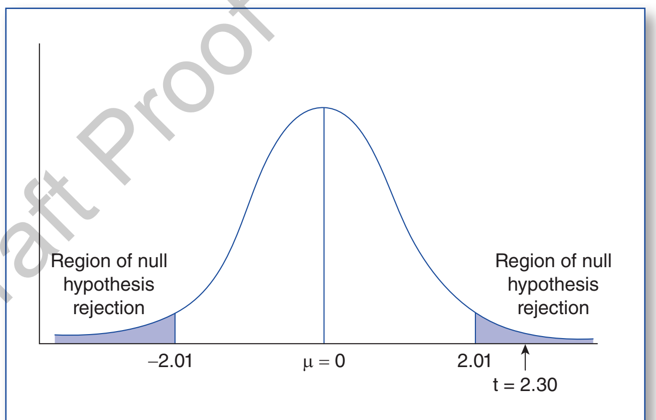
Take a look at Appendix B. It contains the critical values for the distribution of possible *t* statistics. We need to compare our *t* test statistic with the correct critical value. It is unusual in a real research study to be able to locate the precise critical value on a table like the one in Appendix B. Of course, for the purposes of understanding the process of the independent samples *t* test, we can do so here. We have 52.8 degrees of freedom. We must know our alpha level, for which we will continue to use .05.

As we discussed in the previous chapter, statistical tests reported in research are typically two tailed, even with directional hypotheses, as we had; that is, we expected the mortality salience group to indicate higher levels of pleasure spending than the control group. Again, researchers generally prefer two-tailed hypothesis tests because they are more conservative than one-tailed hypothesis tests. To reiterate this point, take a look at the critical values for the *t* distribution. By using our two-tailed test with an alpha of .05 and 52.8 degrees of freedom, we can look to the degrees of freedom that are provided in this appendix. Here, we can estimate our critical value with the degrees of freedom that are given in the appendix of 40 and 60. By using these two degrees of freedom, we see that our critical value is between ± 2.021 and ± 2.000, which are the critical values for 40 and 60 degrees of freedom, respectively. Had we used a one-tailed test, our critical value would have been between ± 1.684 and ± 1.671. Given a smaller critical value with the one-tailed test, there is a great likelihood of rejecting the null hypothesis.

When conducting actual research, it is indeed unusual for researchers to have the exact number of degrees of freedom that appear on these critical value tables. Fret not, because the good news is that the software program Statistical Package for the Social Sciences (SPSS; SPSS Corporation, Chicago, IL) has all the critical values memorized for us, and when conducting research, it is software such as SPSS that we will rely on. In the practice of doing research, no one consults critical value tables, but it is good to understand the role they play in the research process.

Let's go back to our *t* test statistic, which was 2.30. We need to compare this statistic with the critical value for a two-tailed hypothesis test with 52.8 degrees of freedom. Given that critical value is between ± 2.000 and ± 2.021, we see that our *t* statistic is greater than that critical value. Let's look at Figure 7.2. We are testing the hypothesis that there is no (i.e., 0) difference between the group means in that population. To reject this notion, we need a test statistic that is greater than 0. How much greater is "greater than 0"? If the test statistic meets or exceeds the critical value, the test statistic falls in the region of null hypothesis rejection, and we reject the null hypothesis. As you see in Figure 7.2, our test statistic (2.30) is indeed greater than the critical value (which is between ± 2.000 and ± 2.021). Therefore, in this research, Kasser and Sheldon (2000) were able to reject the null hypothesis and conclude that people in the mortality-salience condition indicated a greater likelihood to spend money on pleasure than did people in the control/music condition.

Figure 7.2 A Visual Comparison of the Test Statistic (2.30) With the Critical Value (±2.009); Because the Test Statistic Falls in the Region of Null Hypothesis Rejection, We Reject the Null Hypothesis



We know that our test statistic exceeds our critical value and we reject the null hypothesis. Referring to Kasser and Sheldon's (2000) results, there are other pieces of information it includes that you should be aware of. Specifically, we should discuss the *p* value. Remember from the previous chapter that a *p* value (or "probability" value) tells us how likely a statistical result was obtained because of random variation and is likely to be a Type I error. In other words, was the difference between group means merely a fluke? Researchers generally prefer smaller *p* values because that means there is little chance that the differences between the group means emerged because of chance factors.

We see that the *p* value is $p = .02$. What does that mean in plain English? It means that there is a 2% chance that the difference between the means (i.e., 0.22 and −0.27) was due to random variation. We want this percentage to be as close to 0 as possible. However, there is always a chance that some statistical result was due to random variation. In any *t* test reported, there is almost guaranteed to be some difference between the means. The question is whether that difference is large enough, accounting for variability and sample size, to warrant making conclusions about our larger population. Remember that our "alpha" or "significance level" of 5% means that if there is less than a 5% chance a result in the sample occurred by chance, then we are willing to draw conclusions about our population. We say such results are *statistically significant*. Here, with a *p* value of .02 (i.e., 2%), we can reject the null hypothesis that there will be no difference between the two groups on estimates of pleasure spending in the future. Therefore, we can conclude that participants who wrote essays about their deaths were more likely to think they will spend more money on pleasure in the future than people who did not contemplate their deaths.

Extending our null hypothesis test

In addition to testing the null hypothesis, it is becoming for researchers to provide two additional pieces of information: (1) an effect size and (2) a confidence interval.

The **effect size** is the amount of variability in a dependent variable that can be traced to the independent variable. In essence, the effect size allows us to learn how big an impact the independent variable has on the dependent variable. In our discussion of critical values, remember that the more degrees of freedom you have, the more likely you will be to reject the null hypothesis. Degrees of freedom are contingent on sample size; thus, so too are the critical values that determine how large the region of null hypothesis rejection will be. If you have a large enough sample size, we can reject almost any null hypotheses, no matter how small a relationship there is in the population. For instance, if my 8 a.m. statistics class exam mean is 75% and my 2 p.m. statistics class exam mean is 74.9%, is that a statistically significant difference? If the sample size was large enough, it could be. However, is a difference of 0.1% a particularly strong effect? Reporting an effect size provides such information.

For the independent samples *t* test, the effect size is normally reported in Cohen's *d* (which is typically reported as simply *d*). As you can see in Kasser and Sheldon's (2000) results, the effect size for the dependent variable of pleasure spending is 0.61. What does this mean? There are benchmarks for determining what constitutes a weak, moderate, and strong effect size. Specifically, Cohen (1992) suggested that an effect size, as measured by Cohen's *d*, of less than 0.20 is trivial; an effect size between 0.20 and 0.50 is weak; an effect size between 0.51 and 0.80 is moderate; and an effect size greater than 0.80 is strong. Thus, in our instance, our effect size of 0.61 is moderate. Therefore, the independent variable has a moderate effect on the dependent variable.

The effect size statistic complements the notion of statistical significance. When we ask whether a result is statistically significant, we are asking whether we can take our result obtained in our sample and draw conclusions about our population from which that sample was drawn. When we speak about effect size, we are asking how powerful an effect (in this case, an independent variable) is in affecting behavior. Remember that a result can be statistically significant because it was found using a large sample. That's because large samples will better reflect the populations than will smaller samples. However, just because a result is statistically significant does not mean it has a powerful effect on behavior.

In addition to knowing effect sizes, it will be helpful to know how **confidence intervals** (CIs) are used in reporting statistical results. A confidence interval is a range of possible differences between group means in which we feel confident that the actual group mean difference will lie in the larger population. Stated differently, if we could test an infinite number of samples and calculate the interval for each sample, 95% of those samples would contain the actual population mean. Let's unpack that explanation.

In Kasser and Sheldon's (2000) results, for the dependent variable of pleasure spending, we have a confidence interval of [0.06, 0.92]. What does that mean? Recall that the mean difference between the two groups on this dependent variable was 0.49. If we could draw an unlimited number of samples, 95% of those samples would contain a mean difference between 0.06 and 0.92. Here is the important part, that *the range of possible mean differences in the population does not include 0*. Why is it so important that this range not include 0? If it did, that would mean it is reasonably possible that the mean difference in the population is 0 (i.e., there is no mean difference in the population).

Confidence intervals complement the notion of statistical significance. Again, when we ask whether a result is statistically significant, we are asking whether we can take our result (here, a difference between two means) obtained in our sample and draw conclusions about our population from which that sample was drawn. *When we add a confidence interval to the test of statistical significance, we learn the range of plausible values that our result could take on in the population.*

Effect size: statistical measurement of how powerful the relationship is between variables (e.g., between an independent variable and a dependent variable).

Confidence interval: interval estimate that contains the population mean a certain percentage of time, based on repeated sampling of that population.

LEARNING CHECK

Now that we've seen the independent samples *t* test in action, let's review our conceptual understanding of this tool. In addition to the dependent variable of pleasure spending that we just discussed in great detail, Kasser and Sheldon (2000) also had a dependent variable called "financial worth." That is, participants provided a dollar estimate of how much they would be worth financially 15 years into the future. Here are the statistical results for this dependent variable:

> An independent samples *t* test suggested that participants primed to think about their deaths estimated they would be worth more money (M = 0.16, SD = 0.94) than participants primed to think about music (M = −0.23, SD = 0.38), $t(44.5)$ = 1.99, p = .05, d = 0.54, 95% CI [0.00, 0.77].

Now, answer the following questions. Then check your answers.

1. What is the mean difference between the two groups being examined?
2. What is the *t* test statistic?
3. What is the standard error of the difference between the means? (*HINT:* It is not reported above.)
4. How many degrees of freedom do the researchers have for this analysis?
5. Was the assumption of homogeneity of variance violated or not violated? How do you know?
6. By using Appendix B, approximate the critical value that was used to see whether we reject or do not reject the null hypothesis.
7. What is the probability that the difference between the two groups' means was due to random variation?
8. Did the researchers reject or not reject the null hypothesis?

9. Given your answer to the previous question, what does it mean in plain English?
10. By using Cohen's (1992) guidelines, interpret the effect size.
11. Interpret the 95% confidence interval.

Answers:

1. 0.16 − (−0.23) = 0.39
2. 1.99
3. $$1.99 = \frac{0.39}{\text{Standard error of the difference between the means}}$$

   Standard error of the difference between the means = 0.196
4. 44.5
5. Yes, this assumption was violated because normally, for the independent samples *t* test, degrees of freedom are measured "sample size − 2." Had that been the case, there would have been 60 − 2 = 58 degrees of freedom. This assumption was violated because we have only 44.5 degrees of freedom. We lost degrees of freedom because of this violation.
6. We cannot locate the precise degrees of freedom in this appendix, but we can locate critical values for 40 and 60 degrees of freedom. Our 44.5 degrees of freedom falls in between these two parameters, so let's use the critical value for 40 degrees of freedom (which is ±2.021) and 60 degrees of freedom (which is ±2.000). By using these two critical values, we can approximate our critical value to be about ±2.01.
7. $p = .05$ (5% chance the mean difference was due to random variation).
8. The *t* test statistic of 1.99 falls just below the critical value of 2.01; therefore, we fail to reject the null hypothesis. (*NOTE:* The researchers said $p = .05$; therefore, *p* was *not less than* .05, which it would need to be to reject the null hypothesis.)
9. Writing about one's death or writing about music did not affect college students' estimates of their overall financial worth 15 years into the future.
10. $d = 0.54$ is a moderate effect size.
11. If we could draw an unlimited number of samples from this population, 95% of those samples would contain a mean difference between .00 and .77. Because it contains zero, we cannot be confident that there is a mean difference on this dependent variable in the population.

USING YOUR NEW STATISTICAL TOOL

Now that you have a good understanding of what an independent samples *t* test tells us and how to interpret this tool, let's put it to use!

Hand-Calculating the Independent Samples *t* Test

We will first calculate an independent samples *t* test by hand. Then, you will have the opportunity to practice these calculations in the next Learning Check. Afterward, we will learn how to conduct and interpret the independent samples *t* test with the help of SPSS.

As an extension of Kasser and Sheldon's (2000) experiment, let's add another dependent variable that they could have measured, namely, materialistic values. It seems reasonable, according to terror management theory, that if people just wrote about their deaths, they would likely be more materialistic than if they just wrote about music. One way to fit into Western cultures is to advertise one's status via material possessions. Thus, when reminded that we will in fact die at some point, we might express materialistic values as a means to restore our sense of power over our surroundings.

Step 1: State hypotheses

Remember, we must statistically test the null hypothesis:

$$H_o: \mu_{\text{mortality-salience group}} = \mu_{\text{control group}}$$

In terms of the research hypothesis, we would expect the following:

$$H_r: \mu_{\text{mortality-salience group}} > \mu_{\text{control group}}$$

There are several self-report measures of materialism available. Marsha Richins and Scott Dawson's (1992) 18-item scale has people respond to statements such as "I like to own things that impress people" and "I have all the things I really need to enjoy life" (reverse-coded) using a 1 (*I strongly disagree with the statement*) to 5 (*I strongly agree with the statement*) response range. Thus, scores on this scale can range from 18 to 90. Here are hypothetical data for 12 participants in this experiment:

Experimental Group					
Mortality Salience (Death Essays)			Control (Music Essays)		
72	67	50	51	37	33
44	70	57	78	35	42

Step 2: Calculate the mean for each of the two groups

$$\text{Mortality Salience Group Mean} = \frac{(72 + 67 + 50 + 44 + 70 + 57)}{6} = 60$$

$$\text{Control Mean} = \frac{(51 + 37 + 33 + 78 + 35 + 42)}{6} = 46$$

We now have the numbers needed for the numerator of our *t* statistic:

$$t = \frac{60 - 46}{\text{Standard error of the difference between the means}}$$

Step 3: Calculate the standard error of the difference between the means

These calculations are a little more intricate than calculating the means, but you already know how to do everything because you read Chapter 4, in which we learned about the notions of standard deviation and variance. Recall that the standard deviation for a group of scores is simply the square root of the variance. Or stated another way, the variance is the standard deviation squared.

Here is the formula for the standard error of the difference between the means:

$$s_{x1-x2} = \sqrt{\frac{s_1^{\ 2}}{n_1} + \frac{s_2^{\ 2}}{n_2}}$$

where

$s_1^{\ 2}$ is the variance for the first group (the mortality salience group),
n_1 is the number of people in the first group,
$s_2^{\ 2}$ is the variance for the second group (the control/music group), and
n_2 is the number of people in the second group.

The difficult part is the calculation of the variance for each group (S^2). However, it is not difficult. We did this back in Chapter 4. Here again is what we do:

For each group, take each score and subtract from it the group mean. Doing so will yield a "deviation from the mean" score. So, for the mortality salience group,

Individual Score	Group Mean	(Deviation From Mean)
72	60	12
67	60	7
50	60	−10
44	60	−16
70	60	10
57	60	−3

And for the control (music) group:

Individual Score	Group Mean	(Deviation From Mean)
51	46	5
37	46	−9
33	46	−13
78	46	32
35	46	−11
42	46	−4

As we know from Chapter 4, simply adding together the individual scores' deviations from the mean is pointless as that summation will always be 0. Therefore, we square each deviation from the mean as doing so eliminates the negative numbers.

For the mortality salience group,

Individual Score	Group Mean	(Deviation From Mean)	$(\text{Deviation})^2$
72	60	12	144
67	60	7	49
50	60	−10	100
44	60	−16	256
70	60	10	100
57	60	−3	9

And for the control (music) group,

Individual Score	Group Mean	(Deviation From Mean)	$(\text{Deviation})^2$
51	46	5	25
37	46	−9	81
33	46	−13	169
78	46	32	1,024
35	46	−11	121
42	46	−4	16

Now, we sum the $(\text{deviations})^2$ for each group:

Mortality Salience Group Mean = (144 + 49 + 100 + 256 + 100 + 9) = 658

Control Mean = (25 + 81 + 169 + 1,024 + 121 + 16) = 1,436

We then divide each sum by the number of people in that group, minus 1.

Mortality salience group: $\frac{658}{6-1}$ Control group: $\frac{1{,}436}{6-1}$

Thus, the variance for the mortality salience group is 131.6. The variance for the control group is 287.2. Let's plug these numbers into the formula for the standard error of the difference between the means:

$$\sqrt{\frac{131.6}{n_1}+\frac{287.2}{n_2}}$$

It now gets much easier. The sample size for each group is 6. Let's plug those sample sizes into the formula:

$$\sqrt{\frac{131.6}{6}+\frac{287.2}{6}}$$

Simplifying this equation a little bit gives us

$$\sqrt{21.933+47.867}$$

Simplifying even further gives us

$$\sqrt{69.8}$$

Thus, our standard error of the difference between the means is 8.3546. It's all downhill from here.

Step 4: Calculate the *t* test statistic

Recall from step 2 that our group means were 60 for the mortality salience group and 46 for the control/music group. We have all of our ingredients to perform our independent samples *t* test:

$$t=\frac{\text{Mean difference between the two groups}}{\text{Standard error of the difference between the means}}$$

$$t=\frac{60-46}{8.3546}$$

$$t=1.68$$

What information does this *t* test statistic of 1.68 tell us? Remember, we want to know whether the difference between the means found in our sample is generalizable to the population. Obviously, 60 is greater than 46, but can we conclude that people who contemplated their deaths are more materialistic than those who instead contemplated music?

Step 5: Determine degrees of freedom (*df*s)

We now need to know how much freedom we have to generalize our results to the population. Remember that for an independent samples *t* test, degrees of freedom = sample size – 2. Therefore, given we have data from 12 people, we have 10 degrees of freedom here (i.e., 12 – 2; we are assuming that we have met the four assumptions for the independent samples *t* test). Armed with our degrees of freedom, we must now find our critical value with which to compare our *t* test statistic. If our *t* test statistic exceeds the critical value, we will reject the null hypothesis. If our *t* test statistic is less than the critical value, we do not reject the null hypothesis.

Step 6: Locate the critical value

Consult Appendix B to find our critical value. In returning to our two-tailed test of the null hypothesis with a significance level of .05, our test statistic needs to be greater than ± 2.228 for us to reject the null hypothesis.

Step 7: Make a decision about the null hypothesis

Visually, Figure 7.3 shows how to conceive of using the t test statistic we calculated and the critical value to reject or to fail to reject our null hypothesis. As you can see, our t test statistic of 1.68 falls below our critical value of ± 2.228. Therefore, we cannot reject the null hypothesis. In plain English, we conclude that people express similar levels of materialistic values regardless of whether they had their mortality made salient to them.

Step 8: Calculate an effect size

Remember that we have a fairly small sample size in this example. With a small sample size, it is more difficult to find a statistically significant result than with a larger sample size. However, as we discussed previously, just because we fail to reject the null hypothesis does not mean the effect of the independent variable is nonexistent. Let's calculate the effect size (Cohen's d) for this example. Here is the formula for Cohen's d:

$$d = \frac{\text{Mean difference between the two groups}}{\sqrt{\frac{\text{Variance of group 1}}{2} + \frac{\text{Variance of group 2}}{2}}}$$

We have already calculated all of the information we need to determine Cohen's d:

1. The mean for each group
2. The variance for each group

From step 2 in this section, the mean for the mortality salience group was 60, and the mean for the control (music) group was 46. From step 3, the variance for the mortality salience group was 131.6, and the variance for the control (music) group was 287.2. Let's plug in the numbers for the dependent variable of materialism:

$$d = \frac{60 - 46}{\sqrt{\frac{131.2}{2} + \frac{287.2}{2}}}$$

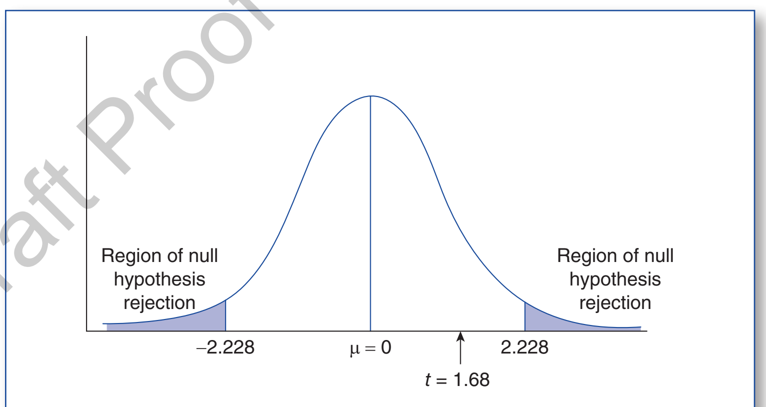
$$d = \frac{14}{\sqrt{65.6 + 143.6}}$$

$$d = \frac{14}{\sqrt{209.2}}$$

$$d = \frac{14}{14.464}$$

$$d = .97$$

Figure 7.3 A Visual Comparison of the Test Statistic (1.68) With the Critical Value (±2.228); Because the Test Statistic Falls Outside the Region of Null Hypothesis Rejection, We Fail to Reject the Null Hypothesis



Keep in mind the benchmarks for interpreting Cohen's d. Specifically, Cohen (1992) suggested that an effect size of less than 0.20 is trivial; an effect size between 0.20 and 0.50 is weak; an effect size between 0.51 and 0.80 is moderate; and an effect size greater than 0.80 is strong. Thus, our effect size of 0.97 is strong.

Step 9: Determine the confidence interval

Finally, we must consider the confidence interval around our mean difference. To calculate a 95% CI for an independent samples *t* test, here is our formula:

> 95% CI = (mean of group 1 − mean of group 2) ± critical value for t × (standard error of the difference between the means)

We have already calculated all of the ingredients needed to calculate our confidence interval:

Mean of group 1 (mortality salience) = 60

Mean of group 2 (control/music) = 46

Critical value for *t* (which we found while determining the region of null hypothesis rejection) = ±2.228

Standard error of the difference between the means = 8.3546

95% CI = 60 − 46 ± 2.228(8.3546)

95% CI = 14 ± 18.614

95% CI = [−4.614, 32.614]

As you can see, our interval contains 0, which means we cannot be confident that the difference we observed in the sample exists in the population.

Now, we have calculated all sorts of numbers, but they are worthless unless we can communicate them concisely. Here is how we would report the results in the text of an article of the analyses we just conducted, using APA style:

> An independent samples *t* test suggested that participants primed to think about their deaths were no more materialistic (M = 60.00, SD = 11.47) than were participants primed to think about music (M = 46.00, SD = 16.95), $t(10)$ = 1.68, p > .05, d = 0.97, 95% CI [−4.61, 32.61].

Remember that the standard deviation (*SD*) is the square root of the variance. We never calculated the *SD* in this section, but we did calculate the variance (in step 3). So here we just took the square root of the variance to get our standard deviation.

LEARNING CHECK

Here are some additional examples to get more practice hand-calculating and interpreting the independent samples *t* tests. The example contains raw data from individual participants, so the calculations are involved. The second example provides summary statistics, so the calculations will be a little less involved than what we've been doing up to this point in this section.

Problem #1

Andrew Christopher and Mark Wojda (2008) conducted a study that examined, among other considerations, sex differences in political attitudes. Specifically, among a sample of 246 working adults, they measured social dominance orientation

(Continued)

(Continued)

(Pratto, Sidanius, Stallworth, & Malle, 1994). Social dominance orientation is a person's preference for social hierarchy and stratification. Participants indicated how negatively or positively they felt about statements such as "Increased economic equality (reverse-coded)" and "Some people are just inferior to others" on a 0 (*very strongly negative*) to 8 (*very strongly positive*) response range. Here are scores from 12 participants on Felicia Pratto and colleagues' (1994) 14-item social dominance measure:

Women	Men
18	51
33	40
42	36
31	60
49	52
25	43

Questions to Answer:

1. What is the hypothesis being tested?
2. What is the mean difference between the two groups being examined?
3. What is the standard error of the difference between the means?
4. What is the *t* test statistic?
5. How many degrees of freedom do the researchers have for this analysis?
6. According to Appendix B, what is the critical value that was used to see whether we reject or fail to reject the null hypothesis?
7. What is the probability that the difference between the two groups' means was due to random variation?
8. Did the research reject or fail to reject the null hypothesis?
9. Given your answer to the previous question, what does that mean in plain English?
10. Calculate the effect size and interpret it according to Cohen's (1992) guidelines.
11. Calculate and interpret the 95% confidence interval.
12. Write these results for the text of an article in proper APA style.

Answers:

1. There is no difference between women and men on social dominance orientation scores.
2. Mean for women = 33; mean for men = 47; mean difference = 14

3. 5.85
4. $t = \frac{47 - 33}{5.85}$

   $t = 2.39$
5. 10
6. ±2.228
7. Less than 5% because the *t* statistic of 2.39 exceeds the critical value of ±2.228.
8. Given the answer to question 7, we reject the null hypothesis.
9. Men scored higher on social dominance orientation than did women.
10. *d* = 1.38; this is a strong effect size. That is, a person's sex was strongly predictive of his or her social dominance orientation score.
11. The 95% confidence interval is −0.97 to −27.03, meaning that if we could draw an unlimited number of samples from this population, 95% of those samples would contain a mean difference between −0.97 and −27.03. That this interval does not contain 0 means we can be confident there is a difference between women and men on social dominance orientation in the population.
12. Here is the proper APA style write-up:

An independent samples *t* test suggested that men tended to score higher in social dominance orientation (*M* = 47.00, *SD* = 8.90) than did women (*M* = 33.00, *SD* = 11.22), *t*(10) = 2.39, *p* < .05, *d* = 1.38, 95% CI [−0.97, −27.03].

This example brings up a methodological point about the independent samples *t* test. In this example, participants were classified as either female or male. So indeed, participants were in only one of the two groups being compared. However, unlike in Kasser and Sheldon's (2000) research, *participants in Christopher and Wojda's (2008) research were not randomly assigned to groups.* It's not possible to randomly assign someone to be a woman or a man. Although there is no random assignment to groups as there would be in a true experiment, you can still use the independent samples *t* test whenever you want to compare two mutually exclusive groups.

Problem #2

A forensic psychologist wants to know whether men convicted of robbery receive longer sentences when they have an ethnic-sounding first name (e.g., Declan) or a more traditional-sounding first name (e.g., David). After sampling 30 male convicts, 15 with an ethnic-sounding first name and 15 with a more traditional-sounding first name, here is what this researcher found:

a) Mean sentence for men with ethnic-sounding first names = 5.5 years with a standard deviation of 2.5 years
b) Mean sentence for men with more traditional-sounding first names = 4.0 years with a standard deviation of 1.75 years

   1. What is the population being studied?
   2. Why couldn't this researcher use random assignment?

(Continued)

(Continued)

3. What is the hypothesis being tested?
4. What is the mean difference between the two groups being examined?
5. What is the standard error of the difference between the means?
6. What is the *t* test statistic?
7. How many degrees of freedom do the researchers have for this analysis?
8. By using Appendix B, find the critical value that was used to see whether we reject or do not reject the null hypothesis.
9. What is the probability that the difference between the two groups' means was due to random variation?
10. Did the researchers reject or fail to reject the null hypothesis?
11. Given your answer to the previous question, what does that mean in plain English?
12. Calculate the effect size and interpret it according to Cohen's (1992) guidelines.
13. Calculate and interpret the 95% confidence interval.
14. Write these results for the text of an article in proper APA style.

Answers:

1. The population is men who have been convicted of robbery.
2. We cannot assign people to be men, nor can we assign them to be robbery convicts. It is simply not practical to use random assignment in this research.
3. There will be no difference between male robbers who have ethnic-sounding first names and male robbers who have more traditional-sounding first names in jail sentences for their crime.
4. Ethnic-sounding first name mean is 5.5 years; more traditional-sounding first name mean is 4.0 years, so the mean difference is 1.5 years.
5. 0.79
6. $t = \frac{1.50}{0.79}$

   $t = 1.90$
7. 28
8. ±2.048
9. Greater than 5% ($p > .05$)
10. Do not reject the null hypothesis because the *t* test statistic is less than the critical value.

11. There was no difference in jail sentences handed down to male burglars who had an ethnic-sounding first name versus those who had a more traditional-sounding first name.

12. $d = 0.69$. This is a moderate effect size. That is, the relationship between a person's first name and his jail term for robbery was moderately strong.

13. The 95% confidence interval is −0.12 to 3.13, meaning that if we could draw an unlimited number of samples from this population, 95% of those samples would contain a mean difference between −0.12 and 3.13. Because this interval contains 0 means, we cannot be confident there is a difference between convicted male robbers with ethnic-sounding and more traditional-sounding first names.

14. Here is the proper APA style write-up:

    An independent samples *t* test suggested that males convicted of robbery were sentenced to similar jail terms, regardless of whether they had an ethnic-sounding first name ($M = 5.50$ years, $SD = 2.50$ years) or a more traditional-sounding first name ($M = 4.00$ years, $SD = 1.75$ years), $t(28) = 1.90$, $p > .05$, $d = 0.69$, 95% CI [−0.12, 3.13].

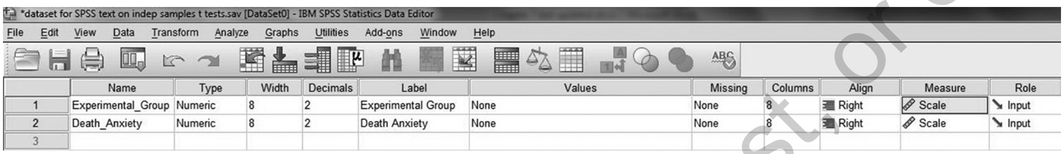
Independent Samples *t* Test and SPSS

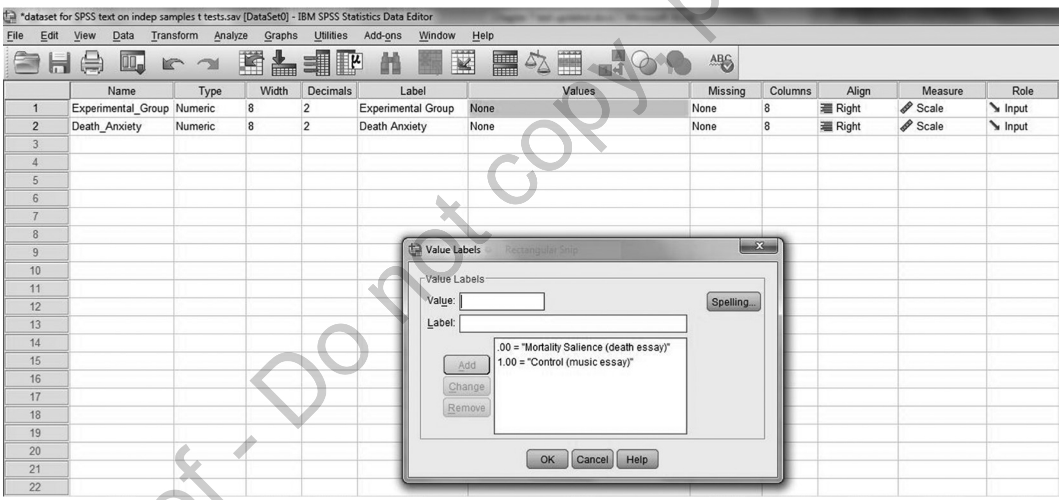
Now that you have a handle on the conceptual nature of the independent samples *t* test and know how to calculate it, let's run our own independent samples *t* test using SPSS. To do so, let's use some hypothetical data. Suppose that Kasser and Sheldon (2000) had included another dependent variable in their experiment. For instance, it might have been logical to include a measure of anxiety about one's death. Thus, it might make sense that writing about one's death would heighten a person's anxiety about his or her ultimate physical demise compared with writing about music. Jon Hoelter (1979) developed the Multidimensional Fear of Death Anxiety (MFOD) scale, in which people respond on a 1 (*strongly disagree*) to 5 (*strongly agree*) response range to items such as "Discovering a dead body would be a horrifying experience" and "I am afraid I will not have time to experience everything I want to." There are 42 items on this scale, so scores could potentially range from 42 to 210. We could easily use this scale and get an aggregate death anxiety score for each participant. Let's suppose we included this death anxiety scale. Here are the aggregate scores on it for a sample of 30 participants (15 in each experimental group):

Experimental Group					
Mortality Salience (Death Essays)			Control (Music Essays)		
112	75	150	100	75	98
187	100	119	82	90	140
152	112	125	65	115	111
136	162	88	112	133	90
87	147	63	84	68	71

Establishing your spreadsheet

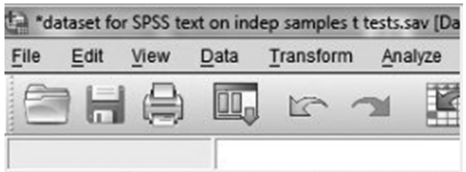
Once you open SPSS, you will need to click on *Variable View* in the lower left corner. You will need to name two variables, namely, the independent variable and the dependent variable. Again, the independent variable is the group to which participants in the sample are randomly assigned, either a mortality salience group (i.e., writing about death) or a control group (i.e., writing about music). Perhaps name this variable something such as `Experimental_Group`. Next, name the dependent variable something such as `Death_Anxiety`. After naming your variables, remember to give them labels. No need to get fancy or creative here. Let's simply label these variables "Experimental Group" and "Death Anxiety."





After naming your variables, you need to let SPSS know whether each participant had been randomly assigned to the mortality salience group or to the control group. To do so, click on *Values for the Experimental Group*. Here, you need to give numerical values to the group participants were randomly assigned to. For the first value, make it a `0` and label it "Mortality Salience" or "Death Essay," or something to that effect. For the next value, make it a `1` and label it "Control" or "Music Essay," or something to that effect. Click *OK* to close the *Values* window. Be sure, as well, that your measure choice is "Scale."

In the left corner of the screen, click on *Data View*. You are now ready to enter the data from these 30 participants. Remember, it makes no difference the order in which you enter data into SPSS. Just be sure to code correctly the experimental group that participants were randomly assigned to (`0` or `1`); then enter each aggregated score on the dependent variable (i.e., `Death_Anxiety`) for that participant. Here is what your file should look like:

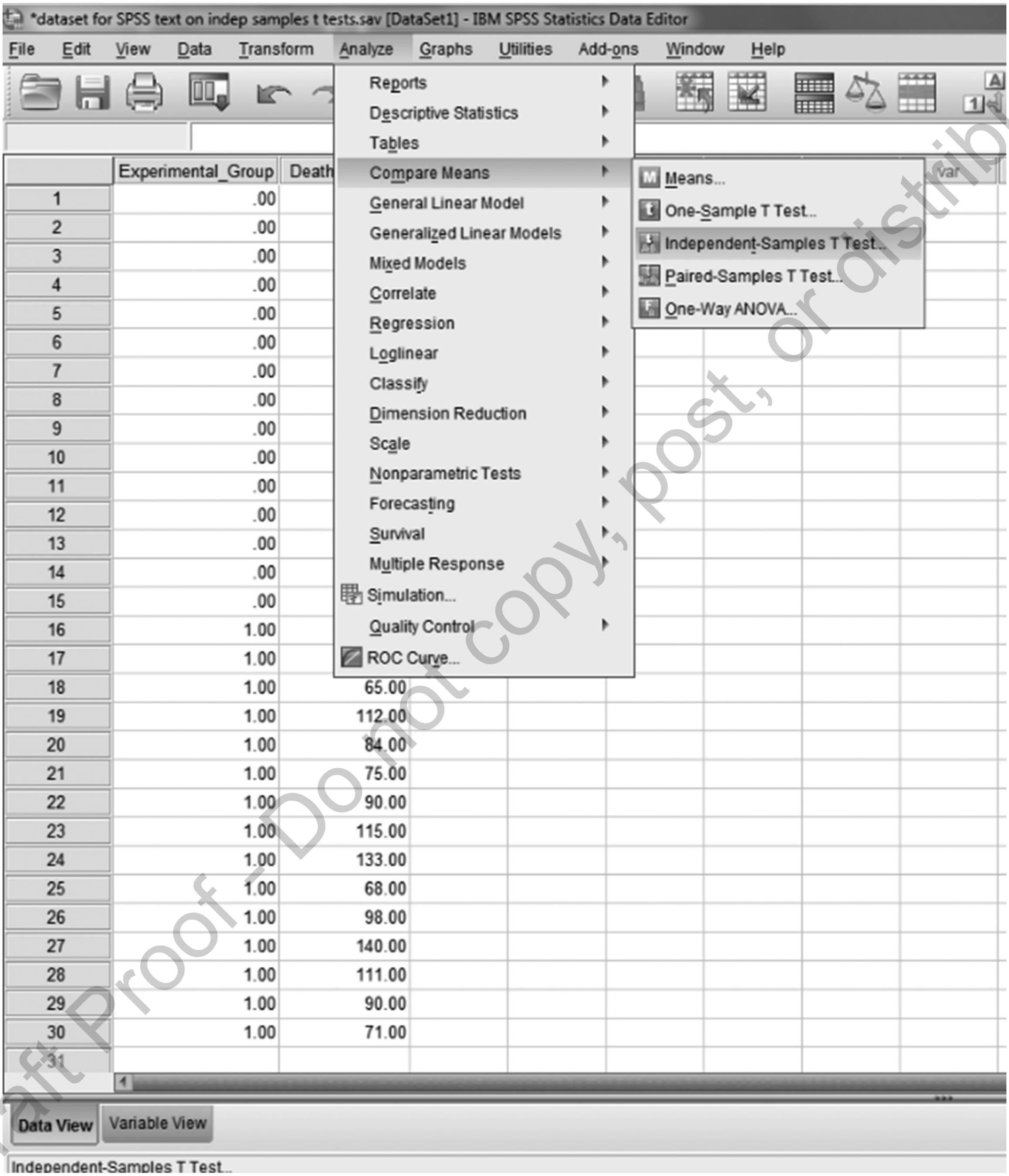


	Experimental_Group	Death_Anxiety
1	.00	112.00
2	.00	187.00
3	.00	152.00
4	.00	136.00
5	.00	87.00
6	.00	75.00
7	.00	100.00
8	.00	112.00
9	.00	162.00
10	.00	147.00
11	.00	150.00
12	.00	119.00
13	.00	125.00
14	.00	88.00
15	.00	63.00
16	1.00	100.00
17	1.00	82.00
18	1.00	65.00
19	1.00	112.00
20	1.00	84.00
21	1.00	75.00
22	1.00	90.00
23	1.00	115.00
24	1.00	133.00
25	1.00	68.00
26	1.00	98.00
27	1.00	140.00
28	1.00	111.00
29	1.00	90.00
30	1.00	71.00
31		

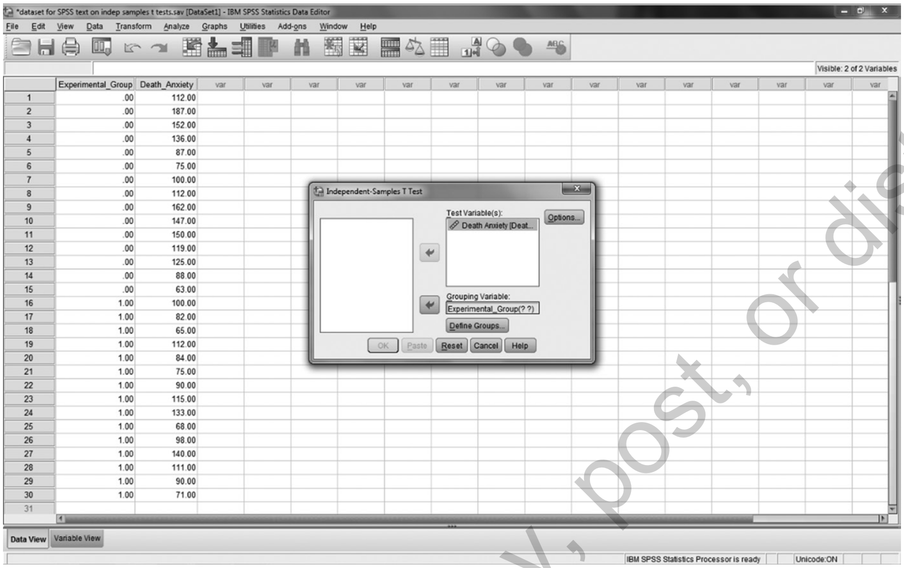
Running your analyses

Once you have entered all of these data into SPSS, you should have 30 rows of data (one row for each participant) and two columns of data (one for the independent variable and one for scores on the dependent variable). You are now set to run your independent samples *t* test. Here's how you run this analysis:

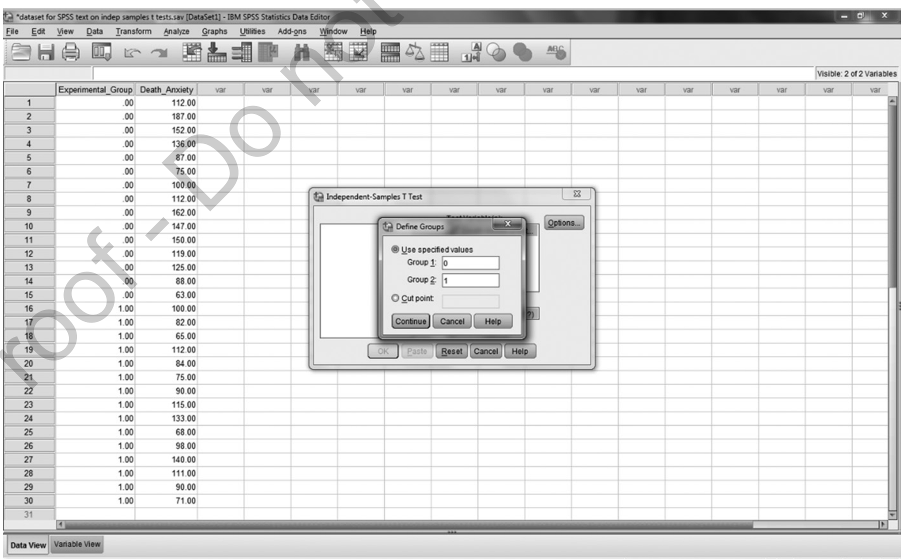
1. Click on *Analyze* → *Compare Means* → *Independent Samples T Test*.


*dataset for SPSS text on indep samples t tests.sav [DataSet1] - IBM SPSS Statistics Data Editor
File
Edit
View
Data
Transform
Analyze
Graphs
Utilities
Add-ons
Window
Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Simulation...
Quality Control
ROC Curve...
Means...
One-Sample T Test...
Independent-Samples T Test...
Paired-Samples T Test...
One-Way ANOVA...
Experimental_Group
Death
var
Data View
Variable View
Independent-Samples T Test...

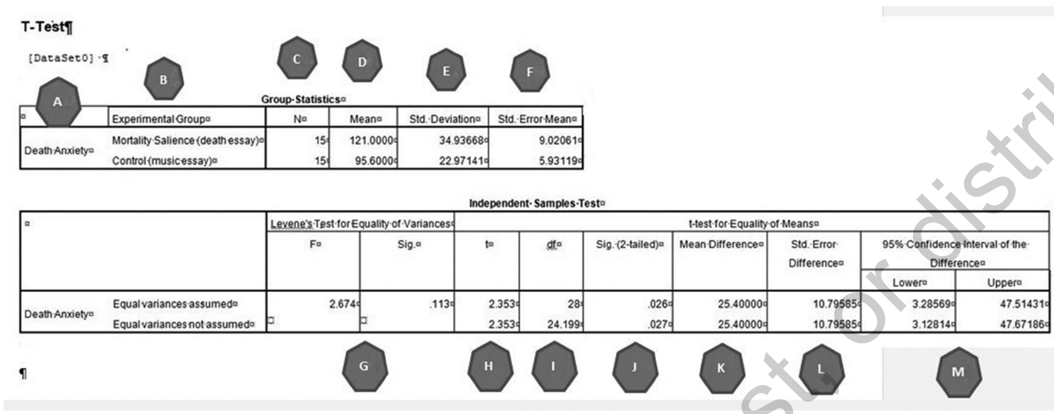
2. Your *Test Variable* is your dependent variable (i.e., `Death_Anxiety`). Your *Grouping Variable* is your independent variable (i.e., `Experimental_Group`).



3. Once you click over your *Grouping Variable*, you will need to *Define Groups*, so click on it. It makes no difference what you call group 1 and what you call group 2. However, remember how the independent variable was coded into SPSS (i.e., with `0`s and `1`s). So, for group 1, give it a `0`, and for group 2, give it a `1`. Click *Continue* and then *OK*, and you should soon have your results!



Hopefully, your output looks like this and contains the same numbers as you see here:



Group Statistics

	Experimental Group	N	Mean	Std. Deviation	Std. Error Mean
Death Anxiety	Mortality Salience (death essay)	15	121.0000	34.93668	9.02061
	Control (music essay)	15	95.6000	22.97141	5.93119

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Death Anxiety	Equal variances assumed	2.674	.113	2.353	28	.026	25.40000	10.79585	3.28569	47.51431
	Equal variances not assumed			2.353	24.199	.027	25.40000	10.79585	3.12814	47.67186

What am I looking at? Interpreting your SPSS output

Now, let's make sense of what we see on our output. There are lots of numbers here, of course, so we'll just take one bite at a time.

A. This is the *dependent variable*, `Death_Anxiety` (yes, it would be helpful if it were labeled as such in SPSS, but consider it a daily hassle). Be aware that the name that appears here is the name that I gave the dependent variable. Had I named the dependent variable something other than `Death_Anxiety`, that's what would have appeared here.

B. This is the *independent variable*, which I named `Experimental_Group`. In addition, and perhaps more importantly, each of the two groups that comprised the independent variable is noted here. Again, these names are the precise names that I provided in SPSS.

C. This is the *number of people in each experimental group*. Ideally, these numbers should be equal.

D. This is the *group mean* for each experimental group. The difference between these two numbers is the numerator for the *t* statistic.

E. This is the *standard deviation* for each group. The standard deviations are assumed to be the same across the two groups (the assumption of homogeneity of variances). Obviously, in this instance, these two numbers are different; what we need to know is whether they are statistically significantly different from each other. We'll deal with that question momentarily.

F. This is the *standard error of the mean* for each group. Recall from Chapter 6 that the standard error of the mean is in fact a standard deviation. Specifically, it is the standard deviation of a sampling distribution of sample means.

G. Here is where we statistically test the homogeneity of variance assumption of the independent samples *t* test. What would our null hypothesis be? We are testing the null hypothesis that there is no difference in variability between the two experimental groups. If we reject this hypothesis, we would need to conclude that there is a difference between the variability in each group and, therefore, that this assumption would be violated.

In this analysis, the *p* value is .113. As this *p* value is greater than .05, we do not reject the null hypothesis and conclude that the variability in the two groups was equivalent (i.e., we did not violate the homogeneity of variances assumption).

Be very careful when examining the Levene's test; it is not the independent samples *t* test! All the Levene's test does is tell us which line to read when interpreting the independent samples *t* test. That is, can the assumption of homogeneity of variances be assumed? If so, as was the case in this instance, read the line "Equal variance assumed." If this assumption has been violated, as it was in Kasser and Sheldon's (2000) experiment, then read the "Equal variances not assumed" line.

H. This is your *t test statistic*. Recall how it was calculated:

$$t = \frac{\text{Mean difference between the two groups}}{\text{Standard error of the difference between the means}}$$

In this example, the *t* statistic is 2.35. How did we get that statistic? Remember the group means under D. We take the difference of those two numbers (i.e., 121.00 and 95.60) and that gives us the numerator of 25.4. The standard error of the difference between the means is, well, hold tight. We'll find that on our output momentarily.

I. Here are our *degrees of freedom*. As we said earlier, the formula for the degrees of freedom is total sample (i.e., 30 in this example) – 2. It's easy enough to figure out when we don't violate the homogeneity of variance assumption. When we do violate this assumption, however, it is like getting a speeding ticket; we have to pay a fine. When we get a speeding ticket, that fine is of course money we must pay. When we violate the homogeneity of variances assumption, that fine is degrees of freedom. We lost some degrees of freedom for violating this assumption. This violation occurred in Kasser and Sheldon's (2000) experiment; hence, that is why the degrees of freedom they reported were not the straightforward "sample size – 2" normally used to calculate degrees of freedom for this statistic.

J. This is the *p value, or significance level,* of the *t* test statistic. Here, our *p* value is .026. In other words, there is a 2.6% chance that our observed difference between the two group means was due to random variation. Because this value is less than .05, we can conclude that the difference between the means was not due to random variation. Therefore, we can generalize this result from the sample to the population from which it was drawn.

K. Here is the *difference between the two group means* that are noted under D previously (in case you didn't feel like doing some simple subtraction).

L. This is the *standard error of the difference between the means*. It is the denominator of the *t* test statistic.

Here is our *t* test statistic and how it was calculated from this printout:

$$2.35 = \frac{121.00 - 95.60}{10.79585}$$

M. Here is *the 95% confidence interval around the mean difference*. As is the case in published research, the critical values are not specified on the SPSS printout, but they are being used to calculate this confidence interval.

Now, SPSS does not provide one piece of information that is essential when reporting and interpreting an independent samples *t* test. Specifically, it does not report the effect size. Why SPSS does not do this is an excellent question. I do not have an equally excellent answer. Fortunately, you're a whiz at effect sizes, so let's calculate the Cohen's *d* using information that SPSS does provide:

$$d = \frac{\text{Mean difference between the two groups}}{\sqrt{\frac{\text{Variance of group 1}}{2} + \frac{\text{Variance of group 2}}{2}}}$$

The variance of each group is not given on the SPSS output; however, the standard deviation is provided. Remember that to obtain the variance for a group, you simply square the standard deviation. Now let's calculate the Cohen's *d*:

$$d = \frac{25.4}{\sqrt{\frac{34.94^2}{2} + \frac{22.97^2}{2}}}$$

$$d = \frac{25.4}{\sqrt{\frac{1220.81}{2} + \frac{527.62}{2}}}$$

$$d = \frac{25.4}{\sqrt{610.41 + 263.81}}$$

$$d = \frac{25.4}{\sqrt{874.22}}$$

$$d = \frac{25.4}{29.57}$$

$$d = 0.86$$

In returning to Cohen's (1992) criteria for interpreting an effect size, we see that our *d* of 0.86 is a strong effect size. In plain English, the independent variable had a powerful effect on people's self-reported levels of death anxiety.

There is one other thing that SPSS does not do for us. It does not report the results in APA style. We need to do that ourselves, and so we shall:

> An independent samples *t* test suggested that participants primed to think about their deaths were more anxious about death (M = 121.00, SD = 34.94) than were participants primed to think about music (M = 95.60, SD = 22.97), $t(28)$ = 2.35, p = .026, d = 0.86, 95% CI [3.29, 47.51].

LEARNING CHECK

Now that we've calculated our independent samples *t* test by hand and know how to run one on SPSS, you know what might be fun? Why not enter the data from our hand calculations (in the previous section of this chapter) into SPSS, run the independent samples *t* test, and interpret our output? This would be a great way to test what you learned in the previous two parts of this chapter and will reinforce this information. Here again are those data (which are materialism scores for a sample of 12 people):

Experimental Group					
Mortality Salience (Death Essays)			**Control (Music Essays)**		
72	67	50	51	37	33
44	70	57	78	35	42

Hopefully your SPSS output looks like this:

T-Test

Group Statistics

	Experimental Group	N	Mean	Std. Deviation	Std. Error Mean
Materialism	Mortality salience (death essay)	6	60.0000	11.47170	4.68330
	Control (music essay)	6	46.0000	16.94698	6.91857

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Materialism	Equal variances assumed	.344	.571	1.676	10	.125	14.00000	8.35464	-4.61530	32.61530
	Equal variances not assumed			1.676	8.787	.129	14.00000	8.35464	-4.96956	32.96956

Questions to Answer:

1. What is the hypothesis being tested?
2. What is the mean difference between the two groups being examined?
3. What is the standard error of the difference between the means?
4. What is the *t* test statistic?
5. Was the assumption of homogeneity of variances violated? How do you know?
6. How many degrees of freedom do the researchers have for this analysis?
7. According to Appendix B, what is the critical value that was used to see whether we reject or fail to reject the null hypothesis?
8. What is the precise probability that the difference between the two groups' means was due to random variation?
9. Did the researchers reject or fail to reject the null hypothesis?
10. Given your answer to the previous question, what does that mean in plain English?
11. What is the effect size?
12. What is the 95% confidence interval?

(Continued)

(Continued)

Answers:

1. There will be no difference between the mortality salience condition and the control condition on materialism scores.
2. 60 − 46 = mean difference of 14
3. 8.3546
4. $t = 1.68$
5. The assumption was not violated. We know this because the Levene's Test for Equality of Variances produced a statistically insignificant result, meaning we cannot reject the null hypothesis that the variability between our two groups is equal.
6. 10
7. ±2.228
8. $p = .125$
9. The researchers failed to reject the null hypothesis because the *t* test statistic of 1.68 is less than the critical value of ±2.228.
10. People expressed similar levels of materialistic values in the mortality salience group and the control group.
11. $d = 0.96$
12. [−4.61, 32.61]

CHAPTER APPLICATION QUESTIONS

1. How does a one-sample *t* test differ from an independent samples *t* test?

   A: A one-sample *t* test is used to compare one group of data with some benchmark, normally the population mean. An independent samples *t* test assesses whether the difference in means between two groups drawn from the same population exists in that population.

3. Why would it be impractical to randomly assign people to a socioeconomic status (SES)?

   A: Socioeconomic status is a participant variable; that is, it is a naturally occurring characteristic that is not possible to assign someone to. A researcher cannot practically take a person who grew up in a middle-class environment and assign him or her to a lower SES status. Such an assignment would likely have no effect on people's behavior or mental processes.

5. Suppose you conducted a two-group study with 11 participants in each group. By using a two-tailed hypothesis test, what would your critical value be?

   A: ±2.086

6. Referring to the previous question, if you used a one-tailed hypothesis test, what would your critical value be?

   A: + 1.725

9. In terms of our ability to reject the null hypothesis, (a) explain how variability can be *a good thing* in an experiment; and (b) explain how variability can be *a bad thing* in an experiment.

   A: With respect to our ability to reject the null hypothesis: (a) Variability is a good thing when there is a large mean difference between the groups in an independent samples *t* test. (b) Variability can be a bad thing when there is a large variance (and, hence, standard deviation) within each group being statistically analyzed.

11. A statistically significant outcome is defined as an outcome that has a _____ probability of occurring if the _____ hypothesis is true.

   a) small; research
   c) small; null
   b) large; research
   d) large; null

   A: c

13. Which of the following *t* test statistics will have the strongest effect size?

   a) $t(28) = 2.16$
   b) $t(38) = 2.44$
   c) $t(48) = 2.72$
   d) it is impossible to know without more information

   A: d

14. Which of the following *t* test statistics is most likely to be statistically significant?

   a) $t(70) = 2.00$
   b) $t(35) = 2.00$
   c) $t(10) = 2.00$
   d) it is impossible to know without more information

   A: a

15. All other things being equal, what will happen to the value of the independent samples *t* test statistic as:

   a) the number of people in each group increases?

   A: *t* statistic increases

   b) the standard error of the difference between the means increases?

   A: *t* statistic decreases

   c) the difference between group means increases?

   A: *t* statistic increases

QUESTIONS FOR CLASS DISCUSSION

2. What is the difference between an independent variable and a dependent variable in an experiment?
4. Why is the independent samples *t* test called an "independent" test?
7. What information does an effect size provide over and above a hypothesis test?
8. What information does a confidence interval provide over and above a hypothesis test?
10. On average, what value is expected for the *t* statistic if the null hypothesis is true?
    a) 0
    b) 1
    c) .05
    d) cannot be determined without sample data
12. You see the following statistical analysis presented from an experiment with two groups to which participants were randomly assigned to one and only one group:

    $$t(48) = 3.21, p < .03, d = 0.70, 95\% \text{ CI } [3.00, 10.00]$$

    What can you conclude from this information?

    I. We can generalize this result from our sample to our population.
    II. There is a 70% chance that the independent variable caused an effect on the dependent variable.
    III. There was a statistically significant difference between the two groups' mean scores.
    IV. A total of 48 people's data were included in these analyses.
    a) I and II only
    b) I and III
    c) I, II, and III only
    d) all of the above are correct statements
16. Use this APA style report of an independent samples *t* test to answer the questions that follow it:

    When using a sample of 50 students, an independent samples *t* test suggested that children and adolescents who had a Big Brother-Big Sister mentor had higher grade-point averages (M = 3.25, SD = 0.51) than did children and adolescents without such a mentor (M = 2.75, SD = 0.60), $t(48) = 2.99$, $p = .002$, $d = 0.91$, 95% CI [0.19, 0.82].

    a) What was the hypothesis being tested?
    b) What is the mean difference between the two groups being examined?
    c) What is the *t* statistic?
    d) What is the standard error of the difference between the means?

e) Was the assumption of homogeneity of variance violated or not violated? How do you know?

f) By using Appendix B, approximate the critical value that was used to see whether we reject or do not reject the null hypothesis. Assume a directional hypothesis.

g) What is the probability that the difference between the two groups' means was due to random variation?

h) Did the researchers reject or not reject the null hypothesis?

i) Given your answer to the previous question, what does that mean in plain English?

j) By applying Cohen's (1992) guidelines, interpret the effect size.

k) Interpret the 95% confidence interval.