جامعة نيويورك أبوظبي

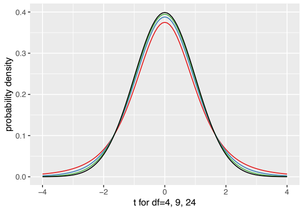
NYU | ABU DHABI

PSYCH-UH 1004Q: Statistics for Psychology

Class 12: independent samples *t*-test

Prof. Jon Sprouse
Psychology

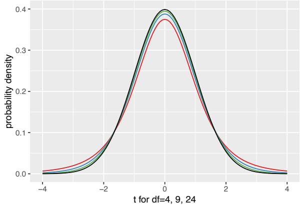
One sample vs two samples!

	one sample *t*-test
Scientific question	Does our sample differ from a population with a known mean (but unknown SD)?
Null Hypothesis	The mean of the population that the sample comes from is equal to the mean of the known population (so, $\mu = \mu_0$)
Equation	$t = \dfrac{\bar{x} - \mu_0}{s_{\bar{x}}}$
Descriptive information	The *t* statistic tells us how much our sample mean differs from the population mean in terms of sample SE (as an estimate)
Null distribution	*t*-distribution based on df 

Our first *t*-test was designed for scientific questions that could be answered using a single sample — that is one condition in our experiment.

One sample vs two samples!

Today we will look at a new version of the *t*-test designed to answer questions that involve **two samples**, which means two conditions in our experiment. Crucially, this version of the *t*-test is for two samples that are <u>independent</u> of each other.

	independent samples *t*-test
Scientific question	Do our two samples come from populations with the same mean (and unknown SDs)?
Null Hypothesis	The means of the populations that two samples come from are identical ($\mu_1 = \mu_2$)
Equation	$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
Descriptive information	The *t* statistic tells us how much the difference between sample means differs from the difference between population means
Null distribution	*t*-distribution based on df 

Remember, for experimental designs with two conditions, we have two options:

Repeated Measures **Independent Measures**

condition 1 condition 2

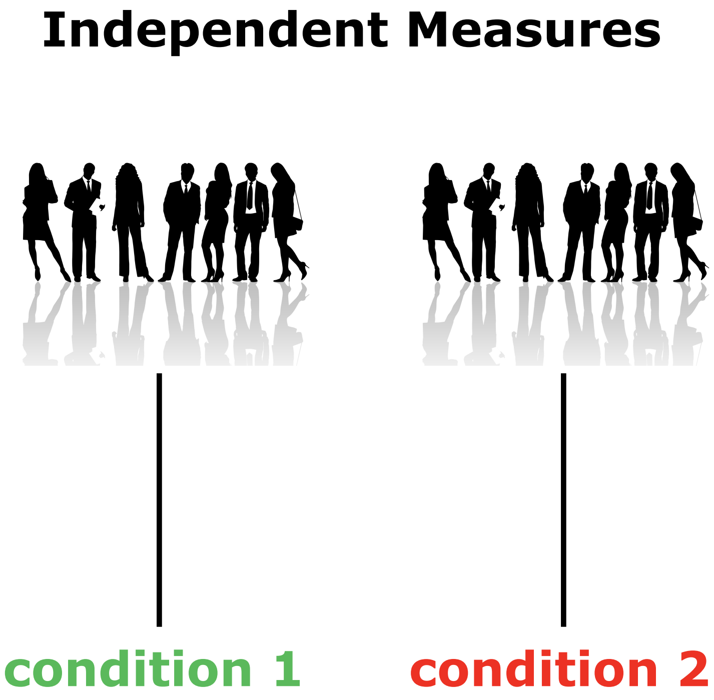
condition 1 condition 2

Repeated Measures:	If each participants sees every condition, we call it repeated measures. It is also called a **within-subjects** design.
Independent Measures:	If each participants sees only one condition, we call it independent measures. It is also called a **between-subjects** design.

Independent samples = Independent measures

If each participants sees only one condition, we call it **independent measures**. It is also called a **between-subjects** design.

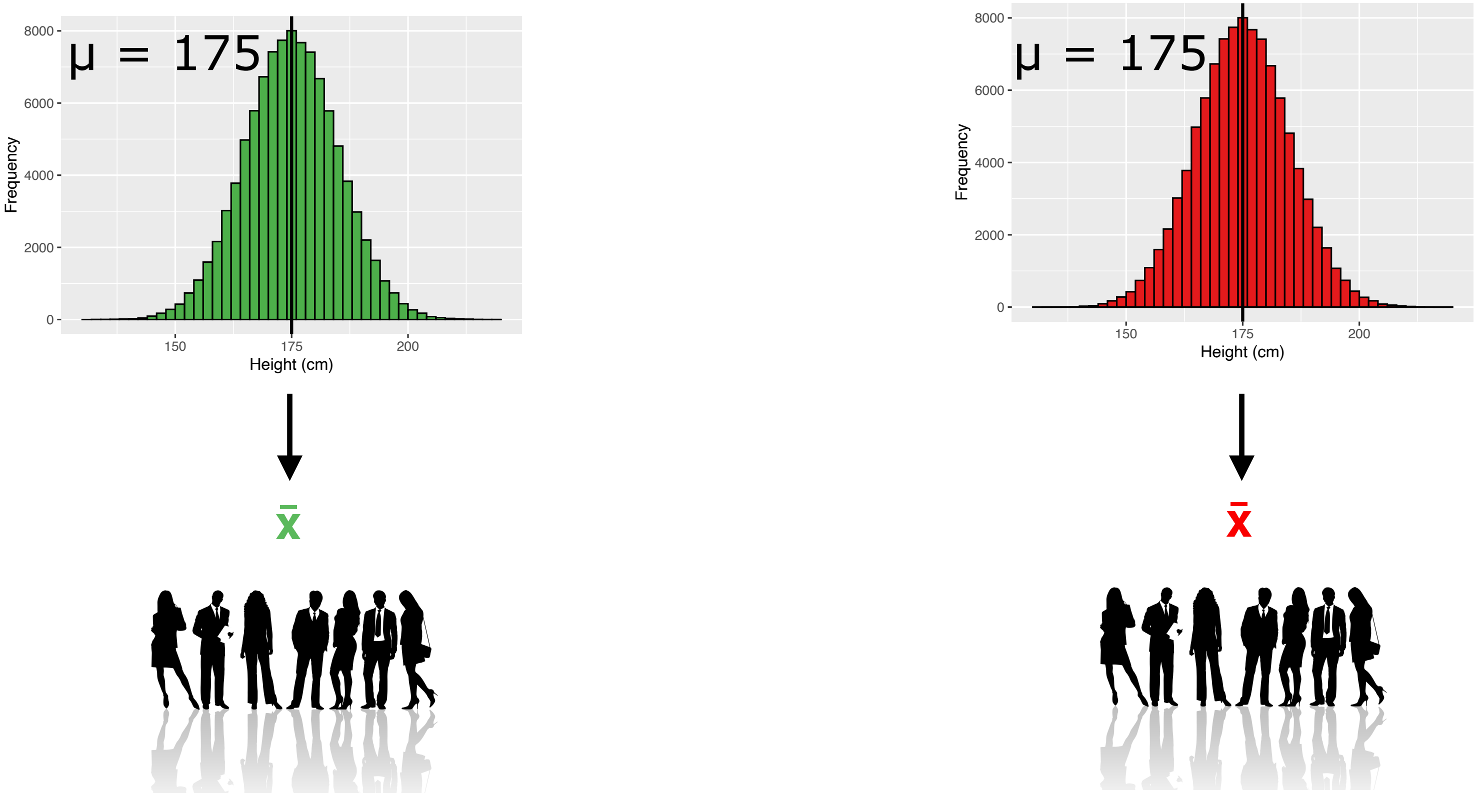
The **independent samples *t*-test** is specifically for a design that has two conditions, with each condition being tested in a different sample of participants.



This is a very common design in science. So this is a *t*-test that you will see and use often. Can you think of some questions that would be tested using this design?

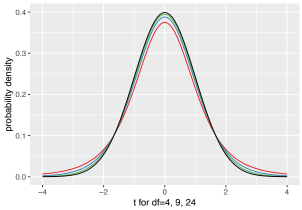
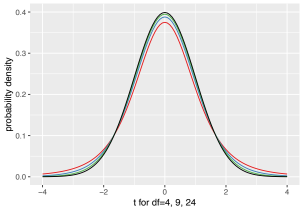
What is our null hypothesis (H_0)?

The null hypothesis is that the two samples come from two populations that have the same mean.



Mathematically, we would say that H_0 is $\mu_1 = \mu_2$. This is "null" because it says that there is no difference between the two population means ($\mu1 - \mu2 = 0$). You should also notice that this means that we would expect no difference between the two sample means ($\bar{x}1 - \bar{x}2 = 0$), except for sampling noise!

one sample vs independent samples *t*-tests

	one sample *t*-test	**independent samples *t*-test**
Scientific question	Does our sample differ from a population with a known mean (but unknown SD)?	Do our two samples come from populations with the same mean (and unknown SDs)?
Null Hypothesis	The mean of the population that the sample comes from is equal to the mean of the known population (so, $\mu = \mu_0$)	The means of the populations that two samples come from are identical ($\mu_1 = \mu_2$)
Equation	$t = \dfrac{\bar{x} - \mu_0}{s_{\bar{x}}}$	$t = \dfrac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$
Descriptive information	The *t* statistic tells us how much our sample mean differs from the population mean in terms of sample SE (as an estimate)	The *t* statistic tells us how much the difference between sample means differs from the difference between population means
Null distribution	*t*-distribution based on df 	*t*-distribution based on df 

Similarity with our previous *t* equation!

First, notice how similar the *t* equations are.

The numerator asks if our **sample mean** is different from the population mean. It is 0 when there is no difference!

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{n}}}$$

The denominator scales that difference by the estimate of the standard error based on the sample standard deviation. I have rewritten the denominator based on simple algebra to really show the similarity:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$$

The numerator asks if the **difference between sample means** is different from the difference between population means. It is 0 when there is no difference!

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

This denominator looks much more complicated. But you can see that it looks like a sum of two estimated standard errors — one for each sample.

Why does the estimated standard error have this equation?

The deepest answer is that it just is. This just is the analytic equation that best estimates the standard error of differences between means.

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

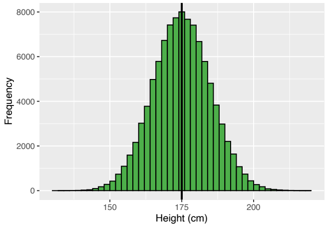
Remember, in mathematical modeling, there is nothing magical about an equation. It just is the equation that gives the value of the real-world thing that we are looking for.

Statisticians found this by looking at real world experiments, and then by using calculus to find the best equation that approximates what they saw in the real world. That’s how mathematical modeling works. You look for the best equation that matches what you see in the real world!

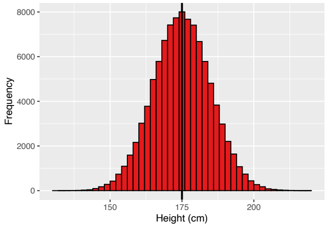
We can prove it to ourselves with a simulation

First, we simulate all of the mean differences that we'd get under the null hypothesis. This is a sampling distribution of mean differences, and its standard deviation is the standard error of mean differences. We can calculate that standard deviation directly from this distribution:

pop1



pop2

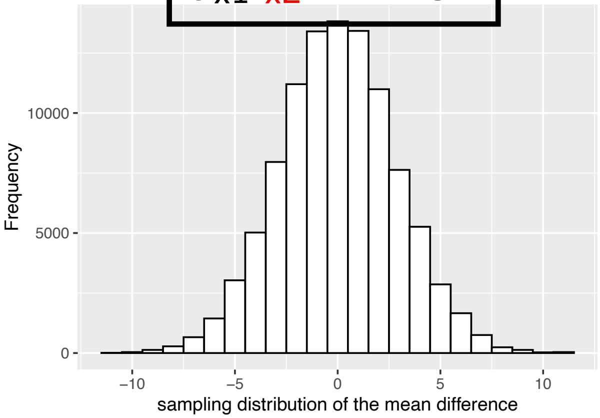


sample 25
10,000 times
from each

$\longrightarrow$

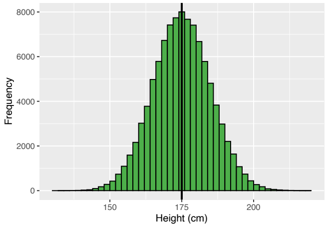
$\bar{x}_1 - \bar{x}_2$

$\sigma_{\bar{x}1-\bar{x}2} = 2.84$

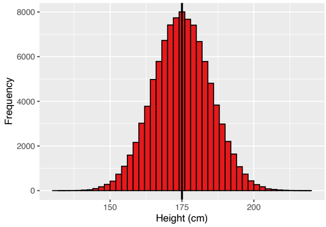


Next we can calculate the estimated standard error using the equation for each of our samples. We will see that the mean of all of those estimates is equal to the standard error that we calculated directly above:

pop1



pop2

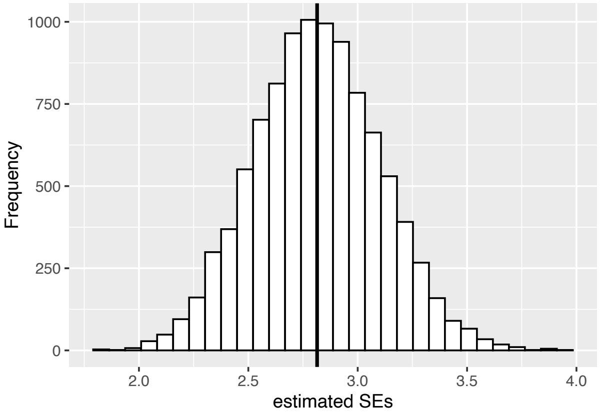


sample 25
10,000 times

$\longrightarrow$

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

mean of SEs = 2.84



How to think about the math of *t*-tests

The **numerator** of your test is determined by your **null hypothesis**:

The null hypothesis is $\mu = \mu_0$. This means that we expect the difference between our sample mean and μ_0 to be 0 plus sampling noise.

The null hypothesis is that $\mu_1 = \mu_2$. With two samples, we just expand each part of the equation. We expect the difference between our sample difference and the population differences to be 0 plus noise.

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The *t* statistic is just how the sample mean differs from the population mean. This is scaled by the standard error of the mean.

The *t* statistic is how differences between sample means differ from differences in population means. This is scaled by the standard error of mean differences.

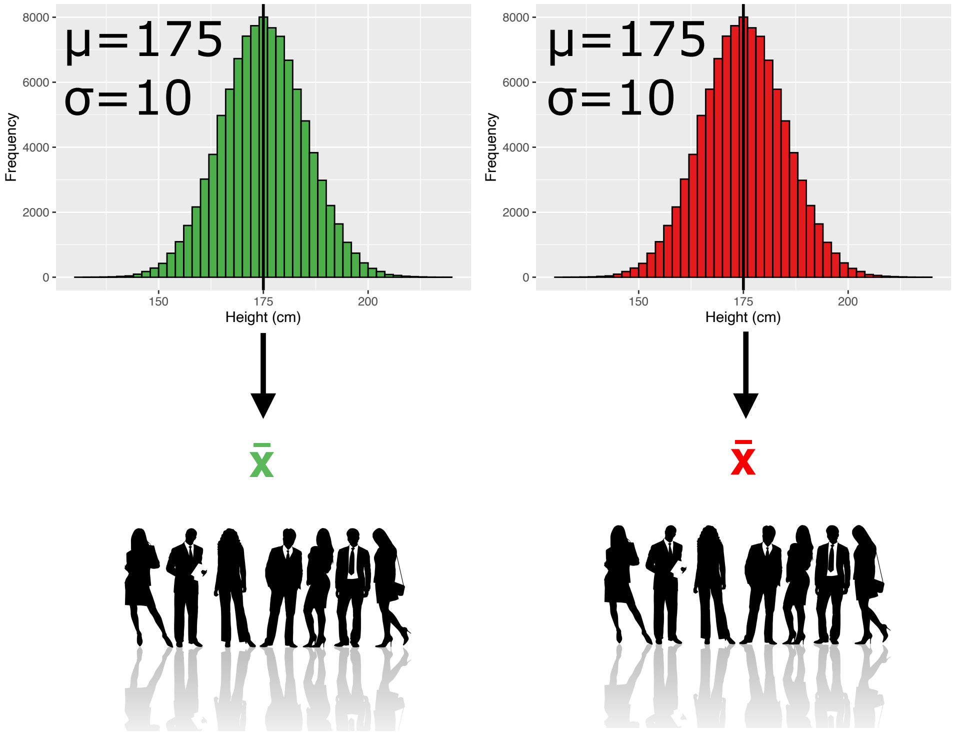
The **denominator** of your test is determined by the **standard error of the measure** in your numerator. It is there to <u>scale</u> the measure!

A complication: equal or unequal variances

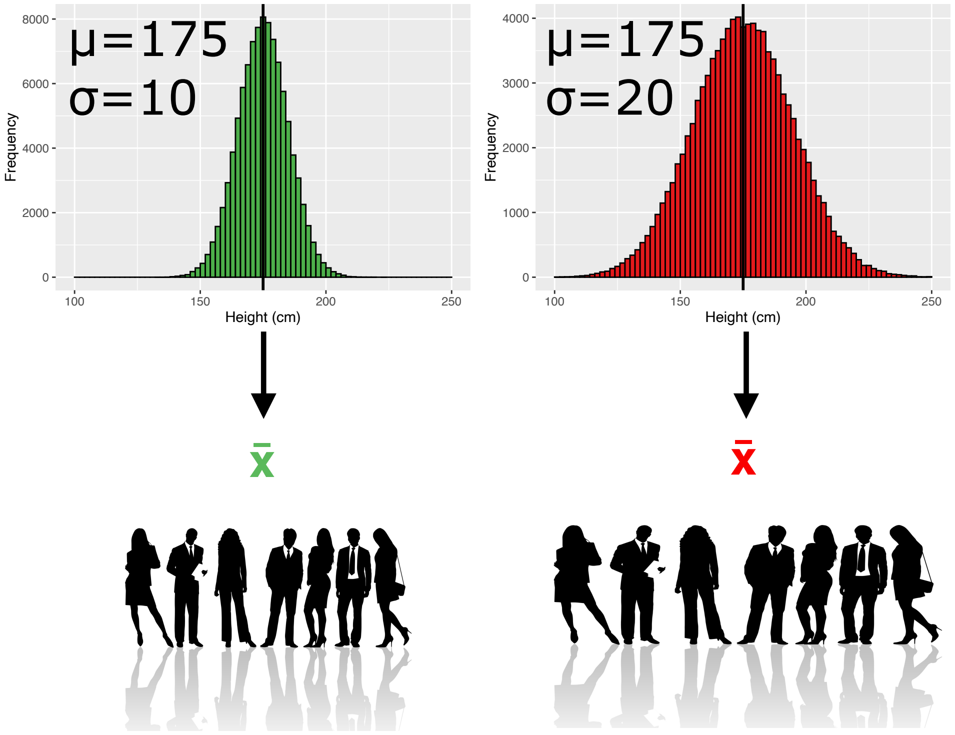
Variances can be equal or unequal

Because we are dealing with two populations, it is possible to have either equal or unequal variances in those populations. Like this:

equal variance



unequal variance

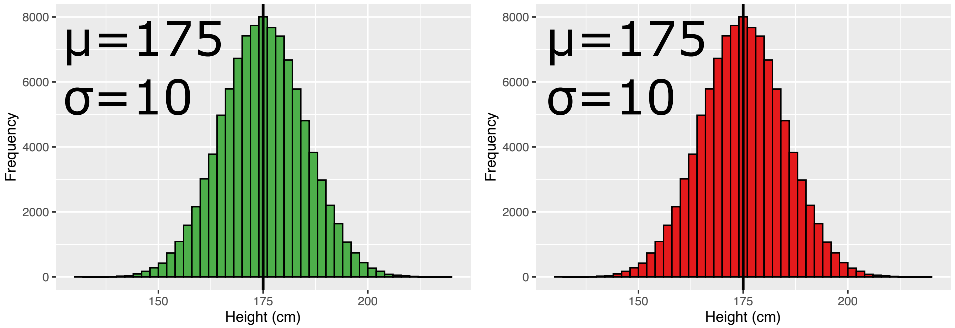


This matters for the calculation of the estimated standard error. And it matters for the shape of the *t*-distribution. Therefore it matters to your *p*-values!

The differences in the equations

The difference in the equations is just about the estimated standard error. So the only change is in the denominator.

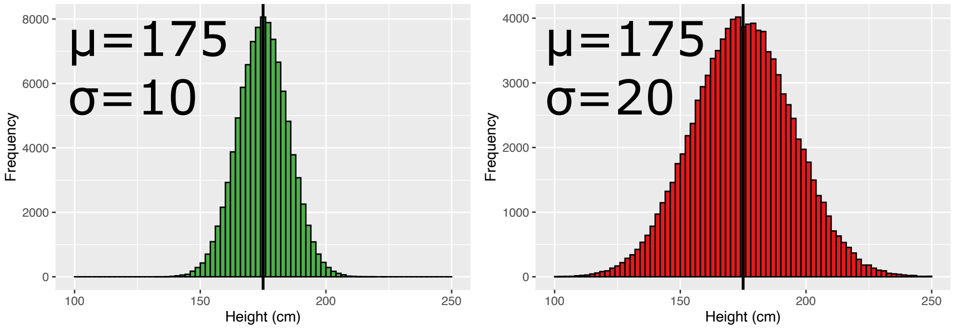
equal variance



$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

The equation for equal variances uses both samples to estimate a single variance number called the pooled variance (s_p^2). The idea is that it combines the two estimates into one much better estimate! The result is a standard *t*-distribution.

unequal variance



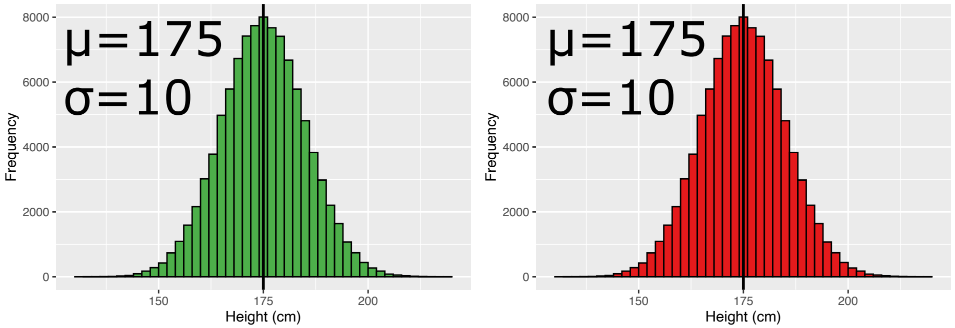
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

We have already seen the equation for unequal variance. It is (perhaps counterintuitively) the simpler equation. But the *t*-distribution is different from our standard *t*-distribution. Because of this, this test is really only used with a computer.

We will only use equal variances

In this course whenever we have an independent samples *t*-test, we will always use the equal variances version.

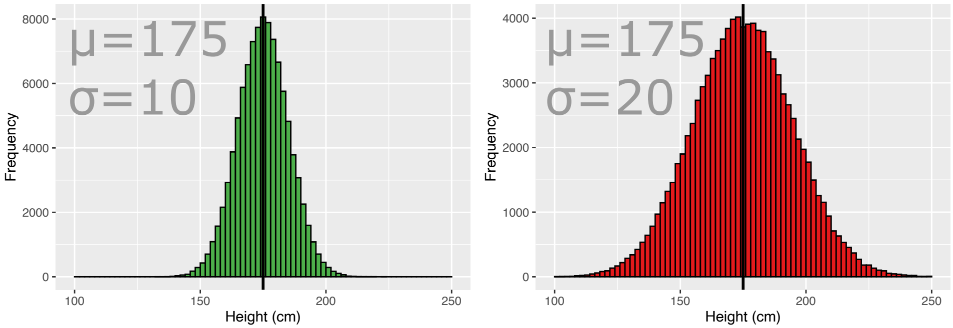
equal variance



$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

The equation for equal variances uses both samples to estimate a single variance number called the pooled variance (s_p^2). The idea is that it combines the two estimates into one much better estimate! The result is a standard *t*-distribution.

unequal variance



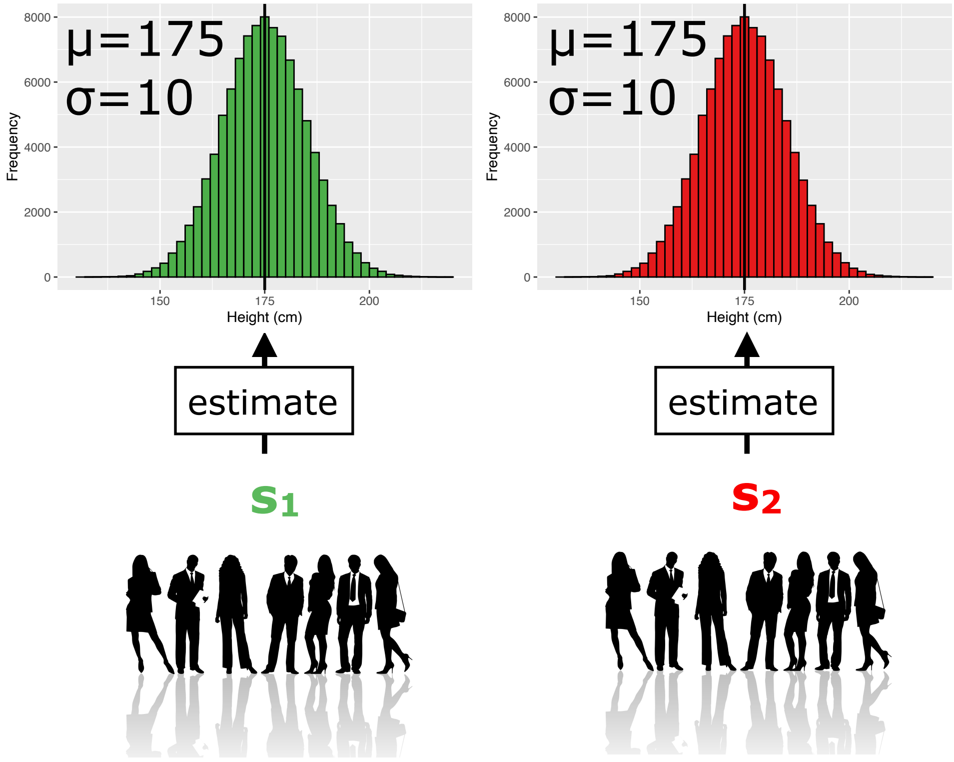
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

We have already seen the equation for unequal variance. It is (perhaps counterintuitively) the simpler equation. But the *t*-distribution is different from our standard *t*-distribution. Because of this, this test is really only used with a computer.

Why do we pool the variance?

The idea behind pooled variance is to calculate the variance for each sample and then calculate a **weighted mean** of the two variances.

The logic behind this *t*-test is that we want to use the variance of the samples as an estimate of the variance of the populations.



But in the equal variance case the the two populations have the same variance! So both samples are estimating the same parameter value!

Logically, this means we can combine these two estimates to make an even better estimate! The way to do this is with a mean. We make it a weighted mean just in case the sample sizes are different - that way we give more weight to the estimate that came from a larger sample!

What is a weighted mean?

To see the logic of a weighted mean, we can compare it to a simple mean.

simple mean

Here is a simple mean of two grades.

$$\frac{70 + 90}{2} = 80$$

You already know this.

weighted mean

To calculate a weighted mean, we multiple each grade by a weight and divide by the total weight.

$$\frac{(25)\,70 + (75)\,90}{25 + 75}$$

This could be because one assignment is worth 25 points and another is worth 75; or one class has 25 people and one has 75.

Notice that we could view simple means as weighted means with the weights equal to 1!

$$\frac{(1)70 + (1)90}{(1+1)}$$

Calculating the pooled variance

To see the logic of a weighted mean, we can compare it to a simple mean.

simple mean

Here is a simple mean of two variances.

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

Remember that s^2 means variance. The subscripts tell you these are different samples.

weighted mean

To calculate a weighted mean, we multiple each variance by a weight and divide by the total weight.

$$\frac{(\text{weight1})\, s_1^2 + (\text{weight2})\, s_2^2}{\text{weight1} + \text{weight2}}$$

Notice that we could view simple means as weighted means with the two weights equal to 1!

$$\frac{s_1^2 + s_2^2}{2} = \frac{(1)s_1^2 + (1)s_2^2}{(1+1)}$$

Calculating the pooled variance

So how do we choose the weight? The answer is that we want to weight by the number of degrees of freedom in each data set (after calculating the mean):

simple mean

Here is a simple mean of two variances.

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

Remember that s^2 means variance. The subscripts tell you these are different samples.

weighted mean

If we place the df for each data set in as the weights, we get the formula for pooled variance:

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

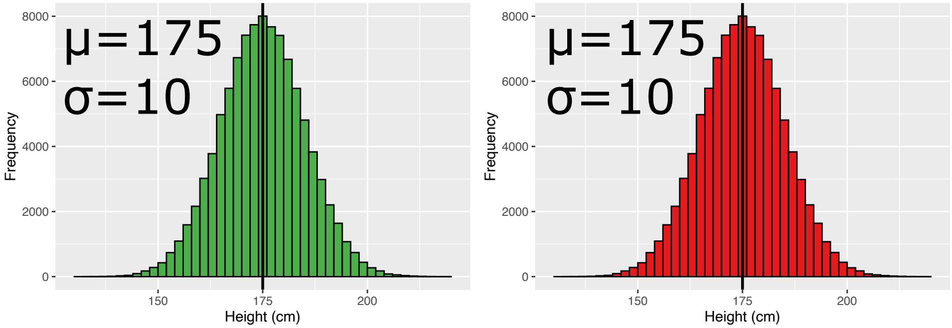
We can simplify the denominator if we want:

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

The full equations for both *t*-tests

Putting this together, we get:

equal variance

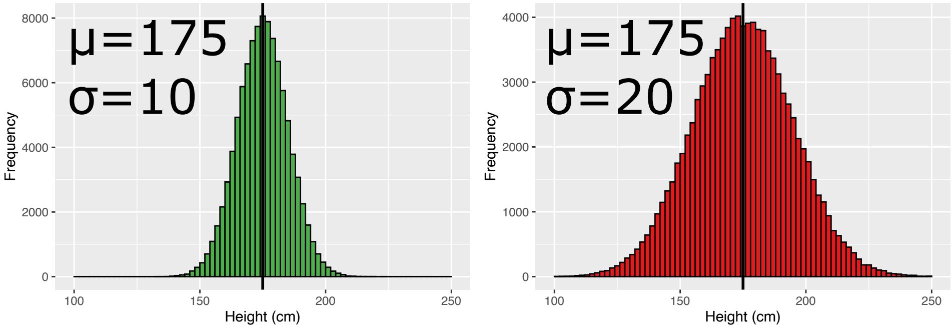


$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

Where the pooled variance is:

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

unequal variance



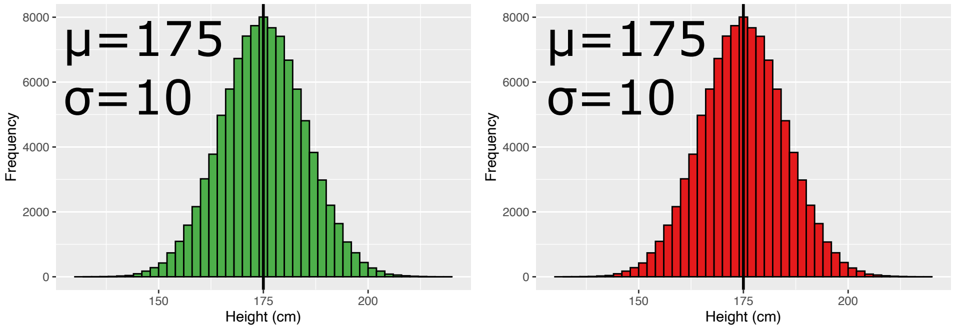
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

But again, this test is really only used with a computer (like with R), because the distribution is not quite identical to the standard *t* distribution so we can't print it in a book easily!

We will only use equal variances

To make our calculations simpler, in this course we will focus only on the equal variance independent samples *t*-test.

equal variance



$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

Where the pooled variance is:

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

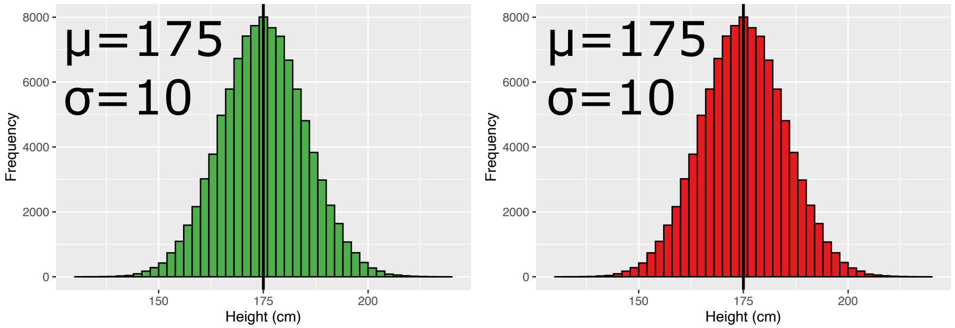
In this course whenever we have an independent samples *t*-test, we will always use the equal variances version.

This means that you will only have to work with pooled variances. This adds a calculation step (you have to calculate it). But it simplifies a number of things, including the shape of the *t*-distribution and the calculation of degrees of freedom for looking up the *t*-distribution.

The df for the *t*-distribution

Remember that the *t*-distribution is a family of distributions. You need to use the df of the test to look up the correct distribution.

equal variance



$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

Where the pooled variance is:

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

The df for the independent samples *t*-test is:

$$(n_1-1) + (n_2-1)$$

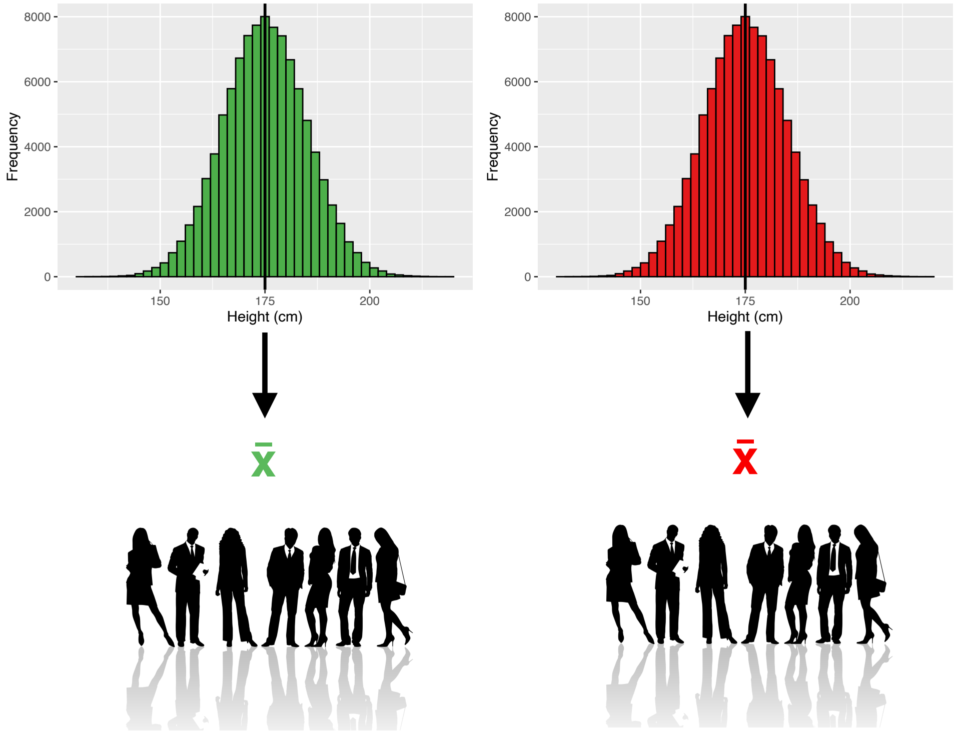
You will recognize this as the sum of the df of the two samples. This makes sense - it is also the df we use to calculate the pooled variance. You can simplify it to: $n_1 + n_2 - 2$

Calculating independent samples *t*-tests using the steps from the book

Step 1: State the hypotheses

The example from the book (section B) is that a researcher wants to test the effect of taking vitamin C on the number of days that people are sick.

The **null hypothesis** is that these two groups come from populations with the same means.



vitamin c

no vitamin c

We can translate H_0 into math, with two equivalent statements:

$$H_0: \mu_1 = \mu_2$$

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

And therefore our samples will have the same mean:

$$\bar{x}_1 - \bar{x}_2 = 0$$

So we know that the expected value of our test statistic will be 0:

$$(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2) = 0$$

Step 2: Select the statistical test and alpha criterion

Our design is two independent samples, with equal variance (because we will always assume that in this course), so the test is the equal variance independent samples *t*-test:

The *t* equation:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

Where the pooled variance is:

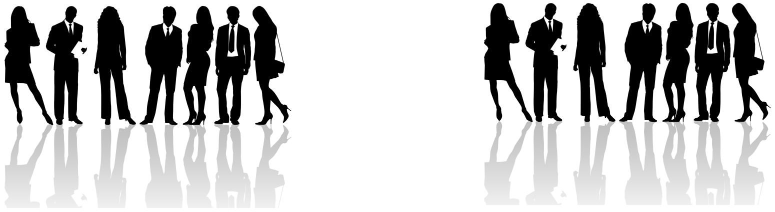
$$s_p^2 = \frac{(n_1-1)\,s_1^2 + (n_2-1)\,s_2^2}{n_1 + n_2 - 2}$$

So we will need to calculate the pooled variance first, and then we can calculate our test statistic.

We will use an alpha of .05 whenever we are running a single statistical test in psychology. It is the standard that has been agreed upon in the field. When we run multiple tests simultaneously in unit 3, we will choose different alphas that will still yield a Type I error rate of .05.

Step 3: Select sample sizes and collect the data

The details of the data are given in the example in the book. I have copied it out here. Notice that he tells us the standard deviation, not the variance. This is an easy way to make a mistake when doing your homework. Pay attention to the variables (they look alike!)



	vitamin c	no vitamin c
n	12	10
$\bar{x}$ (days sick)	4.25	7.75
s (standard deviation)	3	4

Step 4: Find the region of rejection

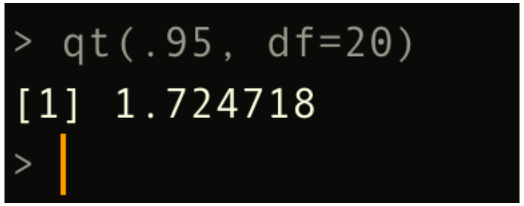
We know we will use a *t*-distribution. The *t*-distribution depends on the df. So we first need to calculate the df for this example:

$$df = (n_1-1) + (n_2-1)$$

$$df = (12-1) + (10-1) = 20$$

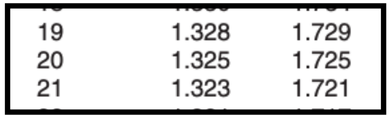
Next we have to decide if our experimental hypothesis is one-tailed or two-tailed. Our H_1 is that vitamin C will lead to fewer days sick, so it is **one-tailed**. This differs from the example in the book. He chooses two-tailed for reasons that I do not agree with. We can look up the critical *t* in R or in our book:

R:



```
> qt(.95, df=20)
[1] 1.724718
>
```

Table A2:



19	1.328	1.729
20	1.325	1.725
21	1.323	1.721

Both options yield 1.725, but we have to flip it to -1.725 because our hypothesis is that vitamin C will reduce the days sick!

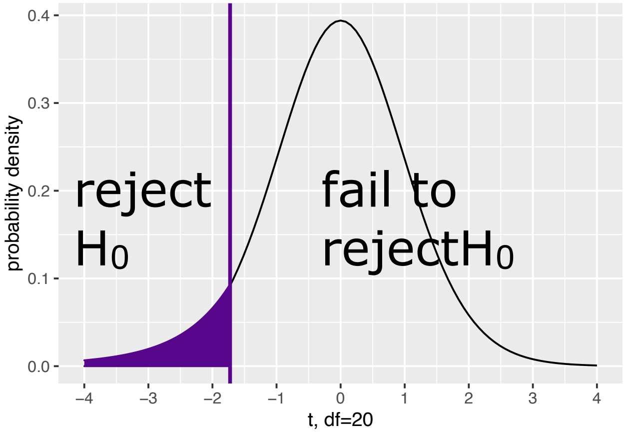
Step 4: Find the region of rejection

We know we will use a *t*-distribution. The *t*-distribution depends on the df. So we first need to calculate the df for this example:

$$df = (n_1-1) + (n_2-1)$$

$$df = (12-1) + (10-1) = 20$$

Next we have to decide if our experimental hypothesis is one-tailed or two-tailed. Our H_1 is that vitamin C will lead to fewer days sick, so it is **one-tailed**. This differs from the example in the book. He chooses two-tailed for reasons that I do not agree with. We can look up the critical *t* in R or in our book:



Both options yield 1.725, but we have to flip it to -1.725 because our hypothesis is that vitamin C will reduce the days sick!

Step 5: Calculate the test statistic

The *t* equation:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

Where the pooled variance is:

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

	vitamin C	no C
n	12	10
x̄	4.25	7.75
s	3	4

$$s_p^2 = \frac{(12-1)3^2 + (10-1)4^2}{12 + 10 - 2}$$

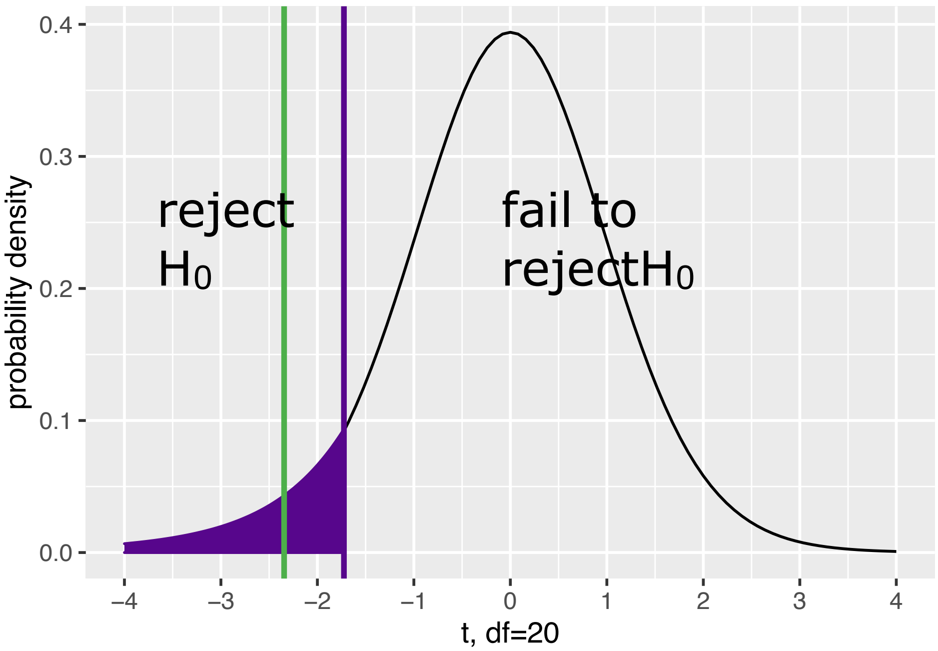
$$t = \frac{(4.25 - 7.75) - 0}{\sqrt{\frac{12.15}{12} + \frac{12.15}{10}}}$$

$$\boxed{t = -2.345}$$

Step 6: Make the statistical decision

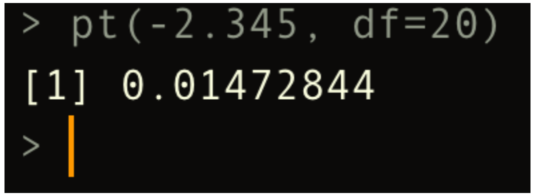
The critical *t* is -1.725 for our one-tailed hypothesis. This is the *t* that leads to *p*=.05. If we want *p*<.05, then we need a *t* that is less than this!

The *t* that we observed in our sample is -2.345. This yields *p*=.015.



Since our *t* is beyond the critical value (which means the *p*-value is less than .05), we decide to reject the null hypothesis.

When we report it, we will want to report the actual *p*-value. We can't get this from the book because it only lists critical *t* values. So we have to use R, and then we can write it up as follows:



```
> pt(-2.345, df=20)
[1] 0.01472844
> 
```

Vitamin C led to fewer days sick (M=4.25, SD=3) than placebo (M=7.75, SD=4). The difference is statistically significant by independent samples *t*-test with equal variances assumed: *t*(20) = -2.345, *p*=.015, one-tailed.

Wait, how did I decide which mean to put in $\bar{x}_1$ and which mean to put in $\bar{x}_2$?

The answer is based on your experimental logic

You always place your interesting condition, sometimes called your experimental condition, in the first slot, and you subtract the boring condition, called your control condition, from it.

$$t = \frac{(\overset{\text{interesting/experimental}}{\overset{\searrow}{\bar{x}_1}} - \overset{\text{boring/control}}{\overset{\swarrow}{\bar{x}_2}}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

The reason is that the control is the stuff you want to remove (=subtract) so you can see the effect in your experimental condition.

Notice that this means the sign of the subtraction tracks the direction of the effect

We can see this by testing different possibilities:

interesting/experimental boring/control

$(7 - 4) = 3$

A positive number means that the effect of the experimental condition is to increase the quantity.

interesting/experimental boring/control

$(4 - 7) = -3$

A negative number means that the effect of the experimental condition is to decrease the quantity.

The *t*-statistic also tracks the direction of the effect

The *t* equation:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

The data:

	vitamin C	no C
n	12	10
$\bar{x}$	4.25	7.75
s	3	4

If we put vitamin C first and no vitamin C second, we get this:

$$t = \frac{(4.25 - 7.75) - 0}{\sqrt{\frac{12.15}{12} + \frac{12.15}{10}}}$$

$$\boxed{t = -2.345}$$

In this example, *t* is negative because vitamin c **decreases** the number of days someone is sick. This is as it should be!

If the effect increased the quantity, *t* would be positive

The data:

	vitamin C	no C
n	12	10
$\bar{x}$	7.75	4.25
s	3	4

Here is hypothetical data with the opposite direction of the effect — vitamin C leads to more days sick.

We do the subtraction in the same way (interesting-control), and now the *t* is positive.

$$t = \frac{(7.75 - 4.25) - 0}{\sqrt{\frac{12.15}{12} + \frac{12.15}{10}}}$$

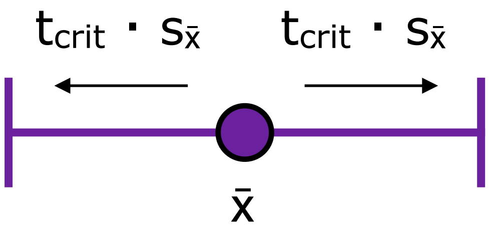
$$\boxed{t = 2.345}$$

In this example, *t* is positive because vitamin c **increases** the number of days someone is sick.

Confidence intervals for differences between two population means

Confidence intervals for differences between population means

Recall that we already know how to create confidence intervals around a single sample mean. We use that CI as are our estimate of the population mean:



$t_{crit} \cdot s_{\bar{x}}$ $t_{crit} \cdot s_{\bar{x}}$

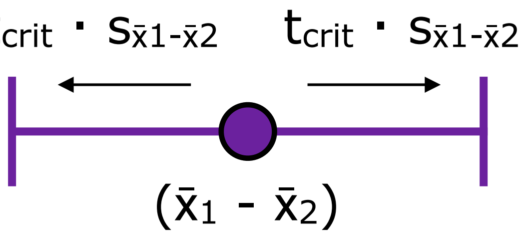
$\mu =$

$\bar{x}$

upper bound: $\bar{x} + (t_{crit} \cdot s_{\bar{x}})$

lower bound: $\bar{x} - (t_{crit} \cdot s_{\bar{x}})$

We can extend this idea to the difference between two population means:



$t_{crit} \cdot s_{\bar{x}1-\bar{x}2}$ $t_{crit} \cdot s_{\bar{x}1-\bar{x}2}$

$\mu_1 - \mu_2 =$

$(\bar{x}_1 - \bar{x}_2)$

upper bound: $(\bar{x}_1 - \bar{x}_2) + (t_{crit} \cdot s_{\bar{x}1-\bar{x}2})$

lower bound: $(\bar{x}_1 - \bar{x}_2) - (t_{crit} \cdot s_{\bar{x}1-\bar{x}2})$

The critical *t* and the estimate of the standard error come from our *t*-test. (But remember it is a two-tailed *t* whenever you create a CI!)

t_{crit} df=20, 95% = 2.09

$$s_{\bar{x}1-\bar{x}2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$