

## 2.4: Chapter 10- Independent Samples

### Key Terms

group mean differences

homogeneity of variance

independent samples

pooled variance

We have seen how to compare a single mean against a given value and how to utilize difference scores to look for meaningful, consistent change via a single mean difference. Now, we will learn how to compare two separate means from groups that do not overlap to see if there is a difference between them. The process of testing hypotheses about two means is exactly the same as it is for testing hypotheses about a single mean, and the logical structure of the formulas is the same as well. However, we will be adding a few extra steps this time to account for the fact that our data are coming from different sources.

### Difference of Means

In Chapter 9, we learned about mean differences, that is, the average value of difference scores. Those difference scores came from one group and two time points (or two perspectives). Now, we will deal with the difference of the means, that is, the average values of separate groups that are represented by separate descriptive statistics. This analysis involves two groups and one time point. As with all of our other tests as well, both of these analyses are concerned with a single variable.

It is very important to keep these two tests separate and understand the distinctions between them because they assess very different questions and require different approaches to the data. When in doubt, think about how the data were collected and where they came from. If they came from two time points with the same people (sometimes referred to as “longitudinal” data), you know you are working with two data points from the same participant (the measurement was repeated) and will use a related samples t test (see Chapter 9). If it came from a single time point that used separate groups, you need to look at the nature of those groups and decide whether they are related. Can individuals in one group being meaningfully matched up with one and only one individual from the other group? For example, are they a romantic couple? If so, we call those data matched pairs and we use a related samples t test (Chapter 9). However, if there’s no logical or meaningful way to link individuals across groups, or if there is no overlap between the groups, then we say the groups are independent and use the independent samplest test, the subject of this chapter.

### Research Questions about Independent Means

Many research ideas in the behavioral sciences and other areas of research are concerned with whether or not two means are the same or different. Logically, we therefore say that these research questions are concerned with group mean differences. That is, on average, do we expect a person from Group A to be higher or lower on some variable than a person from Group B. In any research design looking at group mean differences, there are some key criteria we must consider: the groups must be mutually exclusive (i.e., you can only be part of one group at any given time), and the groups have to be measured on the same variable (i.e., you can’t compare personality in one group to reaction time in another group since those values would not be the same anyway).

Let’s look at one of the most common and logical examples: testing a new medication. When a new medication is developed, the researchers who created it need to demonstrate that it effectively treats the symptoms they are trying to alleviate. The simplest design that will answer this question involves two groups: one group that receives the new medication (the “treatment” group) and one group that receives a placebo (the “control” group). Participants are randomly assigned to one of the two groups (remember that random assignment is the hallmark of a true experiment), and the researchers test the symptoms in each person in each group after they received either the medication or the placebo. They then calculate the average symptoms in each group and compare them to see if the treatment group did better (i.e., had fewer or less severe symptoms) than the control group.

In this example, we had two groups: treatment and control. Membership in these two groups was mutually exclusive—each individual participant received either the experimental medication or the placebo. No one in the experiment received both, so there was no overlap between the two groups. Additionally, each group could be measured on the same variable: symptoms related to the disease or ailment being treated. Because each group was measured on the same variable, the average scores in each group could be meaningfully compared. If the treatment was ineffective, we would expect that the average symptoms of someone receiving the treatment would be the same as the average symptoms of someone receiving the placebo (i.e., there is no difference between the

groups). However, if the treatment was effective, we would expect fewer symptoms from the treatment group, leading to a lower group average.

Now let's look at an example using groups that already exist. A common, and perhaps salient, question is how students feel about their job prospects after graduation. Suppose that we have narrowed our potential choice of college down to two universities and, in the course of trying to decide between the two, we come across a survey that has data from each university on how students at those universities feel about their future job prospects. As with our last example, we have two groups: University A and University B, and each participant is in only one of the two groups (assuming there are no transfer students who were somehow able to rate both universities). Because students at each university completed the same survey, they are measuring the same thing, so we can use a t test to compare the average perceptions of students at each university to see if they are the same. If they are the same, then we should continue looking for other things about each university to help us decide on where to go. But, if they are different, we can use that information in favor of the university with higher job prospects.

As we can see, the grouping variable we use for an independent samples t test can be a set of groups we create (as in the experimental medication example) or groups that already exist naturally (as in the university example). There are countless other examples of research questions relating to two group means, making the independent samples t test one of the most widely used analyses around.

### Hypotheses and Decision Criteria

The process of testing hypotheses using an independent samples t test is the same as it was in the last three chapters, and it starts with stating our hypotheses and laying out the criteria we will use to test them.

Our null hypothesis for an independent samples t test is the same as all others: there is no difference. The means of the two groups are the same under the null hypothesis, no matter how those groups were formed. Mathematically, this takes on two equivalent forms:

$$H_0: \mu_1 = \mu_2$$

or

$$H_0: \mu_1 - \mu_2 = 0$$

Both of these formulations of the null hypothesis tell us exactly the same thing: that the numerical value of the means is the same in both groups. This is more clear in the first formulation, but the second formulation also makes sense (any number minus itself is always zero) and helps us out a little when we get to the math of the test statistic. Either one is acceptable and you only need to report one. The English interpretation of both of them is also the same:

H0: There is no difference between the means of the two groups

Our alternative hypotheses are also unchanged: we simply replace the equal sign (=) with one of the three inequalities (>, <, ≠):

$$H_A: \mu_1 > \mu_2$$

$$H_A: \mu_1 < \mu_2$$

$$H_A: \mu_1 \neq \mu_2$$

or

$$H_A: \mu_1 - \mu_2 > 0$$

$$H_A: \mu_1 - \mu_2 < 0$$

$$H_A: \mu_1 - \mu_2 \neq 0$$

Whichever formulation you chose for the null hypothesis should be the one you use for the alternative hypothesis (be consistent), and the interpretation of them is always the same:

HA: There is a difference between the means of the two groups

Notice that we are now dealing with two means instead of just one, so it will be very important to keep track of which mean goes with which population and, by extension, which dataset and sample data. We use subscripts to differentiate between the

populations, so make sure to keep track of which is which. If it is helpful, you can also use more descriptive subscripts. To use the experimental medication example:

$H_0$ : There is no difference between the means of the treatment and control groups

$$H_0: \mu_{\text{treatment}} = \mu_{\text{control}}$$

$H_A$ : There is a difference between the means of the treatment and control groups

$$H_A: \mu_{\text{treatment}} \neq \mu_{\text{control}}$$

Once we have our hypotheses laid out, we can set our criteria to test them using the same three pieces of information as before: significance level ( $\alpha$ ), directionality (left, right, or two-tailed), and degrees of freedom, which for an independent samples t test are:

$$df = n_1 + n_2 - 2$$

This looks different than before, but it is just adding the individual degrees of freedom from each group ( $n - 1$ ) together. Notice that the sample sizes,  $n$ , also get subscripts so we can tell them apart.

For an independent samples t test, it is often the case that our two groups will have slightly different sample sizes, either due to chance or some characteristic of the groups themselves. Generally, this is not an issue, so long as one group is not massively larger than the other group. What is of greater concern is keeping track of which is which using the subscripts.

### Independent Samples t Statistic

The test statistic for our independent samples t test takes on the same logical structure and format as our other t tests: our observed effect minus our null hypothesis value, all divided by the standard error:

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{M_1 - M_2}}$$

This looks like more work to calculate, but remember that our null hypothesis states that the quantity  $\mu_1 - \mu_2 = 0$ , so we can drop that out of the equation and are left with:

$$t = \frac{(M_1 - M_2)}{s_{M_1 - M_2}}$$

Our standard error in the denomination is still standard deviation ( $s$ ) with a subscript denoting what it is the standard error of. Because we are dealing with the difference between two separate means, rather than a single mean or single mean of difference scores, we put both means in the subscript. Calculating our standard error, as we will see next, is where the biggest differences between this t test and other t tests appears. However, once we do calculate it and use it in our test statistic, everything else goes back to normal. Our decision criteria are still comparing our obtained test statistic to our critical value, and our interpretation based on whether or not we reject the null hypothesis is unchanged as well.

### Standard Error and Pooled Variance

Recall that the standard error is the average distance between any given sample mean and the center of its corresponding sampling distribution, and it is a function of the standard deviation of the population (either given or estimated) and the sample size. This definition and interpretation hold true for our independent samples t test as well, but because we are working with two samples drawn from two populations, we have to first combine their estimates of standard deviation—or, more accurately, their estimates of variance—into a single value that we can then use to calculate our standard error.

The combined estimate of variance using the information from each sample is called the pooled variance and is denoted  $s_p^2$ ; the subscript  $p$  serves as a reminder indicating that it is the pooled variance. The term “pooled variance” is a literal name because we are simply pooling or combining the information on variance—the sum of squares and degrees of freedom—from both of our samples into a single number. The result is a weighted average of the observed sample variances, the weight for each being

determined by the sample size, and will always fall between the two observed variances. The computational formula for the pooled variance is:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

This formula can look daunting at first, but it is in fact just a weighted average. Even more conveniently, some simple algebra can be employed to greatly reduce the complexity of the calculation. The simpler and more appropriate formula to use when calculating pooled variance is:

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

Using this formula, it's very simple to see that we are just adding together the same pieces of information we have been calculating since Chapter 3. Thus, when we use this formula, the pooled variance is not nearly as intimidating as it might have originally seemed.

Once we have our pooled variance calculated, we can drop it into the equation for our standard error:

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

Once again, although this formula may seem different than it was before, in reality it is just a different way of writing the same thing. An alternative but mathematically equivalent way of writing our old standard error is:

$$s_M = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$$

Looking at that, we can now see that, once again, we are simply adding together two pieces of information—no new logic or interpretation required. Once the standard error is calculated, it goes in the denominator of our test statistic, as shown above and as was the case in all previous chapters. Thus, the only additional step to calculating an independent samples t statistic is computing the pooled variance. Let's see an example in action.

#### Example Movies and Mood

We are interested in whether the type of movie someone sees at the theater affects their mood when they leave. We decide to ask people about their mood as they leave one of two movies: a comedy (Group 1,  $n = 35$ ) or a horror film (Group 2,  $n = 29$ ). Our data are coded so that higher scores indicate a more positive mood. We have good reason to believe that people leaving the comedy will be in a better mood, so we use a one-tailed test at  $\alpha = .05$  to test our hypothesis.

#### Step 1: State the Hypotheses

As always, we start with hypotheses:

$H_0$ : There is no difference in average mood between the two movie types

$$H_0: \mu_1 - \mu_2 = 0$$

or

$$H_0: \mu_1 = \mu_2$$

$H_A$ : The comedy film will produce a better average mood than the horror film

$$H_A: \mu_1 - \mu_2 > 0$$

or

$$H_A: \mu_1 > \mu_2$$

Notice that in the first formulation of the alternative hypothesis we say that the first mean minus the second mean will be greater than zero. This is based on how we code the data (higher is better), so we suspect that the mean of the first group will be higher. Thus, we will have a larger number minus a smaller number, which will be greater than zero. Be sure to pay attention to which

group is which and how your data are coded (higher is almost always used as better outcomes) to make sure your hypothesis makes sense!

### Step 2: Find the Critical Values

Just like before, we will need critical values, which come from our t table. In this example, we have a one-tailed test at  $\alpha = .05$  and expect a positive answer (because we expect the difference between the means to be greater than zero). Our degrees of freedom for our independent samples t test is just the degrees of freedom from each group added together:  $35 + 29 - 2 = 62$ . From our t table, we find that our critical value is  $t^* = 1.671$ . Note that because 62 does not appear on the table, we use the next lowest value, which in this case is 60.

### Step 3: Compute the Test Statistic and Effect Size

The data from our two groups are presented in the tables below. [Table 10.1](#) shows the values for the Comedy group, and [Table 10.2](#) shows the values for the Horror group. Values for both have already been placed in the sum of squares tables since we will need to use them for our further calculations. As always, the column on the left is our raw data.

Table 10.1. Raw scores and sum of squares for Group 1 (comedy).

X	X - M	(X - M) <sup>2</sup>
39.10	15.10	228.01
38.00	14.00	196.00
14.90	-9.10	82.81
20.70	-3.30	10.89
19.50	-4.50	20.25
32.20	8.20	67.24
11.00	-13.00	169.00
20.70	-3.30	10.89
26.40	2.40	5.76
35.70	11.70	136.89
26.40	2.40	5.76
28.80	4.80	23.04
33.40	9.40	88.36
13.70	-10.30	106.09
46.10	22.10	488.41
13.70	-10.30	106.09
23.00	-1.00	1.00
20.70	-3.30	10.89
19.50	-4.50	20.25
11.40	-12.60	158.76
24.10	0.10	0.01
17.20	-6.80	46.24
38.00	14.00	196.00
10.30	-13.70	187.69

X	$X - M$	$(X - M)^2$
35.70	11.70	136.89
41.50	17.50	306.25
18.40	-5.60	31.36
36.80	12.80	163.84
54.10	30.10	906.01
11.40	-12.60	158.76
8.70	-15.30	234.09
23.00	-1.00	1.00
14.30	-9.70	94.09
5.30	-18.70	349.69
6.30	-17.70	313.29
$\Sigma = 840$	$\Sigma = 0$	$\Sigma = 5061.60$

Table 10.2. Raw scores and sum of squares for Group 2 (horror).

X	$X - M$	$(X - M)^2$
24.00	7.50	56.25
17.00	0.50	0.25
35.80	19.30	372.49
18.00	1.50	2.25
-1.70	-18.20	331.24
11.10	-5.40	29.16
10.10	-6.40	40.96
16.10	-0.40	0.16
-0.70	-17.20	295.84
14.10	-2.40	5.76
25.90	9.40	88.36
23.00	6.50	42.25
20.00	3.50	12.25
14.10	-2.40	5.76
-1.70	-18.20	331.24
19.00	2.50	6.25
20.00	3.50	12.25
30.90	14.40	207.36
30.90	14.40	207.36
22.00	5.50	30.25

X	X - M	(X - M) <sup>2</sup>
6.20	-10.30	106.09
27.90	11.40	129.96
14.10	-2.40	5.76
33.80	17.30	299.29
26.90	10.40	108.16
5.20	-11.30	127.69
13.10	-3.40	11.56
19.00	2.50	6.25
-15.50	-32.00	1024.00
$\Sigma = 478.6$	$\Sigma = 0.10$	$\Sigma = 3896.45$

Using the sum of the first column for each table, we can calculate the mean for each group:

$$M_1 = \frac{840}{35} = 24.00$$

and

$$M_2 = \frac{478.60}{29} = 16.50$$

These values were used to calculate the middle rows of each table, which sum to zero as they should (the middle column for Group 2 sums to a very small value instead of zero due to rounding error—the exact mean is 16.50344827586207, but that's far more than we need for our purposes). Squaring each of the deviation scores in the middle columns gives us the values in the third columns, which sum to our next important value: the sum of squares for each group:  $SS_1 = 5061.60$  and  $SS_2 = 3896.45$ . These values have all been calculated and take on the same interpretation as they have since Chapter 3—no new computations yet. Before we move on to the pooled variance that will allow us to calculate standard error, let's compute our standard deviation for each group; even though we will not use them in our calculation of the test statistic, they are still important descriptors of our data:

$$s_1 = \sqrt{\frac{5061.60}{34}} = 12.20$$

and

$$s_2 = \sqrt{\frac{3896.45}{28}} = 11.80$$

Now we can move on to our new calculation, the pooled variance, which is just the sums of squares that we calculated from our table and the degrees of freedom, which is just  $n - 1$  for each group:

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{5061.60 + 3896.45}{34 + 28} = \frac{8958.05}{62} = 144.48$$

As you can see, if you follow the regular process of calculating standard deviation using the sum of squares table, finding the pooled variance is very easy. Now we can use that value to calculate our standard error, the last step before we can find our test statistic:

$$s_{M_1-M_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{\frac{144.48}{35} + \frac{144.48}{29}} = \sqrt{4.13 + 4.98} = \sqrt{9.11} = 3.02$$

Finally, we can use our standard error and the means we calculated earlier to compute our test statistic. Because the null hypothesis value of  $\mu_1 - \mu_2$  is 0.00, we will leave that portion out of the equation for simplicity:

$$t = \frac{(M_1 - M_2)}{s_{M_1-M_2}} = \frac{24.00 - 16.50}{3.02} = 2.48$$

The process of calculating our obtained test statistic  $t = 2.48$  followed the same sequence of steps as before: use raw data to compute the mean and sum of squares (this time for two groups instead of one), use the sum of squares and degrees of freedom to calculate standard error (this time using pooled variance instead of standard deviation), and use that standard error and the observed means to get  $t$ .

### Effect Sizes and Confidence Intervals

We have seen in previous chapters that even a statistically significant effect needs to be interpreted along with an effect size to see if it is practically meaningful. We have also seen that our sample means, as a point estimate, are not perfect and would be better represented by a range of values that we call a confidence interval. As with all other topics, this is also true of our independent samples  $t$  tests.

Our effect size for the independent samples  $t$  test is still Cohen's  $d$ , and it is still just our observed effect divided by the standard deviation. Remember that standard deviation is just the square root of the variance, and because we work with pooled variance in our test statistic, we will use the square root of the pooled variance as our denominator in the formula for Cohen's  $d$ . This gives us:

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

For our example above, we can calculate the effect size to be:

$$d = \frac{24.00 - 16.50}{\sqrt{144.48}} = \frac{7.50}{12.02} = 0.62$$

We interpret this using the same guidelines as before, so we would consider this a moderate or moderately large effect.

Our confidence intervals also take on the same form and interpretation as they have in the past. The value we are interested in is the difference between the two means, so our point estimate is the value of one mean minus the other, or  $M_1 - M_2$ . Just like before, this is our observed effect and is the same value as the one we place in the numerator of our test statistic. We calculate this value then place the margin of error—still our critical value multiplied by our standard error—above and below it. That is:

$$\text{Confidence interval} = (M_1 - M_2) \pm t^*(s_{M_1-M_2})$$

Because our hypothesis testing example used a one-tailed test, it would be inappropriate to calculate a confidence interval on those data (remember that we can only calculate a confidence interval for a two-tailed test because the interval extends in both directions). Let's say we find summary statistics on the average life satisfaction of people from two different towns and want to create a confidence interval to see if the difference between the two might actually be zero.

Our sample data are  $M_1 = 28.65$ ,  $s_1 = 12.40$ ,  $n_1 = 40$  and  $M_2 = 25.40$ ,  $s_2 = 15.68$ ,  $n_2 = 42$ . At face value, it looks like the people from the first town have higher life satisfaction (28.65 vs. 25.40), but it will take a confidence interval (or complete hypothesis testing process) to see if that is true or just due to random chance. First, we want to calculate the difference between our sample means, which is  $28.65 - 25.40 = 3.25$ . Next, we need a critical value from our  $t$  table. If we want to test at the normal 95% level of confidence, then our sample sizes will yield degrees of freedom equal to  $40 + 42 - 2 = 80$ . From our table, that gives us a critical value of  $t^* = 1.990$ . Finally, we need our standard error. Recall that our standard error for an independent samples  $t$  test uses pooled variance, which requires the sum of squares and degrees of freedom. Up to this point, we have calculated the sum of squares using raw data, but in this situation, we do not have access to it. So, what are we to do?



If we have summary data like standard deviation and sample size, it is very easy to calculate the pooled variance, and the key lies in rearranging the formulas to work backward through them. We need the sum of squares and degrees of freedom to calculate our pooled variance. Degrees of freedom is very simple: we just take the sample size minus 1.00 for each group. Getting the sum of squares is also easy: remember that variance is standard deviation squared and is the sum of squares divided by the degrees of freedom. That is:

$$s^2 = (s)^2 = \frac{SS}{df}$$

To get the sum of squares, we just multiply both sides of the above equation to get:

$$(s)^2(df) = SS$$

which is the squared standard deviation multiplied by the degrees of freedom ( $n - 1$ ) equals the sum of squares.

Using our example data:

$$(s_1)^2(df_1) = SS_1 \quad (s_2)^2(df_2) = SS_2$$

$$(12.40)^2(40 - 1) = 5996.64 \quad (15.68)^2(42 - 1) = 10080.36$$

And thus our pooled variance equals:

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{5996.64 + 10080.36}{39 + 41} = \frac{16077}{80} = 200.96$$

And our standard error equals:

$$s_{M_1 - M_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{\frac{200.96}{40} + \frac{200.96}{42}} = \sqrt{5.02 + 4.78} = \sqrt{9.89} = 3.14$$

All of these steps are just slightly different ways of using the same formulas, numbers, and ideas we have worked with up to this point. Once we get our standard error, it's time to build our confidence interval.

$$\begin{aligned} \text{Upper bound} &= 3.25 + 1.990(3.14) & \text{Lower bound} &= 3.25 - 1.990(3.14) \\ &= 3.25 + 6.25 & &= 3.25 - 6.25 \\ 95\% \text{ CI} &= 3.25 \pm 1.990(3.14) & &= 9.50 & &= -3.00 \\ 95\% \text{ CI} &= (-3.00, 9.50) \end{aligned}$$

Our confidence interval, as always, represents a range of values that would be considered reasonable or plausible based on our observed data. In this instance, our interval  $(-3.00, 9.50)$  does contain zero. Thus, even though the means look a little bit different, it may very well be the case that the life satisfaction in both of these towns is the same. Proving otherwise would require more data.

Now we can move on to the final step of the hypothesis-testing procedure.

#### Step 4: Make the Decision

Our test statistic has a value of  $t = 2.48$ , and in Step 2 we found that the critical value is  $t^* = 1.671$ . Because  $2.48 > 1.671$ , we reject the null hypothesis:

Reject  $H_0$ . Based on our sample data from people who watched different kinds of movies, we can say that the average mood after a comedy movie ( $M_1 = 24.00$ ,  $SD_1 = 12.20$ ) is better than the average mood after a horror movie ( $M_2 = 16.50$ ,  $SD_2 = 11.80$ ),  $t(62) = 2.48$ ,  $p < .05$ ,  $d = 0.62$ .

Figure 10.1 shows the output from JASP for this example.

Figure 10.1. Output from JASP for the independent-samples  $t$  test described in the Movies and Mood example. The output provides the  $t$  value (2.478), degrees of freedom (62), the exact  $p$  value (.008, which is less than .05), and Cohen's  $d$  for effect size (0.622). Note that the means and standard deviations for both samples are also provided. Based on our sample data from people who

watched different kinds of movies, we can say that the average mood after a comedy movie ( $M = 24.0$ ,  $SD = 12.2$ ) is better than the average mood after a horror movie ( $M = 16.5$ ;  $SD = 11.8$ ),  $t(62) = 2.48$ ,  $p = .008$ ,  $d = 0.62$ . (“JASP independent-samples t test” by Rupa G. Gordon/Judy Schmitt is licensed under [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/).)

## Independent Samples T-Test

Independent Samples T-Test

	t	df	p	Cohen's d
Mood	2.478	62	0.008	0.622

Note. For all tests, the alternative hypothesis specifies that group Comedy is greater than group Horror.

Note. Student's t-test.

## Descriptives

Group Descriptives

	Group	N	Mean	SD
Mood	Comedy	35	24.000	12.201
	Horror	29	16.528	11.775

## Homogeneity of Variance

Before wrapping up the coverage of independent samples t tests, there is one other important topic to cover. Using the pooled variance to calculate the test statistic relies on an assumption known as homogeneity of variance. In statistics, an assumption is some characteristic that we assume is true about our data, and our ability to use our inferential statistics accurately and correctly relies on these assumptions being true. If these assumptions are not true, then our analyses are at best ineffective (e.g., low power to detect effects) and at worst inappropriate (e.g., too many Type I errors). A detailed coverage of assumptions is beyond the scope of this course, but it is important to know that they exist for all analyses.

For the current analysis, one important assumption is homogeneity of variance. This is fancy statistical talk for the idea that the true population variance for each group is the same and any difference in the observed sample variances is due to random chance. (If this sounds eerily similar to the idea of testing the null hypothesis that the true population means are equal, that's because it is exactly the same!) This notion allows us to compute a single pooled variance that uses our easily calculated degrees of freedom. If the assumption is shown to not be true, then we have to use a very complicated formula to estimate the proper degrees of freedom. There are formal tests to assess whether or not this assumption is met, but we will not discuss them here.

Many statistical programs incorporate the test of homogeneity of variance automatically and can report the results of the analysis assuming it is true or assuming it has been violated. You can easily tell which is which by the degrees of freedom: the corrected degrees of freedom (which is used when the assumption of homogeneity of variance is violated) will have decimal places. Fortunately, the independent samples t test is very robust to violations of this assumption (an analysis is “robust” if it works well even when its assumptions are not met), which is why we do not bother going through the tedious work of testing and estimating new degrees of freedom by hand.

## Exercises

1. What is meant by “the difference of the means” when talking about an independent samples t test? How does it differ from the “mean of the differences” in a related samples t test?
2. Describe three research questions that could be tested using an independent samples t test.
3. Calculate pooled variance from the following raw data:

Group 1	Group 2

Group 1	Group 2
16	4
11	10
9	15
7	13
5	12
4	9
12	8

4. Calculate the standard error from the following descriptive statistics.

1.  $s_1 = 24$ ,  $s_2 = 21$ ,  $n_1 = 36$ ,  $n_2 = 49$
2.  $s_1 = 15.40$ ,  $s_2 = 14.80$ ,  $n_1 = 20$ ,  $n_2 = 23$
3.  $s_1 = 12$ ,  $s_2 = 10$ ,  $n_1 = 25$ ,  $n_2 = 25$

5. Determine whether to reject or fail to reject the null hypothesis in the following situations:

1.  $t(40) = 2.49$ ,  $\alpha = .01$ , one-tailed test to the right
2.  $M_1 = 64$ ,  $M_2 = 54$ ,  $n_1 = 14$ ,  $n_2 = 12$ ,  $s_{M_1 - M_2} = 9.75$ ,  $\alpha = .05$ , two-tailed test
3. 95% CI: (0.50, 2.10)

6. A professor is interested in whether the type of software program used in a statistics lab affects how well students learn the material. The professor teaches the same lecture material to two classes but has one class use a point-and-click software program in lab and has the other class use a basic programming language. The professor collects final exam scores for students in each class. Conduct a hypothesis test to answer the research question.

Point-and-Click	Programming
83	86
83	79
63	100
77	74
86	70
84	67
78	83
61	85
65	74
75	86
100	87
60	61
90	76
66	100
54	

7. A researcher wants to know if there is a difference in how busy someone is based on whether that person identifies as an early bird or a night owl. The researcher gathers data from people in each group, coding the data so that higher scores represent

higher levels of being busy, and tests for a difference between the two at the .05 level of significance. Conduct a hypothesis test to answer the research question.

Early Bird	Night Owl
23	26
28	10
27	20
33	19
26	26
30	18
22	12
25	25
26	

8. Lots of people claim that having a pet helps lower their stress level. Use the following summary data to test the claim that there is a lower average stress level among pet owners (Group 1) than among non-owners (Group 2) at the .05 level of significance.

◦  $M_1 = 16.25$ ,  $M_2 = 20.95$ ,  $s_1 = 4.00$ ,  $s_2 = 5.10$ ,  $n_1 = 29$ ,  $n_2 = 25$

9. Administrators at a university want to know if students in different majors are more or less extroverted than others. They provide you with descriptive statistics they have for English majors (coded as 1) and History majors (coded as 2) and ask you to create a confidence interval of the difference between them. Does this confidence interval suggest that the students from the majors differ?

◦  $M_1 = 3.78$ ,  $M_2 = 2.23$ ,  $s_1 = 2.60$ ,  $s_2 = 1.15$ ,  $n_1 = 45$ ,  $n_2 = 40$

10. Researchers want to know if people's awareness of environmental issues varies as a function of where they live. Use the following summary data from two states, Alaska and Hawaii, to test for a difference.

•  $M_H = 47.50$ ,  $M_A = 45.70$ ,  $s_H = 14.65$ ,  $s_A = 13.20$ ,  $n_H = 139$ ,  $n_A = 150$

### Answers to Odd-Numbered Exercises

1)

The difference of the means is one mean calculated from a set of scores compared to another mean calculated from a different set of scores; the independent samples t test looks for whether the two separate values are different from one another. This is different than the "mean of the differences" because the latter is a single mean computed on a single set of difference scores that come from one data collection of matched pairs. So, the difference of the means deals with two numbers but the mean of the differences is only one number.

3)

$SS_1 = 106.86$ ,  $SS_2 = 78.86$ ,  $s_p^2 = 15.48$

5)

- a) Reject
- b) Fail to reject
- c) Reject

7)

Step 1:  $H_0: \mu_1 - \mu_2 = 0$  "There is no difference in the average busyness of early birds versus night owls,"  $H_A: \mu_1 - \mu_2 \neq 0$  "There is a difference in the average busyness of early birds versus night owls."

Step 2: Two-tailed test,  $df = 15$ ,  $t^* = 2.131$

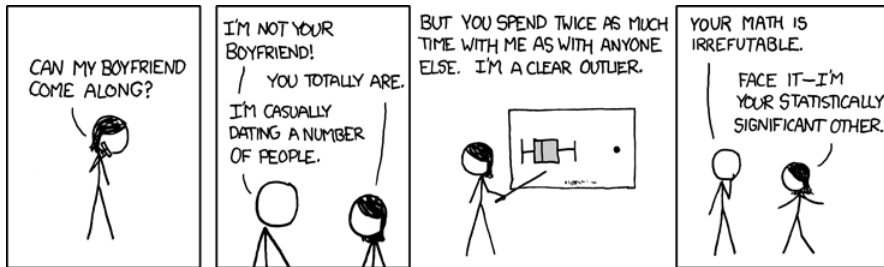
Step 3:  $M_1 = 26.67$ ,  $M_2 = 19.50$ ,  $s_p^2 = 27.73$ ,  $s_{M_1-M_2} = 2.37$ ,  $t = 3.03$

Step 4:  $t > t^*$ , Reject  $H_0$ . Based on our data of early birds and night owls, we can conclude that early birds are busier ( $M_1 = 26.67$ ) than night owls ( $M_2 = 19.50$ ),  $t(15) = 3.03$ ,  $p < .05$ ,  $d = 1.47$ .

9)

$M_1 - M_2 = 1.55$ ,  $t^* = 1.990$ ,  $s_{M_1-M_2} = 0.45$ ,  $CI = (0.66, 2.44)$ . This confidence interval does not contain zero, so it does suggest that there is a difference between the extroversion of English majors and History majors.

“Boyfriend” by Randall Munroe/xkcd.com is licensed under [CC BY-NC 2.5](https://creativecommons.org/licenses/by-nc/2.5/).)



This page titled [2.4: Chapter 10- Independent Samples](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Linda R. Cote, Rupa G. Gordon, Chrislyn E. Randell, Judy Schmitt, and Helena Marvin](#).