MAT220 Lecture 16: Hypothesis Testing Paired Data

Contents

In the previous lecture, we learned how to test claims about a single population parameter. We saw tests of two parameters: the population mean μ with either known or unknown σ, and the binomial success probability p. In this lecture and the next, we will discuss the prospect of testing whether or not *two* populations have the same mean. In fact, this is generally quite a complicated problem, so today we will assume for simplicity that our two populations are paired in some natural way. As we will see, this assumption can be leveraged to create an appropriate sampling distribution from which we can test whether or not two population means are the same.

1 A Review of Hypothesis Tests for the Population Mean

1.1 The workflow of a test

Generally speaking, the goal of a hypothesis test is to make a claim and then test the evidence for the claim. The baseline claim is called the *null hypothesis*, and is written H_0. We then test this statement against something to the contrary by formulating some kind of *alternate hypothesis*, written H_1. From there, we take a sample, and then answer the following question.

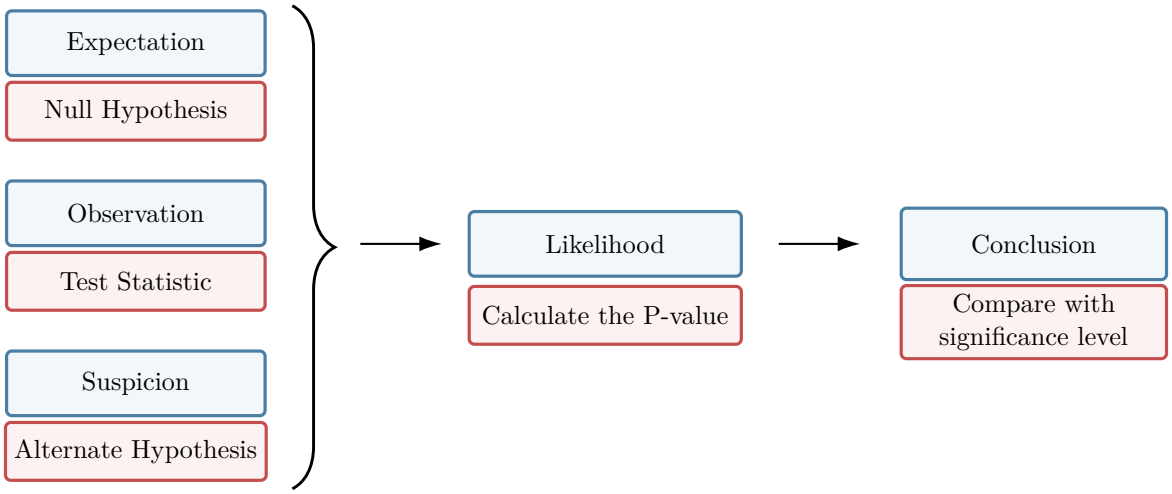
The main question of hypothesis tests

Every hypothesis test asks the same basic question:

Assuming H_0 is true, how unusual is the observed sample?

The probability calculation will generally change based on the setting of the problem.

Generally, hypothesis tests follow the same overall workflow:



Alternatively, the process of a hypothesis test can be broken down into four major steps:

Step 1: Get the information — Identify H_0, H_1, the sample data, and α.

Step 2: Calculate the P-value — Choose the correct test statistic and tail.

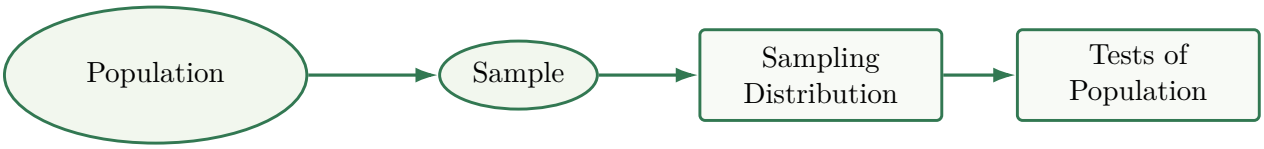
Step 3: Compare P-value to α — Reject H_0 if the P-value is smaller than α.

Step 4: Give an interpretation — Write the conclusion in the language of the original problem.

2 What are Paired Populations?

2.1 The motivation

So far, our statistical tests have mostly followed this pattern:

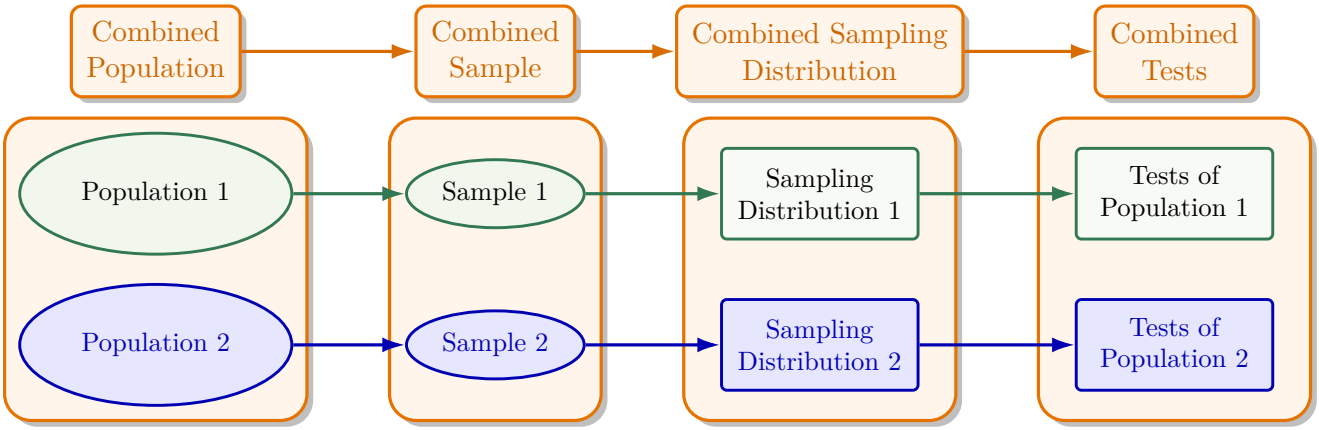


That is, we make a claim about some unknown population parameter of a given population, and then test the legitimacy of that claim by taking a random sample and evaluating the likelihood of our observation.

We will now explore the prospect of comparing two different populations via a hypothesis test. The main claim that we will try to test is:

Are these two populations the same, on average?

In symbols, if the first population has mean μ_1 and the second population has mean μ_2, we are going to try and test whether or not $\mu_1 = \mu_2$. In order to do so, we are going to combine two populations into a third population, take a sample, and then derive the relevant sampling distribution. Schematically, it will look something like this:



As a matter of fact, the act of combining two populations into a third is rather complicated. So, to make our lives easier, we will assume that our populations are naturally paired together in some meaningful way.

2.2 Defining Paired Populations

Throughout the remainder of this lecture, we will only consider populations that are *paired*. Heuristically speaking, populations are paired when there is a natural connection between points in one population and points in the other.

Paired populations

We say that two populations are **paired** when every observation in one data set has a unique counterpart in the other data set.

There are various situations in which we may want to study paired data. For example, we may find paired data in any of the following contexts:

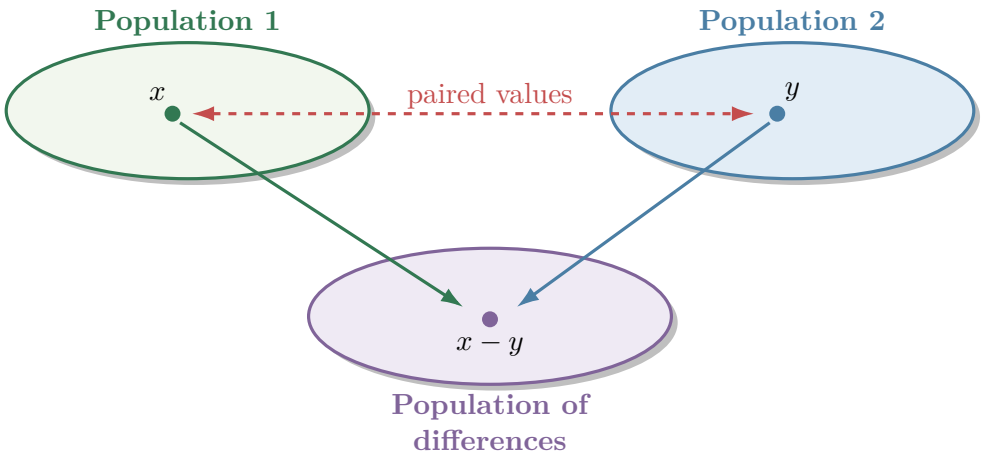
- studying the effects of an action on the same subjects, for instance the effects of caffeine on reaction time.
- studying the accuracy of multiple measuring devices, for instance the accuracy of two different heart rate monitors used on the same person.
- studying groups of naturally matched objects, for instance twins, or left/right feet on the same person.

2.3 Combining paired populations

Suppose that our two paired populations are defined by random variables X and Y. Because the populations are paired, any observation x from X has a unique partner y in Y. That means that we can consider the quantity $x - y$, and this will be some meaningful number that represents the difference between the value of X and the value of Y. As a shorthand, we will write:

$$D = X - Y.$$

We can interpret this as a new random variable – the difference of X and Y.[1] Schematically, we are doing something like the following.



3 The Sampling Distribution for Differences of Pairs

In order to perform hypothesis tests on the difference $X - Y$, we are going to describe the sampling distribution of $\bar{D}$, the random variable of sample means taken from the population D. At the

[1] Hence the letter "D" for "difference".

very least, we know from the Central Limit Theorem that, for sufficiently large n, $\bar{D}$ should be approximately normally distributed with parameters given by:

$$\bar{D} \approx \mathcal{N}\left(\mu_D, \frac{\sigma_D^2}{n}\right).$$

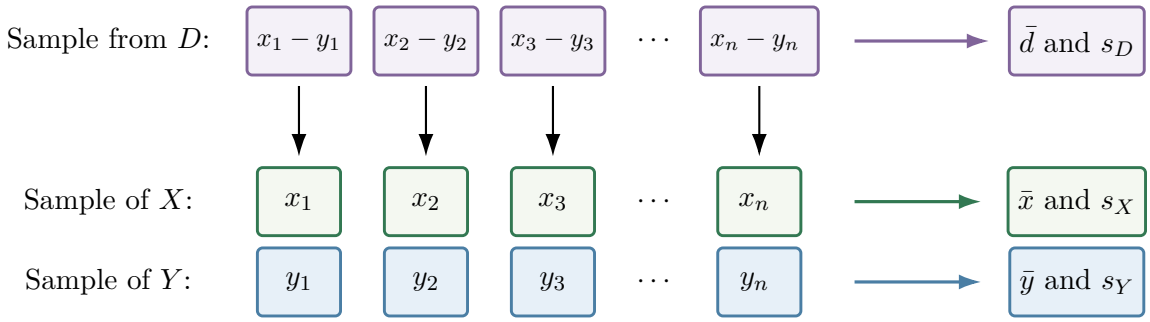
To understand this distribution, we need to understand both the mean μ_D and the variance σ_D^2 of the random variable $D = X - Y$.

3.1 Samples of paired data

Suppose we take a sample of n observations from D, say $d_1, \ldots, d_n$. Since each of the observations is a difference of the form $x - y$, taking n of these will look like:

$$d_1 = x_1 - y_1, \qquad d_2 = x_2 - y_2, \qquad \ldots, \qquad d_n = x_n - y_n.$$

These differences can be broken apart to recover samples of X and Y of size n:



It would be helpful if we could somehow describe the value of the sample statistics $\bar{d}$ and s_D in terms of the information coming from x_i and y_i. At the very least, this should give us clues about how averages and variances interact between X, Y and D.

3.2 The Mean of $D = X - Y$

If we wanted to compute the sample mean of our sample $d_1, \ldots, d_n$, we can use the fact that each $d_i = x_i - y_i$ and break apart the formula into two separate pieces:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) \overset{(*)}{=} \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{x} - \bar{y}.$$

Here, in the step labelled $(*)$, we have broken apart and rearranged the sum using the property:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) &= \frac{1}{n}\Big((x_1 - y_1) + (x_2 - y_2) + \ldots + (x_n - y_n)\Big) \\ &= \frac{1}{n}\Big(x_1 + x_2 + \ldots + x_n - y_1 - y_2 - \ldots - y_n\Big) \\ &= \frac{1}{n}\Big((x_1 + x_2 + \ldots + x_n) - (y_1 + y_2 + \ldots + y_n)\Big) \\ &= \frac{1}{n}(x_1 + x_2 + \ldots + x_n) - \frac{1}{n}(y_1 + y_2 + \ldots + y_n) = \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} y_i.\end{aligned}$$

Sample mean of the differences

For paired data,

$$\bar{d} = \bar{x} - \bar{y}.$$

In other words, the sample mean of the differences is the difference of the sample means.

As a matter of fact, the same phenomenon happens at the population level. Recall that the population mean of a random variable is really given by the *expected value*. In the discrete case, we saw that the formula for the expected value is the weighted average. In the case of a continuous random variable, the definition is a little bit complicated and involves some calculus that we have not seen before. What is important for us is the following property:

$$E(X - Y) = E(X) - E(Y),$$

which holds for all random variables with finite expected values, continuous or otherwise. We can use this property to describe the population mean μ_D in terms of μ_X and μ_Y:

$$\mu_D = E(D) = E(X - Y) = E(X) - E(Y) = \mu_X - \mu_Y.$$

As with the sample statistic $\bar{d}$, we see that the mean of the differences is the difference of the means.

Population mean of the differences

If $D = X - Y$, then

$$\mu_D = \mu_X - \mu_Y.$$

3.3 The Variance of $D = X - Y$

Unfortunately, the variance of a difference is not as well-behaved as the mean. According to the formula for the sample variance, we have:

$$s_D^2 = \frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n - 1}.$$

Now, we know from the previous section that $\bar{d} = \bar{x} - \bar{y}$. This means that we can rewrite the terms in the sum as:

$$(d_i - \bar{d})^2 = \Big((x_i - y_i) - (\bar{x} - \bar{y})\Big)^2 = \Big((x_i - \bar{x}) - (y_i - \bar{y})\Big)^2.$$

Unfortunately, when we expand out this bracket, we gain an extra cross-term:

$$\begin{aligned}(d_i - \bar{d})^2 &= \Big((x_i - \bar{x}) - (y_i - \bar{y})\Big)^2 \\ &= (x_i - \bar{x})^2 - 2(x_i - \bar{x})(y_i - \bar{y}) + (y_i - \bar{y})^2.\end{aligned}$$

This means that if we were to try to break down the formula for s_D^2 and rearrange it into smaller parts, we would obtain:

$$\begin{aligned}s_D^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\Big((x_i - \bar{x})^2 - 2(x_i - \bar{x})(y_i - \bar{y}) + (y_i - \bar{y})^2\Big) \\ &= \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}_{s_X^2} + \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}_{s_Y^2} - \frac{2}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).\end{aligned}$$

Although our algebra has uncovered the variances s_X^2 and s_Y^2, we also have an extra quantity. This extra piece is known as the *sample covariance*, and is often written either s_{XY} or $\operatorname{cov}(x, y)$.

The Sample Variance of the Differences

For paired data, the sample variance can be written

$$s_D^2 = s_X^2 + s_Y^2 - 2\operatorname{cov}(x, y).$$

Compute the differences first:

$$d_i = x_i - y_i,$$

and then compute the sample standard deviation s_D from the list of differences.

At the population level, the same issue appears. We have, in general,

$$\operatorname{Var}(X - Y) \neq \operatorname{Var}(X) - \operatorname{Var}(Y).$$

This can be verified using the definition of the variance Var of a random variable. Again using that the expected value E behaves well with sums and differences, we find that

$$\begin{aligned}\operatorname{Var}(X - Y) &= E\left[((X - Y) - E(X - Y))^2\right] \\ &= E\left[((X - Y) - (E(X) - E(Y)))^2\right] \\ &= E\left[((X - E(X)) - (Y - E(Y)))^2\right] \\ &= E\left[(X - E(X))^2\right] + E\left[(Y - E(Y))^2\right] - 2E\left[(X - E(X))(Y - E(Y))\right] \\ &= \operatorname{Var}(X) + \operatorname{Var}(Y) - 2\operatorname{Cov}(X, Y),\end{aligned}$$

which directly mimics the result for sample variances.

Roughly, the population covariance tells us how X and Y are linearly related to each other – we will explore this in depth during Lecture 18. For now, we summarise our discussion with regard to the population variance σ_D^2.

The Population Variance of the Differences

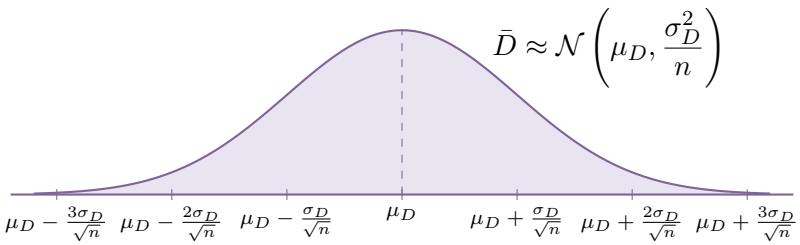
For paired random variables, the population variance can be written

$$\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2\,\mathrm{Cov}(X, Y).$$

In particular, it is generally not the case that $\sigma_D^2 = \sigma_X^2 + \sigma_Y^2$.

3.4 The sampling distribution of differences

If we take a random sample of n independent pairs and compute their differences, then the Central Limit Theorem tells us that the sampling distribution should be approximately normal:



Using the results of the previous section, we now see that if X has mean μ_X and variance σ_X^2 and Y has mean μ_Y and variance σ_Y^2, then the difference $D = X - Y$ has parameters $\mu_D = \mu_X - \mu_Y$ and $\sigma_D^2 = \sigma_X^2 + \sigma_Y^2 - 2\,\mathrm{Cov}(X, Y)$.

However, there is one problem: in practice, we almost never know the value of the extra covariance term. This means that, even if we happened to know the variances σ_X^2 and σ_Y^2, it is not enough information to calculate the value of σ_D^2. Therefore, we replace σ_D with the *sample* standard deviation s_D, and use a Student's t-distribution.

$$z = \frac{\bar{d} - \mu_D}{\sigma_D/\sqrt{n}} \qquad \xrightarrow{\text{replace } \sigma_D \text{ by } s_D} \qquad t = \frac{\bar{d} - \mu_D}{s_D/\sqrt{n}}$$

σ_D is unknown — Student's t, $df = n - 1$

When performing hypothesis tests, the latter becomes the relevant test statistic.

4 Hypothesis Testing Paired Data

4.1 The null hypothesis

Usually, when testing paired data, we are trying to answer:

Are these two populations the same, on average?

If μ_X is the mean of Population 1 and μ_Y is the mean of Population 2, mathematically we can rewrite this as:

$$\mu_X = \mu_Y \qquad \text{or equivalently,} \mu_X - \mu_Y = 0.$$

Since $\mu_D = \mu_X - \mu_Y$, the null hypothesis therefore becomes

$$H_0 : \mu_X - \mu_Y = 0.$$

Whenever we test whether two means are equal, this is always the correct form of the null hypothesis. The alternate hypothesis depends on the suspicion:

$$H_1 : \begin{cases} \mu_X - \mu_Y < 0, & \text{left-tailed test,} \\ \mu_X - \mu_Y > 0, & \text{right-tailed test,} \\ \mu_X - \mu_Y \neq 0, & \text{two-tailed test.} \end{cases}$$

In fact, selecting the correct alternate hypothesis is a challenge.

Choosing the correct alternate hypothesis

If you suspect that the average of Population 1 is *less* than the average of Population 2, then:

$$\mu_X < \mu_Y \qquad \Longrightarrow \qquad \mu_X - \mu_Y < 0$$

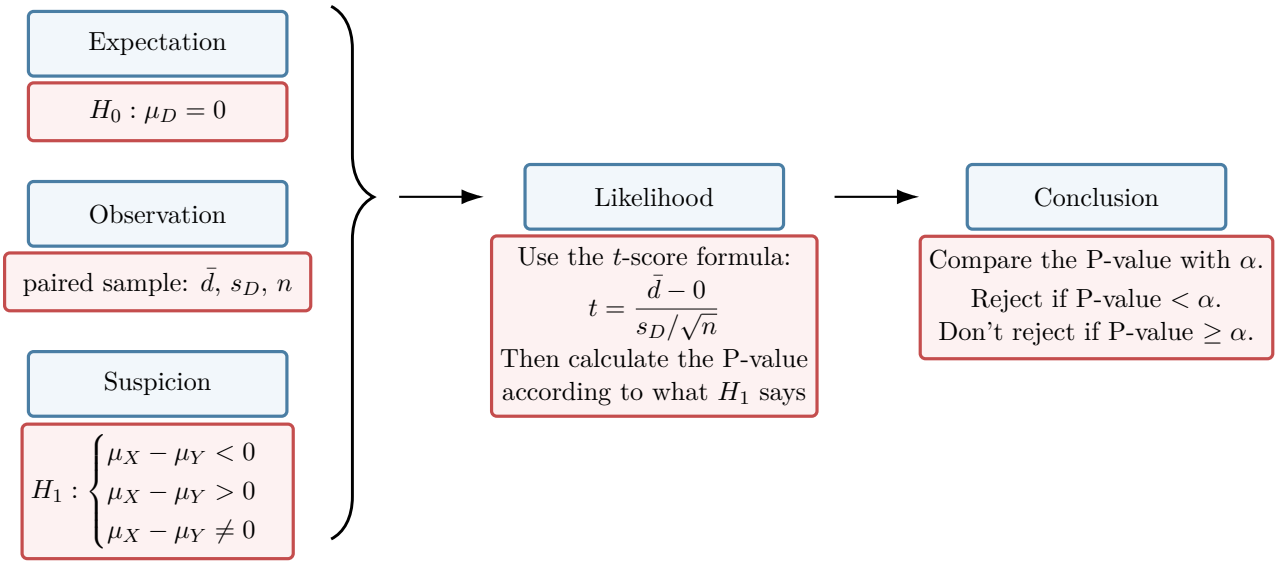
and the alternate hypothesis should read $H_1 : \mu_X - \mu_Y < 0$. Conversely, if you suspect that the average of Population 1 is *greater* than the average of Population 2, then:

$$\mu_X > \mu_Y \qquad \Longrightarrow \qquad \mu_X - \mu_Y > 0$$

and the alternate hypothesis should read: $H_1 : \mu_X - \mu_Y > 0$.

4.2 The paired-data workflow

According to our discussion in Section 3.4, we will generally not have access to the covariance of the two populations, and therefore we treat σ_D^2 as unknown. This means that, under the null hypothesis, the test statistic follows a Student's t-distribution instead of a normal distribution. Putting this all together, the total workflow for a hypothesis test for paired data is:



5 Example: Movie Ratings

Suppose that Dr. O'Connell wants to test whether the public think that *Twilight* and *The Twilight Saga: New Moon* are equally good. Let

$$\mu_X = \text{the average rating of Twilight out of 10, and}$$

$$\mu_Y = \text{the average rating of The Twilight Saga: New Moon out of 10.}$$

After watching both movies himself, he suspects that, on average, people think *Twilight* is probably better than *The Twilight Saga: New Moon*. To test this claim, he takes a simple random sample of 30 people and asks each person to rate both movies. For each person, he calculates the difference $d_i = x_i - y_i$ of their ratings, where x_i is the rating for *Twilight* and y_i is the rating for *The Twilight Saga: New Moon*. From the paired data, he finds

$$\bar{d} = 0.6 \qquad \text{and} \qquad s_D = 1.5.$$

We will perform a test with significance level $\alpha = 0.03$.

Step 1: Extracting the information

According to the setup of the test, the important values are:

$$\mu_{D,0} = 0, \qquad n = 30, \qquad \bar{d} = 0.6, \qquad s_D = 1.5, \qquad \alpha = 0.03.$$

The null hypothesis says that the two movies have the same average rating:

$$H_0 : \mu_X - \mu_Y = 0.$$

Since we defined $D = X - Y$, the suspicion that *Twilight* is rated higher than *The Twilight Saga: New Moon* becomes

$$H_1 : \mu_X - \mu_Y > 0.$$

So, this is a right-tailed test.

Step 2: Calculate the P-value

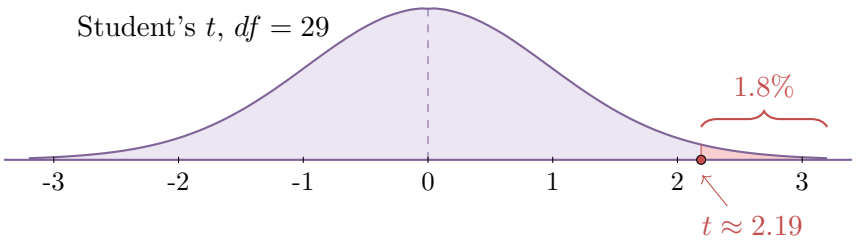
Since we don't know σ_D, we cannot use a z-score. Instead, we use the test statistic:

$$t = \frac{\bar{d} - 0}{s_D/\sqrt{n}} = \frac{0.6 - 0}{1.5/\sqrt{30}} \approx 2.19.$$

Since $n = 30$, the degrees of freedom are $df = n - 1 = 29$. Therefore, the P-value is

$$\text{P-value} = P(T > 2.19) \approx P(T > 2.2) = 1 - P(T < 2.2) \approx 1 - 0.98204 = 0.01796.$$

Our usual table only gives values to one decimal place, so in the above approximation, we have used $t = 2.2$ instead. Graphically, the P-value is the shaded area below:



Put differently, if the two movies truly had equal average ratings, then there would be about a 1.8% chance of getting a t-statistic this large (or larger) from a random sample of 30 people.

Step 3: Compare the P-value to α

In this example, we are asked to perform the test with a significance level of $\alpha = 0.03$. Our P-value is approximately 0.018, which is less than 0.03. Therefore, we reject the null hypothesis.

Step 4: Give an interpretation

Interpretation

At the 3% significance level, the sample provides convincing evidence to suggest that the public rate *Twilight* higher than *The Twilight Saga: New Moon*, on average.

Here, we do not make any separate claims about the average rating of either film. Instead, we tested the difference between their average ratings.