

# The $p$ -value and rejecting the null

Technically, we define the  **$p$ -value** (or the observed level of significance) as the smallest level of significance at which we can reject the null hypothesis, assuming the null hypothesis is true.

We can also think about the  $p$ -value as the total area of the region of rejection. Remember that in a one-tailed test, the region of rejection is consolidated into one tail, whereas in a two-tailed test, the region of rejection is split between two tails.

So, as we might expect, calculating the  $p$ -value as the area of the rejection region will be slightly different depending on whether we're using a one-tailed test or a two-tailed test, and whether the one-tailed test is an upper-tailed test or lower-tailed test.

## Calculating the $p$ -value

### For a one-tailed, lower-tailed test

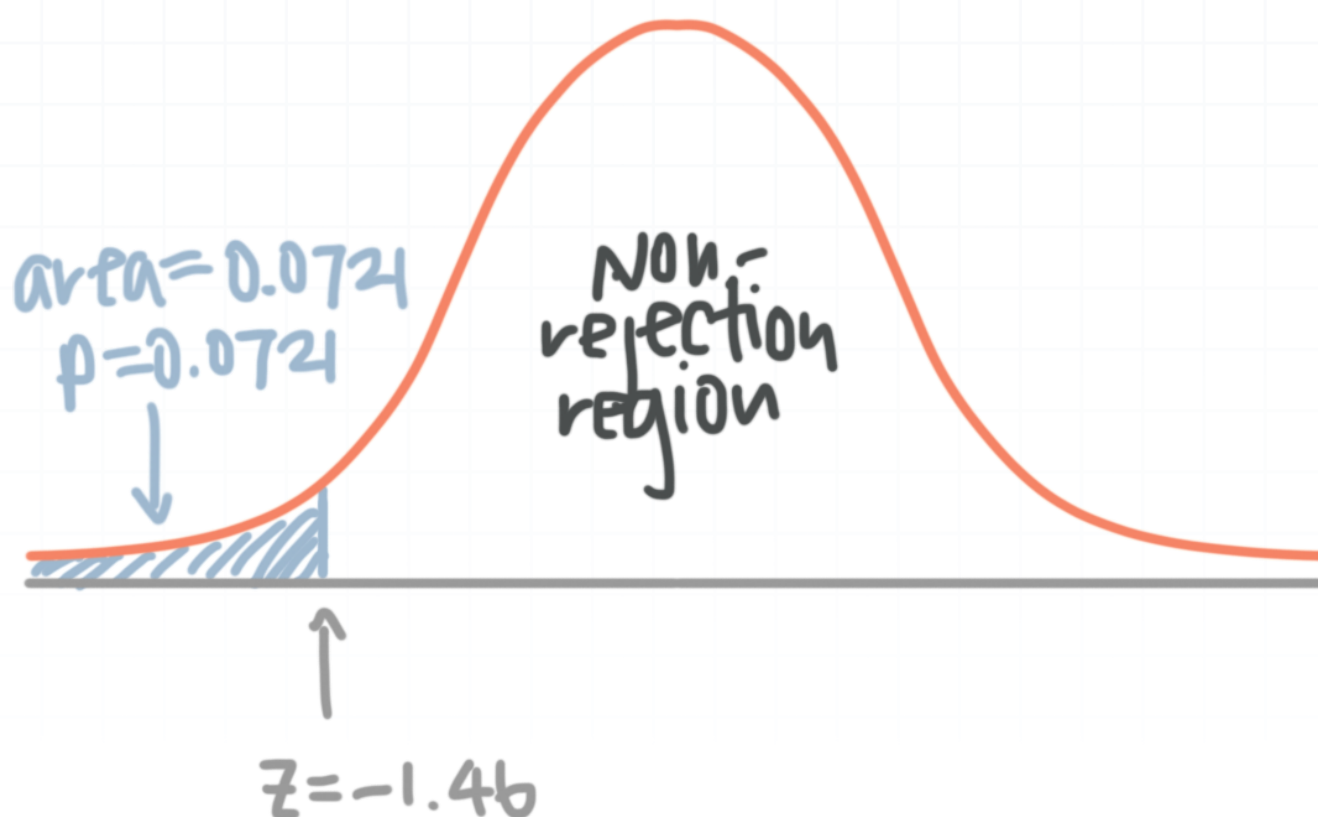
For a one-tailed test, we'll first calculate the  $z$ -test statistic. For a lower-tailed test,  $z$  will be negative. Then the value we find from the negative  $z$ -table represents the area under the probability distribution to the left of the negative  $z$ -value.

For instance, let's assume we calculated a test statistic of  $z = -1.46$ . In a  $z$ -table, we'd find



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	<b>.0721</b>	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823

So 0.0721 is the area under the curve to the left of  $z = -1.46$ , and this is the  $p$ -value also. So  $p = 0.0721$ .



### For a one-tailed, upper-tailed test

For a one-tailed test, we'll first calculate the  $z$ -test statistic. For an upper-tailed test,  $z$  will be positive. Then the value we find from the positive  $z$ -table represents the area under the probability distribution to the left of the positive  $z$ -value.

For instance, let's assume we calculated a test statistic of  $z = 1.46$ . In a  $z$ -table, we'd find

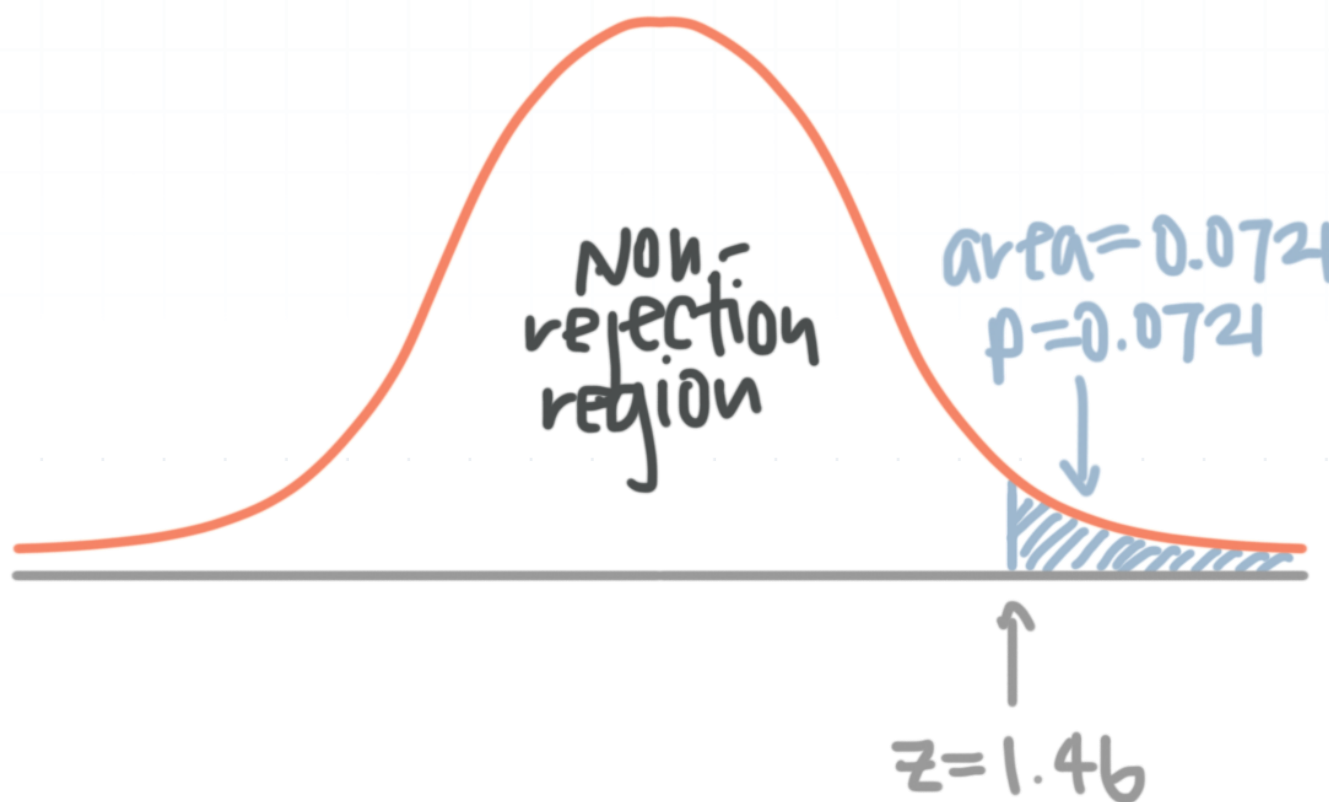


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441

But in an upper-tailed test, we're interested in the area to the right of the  $z$ -value, not the area to the left. To find the area to the right, we need to subtract the value in the  $z$ -table from 1.

$$1 - 0.9279 = 0.0721$$

So 0.0721 is the area under the curve to the right of  $z = 1.46$ , and this is the  $p$ -value also. So  $p = 0.0721$ .



### For a two-tailed test

For a two-tailed test, we'll first calculate the  $z$ -test statistic. For a two-tailed test,  $z$  could be either positive or negative. Then the value we find from the  $z$ -table represents the area under the probability distribution to the left of the  $z$ -value.



For instance, assume we found  $z = 1.23$ . In a  $z$ -table, we'd find

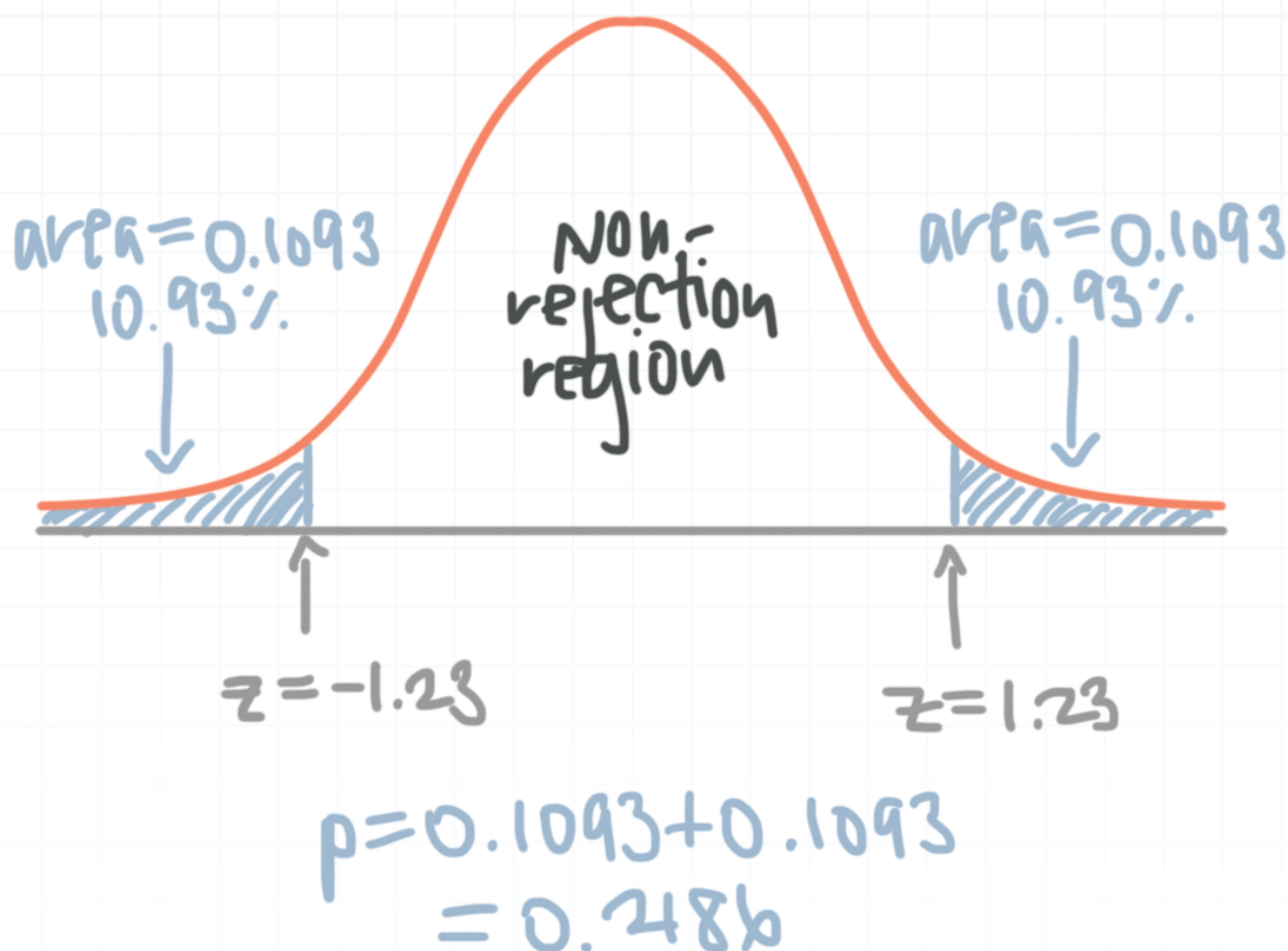
$z$	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	<b>.8907</b>	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177

But for a positive  $z$ -value, we're interested in the area to the right of the  $z$ -value, not the area to the left. To find the area to the right, we need to subtract the value in the  $z$ -table from 1.

$$1 - 0.8907 = 0.1093$$

So 0.1093 is the area under the curve to the right of  $z = 1.23$ . Because this is a two-tailed test, the region of rejection is not only the 10.93 % of area in the upper tail, but also the symmetrical 10.93 % of area in the lower tail. So we'll double 0.1093 to get  $2(0.1093) = 0.2186$ , and this is the  $p$ -value also. So  $p = 0.2186$ .





## Rejecting the null using the $p$ -value

The reason we've gone through all this work to understand the  $p$ -value is because using a  $p$ -value is a really quick way to decide whether or not to reject the null hypothesis.

Whether or not we should reject the null hypothesis  $H_0$  can be determined by the relationship between the  $\alpha$  level and the  $p$ -value.

If  $p \leq \alpha$ , reject the null hypothesis

If  $p > \alpha$ , do not reject the null hypothesis

In our earlier examples, we found

$p = 0.0721$  for the lower-tailed one-tailed test



$p = 0.0721$  for the upper-tailed one-tailed test

$p = 0.2186$  for the two-tailed test

With these in mind, let's say for instance we set the confidence level of our hypothesis test at 90 %, which is the same as setting the  $\alpha$  level at  $\alpha = 0.10$ . In that case,

$$p = 0.0721 \leq \alpha = 0.10$$

$$p = 0.2186 > \alpha = 0.10$$

So we would have rejected the null hypothesis for both one-tailed tests, but we would have failed to reject the null hypothesis in the two-tailed test. If, however, we'd picked a more rigorous  $\alpha = 0.05$  or  $\alpha = 0.01$ , we would have failed to reject the null hypothesis every time. So to summarize the  $p$ -value approach when  $\sigma$  is known,

Lower-tailed test

Reject  $H_0$  when  $p \leq \alpha$

Upper-tailed test

Reject  $H_0$  when  $p \leq \alpha$

Two-tailed test

Reject  $H_0$  when  $p \leq \alpha$

Lastly, remember that this entire  $p$ -value approach applies in exactly the same way, whether we're using a  $z$ -test statistic or a  $t$ -test statistic. So when we our test statistic is a  $t$ -value, we we look up the  $t$ -value in the  $t$ -table. For a lower-tailed test, the value we find becomes the  $p$ -value; for an upper-tailed test, we subtract the value we find from 1 to get the  $p$ -value; and for a two-tailed test, we double the value we find to get the  $p$ -value. Then we use the same rules as the  $z$ -test statistic to compare  $p$  to  $\alpha$



to determine whether or not to reject the null hypothesis. So to summarize the  $p$ -value approach when  $\sigma$  is unknown and/or when we have a small sample,

Lower-tailed test	Reject $H_0$ when $p \leq \alpha$
Upper-tailed test	Reject $H_0$ when $p \leq \alpha$
Two-tailed test	Reject $H_0$ when $p \leq \alpha$

## Rejecting the null using the critical value

When deciding whether or not to reject the null hypothesis, we can also determine **critical values**, instead of taking the  $p$ -value approach. Critical values are like cut-off values that bound the rejection region(s). The critical values depend on the test statistic itself,  $z$  or  $t$ , and the level of significance  $\alpha$ .

For example, if we perform an upper-tailed  $z$ -test and the level of significance is  $\alpha = 0.05$ , then we can first look up the probability that's closest to 0.95 in the  $z$ -table (since the test is upper-tailed), and then find the corresponding critical  $z$ -value from the table that's equivalent to 1.96. The region of rejection is to the right of 1.96, so if the test statistic is larger than 1.96, we need to reject the null hypothesis. Otherwise, we accept the null.

So to summarize the critical value approach,

Lower-tailed test	Reject $H_0$ when $z \leq -z_\alpha$
-------------------	--------------------------------------





Upper-tailed test

Reject  $H_0$  when  $z \geq z_\alpha$ 

Two-tailed test

Reject  $H_0$  when  $z \leq -z_{\frac{\alpha}{2}}$  or  $z \geq z_{\frac{\alpha}{2}}$ 

## Significance

The **significance** (or **statistical significance**) of a test is the probability of obtaining the result by chance. The less likely it is that we obtained a result by chance, the more significant our results.

Hopefully by now it's not too surprising that all of these are equivalent statements:

- The finding is significant at the 0.01 level
- The confidence level is 99 %
- The Type I error rate is 0.01
- The alpha level is 0.01,  $\alpha = 0.01$
- The area of the rejection region is 0.01
- The  $p$ -value is 0.01,  $p = 0.01$
- There's a 1 in 100 chance of getting a result as extreme, or more extreme, as/than this one

The smaller the  $p$ -value, or the smaller the alpha value, or the lower the Type I error rate, and the smaller the region of rejection, the higher the confidence level, and the less likely it is that we got our result by chance.





So, to take a different example, an alpha level of 0.10 (or a  $p$ -value of 0.10, or a confidence level of 90 %) is a pretty low bar to clear. At that significance level, there's a 1 in 10 chance that the result we got was just by chance. And therefore there's a 1 in 10 chance that we'll reject the null hypothesis when we really shouldn't have, thinking that we provided support for the alternative hypothesis when we shouldn't have.

But a stricter alpha level of 0.01 (or a  $p$ -value of 0.01, or a confidence level of 99 %) is a higher bar to clear. At that significance level, there's only a 1 in 100 chance that the result we got was just by chance. And therefore there's only a 1 in 100 chance that we'll reject the null hypothesis when we shouldn't, thinking that we provided support for the alternative hypothesis when we shouldn't have.

If we find a result that clears the bar we've set for ourselves, then we reject the null hypothesis and we say that the finding is significant at the  $p$ -value that we find. Otherwise, we fail to reject the null.

