Test statistics for one- and two-tailed tests

Remember that our steps for any hypothesis test are

1. State the null and alternative hypotheses.

2. Determine the level of significance.

3. Calculate the test statistic.

4. Find critical value(s) and determine the regions of acceptance and rejection.

5. State the conclusion.

We've already covered the first two steps, and now we want to talk about how to calculate the test statistic. But any test statistic we calculate will depend on whether we're running a two-tailed test or a one-tailed test. Whether we run a one- or two-tailed test is dictated by the hypothesis statements we wrote in the first step.

So let's define one- and two-tailed tests, and start over with the hypothesis statements to show when we'll use each test type.

One- and two-tailed tests

We've already learned that, for both means and proportions, we can write the null and alternative hypothesis statements in three ways:

H_0 with an $=$ sign and H_a with a $\neq$ sign

H_0 with a	$\leq$	sign and H_a with a	$>$	sign
H_0 with a	$\geq$	sign and H_a with a	$<$	sign

When the null and alternative hypotheses use the $=$ and $\neq$ signs, we'll use a **two-tailed test** (also called a two-sided or nondirectional test). But in the other two cases, with either $\leq$ and $>$, or $\geq$ and $<$, we'll use a **one-tailed test** (also called a one-sided test or direction test).

Here's the way to understand one- and two-tailed tests. When we state in the null hypothesis that the population mean or population proportion is equal to some value, and then state with the alternative hypothesis that the mean or proportion is not equal to that value, we're not predicting any direction between the variables. We're not saying that one value is greater or less than the other, we're just saying that they're different. Which means the difference could be in either direction (less than or greater than), which means we'll have a region of rejection in each tail of the distribution.

We call it a "two-tailed test" because we have two regions of rejection, one in each tail.

On the other hand, if we hypothesize that the mean or proportion is greater than or less than the stated value, then it means we'll be performing a one-tailed test. If we predict that the population parameter is greater than the stated value, then the only rejection region will be in the upper tail, so we call this an **upper-tailed test**, or a **right-tailed test**.

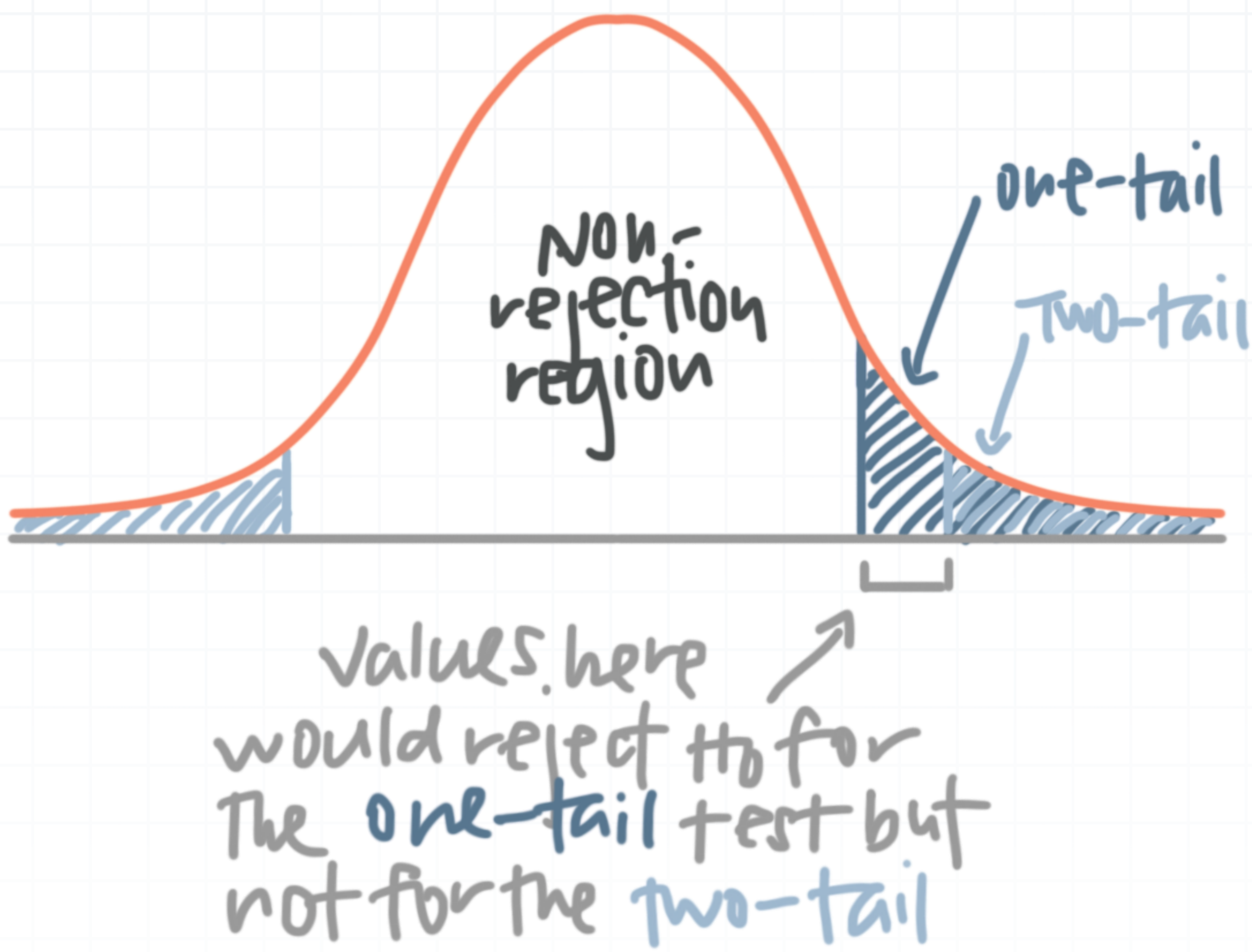
But if we predict that the population parameter is less than the stated value, then the only rejection region will be in the lower tail, so we call this a **lower-tailed test**, or a **left-tailed test**.

Choosing a one-tailed or two-tailed test

Whether we use a one- or two-tailed test is determined by the hypothesis statements we choose. With that in mind, we really want to think ahead when we're writing our hypothesis statements, and consider which kind of test we want to set ourselves up for.

A one-tailed test has a larger region of rejection, because all of the area that represents the region of rejection is consolidated into one tail. A two-tailed test, on the other hand, has the region of rejection split into two tails, which means each individual rejection region for the two-tailed test is smaller than the single rejection region from the one-tailed test.

Inherently, this means that a two-tailed test is always more conservative than a one-tailed test. Looking at this last figure, we could get a whole range of results that fall within the rejection region of the one-tailed test, but fail to reach the rejection region of the two-tailed test.

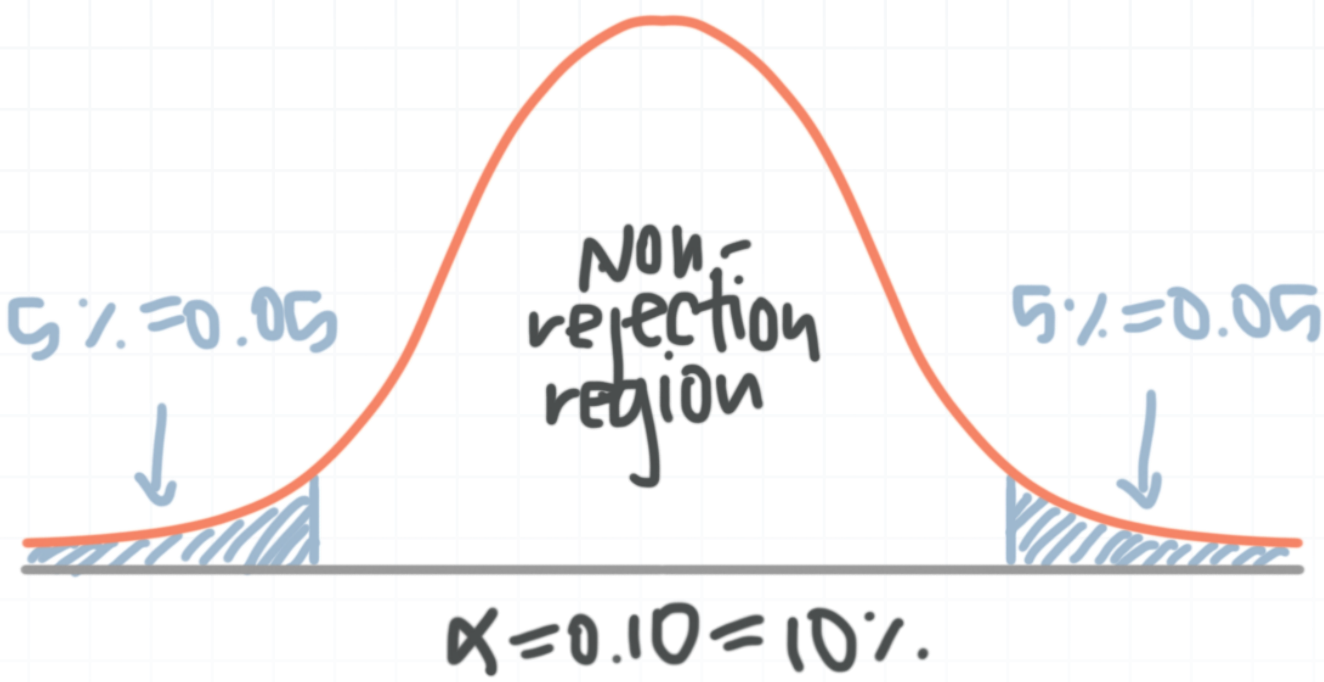


The two-tailed test really forces us to find a more extreme result in order to reach the region of rejection and reject the null hypothesis.

That being said, we should only use a one-tailed test when we have good reason to believe that the difference between the means or proportions is in the specific direction that we think it's in. If we're not extremely confident about directionality, we should play it safe and use the more conservative two-tailed test.

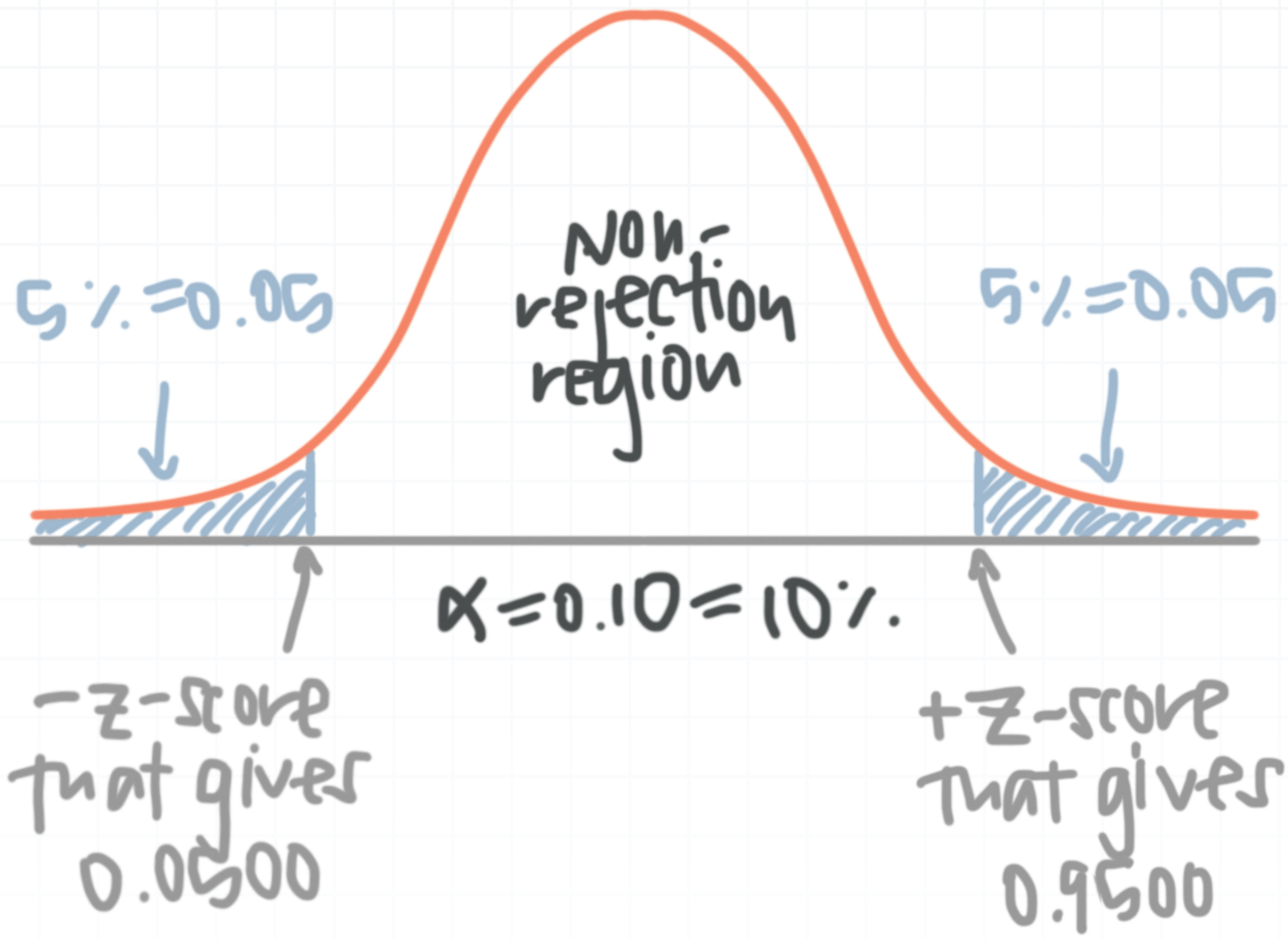
The α value for one- and two-tailed tests

Let's say we're running two different tests. One is two-tailed, and the other is one-tailed. If we set the significance level at $\alpha = 0.10$, then in the two-tailed test we'll split that 10% evenly into two tails, 5% in the lower tail and 5% in the upper tail.



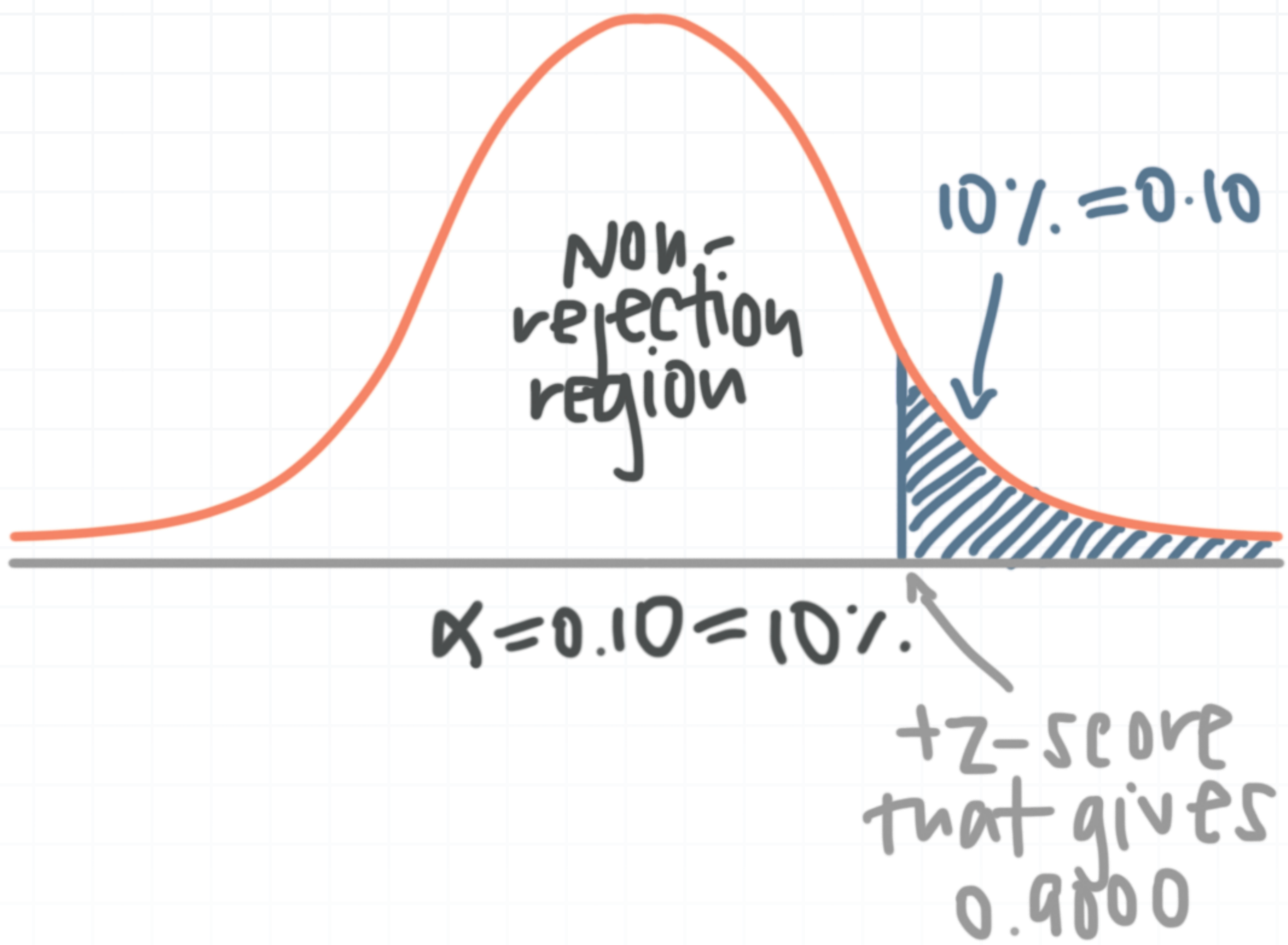
Here's something that's a little tricky: When we go to look up the z-score that corresponds to the boundary of the upper rejection region, realize that only 5% of the rejection area is in that upper tail, not 10%. Which means that, when we look for the z-score, we need to look for a z-score that corresponds to 0.9500, not 0.9000, even though $\alpha = 0.10$.

And the same goes for the lower tail. When we look for the negative z-score that corresponds to the boundary of the rejection region in the lower tail, we need to look for a z-score that corresponds to 0.0500, not 0.1000, even though $\alpha = 0.10$.



Many people get tripped up here, because they assume that, with $\alpha = 0.10$, they're looking for the z-score associated with 0.9000. But because the rejection region is split into both tails, we only have $10\,\%/2 = 5\,\%$ of the area in each tail, and that $5\,\%$ corresponds to 0.9500 in the upper tail and 0.0500 in the lower tail.

But if instead we're using a one-tailed test with the same $\alpha = 0.10$, the entire $10\,\%$ of area that represents the rejection region is consolidated into one tail of the distribution. So if we're running an upper-tailed test, then we'll find the boundary of the rejection region at a z-score for 0.9000.



For a lower-tailed test, we'll find the boundary of the rejection region at a z-score for 0.1000.

Calculating the test statistic

At this point, we know how to write the hypothesis statements, determine the alpha level we want to use, and set up a one- or two-tailed test based on the hypothesis statements. We can also find the boundary or boundaries of the rejection region based on the alpha value and which test type we're using.

The next step is always to calculate the test statistic, and then determine whether that value lies inside or outside the region of rejection.

The formula we'll use to calculate the test statistic depends on the information we have and the size of our sample, but the general formula for the test statistic is

$$\text{test statistic} = \frac{\text{observed} - \text{expected}}{\text{standard deviation}}$$

The specific formula for the test statistic will depend on whether or not the population standard deviation σ is known or unknown. When σ is known, we'll use

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

When σ is unknown, and/or if we're using a small sample $n < 30$, we'll use the sample standard deviation $s_{\bar{x}}$ instead of the population standard deviation σ, but we'll use the more conservative t-table to compensate for using $s_{\bar{x}}$.

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

These are test statistics for the mean, but the test statistic for the proportion, as long as $n\hat{p} \geq 5$ and $n(1 - \hat{p}) \geq 5$, will be

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

Let's do an example where we calculate the test statistic for the mean, where population standard deviation is unknown (which will usually be the case).

Example

We're testing the claim that a car dealership averages $\$1,000,000$ in monthly sales. We take a random sample of 40 months of sales data and find $\bar{x}=\$985,000$ and $s=\$200,000$. Find the hypothesis statements, choose a one- or two-tailed test, and calculate a test statistic.

There's nothing in the problem to indicate we're very confident about directionality, so we should choose a two-tailed test, and the hypothesis statements will therefore be

H_0: The dealership sells $\$1,000,000$

$$\mu=\$1,000,000$$

H_a: The dealership makes sales other than $\$1,000,000$

$$\mu\neq\$1,000,000$$

Because population standard deviation is unknown, the test statistic will be

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

$$t = \frac{\$985{,}000 - \$1{,}000{,}000}{\frac{\$200{,}000}{\sqrt{40}}}$$

$$t = -\$15{,}000 \cdot \frac{\sqrt{40}}{\$200{,}000}$$

$$t = -\frac{3\sqrt{40}}{40}$$

$$t = -\frac{3\sqrt{10}}{20}$$

$$t \approx -0.4743$$

In the next section, we'll look at what to do with this test statistic once we have it.