

8.3: The Independent Samples t-Test Formula

Before we can use the formula, it is important to understand what it can tell us and how it gets there. The independent samples t -test formula looks quite complicated but is meant to tell us something that is conceptually fairly simple: The obtained t -value tells us far apart the two group means are using standard error. Another way to say this is that it tells us how many standard errors apart the two group means are. It does this by taking the difference in the sample means and dividing it by the standard error. The difference in means is calculated in the numerator of the formula and the pooled standard error is calculated in the denominator of the formula. The standard error (SE) used is known as a pooled standard error because it is calculated by putting information from both groups together (also known as *pooling information*). Therefore, we can understand the formulas main construction and outcomes as follows:

$$t = \frac{\text{difference in sample means}}{\text{pooled standard error}} = \text{how many SEs apart the two sample means are}$$

The numerator focuses on the difference between the mean of Group 1 and the mean of Group 2. This is the core of the formula because the hypotheses tested using an independent samples t -test are specifically asking whether these two means differ. Therefore, you can think of the numerator as the main focus of the formula and the denominator as taking into account other necessary information and adjustments. This will be true for the main format used for all three versions of the t -test in this book. The denominator of the independent samples t -test formula is used to take into account the error in the samples (via the pooled standard error).

The formula contains six symbols which represent and, thus, must be replaced with specific values. The independent samples t -test formula is as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \right] \left[\frac{n_1 + n_2}{n_1 \times n_2} \right]}}$$

Note

The df is the sum of $n-1$ for each group (or total n minus 2) and appears in the formula's denominator.

The numerator asks for the mean of Group 1 and the mean of Group 2. These will be used to calculate the difference in means. The denominator asks for the sample size for Group 1, the sample size for Group 2, the variance for Group 1, and the variance for Group 2. These four things will be used to calculate the pooled standard error. Thus, the formula requires we know three basic things about each of the two groups so that we can plug them in and solve: mean, variance, and sample size.

Though the formula looks very complex, focusing on a few things can help us see that it is manageable. First, though there appear to be many symbols in the formula there are only six and the simplest one of them, sample size, appears the most often. Thus, the majority of the formula is just asking us to input sample sizes. Second, the mathematical operations required are ones most of us are familiar with. Operations refer to mathematical actions we take. The operations this formula requires include: adding, subtracting, multiplying, dividing, squaring (if converting SD to variance), and square rooting. Though the formula requires us to do many steps, each step only includes one of these six basic operations (and we can use calculators to make the more challenging steps of squaring and square rooting easier).

Notice that the formula includes all three components that impact statistical power (see Chapter 6 for a review of the components of power):

1. The size of the change, difference, or pattern observed in the sample
2. The sample size
3. The size of the error in the sample statistic(s)

The first component, size of difference, is being incorporated in the numerator. The denominator is a bit more complex and it can be hard to see what it is doing so let's take a moment to consider its role. The denominator allows necessary adjustments to be made that incorporate the other two components of statistical power: sample size and error. Error is being measured using the variance (which is the squared version of standard deviation) of each group. These components are both incorporated into the

formula. The way sample sizes and variances are put together in the denominator is used to find the pooled standard error. Pooled standard error, therefore, takes into account two of the three components of power: the size of the sample and the size of the error.

Let's take a moment to see the connection between standard error for one group (which we reviewed in Chapter 6) and standard error when it is pooled so that it takes into account two groups. Standard error for a single sample is found using this formula:

$$SE = \frac{\sigma}{\sqrt{n}}$$

The independent samples *t*-test needs to pool this from two separate groups. It does so by first finding the pooled standard deviation (because there are two groups so there are two standard deviations) and adjusting that by a calculation of samples sizes. The left section of the denominator is calculating the pooled standard deviation. The **pooled standard deviation** (S_{pooled} or S_p) is the weighted average of the standard deviation of two or more groups. When we say an average is "weighted" it means that proportionally more "weight" is given to groups with larger sample sizes. Another way of saying this is that a weighted average takes into account any differences in sample size so that the group(s) with more cases (i.e. larger sample sizes) more heavily impact the value being calculated. This is how pooled standard deviation can be found when group variances are approximately even:

$$S_p = \sqrt{\left[\frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} \right]}$$

You may wonder why pooled standard deviation is calculated using variances and sample sizes rather than standard deviations and sample sizes. Recall from Chapter 4 that variance is used when we actually want the standard deviation but still have some work to do before we do the final step of square rooting. Notice that the S_p formula asks us to do calculations with variances and sample sizes but that those are all under a square root sign. Thus, we are using variance in route to finding the pooled standard deviation (which we will get to by square rooting as the last step of the S_p formula).

This needs to be further adjusted to go from pooled standard deviation to pooled standard error. Within the brackets on the right side of the denominator, we see the use of just the sample sizes. That piece is also under the square root sign so, on its own, it looks like this:

$$\sqrt{\left[\frac{n_1 + n_2}{n_1 \times n_2} \right]}$$

The presence of this piece changes the denominator of the formula from just finding the S_p into a formula for finding pooled standard error (SE_p). It is doing the same work that is done by the denominator of the SE formula for one group. Therefore, the formula for pooled standard error for an independent samples *t*-test looks like this:

$$SE_p = \left(\sqrt{\left[\frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} \right]} \right) \left(\sqrt{\left[\frac{n_1 + n_2}{n_1 \times n_2} \right]} \right)$$

This can rewritten for (slightly more) simplicity (though, admittedly, it still looks a bit complex) as follows:

$$SE_p = \sqrt{\left[\frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} \right] \left[\frac{n_1 + n_2}{n_1 \times n_2} \right]}$$

Therefore, the denominator of the independent samples *t*-test formula is the formula for pooled standard error for the test.

Let's put the numerator and denominator together. The numerator is used to find the difference in the means. The denominator is used to find the relevant standard error for the formula. Finally, the formula tells us to divide the numerator by the denominator. When we do, we get a *t*-value which tells us how many standard errors the two means are apart.

Thus, we have come full circle and can now (hopefully) see why the formula can be thought of as follows:

$$t = \text{how many standard errors apart the two sample means are}$$

Interpreting Obtained t -Values

Obtained t -values have two components: a magnitude and a direction. The magnitude is the absolute value of t and it represents how many standard errors the mean of one sample is from the mean of the other sample. The larger the value, the farther apart the two means are. As the t -value increases, the evidence for the research hypothesis and against the null hypothesis also increases. Conversely, as the t -value decreases, the evidence for the research hypothesis and against the null hypothesis also decreases. Thus, researchers are generally hoping for larger t values. The other component of t is its direction. When t is positive, it indicates that Group 1 had the higher mean than Group 2. Conversely, when t is negative, it indicates that Group 1 had the lower mean than Group 2. If the obtained t -value was 0.00, it would indicate that the means were zero standard errors apart (which is the same as saying the means were equal). If, for example, the obtained t -value was 3.00 it would indicate that the means of the two groups were three standard errors apart and that Group 1 had the higher mean. However, if the obtained t -value was -3.00 it would indicate that the two means were three standard errors apart and that Group 2 had the higher mean.

When testing a two-tailed (non-directional) hypothesis, only the magnitude needs to be considered to determine whether a result is significant. This is because a two-tailed hypothesis will be significantly supported if the difference in the means is sufficiently large regardless of which group mean was higher. However, when testing a one-tailed (directional) hypothesis, both magnitude and direction need to be considered. When it is hypothesized that Group 1 will have the higher mean than Group 2, the hypothesis will be significantly supported if the difference in the means is sufficiently large and the result is positive. When it is hypothesized that Group 2 will have the higher mean than Group 1, the hypothesis will be significantly supported if the difference in the means is sufficiently large and the result is negative. Thus, the direction of the results must match the direction of the hypothesis when using a one-tailed test of significance.

Alternative Ways to Write the Independent Samples t -Test Formula.

You may see the independent samples t -test written other ways. One common version is this:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{SS_1 + SS_2}{n_1 + n_2 - 2} \right] \left[\frac{n_1 + n_2}{n_1 \times n_2} \right]}}$$

Another common version is this:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left[\frac{SS_1 + SS_2}{n_1 + n_2 - 2} \right] \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$$

These formulas are just different ways of getting to the same end result which some people find more or less intuitive. Let's take a quick look at the variations in these formulas and how they are equivalent to what is in the version of the formula shown earlier. First, both of these versions use sum of squares (SS) in place of adjusted sample size multiplied by variance for each group. This is because variance is calculated by dividing SS by adjusted sample size (calculated as $n-1$ when working with samples). Thus, if we multiple variance by $n-1$ it turns it back into SS .

$$\text{Variance Formula: } s^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1}$$

$$\text{Sum of Squares Formula: } SS = \Sigma(X - \bar{X})^2$$

We can write it out to demonstrate. Variance times adjusted sample size is equal to SS because the adjusted sample sizes cancel out like so:

$$\begin{aligned} s^2 &= \frac{\Sigma(X - \bar{X})^2}{n - 1} \\ s^2(n - 1) &= \frac{\Sigma(X - \bar{X})^2}{\cancel{n - 1}} (\cancel{n - 1}) \\ SS &= \Sigma(X - \bar{X})^2 \end{aligned}$$

For this reason, SS can be used in place of $(n-1)s^2$ for each group in the denominator of the formula.

The other option for rewriting the formula is to use $\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$ in place of $\left[\frac{n_1 + n_2}{n_1 \times n_2}\right]$ in the right side of the denominator of the formula. These are just two ways one can write the same things.

$$\left[\frac{1}{n_1} + \frac{1}{n_2}\right] = \left[\frac{n_1 + n_2}{n_1 \times n_2}\right]$$

Therefore, though we have seen three different ways to write the independent samples t -test formula, they are all simply different ways of writing the same mathematical concepts and steps and, thus, will all yield the same result.

Reading Review 8.2

1. What is being calculated and represented by the numerator of the independent samples t test formula?
2. What is being calculated and represented by the denominator of the independent samples t -test formula?
3. Without making a determination of significance, how might $t = 0.00$ be interpreted?
4. Without making a determination of significance, how might $t = 1.46$ be interpreted?
5. Without making a determination of significance, how might $t = -3.75$ be interpreted?

Formula Components

Now that we have taken some time to understand the construction of the independent samples t -test formula, let's focus on how to actually use it, starting with identifying all of its parts.

In order to solve for t , six things must first be known:

\bar{X}_1 = the mean for Group 1

\bar{X}_2 = the mean for Group 2

S_1^2 = the variance for Group 1

S_2^2 = the variance for Group 2

n_1 = the sample size for Group 1

n_2 = the sample size for Group 2

We use the raw scores from each group to calculate their respective means, variances, and sample sizes.

Formula Steps

The steps are shown in order and categorized into two sections: A) preparation and B) solving. I recommend using this categorization to help you organize, learn, and properly use all inferential formulas. Preparation steps refer to any calculations that need to be done before values can be plugged into the formula. For the independent samples t -test this includes finding the descriptive statistics that make up the six components of the formula for each group (listed in the section above). Once those are known, the steps in section B can be used to yield the obtained value for the formula. The symbol for the obtained value for each t -test is t . Follow these steps, in order, to find t .

Section A: Preparation

1. Find n for Group 1.
2. Find \bar{x} for Group 1.
3. Find s^2 for Group 1.
4. Find n for Group 2.
5. Find \bar{x} for Group 2.
6. Find s^2 for Group 2.

Section B: Solving

1. Write the formula with the values found in section A plugged into their respective locations.
2. Solve the numerator by subtracting the mean of Group 2 from the mean of Group 1.
3. Solve for the left side of the denominator as follows:

- a. Multiply variance for Group 1 by $n-1$ for Group 1 to get the SS for Group 1.
 - b. Multiply variance for Group 2 by $n-1$ for Group 2 to get the SS for Group 2.
 - c. Add the SS for Group 1 (which is the result of step 3a) to the SS for Group 2 (which is the result of step 3b). This gives you total SS .
 - d. Find the degrees of freedom (df) by adding the sample size for Group 1 to the sample size for Group 2 and then subtracting 2 from that total.
 - e. Divide the total SS (which is result of step 3c) by the df (which is the result of step 3d).
4. Solve right side of the denominator as follows:
1. Add the sample size of Group 1 to the sample size of Group 2 to get the total sample size.
 2. Multiply the sample size of Group 1 by the sample size of Group 2.
 3. Divide the total sample size (which is the result of step 4a) by the product of sample sizes (which is the result of step 4b).
5. Multiply the left side of the denominator (which is the result of step 3e) by the right side of the denominator (which is the result of step 4c).
6. Square root the denominator (which means square root the result of step 5) to get the pooled standard error for the formula.
7. Finally, divide the numerator (which is the result of step 2) by the pooled standard error (which is the result of step 6) to get the obtained t -value.

This page titled [8.3: The Independent Samples t-Test Formula](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by .