CHAPTER 10

Independent Samples $t$

LEARNING OBJECTIVES

After reading this chapter, you should be able to do the following:

- Explain when to use an independent samples $t$ test
- Explain the logic of the independent samples $t$ test
- Write null and research hypotheses using symbols and words for both one- and two-tailed tests
- Compute degrees of freedom and define a critical region for both one- and two-tailed tests
- Compute an independent samples $t$ using a calculator and SPSS
- Compute and interpret an effect size ($d$)
- Summarize the results of the analysis using American Psychological Association (APA) style

INDEPENDENT SAMPLES $t$

Thus far, you have learned to use two different $t$ tests: the single-sample $t$ and the related samples $t$. You should use the single-sample $t$ whenever you want to compare a sample mean to a population mean (or a value of theoretical interest) but you do not know the population’s standard deviation. You should use the related samples $t$ test when comparing the sample means from either the same individuals measured at two different times or pairs of matched people measured under different conditions. When you need to compare two sample means that are unrelated, you must use a third $t$ test: an independent (samples) $t$ test.

An example may help illustrate the difference between these three $t$ tests. Suppose that you are interested in the effect of a drug on weight loss. To test the efficacy of this drug, you could give the drug to a sample of people and, after they are on the drug for 1 month, compare the mean pounds lost by the sample to zero pounds. In this case, you would use a single-sample $t$. However, if you measured the sample’s mean weight before taking the drug and then again a month later after taking the drug, you would use a related samples $t$. Finally, you could also give one sample of people the drug and another sample a placebo. After both samples had taken their respective “drugs” for 1 month, you could compare the mean weight loss of the two samples. In this final option, you would use an independent samples $t$ test because the two samples contain different and unmatched people.
The independent $t$ test uses two samples from the population to represent two different conditions. As in the example earlier, it is often the case that one sample is intended to represent what the population would be like if nothing were done to it (i.e., a control condition), and another sample is intended to represent what the population would be like if it were given some treatment (i.e., experimental condition). The objective of the independent samples $t$ test is to determine if the difference between the two sample means is likely or unlikely to be due to sampling error.

The logic of the independent $t$ test is similar to that of the $z$ for the sample mean, the single-sample $t$ test, and the related samples $t$ test. All four tests compute a ratio, specifically, the observed deviation between two means over the deviation expected due to sampling error:

\[
\text{Obtained } t \text{ or } z = \frac{\text{Observed difference between the means}}{\text{Mean difference expected due to sampling error}}.
\]

For all of these tests, if the null hypothesis is true, the obtained $t$ or $z$ should be zero. However, if the null hypothesis is false, the obtained $t$ or $z$ should be far from zero.

1. Which significance test should you use to determine if the difference between two unrelated samples is likely to be due to sampling error?
   a. $z$ for a sample mean
   b. Single-sample $t$ test
   c. Independent $t$ test

2. If the null hypothesis is true, the $z$ for a sample mean, single-sample $t$ test, related samples $t$ test, and the independent $t$ test all expect an obtained value close to
   a. $+1.65$.
   b. $+1.96$.
   c. $+1.00$.
   d. $0$.

The independent $t$ test compares two sample means from two unrelated (i.e., independent) samples/groups. For example, suppose you and your friend Bill team up on a research project for your human cognition course. The two of you want to test how verbal labels influence participants’ memory of pictures. You investigate this question by showing all participants a series of 25 simple line drawings like those shown in Figure 10.1 and asking them to recall the drawings 10 minutes later. However, half of the participants only saw the drawings while the other half also saw a verbal description of each drawing similar to that in Figure 10.1. Your study is similar to one conducted by Bower, Karlin, and Dueck (1975).

You need to compare the memory of those who saw the drawings and verbal descriptions (i.e., the experimental group) to the memory of those who only saw the drawings (i.e., the control group). In this situation, you created two samples and gave each a different treatment. You then measured the mean number of drawings each sample recalled. The sample mean from the control group estimates what the population’s mean memory score would be if the population only saw the drawings with no verbal labels. In contrast, the other sample estimates what the population’s mean memory score would be if everyone saw the drawings and verbal labels. You can use an independent samples $t$ test.
to determine if the difference between these two sample means was likely or unlikely to have occurred due to sampling error. If the experimental group had a higher mean and the obtained $t$ value is in the critical region, you could conclude that the verbal descriptions increased memory scores. If, however, the verbal description group had a significantly lower mean, you could conclude that the verbal descriptions decreased memory scores. Figure 10.2 illustrates this research scenario.

In your study, you took one sample from a single population and then divided that sample to create two different groups; one received no verbal labels (i.e., control condition) and the other received verbal labels (i.e., experimental condition). You essentially took people who were similar and made them different by giving them different treatments (i.e., you gave them different levels of the IV [independent variable]). You then used the independent $t$ test to determine if the different IV levels affected memory differently.
An independent \( t \) test can also be used to compare two distinct populations of people who are already different in some way. In other words, an independent \( t \) test can compare groups with a pre-existing difference. For example, suppose you and Bill decide to design another memory experiment to test a different research question. You want to test a theory of learning called “learning styles.” The theory proposes that some people are “visual learners” and others are “verbal learners.” According to learning styles theory, “visual learners” should learn visual material better than “verbal learners” learn visual material. You want to test this prediction by determining if the mean visual memory score of “visual learners” is significantly higher than that of “verbal learners.” To test this hypothesis, you take a sample of “visual learners” and a sample of “verbal learners,” show both groups 25 simple line drawings (i.e., visual information), and measure their ability to recall the drawings 2 days later by asking them to re-create the 25 drawings. You then use the independent \( t \) test to determine if the two samples’ mean visual memory scores are significantly different. Figure 10.3 illustrates the “visual learner” versus “verbal learner” scenario.
3. An independent *t* test can be used to compare differences between people that
   a. are created by the researcher by providing different IV levels.
   b. already exist in different populations of people.
   c. Both of the above

**CONCEPTUAL FORMULA FOR THE INDEPENDENT SAMPLES *t***

Regardless of whether you have means from two groups with a preexisting difference or from two groups in which you created a difference, the same independent *t* test formula is used. The independent *t* test is the ratio of the observed mean difference over the difference expected due to sampling error:

\[
t = \frac{\text{Samples’ mean difference} - \text{Populations’ mean difference expected if } H_0 \text{ is true}}{\text{Mean difference expected due to sampling error}}.
\]

As indicated in the earlier “logical formula,” there are three terms in this *t* test. The two samples’ mean difference is determined by the data. In the “visual learners” versus “verbal learners” scenario, this term is the actual difference between the memory scores of “visual” and “verbal” learners. The null hypothesis determines the other term in the numerator. For example, if learning style (visual vs. verbal) has no impact on memory scores, we would expect the population mean memory score of the two groups to be the same and their mean difference to be zero. In most research situations (all situations in this book), the populations’ mean difference that is expected if the null is true is zero. In other words, the numerator is simply the difference between the two sample means. The term in the denominator represents the amount of sampling error expected. In the following formula, *M*₁ and *M*₂ are the sample means of Group 1 and Group 2, respectively.

\[
t = \frac{(M_1 - M_2)}{SEM}.
\]

4. The independent *t* test is a ratio of the difference between two sample means over an estimate of
   a. sampling error.
   b. the standard deviation of the scores.
   c. variability.

5. The numerator of the independent samples *t* test is the difference between
   a. two sample means.
   b. a sample mean and a population mean.
   c. a sample mean and the null hypothesis.
TWO-TAILED INDEPENDENT t TEST EXAMPLE

After designing your study investigating the effect of verbal descriptions on memory for simple line drawings, you and your friend Bill discuss your predictions. Bill thinks that providing the verbal descriptions along with the line drawings will distract the participants from attending to the line drawings. So, Bill thinks that the mean memory score for the verbal description group will be lower than the mean memory score for the no verbal description group. You disagree. You think that the verbal descriptions will help give meaning to the otherwise abstract line drawings, and this greater meaning should increase memory scores. So, you think the verbal description group will have the higher mean memory score. Because there are two competing theories that make opposing predictions, you wisely choose to conduct a two-tailed t test. Twelve students volunteer to participate in your study. Half of them see the 25 line drawings with verbal descriptions and the other half see the 25 line drawings without verbal descriptions. You record each person’s memory score. Memory scores form a normal distribution. You use a two-tailed t test with an alpha of .05 to test your research question.

Group 1: With Verbal Descriptions Group:
21, 22, 20, 20, 18, 20

Group 2: Without Verbal Descriptions Group:
19, 20, 19, 18, 16, 20

Step 1: Examine the Statistical Assumptions

You collected your data carefully, thereby satisfying the data independence assumption. The DV, number of line drawings correctly recalled, is on an interval/ratio scale, and the IV, presence or absence of verbal descriptions, identifies two different groups/conditions. Therefore, the study meets the appropriate measurement of variables assumption. When it comes to the normality assumption, distributions of memory scores tend to be normally shaped and therefore this assumption is likely met. This study will probably have very low statistical power and a lot of sampling error because of its very small sample sizes, and you will need to be very cautious interpreting the results. But you and Bill decide to analyze the data anyway.

The last assumption to consider is the homogeneity of variances assumption. For the independent t test, this assumption is that the two groups have similar variability in their memory scores. In previous chapters, we used the general rule that if the standard deviation in one condition is double that of another condition, this assumption might be violated. This is still a good guide to follow for the independent t test, but for the independent t test, there is a more precise way of assessing this assumption. As you know from the previous section, the obtained t value of an independent t test is the mean difference between the two conditions divided by expected sampling error. There are actually two ways to compute expected sampling error when doing an independent t test. One way assumes homogeneity of variance (i.e., that the two conditions have similar variances) and the other way does not. Which way is best depends on the data. If the variances are in fact similar, then computing sampling error by assuming equal variances is best. If, however, the variances are in fact very different, computing sampling error without assuming equal variances is best. In most cases, assuming equal variances will be the best, so that is what this book teaches you to do when doing hand computations. However, there are times when assuming unequal variances is better. Obviously, the amount of sampling error affects the obtained t value and your decision about rejecting the null.
Fortunately, SPSS computes both obtained $t$ values automatically. Therefore, when using SPSS, you should know how to determine which obtained $t$ value is the best for your data. SPSS provides a Levene's test to help you make this decision. We will describe this test in detail in the SPSS section of this chapter, but the main idea is that this test compares the variability in the two conditions, and the results of this test indicate if the two variances are similar enough to satisfy the equal variances assumption. Based on the results of this Levene's test, you will choose between the $t$ test that assumes equal variance or the one that does not. Again, you will see this test and how to interpret its results in the SPSS section of the chapter. The results of Levene's test are generally consistent with the double standard deviation guideline you learned earlier in this text. After assessing the assumptions, you are ready to move on to the next step.

6. You use an independent samples $t$ statistic when
   a. the IV defines two independent samples and the DV is measured on an interval/ratio scale.
   b. the IV defines two matched samples and the DV is measured on an interval/ratio scale.
   c. the IV defines one sample, the DV is measured on an interval/ratio scale, and the DV is measured twice on that same sample.
   d. the IV defines one sample and the DV is measured on an interval/ratio scale, and you do not know the population standard deviation.
   e. the IV defines one sample and the DV is measured on an interval/ratio scale, and you do know the population standard deviation.

7. Levene's test will help you determine
   a. whether or not the two sample means are significantly different from each other.
   b. if you should reject the null or fail to reject the null.
   c. if the two conditions have similar variances or variances that are very different (i.e., if the homogeneity of variance assumption is met or violated).
   d. if the sample size is sufficiently large.

---

Step 2: State the Null and Research Hypotheses Symbolically and Verbally

As mentioned above, you and Bill have opposing predictions about the effects of verbal labels on memory, so you correctly agree to use a two-tailed test. As shown in Table 10.1, the two-tailed research hypothesis states that the two means are different (i.e., are not equal). The null hypothesis is the exact opposite, stating that the two means are not different (i.e., are equal).

8. Which of the following represents the null hypothesis when doing a two-tailed independent test?
   a. $\mu_1 = \mu_2$
   b. $\mu_1 \neq \mu_2$
An Introduction to Statistics

Table 10.1 Symbolic and Verbal Representations of Two-Tailed Research and Null Hypotheses for an Independent t Test

<table>
<thead>
<tr>
<th>Symbolic</th>
<th>Verbal</th>
<th>Mean Difference Created by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research hypothesis (H₁)</td>
<td>H₁: µ₁ ≠ µ₂ OR H₁: µ₁ − µ₂ ≠ 0</td>
<td>If a population were given the verbal descriptions, their mean memory score would be significantly different from a population not given the verbal descriptions.</td>
</tr>
<tr>
<td>Null hypothesis (H₀)</td>
<td>H₀: µ₁ = µ₂ OR H₀: µ₁ − µ₂ = 0</td>
<td>If a population were given the verbal descriptions, their mean memory score would not be significantly different from a population not given the verbal descriptions.</td>
</tr>
</tbody>
</table>

9. Which of the following is the best summary of the two-tailed research hypothesis?
   a. People in the verbal description group will have higher memory scores than people in the no verbal description group.
   b. The people in the verbal description group will have different memory scores than people in the no verbal description group.

Step 3: Compute the Degrees of Freedom and Define the Critical Region

When you computed the degrees of freedom (df) for the single-sample t test, the df formula was df = (N − 1). The df formula for an independent t test is different because the independent t test uses two samples rather than just one sample. Therefore, you have to compute the df for each sample and combine them to get the df for the independent t test. The independent t test df formula is df = (n₁ − 1) + (n₂ − 1), where n₁ and n₂ represent the sample sizes of the two samples, respectively.

In this case, the df is as follows:

\[ df = (n₁ − 1) + (n₂ − 1) = (6 − 1) + (6 − 1) = 10. \]

You then use the correct t table of critical values and the \( \alpha = .05 \) criterion to find the critical value of t for this two-tailed independent t test. In this case, the critical value is 2.2281. This means that the two critical regions are \( t \geq +2.2281 \) and \( t \leq -2.2281 \).

Step 4: Compute the Test Statistic

4a. Compute the Deviation Between the Two Sample Means

As mentioned above, there are two terms in the numerator of the t statistic. The first \( (M₁ − M₂) \) is the difference between the two sample means. The second \( (\mu₁ − \mu₂) \) is the difference between the two
population means, assuming the null hypothesis is true. Although it is possible to test null hypotheses that predict a specific difference other than zero (e.g., $\mu_1 - \mu_2 = 10$), these types of tests are very rare and will not be covered in this text. For our purposes, $\mu_1 - \mu_2$ will always equal 0. Thus, the numerator is simply the difference observed between the two sample means:

$$(M_1 - M_2) = (20.17 - 18.67) = 1.50.$$ 

10. When computing the numerator of the independent samples $t$ test, the population mean difference (i.e., $\mu_1 - \mu_2$) will always be
   a. 0.
   b. 10.
   c. the same as the sample mean difference.

4b. Compute the Expected Sampling Error

Computing the denominator, or expected sampling error, requires several steps. First, compute the standard deviation for Group 1 ($SD_1$), and then compute the standard deviation for Group 2 ($SD_2$). In this problem, the standard deviation for each group of scores is not given to you, so you must compute each of them. In other problems, this information may be provided. As you may recall, the computational formula for sum of squares is $SS = \sum X^2 - \frac{(\sum X)^2}{n}$. Sum all of the scores from one of the groups to find $\sum X$ for that group of scores. Then square every score and sum all of the squared scores to find $\sum X^2$ for that group of scores. In this example, the $\sum X$ for Group 1 (i.e., $\sum X_1$) = 121, and the $\sum X^2$ for Group 1 = 2,449. There were six scores in Group 1, so $n_1 = 6$. Therefore, the $SS_1$ and $SD_1$ are as follows:

$$SS_1 = \sum X^2 - \frac{(\sum X)^2}{n} = 2,449 - \frac{(121)^2}{6} = 2,449 - 2,440.17 = 8.83.$$ 

$$SD_1 = \sqrt{\frac{SS_1}{n-1}} = \sqrt{\frac{8.83}{5}} = 1.33.$$ 

The $\sum X$ for Group 2 (i.e., $\sum X_2$) = 110, and the $\sum X^2$ for Group 2 = 2,026. There were six scores in Group 2, so $n_2 = 6$. Therefore, the $SS_2$ and $SD_2$ are as follows:

$$SS_2 = \sum X^2 - \frac{(\sum X)^2}{n} = 2,102 - \frac{(112)^2}{6} = 2,102 - 2,090.67 = 11.33.$$ 

$$SD_2 = \sqrt{\frac{SS_2}{n-1}} = \sqrt{\frac{11.33}{5}} = 1.51.$$ 

Once you have computed the standard deviation for each group, compute the pooled variance. This method assumes that the variances in the two conditions are similar; that is, it is assuming homogeneity of variance. If the two populations from which these samples were drawn do in fact have similar variances, it is best to pool them when computing sampling error because you are using more data to
estimate the populations’ variances. When appropriate, pooling the variances will make your significance test more accurate. The formula for the pooled variance is as follows:

\[
SD_p^2 = \frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2} = \frac{(6 - 1)(1.33)^2 + (6 - 1)(1.51)^2}{(6 - 1) + (6 - 1)} = 2.02.
\]

The pooled variance formula can be a bit confusing, but it is important that you understand what it represents. The pooled variance is the average of the two sample variances weighted by the sample size. In this particular example, the sample sizes are equal (i.e., six in each group) and so you could compute the pooled variance simply by taking the average of the two variances (remember the variance is \(SD^2\)). \(SD_p^2 = \frac{(1.33^2 + 1.51^2)}{2} = 2.02\). This simple formula gives the same value for the pooled variance as the more complex formula above only when the sample sizes are the same. If the sample sizes are different, the more complex formula will give a different, more accurate value because it is giving more weight to the larger sample. In general, the larger the sample, the more accurately it represents the population, and so when sample sizes are different, use the more complex formula that puts greater weight on the data from the larger sample.

The method of computing sampling error that assumes homogeneity of variance uses the pooled variance to determine the estimated standard error of the mean (\(SEM\)) difference, as illustrated by the following equation. Be sure to use \(n\) and not \(df\) in the denominator and note that the pooled variance (\(SD_p^2\)) is already squared.

\[
SEM_i = \sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}} = \sqrt{\frac{2.02}{6} + \frac{2.02}{6}} = .82.
\]

The estimated standard error of the mean difference, .82, is our estimate of expected sampling error. A standard error of 0.82 indicates that the difference between the two means expected due to sampling error is 0.82.

**Reading Question**

11. The estimated standard error of the mean difference is
   a. used to find the critical value for the \(t\) test.
   b. an estimate of how different the two sample means are expected to be due to sampling error.
   c. Both (a) and (b) are correct.

**4c. Compute the Test Statistic (Independent \(t\) Test)**

The independent \(t\) test is the ratio of the observed deviation between the two sample means divided by the estimated standard error of the mean difference:

\[
t = \frac{(M_1 - M_2)}{SEM_i} = \frac{(20.17 - 18.67)}{.82} = \frac{1.5}{.82} = 1.85.
\]
The difference between the two means (i.e., $20.17 - 18.67 = 1.50$) was 1.83 times larger than the deviation expected due to sampling error (i.e., 0.82). The critical regions of $t$ were $t \geq 2.2281$ and $t \leq -2.2281$, and so the obtained $t$ value did not fall in the critical region. The obtained $t$ value is not sufficiently large to reject the null hypothesis, and so you do not reject the null hypothesis.

12. When the obtained $t$ value is not further from zero, then the critical value the null hypothesis should
   a. be rejected.
   b. not be rejected.

Step 5: Compute an Effect Size and Describe It

The formula for computing the effect size of an independent $t$ test is similar to that of a single-sample $t$. However, when computing the $d$, the denominator is the pooled variance (i.e., $SD^2$), which can be converted to the pooled standard deviation (i.e., $SD_p$) by taking its square root.\(^1\) You computed the pooled variance in Step 3b; it was 2.02. The $d$ computation is as follows:

$$d = \frac{\text{Observed difference between the means}}{\text{Pooled standard deviation}} = \frac{M_1 - M_2}{\sqrt{SD^2_p}} = \frac{20.17 - 18.67}{\sqrt{2.02}} = 1.06.$$

The same effect size cutoffs are used for an independent $t$ test. If $d$ is close to .2, the effect size is small; if it is close to .5, the effect size is medium; and if it is close to .8, it is large. An effect size of 1.06 is considered a large effect.

The results from Steps 4 and 5 may be a bit confusing. In Step 4, you failed to reject the null hypothesis, which meant that the verbal label group did not remember more of the pictures than the no verbal label group. However, the effect size of 1.06 indicates that the difference between the two means is large. Whenever the null is not rejected and yet there was a medium or large effect size, the sample size used in the study was too small for the statistical test or the effect size to be trusted. When interpreting any study’s results, it is also important to consider the results from other similar studies. When Bower et al. (1975) conducted a similar study, they found that the verbal labels did improve memory performance. The combination of small sample sizes and the large effect size of your study and the fact that a similar study rejected the null suggests that you should interpret your null result with caution. In situations like this, you should obtain a larger sample size and then rerun the study.

13. Whenever you failed to reject the null hypothesis and yet the effect size is medium or large, you should conclude
   a. that the treatment really does work.
   b. that the treatment really does not work.
   c. that your sample size was too small and you should rerun the study with a larger sample size.

\(^1\) This is probably the most commonly used formula for computing $d$ for independent measures designs. However, some researchers choose different denominators (Cumming, 2012). Because there are different ways to calculate $d$, it is important to tell the reader how you computed the effect size. In this book, we are always using the same calculation and so we do not need to say repeatedly what denominator we used, but when you present data, you should state how $d$ was calculated.
Step 6: Interpreting the Results of the Hypothesis Test

The following paragraph summarizes these test results:

There was not a significant difference between the memory scores of those who got the verbal descriptions ($M = 20.17$, $SD = 1.33$) and those who did not ($M = 18.67$, $SD = 1.51$), $t(10) = 1.83$, $p > .05$, $d = 1.06$. However, it is important to note that the sample sizes were small, so this study should be repeated with larger sample sizes before conclusions are drawn.

14. When writing your results, if you failed to reject the null hypothesis and your sample size was small, you should
   a. point this out to the readers so that your report does not mislead them.
   b. not include this information in the report.

ONE-TAILED INDEPENDENT t TEST EXAMPLE

You and your friend Bill now turn your research efforts to studying “learning styles.” Do “visual learners” have better memory for visual information than “verbal learners”? If the learning styles theory is accurate, “visual learners” should recall more visual information than “verbal learners.” Given that the theory makes a specific prediction, a one-tailed $t$ test is appropriate in this situation. Twenty-nine “verbal learners” and 31 “visual learners” volunteered to participate in a study investigating this question. All learners were presented with simple line drawings and then were asked to re-create as many of the line drawings as they could remember. The visual memory scores of each group are listed below. You correctly use a one-tailed independent $t$ test with an alpha level of .05 to determine if “visual learners” recall more of the line drawings than “verbal learners.”

Group 1: Verbal Learners Group:

$M_1 = 15.00$; $SD_1 = 1.41$, $n_1 = 29$

Group 2: Visual Learner Group:

$M_2 = 15.25$; $SD_2 = 1.67$, $n_2 = 31$

Step 1: Examine the Statistical Assumptions

As in the previous example, these data meet all of the statistical assumptions. The data within each condition are independent, and the DV is measured on an interval/ratio scale. You also know that memory scores tend to be normally distributed, so the normality assumption is likely met. As before, the homogeneity of variance assumption will be assessed with Levene’s test provided by SPSS.

Step 2: State the Null and Research Hypotheses Symbolically and Verbally

The research hypothesis for a one-tailed test specifies which of the two sample means will be higher if the “treatment” works. For this example, Group 1 was “verbal learners,” and Group 2 was “visual learners.” The learning styles theory predicts that “visual learners” should learn visual information better than “verbal learners” do, and this prediction is represented symbolically by the following
research hypothesis: \( \mu_{\text{verbal}} < \mu_{\text{visual}} \) or \( \mu_{\text{verbal}} - \mu_{\text{visual}} < 0 \). The null hypothesis is the opposite of the research hypothesis, including all other possible outcomes. Thus, the null hypothesis states that \( \mu_{\text{verbal}} \geq \mu_{\text{visual}} \) or \( \mu_{\text{verbal}} - \mu_{\text{visual}} \geq 0 \). In other words, the null hypothesis is that “visual learners” do not learn visual information any better than the “verbal learners.” The research and null hypotheses for this one-tailed independent \( t \) test are shown in Table 10.2.

### Table 10.2
Symbolic and Verbal Representations of One-Tailed Research and Null Hypotheses for an Independent \( t \) Test

<table>
<thead>
<tr>
<th>Symbolic</th>
<th>Verbal</th>
<th>Mean Difference Created by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research hypothesis (( H_1 ))</td>
<td>( H_1: \mu_1 &lt; \mu_2 ) OR ( H_1: \mu_1 - \mu_2 &lt; 0 )</td>
<td>The population of “visual learners” does learn visual information better than the population of “verbal learners” does.</td>
</tr>
<tr>
<td>Null hypothesis (( H_0 ))</td>
<td>( H_0: \mu_1 \geq \mu_2 ) OR ( H_0: \mu_1 - \mu_2 \geq 0 )</td>
<td>The population of “visual learners” does not learn visual information better than the population of “verbal learners” does.</td>
</tr>
</tbody>
</table>

15. Which of the following could represent a null hypothesis when doing a one-tailed independent test? Note that this question is not asking about the study described in the text. Which of the following could possibly be a one-tailed null hypothesis?
   a. \( \mu_1 \geq \mu_2 \)
   b. \( \mu_1 < \mu_2 \)
   c. \( \mu_1 > \mu_2 \)
   d. \( \mu_1 = \mu_2 \)

### Step 3: Compute the Degrees of Freedom and Define the Critical Region

Computing the \( df \) for one-tailed tests is done in exactly the same manner as with a two-tailed test. In this case,

\[
df = (n_1 - 1) + (n_2 - 1) = (29 - 1) + (31 - 1) = 58.
\]

The one-tailed \( t \) table indicates that a study with a \( df = 58 \), when using \( \alpha = .05 \), has a critical value of 1.6716. The research hypothesis predicts that \( \mu_1 \), the verbal group, will be less than \( \mu_2 \), the visual group, so it is predicting a negative obtained \( t \) value, \( \mu_{\text{verbal}} - \mu_{\text{visual}} < 0 \). Thus, the critical region is in the negative side of the distribution. The null should be rejected if \( t \leq -1.6716 \). If “verbal learners” were labeled as Group 2 and “visual learners” as Group 1, we would have expected a positive \( t \), and the critical region would have been in the positive side of the distribution.

Copyright ©2017 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
An Introduction to Statistics

16. When you are doing a one-tailed \( t \) test, the critical region is always on the side of the \( t \) distribution that is predicted by the
   a. null hypothesis.
   b. research hypothesis.

Step 4: Compute the Test Statistic

4a. Compute the Deviation Between the Two Sample Means

Again, the numerator of the test is the difference between the two sample means:

\[
(M_1 - M_2) = (15.00 - 15.25) = -0.25.
\]

4b. Compute the Average Sample Error That Is Expected

As in the previous example, you need the standard deviation for each group of scores. In this problem, they were provided for you, but you should know how to compute them if you need to for future problems. Review the previous example if you are unsure how this is done. Once the standard deviation for Group 1 (i.e., \( SD_1 = 1.41 \)) and the standard deviation for Group 2 (i.e., \( SD_2 = 1.67 \)) have been computed, you compute the pooled variance (i.e., \( SD_p^2 \)) as follows:

\[
SD_p^2 = \frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2} = \frac{(29 - 1)(1.41)^2 + (31 - 1)(1.67)^2}{29 + 31 - 2} = 2.402.
\]

The pooled variance is the average variance for the two samples weighted by the sample size. In this case, the second sample was a bit larger than the second sample, and so the variance associated with that sample was given more weight when computing the pooled variance.

It only makes sense to pool (or average) the two standard deviations if they are estimating the same population parameter. If the standard deviations are significantly different from each other, the pooled variance must be computed in a different way. Fortunately, SPSS does all of these computations automatically, and so we will revisit this issue in the SPSS section that follows.

After you have computed the pooled variance, you will use it to compute the estimated standard error of the mean difference, which is done as follows:

\[
SEM = \sqrt{SD_p^2/n_1 + SD_p^2/n_2} = \sqrt{\frac{2.402}{29} + \frac{2.402}{31}} = 0.40.
\]

Be sure to use \( n \) from each group and not \( df \) in the denominator when computing the estimated standard error of the mean difference. Also, note that the pooled variance (\( SD_p^2 \)) is already squared, and so you should not square it again.

17. When computing the standard error of the mean difference, the equation calls for using
   a. the degrees of freedom for each group.
   b. the sample sizes from each group.
4c. Compute the Test Statistic (Independent t Test)

Again, the t test computation is identical to a two-tailed test:

\[
 t = \frac{(M_1 - M_2)}{SEM_i} = \frac{(15.00 - 15.25)}{0.40} = -0.250 \\
 = -0.63.
\]

The obtained t value of −0.31 is not further from zero than the critical value of −1.6716. Therefore, you do not reject the null hypothesis.

Reading Question

18. Which of the following values can never be negative?
   a. The numerator of a t test
   b. The obtained t value
   c. The standard error of the mean difference

Step 5: Compute an Effect Size and Describe It

Again, computing the effect size for one- and two-tailed tests is identical:

\[
 d = \frac{\text{Observed difference between the means}}{\text{Standard deviation}} = \frac{M_1 - M_2}{\sqrt{SD_p^2}} = \frac{15.00 - 15.25}{\sqrt{2.402}} = -0.16.
\]

An effect size with an absolute value of 0.16 is small. It means that the learning style of the learners had virtually no effect at all on memory scores.

Reading Question

19. When determining if an effect size is small, medium, or large, you should
   a. ignore the sign of the computed effect size and use its absolute value.
   b. recognize that negative effect sizes are always small effect sizes.

Step 6: Interpreting the Results of the Hypothesis Test

When interpreting the results of any study, you should always consider the results of the significance test and the effect size. In this learning styles study, the significance test suggested that the difference in mean memory scores between visual and verbal learners was probably just sampling error. In other words, the data did not support the learning styles theory. The small effect size in this study also did not support the idea that learning styles influence memory performance. This study generated an unambiguous “null result.” If many other studies in the literature report similar null results, then you can be more confident that your null result accurately reflects the true situation.

Even though you may have heard a lot about learning styles and have been told that you are a “visual” or “auditory” or “verbal” learner, you should know that experimental studies have generally found null results when investigating learning styles (Cook, Gelula, Dupras, & Schwartz, 2007; Pashler, McDaniel, Rohrer, & Bjork, 2008). In fact, the collection of null results is sufficiently large that memory researchers generally agree that, currently, there is no evidence supporting learning styles theory (Pasher et al., 2008). Instead, memory researchers suggest that instructors should incorporate
memory strategies supported by evidence. A great deal of memory research supports the effectiveness of using deeper levels of processing or retrieval practice effects. In other words, you should think about the meaning of what you are trying to learn and practice recalling the information from your memory. Research suggests that these studying strategies will increase your academic performance and that relying on learning styles will not.

The results from your and Bill’s learning styles study might be summarized as follows:

Contrary to the prediction of learning styles theory, the visual learners ($M = 15.25$, $SD = 1.41$) did not learn significantly more visual information than the verbal learners ($M = 15.00$, $SD = 1.67$), $t(58) = -0.63$, $p > .05$ (one-tailed), $d = -0.16$. The study’s null result is consistent with the null results found by several other researchers investigating learning styles. Collectively, the null results of several studies on learning styles strongly suggest that there is very little, if any, merit in the learning styles theory of learning.

**Reading Question**

20. Researchers need to be very cautious when interpreting null results because
a. they prove that the IV had no effect on the DV.
b. a null result might occur because of a problem with the study’s experimental procedure.

**OTHER ALPHA LEVELS**

In both of the previous examples, you used an alpha of .05. If you were to use an alpha of .01, rather than .05, it would be harder to reject the null hypothesis. This means that you would have less statistical power but a lower risk of making a Type I error.

21. Which alpha value has a lower risk of making a Type I error?
   a. .05
   b. .01

22. Which alpha value has a lower risk of making a Type II error?
   a. .05
   b. .01

**SPSS Data File**

We are going to use the data from the first example in the chapter to illustrate how to use SPSS to conduct an independent samples $t$ test. The data are reproduced as follows:
To enter data for an independent samples $t$ test, you will need two columns. The first is the IV (i.e., grouping variable). In this case, we need to enter a number to indicate the group that each person was in (i.e., verbal descriptions or no verbal descriptions). You can use any numbers, but below we used a “1” to indicate verbal descriptions and a “2” to indicate no verbal descriptions. The second column is the DV (the dependent variable, i.e., the variable on which the two groups are being compared). In this case, the DV is the visual memory score of each person. When you are done, the data file should look like Figure 10.4.

**Figure 10.4** SPSS Screenshot of Data Entry Screen for an Independent $t$ Test

---

**Computing an Independent Samples $t$ Test**

- Click on the Analyze menu. Choose Compare Means and then Independent Samples $t$ Test (see Figure 10.5).
- Move the Independent Variable (the one that indicates which group someone is in) into the Grouping Variable box, and click on Define. Enter the values you used to designate Group 1 and Group 2 in the appropriate boxes (in the above screenshot, you would enter the values 1 and 2, respectively).
Move the Dependent Variable (the one that indicates the actual scores of the participants) into the Test Variables box (see Figure 10.6).

- Click on the OK button.

**Output**

The SPSS output for this analysis is given in Figure 10.7.

**Levene’s Test (Test for Homogeneity of Variance)**

There are two ways to estimate expected sampling error (i.e., the denominator of the independent t-test). One way assumes homogeneity of variance, and the other does not. This assumption states that the variances in the two populations (e.g., verbal descriptions and no verbal descriptions) are the same. This assumption is important because if the variances are equal, you should compute the amount of expected sampling error as illustrated in the previous examples. However, if this assumption is
violated, the estimated sampling error should be computed differently. Levene’s test will indicate which method for computing sampling error is most appropriate. Just as we do a t test to determine if two means are significantly different, we can do a test to determine if the two variances are significantly different. The standard deviations for the verbal learners and visual learners were 1.41 and 1.67, respectively. Thus, the variance was $1.41^2$ or 1.99 for the verbal learners and $1.67^2$ or 2.79 for the visual learners. SPSS automatically runs Levene’s test for homogeneity of variance to determine if 1.99 is significantly different from 2.79. Note that our double standard deviation rule suggests that the homogeneity of variance assumption is not violated.

23. Levene’s test is automatically run by SPSS to determine if the _____ of the two groups are significantly different.
   a. means
   b. variances

If the variances of the two groups are not significantly different, the proper way to compute the estimate of sampling error is to use the “Equal variances assumed” method. If the variances are significantly different, the proper way to compute the estimate of sampling error is to use the “Equal variances not assumed” method. You can determine if the variances are equal or not by looking at the “Sig.” value under the “Levene’s Test for Equality of Variances” label in the “Independent Samples Test” output. If the Sig. value is less than or equal to .05, the variances are not similar and the equal variance
assumption was violated. If the Sig. value is greater than .05, the variances are similar and the assumption of equal variance was met.

24. Use the “Independent Samples Test” output to find the Sig. value for Levene’s test and then determine if the variances are similar or not similar. The Sig. (p) value for Levene’s test is _____ and therefore the variances for the verbal and visual groups ______.
   a. .676; are similar (i.e., equal)
   b. .097; are not similar (i.e., not equal)
Levene’s test indicated that the variances in the two groups were equal. Therefore, the best way to compute the estimate of sampling error is to use the “Equal variances assumed” method. If you look in the left-most cell of the “Independent Samples Test” output, you will see two labels. The one at the top is “Equal variances assumed,” and the one at the bottom is “Equal variances not assumed.” Levene’s test indicated that we should use the “Equal variances assumed” method. SPSS automatically computes two different $t$ tests—one using the “equal variances” method and the other using the “not equal variances” method. The results of both $t$ tests are shown in the output. In this case, we should choose the obtained $t$ value and degrees of freedom that is across from the “Equal variances assumed” heading (i.e., $t = -1.830$ and $df = 10$).

2. Like the obtained $t$ value, the degrees of freedom are also computed differently when the homogeneity of variance assumption is violated. You will not need to do this by hand in this book, but the formula is different from the one you use when the homogeneity of variance assumption is met. The $df$ formula is $df = \left(\frac{SD_1^2 / n_1 + SD_2^2 / n_2}{\left(\frac{SD_1^2 / n_1}{n_1 + 1}\right) + \left(\frac{SD_2^2 / n_2}{n_2 + 1}\right)}\right)^{-2}$.

When the variances are not similar, the $SEM$ is computed as $\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}$. Finally, when the $t$ test is modified in this way, it is called a Welch $t$ test (Welch, 1947). Fortunately, SPSS will do all of these calculations automatically. All you have to know is which $t$ test to interpret.
OVERVIEW OF THE ACTIVITIES

In Activity 10.1, you will work through all of the steps of hypothesis testing for a one-tailed hypothesis test. While doing so, you will work to understand conceptually what you are doing at each step and the implications of the statistical decisions you make with regards to Type I error, Type II error, and statistical power. In this activity, you will also review information from two studies and compare and contrast the results of the significance test, effect size, and study design. In Activity 10.2, you will review some of the material from Activity 10.1, do a two-tailed hypothesis test, and work with SPSS. In Activity 10.3, you will read different research scenarios and decide which statistic should be used to test the hypothesis. Activity 10.4 is a group exercise giving you the opportunity to collect data using an independent measures design, a repeated measures design, and a matched design. This activity is intended to help you understand the relative advantages/disadvantages of the different types of designs and to review the information from Chapters 9 and 10. Finally, in Activity 10.5, you will compute and interpret confidence intervals for a two-group independent measures design.

Activity 10.1: Hypothesis Testing With the Independent $t$ Test

Learning Objectives

After reading the chapter and completing this activity, you should be able to do the following:

- State null and research hypotheses for an independent $t$ test
- Compute and interpret the results of an independent $t$ test
- Explain how sampling error influences an independent $t$ test
- Compute and interpret the effect size estimate for an independent $t$ test
- Use the distributions of sample means to locate the probabilities of Type I error, Type II error, statistical power, and the probability of rejecting a false null hypothesis

INDEPENDENT $t$ TEST EXAMPLE

It is well known that acetaminophen lessens people’s physical pain. A recent study suggests that acetaminophen can lower people’s psychological pain as well (DeWall et al., 2010). Surprised by these findings, you decided to see if acetaminophen works for different types of psychological pain than those assessed in the original research. You obtained a sample of volunteers and gave half of them acetaminophen and the other half a placebo pill. Both groups of participants read socially painful stories and rated how painful the experience would be for them. Specifically, you want to know if the mean social pain rating of the acetaminophen group is statistically lower than the mean social pain rating of the control group. Stated differently, you want to know if the mean difference in social pain rating between these two samples is likely or unlikely to be due to sampling error. You are
predicting that the acetaminophen will reduce social pain ratings; therefore, you used a one-tailed hypothesis test with $\alpha = .05$.

\[ n_{\text{drug}} = 31, \quad M_{\text{drug}} = 213, \quad SD_{\text{drug}} = 18. \]
\[ n_{\text{control}} = 31, \quad M_{\text{control}} = 222, \quad SD_{\text{control}} = 20. \]

**STATISTICAL ASSUMPTIONS**

1. Match the assumption to the fact that is relevant to that assumption.
   - Independence
   - Appropriate measurement of the IV and the DV
   - Normality
   - Homogeneity of variance
   a. Samples of this size tend to form distributions of sample means with a normal shape.
   b. Data were collected from one participant at a time.
   c. This assumption will be assessed later by Levene’s test.
   d. The IV manipulation is well defined, and the participants’ responses were given on an interval/ratio scale.

2. How do you know that the homogeneity of variance assumption is not violated?
   a. One standard deviation is not double the size of the other.
   b. The standard deviations are different by less than 5.

3. How do you know that the normality assumption is probably not violated?
   a. The sample size in each group is greater than 30.
   b. The research scenario indicates that social pain scores are normally distributed.

**UNDERSTANDING THE NULL AND RESEARCH HYPOTHESES**

4. Write $H_0$ next to the symbolic notations for the null hypothesis and $H_1$ next to the research hypothesis.
   a. $H_0$: $\mu_{\text{drug}} = \mu_{\text{placebo}}$
   b. $H_0$: $\mu_{\text{drug}} \neq \mu_{\text{placebo}}$
   c. $H_0$: $\mu_{\text{drug}} > \mu_{\text{placebo}}$
   d. $H_0$: $\mu_{\text{drug}} > \mu_{\text{placebo}}$
   e. $H_0$: $\mu_{\text{drug}} < \mu_{\text{placebo}}$
   f. $H_0$: $\mu_{\text{drug}} < \mu_{\text{placebo}}$
5. Write \( H_0 \) next to the verbal description of the null hypothesis and \( H_1 \) next to the research hypothesis.
   a. _____If a population of people were given the drug, they would have higher social pain scores than would a population of people given the placebo.
   b. _____If a population of people were given the drug, they would not have higher social pain scores than would a population of people given the placebo.
   c. _____If a population of people were given the drug, they would have lower social pain scores than would a population of people given the placebo.
   d. _____If a population of people were given the drug, they would not have lower social pain scores than would a population of people given the placebo.
   e. _____If a population of people were given the drug, their social pain scores would be different from a population of people given the placebo.
   f. _____If a population of people were given the drug, their social pain scores would not be different from a population of people given the placebo.

6. Assuming that the null hypothesis is true (i.e., that acetaminophen does not affect social pain at all), what precise value would you expect for the mean difference between the mean ratings of social pain for the acetaminophen sample and the placebo sample? Precise expected value for the mean difference between the two sample means if the null is true = _____.

7. In this situation, you obtained two samples of people from the population. One sample was assigned to take acetaminophen, and the other sample was assigned to take a placebo. Even if you assume that the acetaminophen does not affect social pain at all, would you be surprised if the samples’ mean difference was not exactly the value you indicated earlier?
   a. No, I would not be surprised. The two sample means are likely to be different because the population parameters are different.
   b. No, I would not be surprised. The two sample means are likely to be different because of sampling error even if the drug has no effect.
   c. Yes, I would be surprised. The two sample means should be exactly the same if the drug does not affect social pain.
   d. Yes, I would be surprised. The two sample means should be exactly the same because the sizes of the two samples are the same.

8. Just as was the case with the \( z \) for a sample mean, the single-sample \( t \) test, and the related samples \( t \) test, you must compute the amount of sampling error that is “expected by chance or sampling error” before you can determine if the null hypothesis is likely or unlikely to be true. The acetaminophen and control groups each had 31 people with social pain SDs of 18 and 20, respectively. Use the two equations provided below to compute the value for standard error of the mean difference, estimated sampling error.

\[
SD^2_p = \frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}
\]

\[
SEM_d = \sqrt{\frac{SD^2_p}{n_1} + \frac{SD^2_p}{n_2}}
\]
9. What does the \( SEM \) estimate?
   a. The typical amount of sampling error in the two separate samples
   b. The typical distance between the two sample means expected due to sampling error
   c. The typical distance between the sample scores and the means for the two samples

10. Assuming that acetaminophen does not affect social pain at all (i.e., assuming the null hypothesis is true), create a distribution of sample mean differences based on the two sample sizes of \( n_{\text{drug}} = 31 \) and \( n_{\text{control}} = 31 \). This distribution represents the frequency of all possible sample mean differences when both sample sizes are 31. You should label the frequency histogram so that it is centered on the sample mean difference you expect if the null is true. Look back at your answer to Question 6 to determine what the mean of the distribution of sample means should be if the null hypothesis is true.

After you label the mean, label each standard error of the mean to the right and left of the mean. You computed the standard error of the mean in Question 8.

Next, use the independent \( t \) formula to determine the \( t \) score associated with each sample mean difference. In other words, convert each raw mean difference score on the distribution into a \( t \) value.

Finally, label the critical \( t \) value and shade the critical region. You will need to consult a critical \( t \) value table to locate the critical value of \( t \).

Evaluating the Likelihood of the Null Hypothesis

11. Now that the critical region of \( t \) is determined, you can find the \( t \) value associated with the mean difference between the samples and determine if it is in the critical region of the \( t \) distribution (also known as the region of rejection). The sample means were \( M_{\text{drug}} = 213 \) and \( M_{\text{control}} = 222 \). Locate the mean difference on the distribution of sample mean differences above. Is it in the critical region or outside of the critical region?
   a. In the critical region
   b. Outside of the critical region

---

![Critical t Value](image)
12. Compute the obtained \( t \) value.

13. What should you conclude about the null hypothesis and why?
   a. Reject \( H_0 \) because the computed \( t \) value is in the critical region
   b. Reject \( H_0 \) because the computed \( t \) value is outside of the critical region
   c. Fail to reject \( H_0 \) because the computed \( t \) value is in the critical region
   d. Fail to reject \( H_0 \) because the computed \( t \) value is outside of the critical region

**Effect Size**

The process of significance testing you just completed in the previous question tells us whether or not the null hypothesis is likely to be true. It does not indicate how effective the IV was in affecting the DV. In the preceding scenario, the null hypothesis was rejected, but the researchers only know that acetaminophen will probably decrease social pain. They do not know how effective it would actually be. To get this information, they must compute an estimate of the study's effect size.

14. Compute the estimate of effect size (\( d \)) of acetaminophen on social pain. (Note: When computing the \( d \) for an independent \( t \) test, you need to use the “pooled standard deviation” or the square root of the pooled variance. Consult your reading for a reminder.)

15. Is the effect size small, small to medium, medium, medium to large, or large?

**Summarize the Results**

16. Choose the best APA style summary of the results.
   a. The mean social pain ratings of the acetaminophen group (\( M = 213, SD = 18 \)) and the placebo group (\( M = 222, SD = 20 \)) were significantly different, \( t(60) = -1.86, p < .05, d = .47 \). The effect size suggests that acetaminophen is very effective at reducing social pain.
   b. The mean social pain ratings of the acetaminophen group (\( M = 213, SD = 18 \)) were not significantly different from the mean ratings of the placebo group (\( M = 222, SD = 20 \)), \( t(60) = -1.86, p > .05, d = .47 \).
   c. The mean social pain ratings of the acetaminophen group (\( M = 213, SD = 18 \)) were significantly less than the mean ratings of the placebo group (\( M = 222, SD = 20 \)), \( t(60) = -1.86, p < .05, d = .47 \). The effect size suggests that acetaminophen is moderately effective at reducing social pain because it reduces it by about a half a standard deviation.
Type I Errors, Type II Errors, and Statistical Power

17. Before collecting data from the sample that took acetaminophen, you can’t locate the center of the distribution of sample mean differences if the research hypothesis is true. However, after you have the data from the sample, you do have an idea of where the research hypothesis distribution of sample mean differences is located. What can you use to estimate the location of this distribution? Provide a specific value. (Hint: What value might represent the mean difference between the social pain scores of all those who took acetaminophen and all those who took the placebo if the research hypothesis is true?)

Specific value for the center of the research hypothesis distribution of sample mean differences = _____.

A major benefit of locating the center of the research hypothesis distribution of sample mean differences, even if it is an estimated center that is inferred from sample statistics, is that it allows researchers to quantify several other very important statistical concepts. This quantification process can be illustrated by “building a distribution.” You have already built one of the distributions that are necessary in Question 10. As you know from Question 10, the null hypothesis distribution of sample mean differences is centered at a t value of 0. The null t distribution is re-created for you as follows:

![Critical value of t = -1.67

\( df = 60 \)]

18. The second distribution you need to build is the distribution of the sample mean differences if the research hypothesis is true. As mentioned earlier, the center of this distribution is determined by the actual mean difference of the samples. In this case, the actual sample mean difference was −9, meaning that the center of this distribution of sample means is at −9 on the raw score number line. What is the t value associated with a sample mean difference of −9?

Draw a vertical line on the above figure at the t value associated with a mean difference of −9 to represent the center of the distribution of sample means if the research hypothesis is true. Remember that the only reason we can locate this curve’s center is because you know the actual mean difference of the samples. You can’t know this before you collect data.

19. In Activity 6.1, you created the research hypothesis distribution of sample means by sketching a normally shaped distribution. When working with t distributions, the research distribution of sample means can be skewed with smaller sample sizes.3 Here, our sample size is large enough

---

3 The severity of the research curve’s skew depends on the effect size and the sample size, but as sample size increases, the research curve approaches a normal shape (Cummings, 2012).
that the distribution is essentially normal. In addition, we are less concerned with the research curve’s precise shape than we are with using it to locate the Type II and statistical power regions in the research curve. Therefore, sketch in the distribution of sample mean differences if the research hypothesis is true, assuming it has a normal shape. (Just try to duplicate the null curve’s shape and spread but have it centered at −1.86 on the t score number line rather than at 0, where the null curve is centered.)

After you have completed sketching in the research hypothesis distribution of sample mean differences, look at the following figure and confirm that your figure looks like it. If it doesn’t, determine what is wrong and fix it. Be sure you understand why the figure looks as it does.

Just as can be done with the z for the sample mean, single-sample t test, and the related samples t test, we can use these curves to quantify the probability of important statistical concepts. The null distribution of sample mean differences represents all possible outcomes if the null hypothesis is true. The research distribution of sample mean differences represents all possible outcomes if the research hypothesis is true. By “cutting” these curves into different sections at the critical value of t, we can estimate the probability of (a) rejecting a false null (i.e., statistical power), (b) failing to reject a false null (i.e., Type II error), (c) rejecting a true null (i.e., Type I error), and (d) failing to reject a true null. In the following figure, there is a distribution of sample mean differences for the null hypothesis and a distribution of sample mean differences for the research hypothesis. The two curves have been “separated” so that it is easier to see the distinct sections of each curve. The areas under each of these respective curves represent statistical power, Type II errors, Type I errors, and failing to reject a true null. Try to determine which areas under each curve represent what in the following figure.
20. Match the area to the statistical outcome

Type I error [rejecting a true null]: Area ________.
Type II error [failing to reject a false null]: Area ________.
Statistical power [rejecting a false null]: Area ________.
Failing to reject a true null: Area ________.

21. Go back to Activity 6.1. Make note of the fact that the areas under the curve that correspond to each of the above concepts are not the same as in Activity 6.1. What is different about this scenario that changed the locations of the above areas on the curves?
   a. In this scenario, the hypothesis test was one-tailed, while in the previous scenario, the hypothesis test was two-tailed.
   b. In this scenario, the research hypothesis predicted a negative difference, while in the previous scenario, the predicted mean difference was positive.
   c. In this scenario, there were two samples that resulted in two separate curves, while in the previous scenario, there was just one sample and one curve.
22. If a one-tailed, $\alpha = .01$ significance test was used instead of a one-tailed, $\alpha = .05$ significance test, what would happen to the critical value line in the preceding graphs? It would move
a. closer to zero.
b. closer to the left tail.

23. What would happen to the probability of each of the following if a one-tailed, $\alpha = .01$ significance test was used instead of a one-tailed, $\alpha = .05$ significance test? Indicate if each would increase, decrease, or stay the same.

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Increase</th>
<th>Decrease</th>
<th>Stay the Same</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I error</td>
<td>Increase</td>
<td>Decrease</td>
<td>Stay the same</td>
</tr>
<tr>
<td>Type II error</td>
<td>Increase</td>
<td>Decrease</td>
<td>Stay the same</td>
</tr>
<tr>
<td>Statistical power</td>
<td>Increase</td>
<td>Decrease</td>
<td>Stay the same</td>
</tr>
</tbody>
</table>

24. Which two of the following values are estimates and which one value is known precisely?

<table>
<thead>
<tr>
<th>Error Type</th>
<th>Estimate only</th>
<th>Precise value is known</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I error</td>
<td>Estimate only</td>
<td>Precise value is known</td>
</tr>
<tr>
<td>Type II error</td>
<td>Estimate only</td>
<td>Precise value is known</td>
</tr>
<tr>
<td>Statistical power</td>
<td>Estimate only</td>
<td>Precise value is known</td>
</tr>
</tbody>
</table>

25. Why are the two values you identified in Question 24 estimates?

a. Type II error and power are estimated based on the null hypothesis curve.
b. The mean (center) and the standard deviation (spread) of the research distribution curve are estimated from the sample data.

Summary of Key Points in Significance Testing

26. So far in this course, we have gone through several examples illustrating the details of how significance testing works. These details included each of the following key points. Read through these key points and try to fill in the key terms that are missing. You may use some of the terms from page 345 more than once.

- The ________________ allows us to predict the center, shape, and spread of the distribution of sample means (or the distribution of sample mean differences) if the null is true.
  - State the theorem below.

- The ______ hypothesis precisely locates the center of the null distribution of sample means (or the null distribution of sample mean differences). This location is at a $z$ value of ______ or at $t$ value of ______.

- The center of the________ hypothesis distribution of sample means (or mean differences) is estimated from the sample mean (or sample mean difference) and cannot be known before data are collected.
• The probability of a ______________ is set by researchers when they set the $\alpha$ value.
• By building the null and research hypothesis distributions of sample means (or sample mean differences), we can quantify the probability of failing to reject a false null (i.e., ______ ______), as well as the likelihood of rejecting a false null (i.e., ____________).
• The ______________________ of $t$ or $z$ “cuts” likely values if the null is true from unlikely values if the null is true; if a statistic (e.g., $z$ value, $t$ value) is more extreme than this value, the null hypothesis is unlikely to be true.
• If the null is rejected, the ________ hypothesis is considered likely to be true.
• After a __________________________ is performed, researchers know whether or not the null is likely to be true. They do not know how effective the IV was at affecting the DV. To quantify the impact of the IV on the DV, researchers must compute a/an ____________________.

<table>
<thead>
<tr>
<th>Type I error</th>
<th>Type II error</th>
<th>Null</th>
<th>Critical value</th>
<th>Statistical power</th>
<th>Power</th>
<th>Central limit theorem</th>
<th>Significance test</th>
<th>Effect size</th>
<th>Research</th>
</tr>
</thead>
</table>

**Putting It All Together**

27. A researcher obtained the scores for all 1.5 million students who took the SAT this year and found that males scored significantly higher than females ($p < .001$). Given the sample size, what additional information would you want before you decide if this statistically significant difference is large enough to be an important difference?

a. The confidence interval
b. The effect size

Suppose that you have a friend who suffers from debilitating panic attacks that have not been helped by treatments that are currently on the market. Your friend’s psychiatrist asks if he would like to participate in a clinical trial. There are two clinical trials available (Drug A or Drug B), and your friend asks you for help in deciding which drug seems more promising based on the available data.

Although neither of the drugs are yet approved by the Food and Drug Administration (FDA), both are currently in the human trial stage of the approval process. The side effects for both drugs seem to be minor, and if your friend agrees to participate in one of the trials, all costs would be paid by a National Institutes of Health (NIH) research grant. The preliminary results from the trials are described below.

**Drug A information:** In total, 250 participants took Drug A for 3 months, while another 250 participants took a placebo for 3 months in a double-blind study. While taking the “drug,” the participants recorded the number of panic attacks they experienced. The mean number of panic attacks in the Drug A group was $M = 5.12, SD = 4.21, n = 250$, while the mean number of panic attacks in the placebo group was $M = 6.87, SD = 4.62, n = 250, t(498) = 4.43, p < .001, d = .39$. To participate in this trial, your friend would need to go to a local hospital once a week for a 1-hour visit.

**Drug B information:** In total, 150 participants took Drug B for 3 months, while another 150 participants took a placebo for 3 months in a double-blind study. While taking the “drug,” the participants recorded the number of panic attacks they experienced. The mean number of panic attacks...
attacks in the Drug B group was $M = 5.47, SD = 2.89, n = 150$, while the mean number of panic attacks in the placebo group was $M = 8.11, SD = 2.94, n = 150$, $t(298) = 7.84, p < .001, d = .70$. To participate in this trial, your friend would need to go to a hospital that is approximately 45 minutes away once a week for 1.5-hour visits.

28. When deciding which of these two drug studies provides more compelling evidence, you should consider their respective sample sizes. At first glance, it is tempting to assume automatically that studies with larger samples are always better. However, that is not the case. As long as a sample size is sufficiently large, an even larger sample is not appreciably better. Compute the standard error of the mean or expected sampling error for both of these studies below. (You will need to compute the pooled $SD$ first.)

The $SEM$ for Study A =

The $SEM$ for Study B =

Based on the $SEMs$ for the two studies:

a. Study A is much more compelling in terms of its sample size.

b. Study B is much more compelling in terms of its sample size.

c. Studies A and B are approximately equal when it comes to sample size; both have relatively small $SEMs$, so choosing between these studies should probably be based on something other than sample size.

29. Based on the effect sizes for the two studies:

a. Study A has a more impressive effect size.

b. Study B has a more impressive effect size.

c. Studies A and B have equally impressive effect sizes.

30. Compare the results of the significance test for the two studies.

a. Study A had sufficient evidence to reject the null.

b. Study B had sufficient evidence to reject the null.

c. Both studies had sufficient evidence to reject the null.

31. Compare the relative cost in terms of time commitment (e.g., travel time as well as treatment time).

a. Drug A requires a greater time commitment.

b. Drug B requires a greater time commitment.

32. Which drug trial would you suggest to your friend? (You must choose one.) Explain your rationale; explain the relative importance of the respective sample sizes, the results of the hypothesis tests, the effect sizes, and the relative cost in terms of time commitment (e.g., travel time as well as treatment time). You should not just count the number of times that Drug A or B had “better results.” You need to talk about which of these four factors is most important when deciding between these two drug trials and conclude your answer with a recommendation.
Activity 10.2: A Two-Tailed Independent $t$ Test

Learning Objectives

After reading the chapter and completing this activity, you should be able to do the following:

- Compute and interpret an independent samples $t$ test by hand and using SPSS
- Draw a distribution of $t$ scores assuming the null hypothesis is true and a distribution of $t$ scores assuming the research hypothesis is true
- Explain the homogeneity of variance assumption and how it is tested
- Explain what the numerator and the denominator of the $t$ test each measure
- Explain the effect of using a one-tailed versus a two-tailed test when hypothesis testing
- Explain how it is possible to not reject the null even when it is false

Two-Tailed Independent $t$ Example

In a course on reasoning and decision making, you heard about a phenomenon called anchoring and adjustment. The phenomenon suggests that the way a question is asked can influence the answer people give. You want to see if you can replicate this finding. You asked one group of participants, “Is the Nile River longer than 800 miles [1 mile = 1.61 kilometers]?” They answered “yes” or “no.” Then you asked them to estimate the Nile’s length to the nearest mile. You changed the first question slightly for a second group of participants. You asked them, “Is the Nile longer than 3,000 miles?” Then you asked them to estimate the Nile’s length to the nearest mile. The actual length of the Nile is 4,184 miles. You are interested in whether first asking people the 800 or 3,000 “yes–no” question caused the participants to give estimates that were significantly different. You know from previous research that these types of estimates tend to form a normal distribution of scores. You decide to use a $\alpha = .05$. Your data are as follows:

<table>
<thead>
<tr>
<th>Group 1 800 Low Anchor</th>
<th>Group 2 3,000 High Anchor</th>
</tr>
</thead>
<tbody>
<tr>
<td>650</td>
<td>600</td>
</tr>
<tr>
<td>850</td>
<td>1,000</td>
</tr>
<tr>
<td>900</td>
<td>1,999</td>
</tr>
<tr>
<td>1,000</td>
<td>2,000</td>
</tr>
<tr>
<td>1,000</td>
<td>2,700</td>
</tr>
<tr>
<td>1,100</td>
<td>3,000</td>
</tr>
<tr>
<td>1,273</td>
<td>4,000</td>
</tr>
<tr>
<td>1,320</td>
<td>4,000</td>
</tr>
<tr>
<td>2,000</td>
<td>4,300</td>
</tr>
<tr>
<td>2,000</td>
<td>5,000</td>
</tr>
<tr>
<td>2,000</td>
<td>6,600</td>
</tr>
</tbody>
</table>
1. Match the assumption to the fact that is relevant to that assumption.

   _____ Independence
   _____ Appropriate measurement of the IV and the DV
   _____ Normality
   _____ Homogeneity of variance

   a. Estimates like this tend to form a normal distribution of scores.
   b. Data were collected from one participant at a time.
   c. This assumption will be assessed later by Levene’s test.
   d. The IV manipulation is well defined, and the participants’ responses were given on an interval/ratio scale.

2. Write $H_0$ next to the symbolic notations for the null hypothesis and $H_1$ next to the research hypothesis.

   a. _____ $\mu_{\text{high anchor}} > \mu_{\text{low anchor}}$
   b. _____ $\mu_{\text{high anchor}} \geq \mu_{\text{low anchor}}$
   c. _____ $\mu_{\text{high anchor}} < \mu_{\text{low anchor}}$
   d. _____ $\mu_{\text{high anchor}} \leq \mu_{\text{low anchor}}$
   e. _____ $\mu_{\text{high anchor}} = \mu_{\text{low anchor}}$
   f. _____ $\mu_{\text{high anchor}} \neq \mu_{\text{low anchor}}$

The first requirement for using SPSS to analyze data is to enter the data correctly. In SPSS, all the information about a given participant goes on a single row. When conducting an independent $t$ test, as in this case, you will need two columns. One column should indicate the IV condition each person is in, either the low anchor condition or the high anchor condition. You should use a coding system like “1” for those in the low (800) anchor condition and “2” for those in the high (3,000) anchor condition. So, for the IV column, you would have a column of “1s” and “2s.” Create value labels for this coding scheme if you remember how. The second column should be each person’s DV, in this case, each participant’s estimate of the Nile’s length. When you are done entering data, it should look similar to the file shown on in Figure 10.4; it should not look like the two columns of data in the table on page 347.

To run the independent $t$ test:

- Click on the Analyze menu. Choose Compare Means and then Independent Samples $t$ Test.
- Move the Independent Variable (the one that indicates which group someone is in) into the “Grouping Variable” box and click on “Define.” Enter the values you used to designate Group 1 and Group 2 in the appropriate boxes.
- Move the Dependent Variable (the one that indicates the actual scores of the participants) into the “Test Variables” box. Click on the OK button.

3. What is the standard error of the mean difference?
4. Which of the following statements accurately describes what the standard error is measuring in this study? (Note: The standard error is the denominator of the \( t \) statistic if you calculate it by hand.)
   a. The standard error is the difference between the means of the two groups or conditions.
   b. The standard error is a measure of sampling error; we would expect a mean difference of this size to occur due to sampling error.
   c. The standard error is a measure of sampling error; with these sample sizes, we would expect a mean difference of this size to occur due to sampling error.

5. What two values are used to generate the mean difference? (Note: The mean difference is the numerator of the \( t \) statistic if you calculate it by hand).

6. Give the exact mean difference value for this study. Mean difference = 

7. The mean difference (or numerator of the \( t \) statistic) is the actual difference between the mean of those who were given the 800 anchor and the mean of those who were given the 3,000 anchor. The standard error of the mean difference (or the denominator of the \( t \) statistic) is the size of the mean difference we would expect by sampling error. Given these two facts, which of the following is true?
   a. If the mean difference is about the same size as the standard error, the null hypothesis is likely to be false.
   b. If the mean difference is about the same size as the standard error, the null hypothesis is likely to be true.

8. What is the assumption of homogeneity of variance? Select all that apply.
   a. The standard deviations for the two sample means are not significantly different from the population standard deviation.
   b. The standard deviations for the two sample means are not significantly different from each other.
   c. The variances for the two populations being compared are equal.
   d. The variances for the two populations being compared are significantly different from each other.

9. Was the homogeneity of variance assumption violated in this study?
   a. Yes, it was violated. The Sig. level for Levene’s test was greater than .05.
   b. Yes, it was violated. The Sig. level for Levene’s test was less than .05.
   c. No, it was not violated. The Sig. level for Levene’s test was greater than .05.
   d. No, it was not violated. The Sig. level for Levene’s test was less than .05.
10. What do you do if this assumption is violated?
   a. You cannot do an independent samples \( t \) if this assumption is violated.
   b. Use the \( t \), \( df \), and Sig. information from the “Equal variances assumed row.”
   c. Use the \( t \), \( df \), and Sig. information from the “Equal variances not assumed row.”

11. How do you use the information in the Sig. column to determine if you should reject or fail to reject the null when doing a two-tailed test?
   a. You divide the Sig. by 2 and then reject the null hypothesis if that number is less than alpha.
   b. You divide the Sig. by 2 and then reject the null hypothesis if that number is greater than alpha.
   c. You reject the null hypothesis if the Sig. is less than alpha.
   d. You reject the null hypothesis if the Sig. is greater than alpha.
   e. You divide the alpha by 2 and then reject the null hypothesis if that number is less than Sig.
   f. You divide the alpha by 2 and then reject the null hypothesis if that number is greater than Sig.

12. SPSS does not compute \( d \), and so you need to compute it by hand. To do so, you will need to compute the pooled variance using the \( SD \) values in the SPSS output.

13. Summarize the results of this study using APA style.
   Those given the 800 anchor gave estimates of the length of the Nile that were significantly lower \( (M = _____, SD = _____) \) than those given the 3,000 anchor \( (M = _____, SD = _____) \), \( t(______) = _____, p = _____ \) (two-tailed), \( d = _____ \).

   (Note: The \( df \), \( t \), and Sig. value should all come from the “Equal variances not assumed row.”)

14. If we had predicted that the people given 3,000 as an anchor would give higher estimates than those given 800 as an anchor, it would have been reasonable to do a one-tailed test. What would the null and research hypotheses be if we had predicted higher estimates from the 3,000 group? Write \( H_0 \) next to the symbolic notations for the null hypothesis and \( H_1 \) next to the research hypothesis.
   a. \( H_{high\anchor} > H_{low\anchor} \)
   b. \( H_{high\anchor} \geq H_{low\anchor} \)
   c. \( H_{high\anchor} < H_{low\anchor} \)
   d. \( H_{high\anchor} \leq H_{low\anchor} \)
   e. \( H_{high\anchor} = H_{low\anchor} \)
   f. \( H_{high\anchor} \neq H_{low\anchor} \)

15. Would the obtained \( t \) value change if we did a one-tailed test?
   Yes \hspace{1cm} No
16. How do you use the Sig. column of the SPSS output to determine if you should reject or fail to reject the null hypothesis when using a one-tailed test?
   a. You divide the Sig. by 2 and then reject the null hypothesis if that number is less than the alpha.
   b. You divide the Sig. by 2 and then reject the null hypothesis if that number is greater than the alpha.
   c. You reject the null hypothesis if the Sig. is less than the alpha.
   d. You reject the null hypothesis if the Sig. is greater than the alpha.
   e. You divide the alpha by 2 and then reject the null hypothesis if that number is less than Sig.
   f. You divide the alpha by 2 and then reject the null hypothesis if that number is greater than Sig.

17. In this case, would we reject the null when doing a one-tailed test with an alpha of .01?

After learning that people tend to anchor on the initial numbers they are given when making estimates, you wonder if this same principle could be used to increase sales in a grocery store. Specifically, you want to know if providing a high anchor for how many of something shoppers should purchase increases the number of things shoppers buy. You enlist the help of two branches of a neighborhood grocery store chain. The stores are in similar locations with similar demographics and with similar sales figures. One store posts signs for a sports drink that advertise the price as 10 for $10 (high anchor). The other store posts signs as 5 for $5 (low anchor). Both signs also indicate that the price for one bottle is $1. The computerized cash registers recorded the number of bottles of the sports drink purchased by each shopper who purchased at least one bottle of the sports drink. The people who saw the low anchor signs purchased an average of 6 bottles ($SD = 2.39, n = 15$), while the people who saw the high anchor signs purchased an average of 9.47 bottles ($SD = 3.29, n = 15$). The daily sales of sports drinks across a year do form a normal distribution. The number of bottles sold to each shopper who bought any one a given weekend are as follows:

Group 1: Low anchor (5 for $5): 5, 10, 2, 5, 5, 5, 5, 10, 8, 10, 5, 5, 4, 6

Group 2: High anchor (10 for $10): 10, 10, 15, 6, 8, 10, 15, 10, 10, 5, 10, 6, 8, 5, 14

18. Match the assumption to the fact that is relevant to that assumption.

   ____ Independence
   ____ Appropriate measurement of the IV and the DV
   ____ Normality
   ____ Homogeneity of variance

   a. The signs in the different stores were different, and the number of bottles sold to each shopper was recorded.
   b. This assumption will be assessed later with Levene’s test.
   c. The population of number of sport drinks bought by individual shoppers has a normal shape.
   d. The purchases of individual shoppers was recorded at both stores.
19. Write H₀ next to the null hypothesis and H₁ next to the research hypothesis.
   _____a. µ₁ = µ₂
   _____b. µ₁ ≠ µ₂
   _____c. µ₁ > µ₂
   _____d. µ₁ < µ₂
   _____e. µ₁ > µ₂
   _____f. µ₁ < µ₂

20. What is the critical region for this one-tailed test (use α = .01)?
   a. t > 2.4671
   b. t < -2.4671
   c. t > 2.4671 or t < -2.4671

21. Compute the test statistic (t).
   a. -3.30
   b. 1.05
   c. -3.47
   d. .003
   e. 1.31

22. Should you reject or fail to reject the null hypothesis?
   a. Reject
   b. Fail to reject

23. Compute the effect size (d).
   a. -3.30
   b. 1.05
   c. -3.47
   d. .003
   e. 1.21

24. How large is the effect?

25. Read and evaluate the following APA style summary of the anchoring study at the grocery store.
Then determine if there is an error or omission and, if so, identify the problem. (Select all that apply.)

The low anchor group (M = 6.00, SD = 2.39) bought significantly less than the high anchor group
(M = 9.47, SD = 3.29), t(28) = -3.30, d = 1.21.
   a. There are no errors or omissions in the above APA summary.
   b. The means and standard deviations for each condition are missing.
   c. Some of the statistical t test information is missing.
   d. The sentence implies that the means are significantly different when they are not
      significantly different.
26. Read and evaluate the following APA-style summary of the anchoring study at the grocery store. Then determine if there is an error or omission and, if so, identify the problem. (Select all that apply.)

The low anchor group bought significantly less than the high anchor group, \( t(28) = -3.30, p = .003 \) (one-tailed), \( d = 1.21 \).

- a. There are no errors or omissions in the above APA summary.
- b. The means and standard deviations for each condition are missing.
- c. Some of the statistical \( t \) test information is missing.
- d. The sentence implies that the means are significantly different when they are not significantly different.

**Activity 10.3: How to Choose the Correct Statistic**

**Learning Objectives**

After reading the chapter and completing the homework and this activity, you should be able to do the following:

- Read a research scenario and determine which statistic should be used
- One of the more challenging aspects of this course is choosing the correct statistic for a given scenario. A common strategy is to match the goal described in the scenario with the statistic that accomplishes that goal. Obviously, using this strategy requires you to first identify the research goal. Look at Appendix J to see a decision tree and a table that may help you choose the correct statistic for a given research situation.

**Very Basic Example Problems**

Determine which statistic should be used in each of the following situations. When you are assessing a given situation, you need to recognize that if a problem does not give you the sample mean (\( M \)) or the sample standard deviation (\( SD \)), you can always compute these values from the data. However, if a problem does not give you the population mean (\( \mu \)) or the population standard deviation (\( \sigma \)), you should assume that these values are not known.

1. Do male teachers make more money than female teachers?
2. Do people have less body fat after running for 6 weeks than before they started running?
3. Intelligence quotient (IQ) scores have a population mean of 100 and a standard deviation of 15. Does the college football team have a mean IQ that is significantly greater than 100?
4. Is the mean height of a sample of female volleyball players taller than 68 inches?
More Detailed Example Problems

Determine which statistic should be used in each of the following research scenarios: $z$ for sample mean, single-sample $t$, related samples $t$, or independent samples $t$.

5. Previous studies have shown that exposure to thin models is associated with lower body image among women. A researcher designs a study to determine if very young girls are similarly affected by thin images. Forty kindergarteners are randomly assigned to one of two groups. The first group plays with Barbie dolls for 30 minutes. The second group plays with a doll with proportions similar to the average American woman. After the 30-minute play period, the researcher measures each girl’s body image using a graphic rating scale that yields an interval scaled measure of body image. Which statistic should this researcher use to determine if girls who played with Barbie dolls reported lower body image than girls who played with dolls with proportions similar to the average American woman?

6. A teacher of an art appreciation course wants to know if his course actually results in greater appreciation for the arts. On the first day of class, the teacher asks students to complete an art appreciation survey that assesses attitudes toward a variety of forms of art (e.g., painting, theater, sculpture). Scores on the survey range from $1 = \text{strongly disagree}$ to $5 = \text{strongly agree}$, and responses to all of the questions are averaged to create one measure of art appreciation. The same survey was given on the last day of class. The teacher analyzes the survey data and finds that scores were significantly higher on the last day of class than on the first day of class. Which statistic should this researcher use to determine if students’ art appreciation scores were higher after the class than before the class?

7. An insurance company keeps careful records of how long all patients stay in the hospital. Analysis of these data reveals that the average length of stay in the maternity ward for women who have had a caesarean section is $\mu = 3.9$ days with a standard deviation of $\sigma = 1.2$. A new program has been instituted that provides new parents with at-home care from a midwife for 2 days after the surgery. To determine if this program has any effect on the number of days women stay in the hospital, the insurance company computes the length of stay of a sample of 100 women who participate in the new program and find that their mean length of stay is 3.4 with a standard deviation of 1.4. Which statistic would help determine if the new program is effective at lowering the average length of mothers’ hospital stay?

8. Abel and Kruger (2010) recently analyzed the smiles of professional baseball players listed in the Baseball Register. The photos of players were classified as either big smiles or no smiles. The age of death for all players was also recorded. The results revealed that players with big smiles lived longer than those with no smiles. Which statistic could be used to determine if there was a significant difference in life span between those with big versus no smiles?

9. A questionnaire that assesses the degree to which people believe the world is a fair and just place has a mean of $\mu = 50$. A researcher wonders if this belief is affected by exposure to information, suggesting that the world is not a fair and just place. To answer this research question, he conducts a study with 73 students and has them watch a series of videos where bad things happen to good people. After watching these videos, he gives them the questionnaire and finds
that the average score after watching the videos was 48.1 with a standard deviation of 16.2. Which statistic should the researcher use to determine if watching the video significantly reduced endorsement of the view that the world is fair and just?

10. It is well known that acetaminophen reduces physical pain. DeWall et al. (2010) found that the drug can also reduce psychological pain. Another researcher wonders if the same is true of aspirin. To test the efficacy of aspirin in treating psychological pain, they measured participants’ psychological pain, gave them the drug, and then again measured their psychological pain. Psychological pain was measured using an interval scale of measurement. Which statistic should be used to determine if aspirin reduced psychological pain?

11. A recent study revealed that the brains of new mothers grow bigger after giving birth. The researchers performed magnetic resonance imaging on the brains of 19 women and found that the volume of the hypothalamus was greater after giving birth than prior to giving birth. Which statistic would researchers use to determine if the volume of the hypothalamus was greater after giving birth than before?

12. A plastic surgeon notices that most of his patients think that plastic surgery will increase their satisfaction with their appearance and, as a result, make them happier. To see if this is actually the case, he asks 52 of his patients to complete a survey 1 week prior to and then again 1 year after the surgery. The survey consisted of 10 questions such as “I am happy” and “Life is good.” Responses to each item were scored with 1 = strongly agree and 5 = strongly disagree. Scores on all survey items summed into one index of happiness. Which statistic should the surgeon use to determine if patients are happier after having plastic surgery than before?

Activity 10.4: Comparing Independent, Matched, and Related Research Designs

Learning Objectives

After reading the chapter and completing this activity, you should be able to do the following:

- Compute and interpret independent and related samples t tests using SPSS
- Explain how the numerator of a t could be increased and the consequences
- Explain how the denominator of a t could be decreased and the consequences
- Describe the differences between independent samples and related samples designs
- Explain why related samples designs are generally preferable to independent samples designs

Today, we are going to investigate the effect of an IV on a DV using an independent, matched, and related samples design. Specifically, we are going to measure your heart rate while sitting and standing.
1. What effect do you expect people’s physical position of sitting versus standing to have on their heart rate? Will these two conditions produce different heart rates? If so, which condition do you expect to have a higher heart rate?

2. We are going to investigate the relationship between physical position and heart rate using three different designs: (1) an independent design, (2) a matched design, and (3) a related design. Our IV for each design will be whether you are sitting or standing, and the DV will be your heart rate. Once these data are collected, we will analyze the data in SPSS and compare the statistical estimates and conclusions generated by each research design. After we run each $t$ test, use the results to complete the following table.

<table>
<thead>
<tr>
<th></th>
<th>Independent</th>
<th>Matched</th>
<th>Related</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of people in the study (and number of data points)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$df$</td>
<td>$(n_1 - 1) + (n_2 - 1)$</td>
<td>$(N - 1)$</td>
<td>$(N - 1)$</td>
</tr>
<tr>
<td>Observed mean difference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimate of sampling error</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obtained $t$ value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig. ($p$ value)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reject or fail to reject (use $\alpha = .05$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Compare the estimates of sampling error for each type of design. Which design has the smallest estimate of sampling error? Explain why this type of design has less sampling error than the other two designs. In your explanation, be sure to use the words *individual differences, sample size, and sampling error*.

3. Compare the obtained $t$ values for each type of design. Which design produced the largest obtained $t$ value? Explain why this happened. In your explanation, be sure to use the words *individual differences, sample size, and sampling error*.

4. Which type of design is probably the best choice for studying the effect of physical position on heart rate and why?
Activity 10.5: Confidence Intervals for Mean Differences Between Independent Samples

Learning Objectives

After reading the chapter and completing this activity, you should be able to do the following:

- Compute and interpret CIs for means and mean differences between independent samples
- Write complete APA-style statistical reports that include independent t tests and CIs
- Identify the distinct purposes of hypothesis testing, effect sizes, and CIs

1. Previously you learned to compute confidence intervals (CIs) for a population mean, a mean difference between a sample mean and a population mean, and a mean difference between two related sample means. In this activity, you will compute a fourth CI for the mean difference between two independent samples. As with all of the previous CIs, you need three numbers to compute the confidence interval. What are the three numbers you need to compute all CIs? (Choose three.)
   a. An obtained t value
   b. A point estimate based on sample data
   c. A critical t value
   d. An estimate of sampling error (i.e., some kind of SEM)
   e. A Type II error rate

2. When the research design involves two independent samples (e.g., a sample of men and a sample of women), the point estimate is the actual difference observed between the independent sample means. Then a margin of error is added to and subtracted from the point estimate to create the CI. What two values are multiplied together to produce a CI’s margin of error?
   a. An obtained t value
   b. A point estimate based on sample data
   c. A critical t value
   d. An estimate of sampling error (i.e., some kind of SEM)
   e. A Type II error rate

3. Why would researchers want to know a CI for a mean difference?
   a. CIs reveal the range of ways a study’s results might manifest in a population
   b. CIs determine the magnitude of an IV’s effect on a DV
INVESTIGATING ANXIETY DISORDERS

4. A clinical psychologist specializing in treating those with anxiety disorders wonders if the average severity of symptoms differs for men and women. If there were a gender difference, it might hint at potential situational or biological factors contributing to the development of anxiety disorders. On the other hand, if anxiety disorders were present equally in men and women, it might suggest other contributing factors. So, her research goal is to determine if men and women differ in anxiety severity. Place an H₀ and H₁ in front of the null and research hypotheses, respectively. She wants α = .05, two-tailed test.
   a. _____ Women will be higher in anxiety than men.
   b. _____ Men will be higher in anxiety than women.
   c. _____ Men and women will not differ in anxiety.
   d. _____ Men and women will differ in anxiety.

5. She started her investigation by analyzing data from a sample of intake questionnaires her clients completed. She had questionnaires from eight men and eight women. Determine the critical value for this t-test.
   a. 2.9768
   b. 2.1448
   c. 1.9600
   d. 2.4441

6. The level of anxiety from each person is presented in the following table. Use SPSS to compute the obtained t-value for this preliminary study. A higher score represents higher anxiety.

<table>
<thead>
<tr>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
</tr>
</tbody>
</table>

The obtained t-value is
   a. −.50
   b. .57
   c. 1.01
   d. .87

7. Should the null hypothesis be rejected?
   a. Yes, the obtained value is less than the critical value.
   b. No, the obtained value is less than the critical value.
   c. Yes, the obtained value is more than the critical value.
   d. No, the obtained value is more than the critical value.
8. Compute the effect size for this study.

9. Compute the 95% CI for the mean difference observed in this preliminary study.

   The point estimate is = ________.
   The critical $t$ value is = ________.
   The $SEM$ is = ________.
   The UB = ________.
   The LB = ________.

10. To compose the APA results summary, you will need the 95% CI for the mean anxiety of men and the 95% CI for the mean anxiety of women. Use the mean for each group as the respective point estimates and the $SD$s and $N$s to compute the respective $SEM$s. You will also need to look up the $t_{ci}$ for each group. Always use the two-tailed .05 $t$ value table to find the $t_{ci}$ for a 95% CI.

   Women point estimate = __________
   Women $SD$ = __________
   Women $N$ = __________
   Women $SEM$ = __________
   Women UB = __________
   Women LB = __________

   Men point estimate = __________
   Men $SD$ = __________
   Men $N$ = __________
   Men $SEM$ = __________
   Men UB = __________
   Men LB = __________

11. Below is the clinician’s rough draft of a brief summary of the results. Fill in the blanks so that it complies with APA recommendations.

   The mean anxiety level for women ($M = 8.38, SD = 1.85$), 95% CI [_______, _______] was not significantly different from that for men ($M = 7.88, SD = 1.64$), CI [_______, _______]. $t$(____) = ______, $p < .05$, $d =$ ______. CI [_______, _______].

12. What should this clinician conclude about anxiety levels of men and women?
   
   a. Based on this study, it looks like the average anxiety levels do not differ for men and women. Although the sample sizes are small, the effect size is also small and so it is very unlikely that the result is a Type II error.
   
   b. Based on this study, it looks like the average anxiety levels do not differ for men and women. However, the sample size is very small, and it is possible that this result is a Type II error. The study should be replicated with a larger sample.
   
   c. Based on this study, it looks like the average anxiety levels do not differ for men and women. However, there is quite a bit of sampling error, and it is possible that this result is a Type I error. The study should be replicated with a larger sample.
EVALUATING CELL PHONES’ EFFECT ON SLEEP

A public health advocacy organization hires your firm to investigate if cell phones contribute to sleep deprivation in teenagers. Your firm recruits 46 students between the ages of 16 and 18. Half of them are randomly assigned to a cell phone group and are told to use their cell phone for 1 hour right before going to bed. They are allowed to do whatever they like on the phone (i.e., talk, games, etc.) as long as they are using the phone for a full hour right before bed. The other half are told that they can use their cell phone during the day, but they must not look at the phone for the last hour before they go to sleep. Furthermore, students in both groups agree that they will go to bed between 10 and 11 p.m. each night. Every night for 2 weeks, the students wear a sleep monitor that records the number of hours they slept. The average number of hours of sleep obtained over the course of the week was computed for each student. Again, an intern entered the data into an SPSS data file called “CellPhoneSleep.sav.” Use SPSS to run the appropriate analyses. By now you know that you will need to produce a significance test, a CI for each mean, an effect size, and a CI for the mean difference. Use a one-tailed test with $\alpha = .05$ and 95% CIs. The instructions for running an independent measures $t$ are in the reading for Chapter 10. This analysis will also provide the confidence interval around the mean difference. To obtain the confidence interval around the two sample means (i.e., cell phone and no cell phone), you need to follow these instructions:

- Click on Analyze, Descriptive Statistics, and Explore.
- Move the dependent variable into the Dependent List box.
- Move the independent variable (grouping variable) into the Factor list box.
- Click on the Statistics button.
- In the Explore:Statistics box, select Descriptives and make sure the Confidence Interval for Mean is set at 95%.
- Click on the Continue button and then on the OK button to run the analysis.

13. Now, fill in the blanks in the APA-style report of the results. When reporting numbers in APA style, round to the second decimal place.

The participants who used a cell phone before bed ___(did/did not)___ sleep significantly less ($M = $____, $SD = $____), 95% CI [____,____] than participants who did not use a cell phone before bed ($M = $____, $SD = $____), CI [____,____], $t$(____) = $$, $p = $$ (one-tailed), $d = $$, CI [____,____].

14. The manager of the public health firm that hired you is surprised that the confidence interval around the mean difference is so large (i.e., over 1 hour) and wants to know what could be done to make that interval smaller. Which of the following would help reduce the width of the interval?

a. Compute a 99% confidence interval rather than a 95% confidence interval
b. Redo the study with a larger sample size to reduce sampling error
c. Do a two-tailed test rather than a one-tailed test
15. The APA publication manual recommends that researchers include confidence intervals, significance tests, and effect sizes. Match the statistical procedure to the distinct information it provides.

_____ Confidence intervals provide . . .
_____ Significance tests provide . . .
_____ Effect sizes provide . . .

a. an index with which to compare how well different treatments work.
b. if the null hypothesis is true, the probability that a result is due sampling error.
c. a range of plausible values for a population parameter.
Chapter 10 Practice Test

To determine if a regular bedtime helps children academically, a researcher obtains a sample of 50 children in kindergarten who have a regular bedtime every night and a separate sample of 50 children who do not have a regular bedtime every night. Each child’s scores on a variety of math tests are combined into a single measure of math performance with higher scores indicating better performance. The mean score for the kindergarteners with a regular bedtime was 68.00 \((SD = 10.50)\) while the mean score for the kindergarteners without a regular bedtime was 63.00 \((SD = 10.30)\). The researcher is not sure if a regular bedtime will help with school performance, and so she decides to do a two-tailed test.

1. Match the assumption to the fact that is relevant to that assumption.
   
   ____ Independence  
   ____ Appropriate measurement of the IV and the DV  
   ____ Normality  
   ____ Homogeneity of variance

   a. Samples of 30 or more participants increase confidence that this assumption was not violated.
   b. Data were collected from one participant at a time.
   c. This assumption will be assessed later by Levene’s test; you can also be confident that this assumption was not violated if the standard deviations for the two groups are similar.
   d. One variable is a grouping variable, and the other variables is an interval/ratio variable.

2. What is the two-tailed research hypothesis?
   
   a. The children with a regular bedtime will have better math scores than the children without a regular bedtime.
   b. The children with a regular bedtime will have worse math scores than the children without a regular bedtime.
   c. The children with a regular bedtime will have significantly different math scores than the children without a regular bedtime.
   d. The children with a regular bedtime will not have significantly different math scores than the children without a regular bedtime.

3. What is the two-tailed null hypothesis?
   
   a. The children with a regular bedtime will have better math scores than will the children without a regular bedtime.
   b. The children with a regular bedtime will have worse math scores than will the children without a regular bedtime.
   c. The children with a regular bedtime will have significantly different math scores than will the children without a regular bedtime.
   d. The children with a regular bedtime will not have significantly different math scores than will the children without a regular bedtime.
4. Compute the df for this study.
   a. 100
   b. 50
   c. 99
   d. 98
   e. 49

5. Locate the critical value for this t test ($\alpha = .05$).
   a. 1.9845
   b. 1.6606
   c. 2.0096
   d. 1.6766

6. Which of the following statements describes the critical region?
   a. Scores that are unlikely if the null hypothesis is true
   b. Sample mean differences that are unlikely if the null hypothesis is true
   c. Scores that are impossible if the null hypothesis is true
   d. $t$ statistics that are likely to occur if the research hypothesis is true

7. Compute the $t$ statistic.
   a. .48
   b. 5
   c. 2.40
   d. 2.10
   e. 1.86

8. Should the researcher reject the null hypothesis?
   a. Reject
   b. Not reject

9. Compute the effect size ($d$).
   a. .48
   b. 5
   c. 2.40
   d. 2.10
   e. 1.86

10. How large is the effect?
    a. Small
    b. Small to medium
    c. Medium
    d. Medium to large
    e. Large
11. Are the two groups significantly different from each other?
   a. Yes, because the researcher rejected the null hypothesis.
   b. No, because the researcher failed to reject the null hypothesis.
   c. Yes, because the effect size was in the critical region.
   d. No, because the effect size was not in the critical region.

12. Compute the 95% confidence interval around the mean difference.
   a. LB = 1.2, UB = 1.80
   b. LB = −3.10, UB = 6.90
   c. LB = .87, UB = 9.13
   d. LB = 2.92, UB = 7.08

13. Fill in the blanks:
   The children who had a regular bedtime had significantly higher math test scores (M = ______, SD = ______) than children who did not have a regular bedtime (M = ______, SD = ______).
   t(______) = ______, p < ______ (two-tailed), d = ______. 95% CI [_____, _____].

The researcher who did the earlier study on regular bedtimes is not sure why children with regular bedtimes do better on math tests than other children. It is possible that the children with regular bedtimes are somehow different from children without regular bedtimes (e.g., personality differences, parenting differences). To determine if it is the regular bedtime that is important rather than these extraneous factors, the researcher recruits a sample of 50 children who do not currently have a regular bedtime to participate in a study. The children are randomly assigned to two groups. For children in the regular bedtime group, the parents attend a class where they are told about the benefits of a regular bedtime and are given tips about how to implement a regular bedtime at home. For children in the control group, the parents also attend a class, but rather than learning about the benefits of a regular bedtime, they are told about the benefits of reading.

Over the course of the next 3 months, the parents are sent several reminders about the information in the classes they attended. At the end of the 3 months, all students took a math test.

14. What is the one-tailed research hypothesis for this study?
   a. \( \mu_{\text{regular bedtime}} > \mu_{\text{control}} \)
   b. \( \mu_{\text{regular bedtime}} \geq \mu_{\text{control}} \)
   c. \( \mu_{\text{regular bedtime}} < \mu_{\text{control}} \)
   d. \( \mu_{\text{regular bedtime}} \leq \mu_{\text{control}} \)
   e. \( \mu_{\text{regular bedtime}} = \mu_{\text{control}} \)
   f. \( \mu_{\text{regular bedtime}} \neq \mu_{\text{control}} \)

15. What is the one-tailed null hypothesis for this study?
   a. \( \mu_{\text{regular bedtime}} > \mu_{\text{control}} \)
   b. \( \mu_{\text{regular bedtime}} \geq \mu_{\text{control}} \)
   c. \( \mu_{\text{regular bedtime}} < \mu_{\text{control}} \)
   d. \( \mu_{\text{regular bedtime}} \leq \mu_{\text{control}} \)
   e. \( \mu_{\text{regular bedtime}} = \mu_{\text{control}} \)
   f. \( \mu_{\text{regular bedtime}} \neq \mu_{\text{control}} \)
16. Use the SPSS output from the independent samples $t$ to determine how many children participated in the study.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
Group & N & Mean & Std. Deviation & Std. Error Mean \\
\hline
MathTest & control group & 30 & 63.5667 & 16.80232 & 3.08220 \\
regular sleep group & 30 & 71.4667 & 18.66166 & 3.40714 \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|c|c|}
\hline
Levene's Test for Equality of Variances & & & & & & & & \\
\hline
F & Sig. & t & df & Sig. (2-tailed) & Mean Difference & Std. Error Difference & 95\% Confidence Interval of the Difference \\
\hline
Equal variances not assumed & & & & & & & & \\
\hline
\end{tabular}
\end{table}

a. 30 

b. 60 

c. 58 

17. Is the homogeneity of variance assumption violated?
   a. Yes, the $p$ value for Levene’s test is less than .05. 
   b. No, the $p$ value for Levene’s test is greater than .05. 

18. Which line of output should you use to interpret the independent samples $t$ test?
   a. Equal variances assumed 
   b. Equal variances not assumed 

19. What is the one-tailed $p$ value?
   a. .351 
   b. .1755 
   c. .091 
   d. .0455 

20. Should the researcher reject or fail to reject the null hypothesis?
   a. Reject 
   b. Fail to reject 

21. Compute the effect size ($d$).
   a. .44 
   b. 1.72 
   c. 7.90 
   d. .091
22. Did the regular sleep group receive significantly higher scores on the math test than the control group?
   a. Yes
   b. No

23. Choose the best APA-style summary of these results.
   a. Children assigned to the regular sleep group ($M = 71.47, SD = 18.66$) were significantly different from the children assigned to the control group ($M = 63.57, SD = 16.88$), $t(58) = -1.72$, $p = .045$ (one-tailed), $d = .44$.
   b. Children assigned to the regular sleep group ($M = 71.47, SD = 18.66$) were not significantly different from the children assigned to the control group ($M = 63.57, SD = 16.88$), $t(58) = -1.72$, $p = .045$ (one-tailed), $d = .44$.
   c. Children assigned to the regular sleep group had significantly higher scores on the math test ($M = 71.47, SD = 18.66$) than children assigned to the control group ($M = 63.57, SD = 16.88$), $t(58) = -1.72$, $p = .045$ (one-tailed), $d = .44$.
   d. Children assigned to the regular sleep group did not have significantly higher scores on the math test ($M = 71.47, SD = 18.66$) than children assigned to the control group ($M = 63.57, SD = 16.88$), $t(58) = -1.72$, $p = .045$ (one-tailed), $d = .44$.

24. For the independent samples $t$ test, what is the numerator?
   a. The observed difference between means
   b. The difference between the means expected due to sampling error
   c. The difference between the null and research hypotheses

25. For the independent samples $t$ test, what is the denominator?
   a. The observed difference between means
   b. The difference between the means expected due to sampling error
   c. The difference between the null and research hypotheses

REFERENCES

Welch, B. L. (1947). The generalization of “Student’s” problem when several different population variances are involved. Biometrika, 34(1–2), 28–35.