

Introduction to Statistics - Young Researchers Fellowship Program

Lecture 5 - Hypothesis Testing - One Sample Tests

Daniel Sánchez Pazmiño

Table of contents

Introduction to hypothesis testing	1
Null and alternative hypotheses	2
Case 1: A left-tailed test	2
Case 2: A right-tailed test	2
Case 3: A two-tailed test	3
Errors	3
The z-test (one sample)	4
Left-tailed test	5
Right-tailed test	7
Two-tailed test	8
Summary	10

Introduction to hypothesis testing

In the previous sessions, it was explained how we can use a sample to make inference (guesses) about the values of the population. We can expand our understanding of the population by using the sample to make statements about the specific value of a population parameter.

A confidence interval tells you whether or not the population parameter is likely to be in the range of values. A hypothesis test is a statistical test that is used to determine whether there is enough evidence in a sample of data to infer that a certain condition is true for the entire population.

Null and alternative hypotheses

The hypothesis test is a process that leads to the rejection or acceptance of a particular statement about the population, which is called the null hypothesis. This null hypothesis is some kind of assumption about the population parameter, which we often consider to be the baseline scenario or the “safe” option.

The alternative hypothesis is the opposite of the null hypothesis. Basically, if the null hypothesis assumes that the population parameter is equal to a certain value, the alternative hypothesis assumes that the population parameter is not equal to that value.

There are three main cases where we define the null and alternative hypotheses in a specific way.

Case 1: A left-tailed test

In a hypothesis test which is called the “left-tailed test”, the null hypothesis is that the population parameter is equal to a certain value, and the alternative hypothesis is that the population parameter is, in truth, less than that value.

An example considers the average GPA of students in a university. The null hypothesis is that the average GPA is greater or equal to 3.0, and the alternative hypothesis is that the average GPA is less than 3.0. This numeric value that we’re trying to compare is called the “null value”, and it is denoted as μ_0 . Remembering that the unknown population parameter is denoted as μ , we can write the null and alternative hypotheses as follows:

$$\begin{aligned}H_0 : \mu &\geq \mu_0 \\H_1 : \mu &< \mu_0\end{aligned}$$

Case 2: A right-tailed test

In a hypothesis test which is called the “right-tailed test”, the null hypothesis is that the population parameter is equal to a certain value, and the alternative hypothesis is that the population parameter is, in truth, greater than that value.

An example considers the average GPA of students in a university. The null hypothesis is that the average GPA is less or equal to 3.0, and the alternative hypothesis is that the average GPA is greater than 3.0. This numeric value that we’re trying to compare is called the “null value”, and it is denoted as μ_0 . Remembering that the unknown population parameter is denoted as μ , we can write the null and alternative hypotheses as follows:

$$H_0 : \mu \leq \mu_0$$

$$H_1 : \mu > \mu_0$$

Case 3: A two-tailed test

In a hypothesis test which is called the “two-tailed test”, the null hypothesis is that the population parameter is equal to a certain value, and the alternative hypothesis is that the population parameter is, in truth, different than that value.

An example considers the average GPA of students in a university. The null hypothesis is that the average GPA is equal to 3.0, and the alternative hypothesis is that the average GPA is different than 3.0. This numeric value that we’re trying to compare is called the “null value”, and it is denoted as μ_0 . Remembering that the unknown population parameter is denoted as μ , we can write the null and alternative hypotheses as follows:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

This is a very popular case, as it is a clear example of a situation where we want to test whether a population parameter is equal to a certain value.

Errors

It is important to always keep in mind that the null hypothesis implies that the alternative hypothesis is false, and the alternative hypothesis implies that the null is false. We then use sample data to determine whether we can reject the null hypothesis or not.

However, given that we know that the sampling distribution of the sample mean allows for us getting a “extreme” value of the sample mean which might lead us to reject the null hypothesis, even though it is true, we need to be careful when making decisions. The table below shows how we can make mistakes when making decisions about the null hypothesis.

Decision	H_0 is true	H_0 is false
Reject H_0	Type I error	Correct decision
Reject H_0	Correct decision	Type II error

We know that there is a chance that we reject the null hypothesis even though it is true. This is called a *Type I error*, and it is often denoted as α . As an example, consider the GPA example. We might make the following hypothesis test:

$$H_0 : \mu \geq 3.0$$

$$H_1 : \mu < 3.0$$

Suppose you get a sample and acquire a sample mean of 2.8. You might be tempted to reject the null hypothesis, but you know that there is a chance that the null hypothesis is true and that the 2.8 sample mean is just an extreme value of the sample mean in its sample distribution. If you were to actually reject the null hypothesis, you would be making a Type I error.

We can define the *probability* of making a Type I error as α , which is often called the *significance level*. This is the probability of rejecting the null hypothesis when it is true. As you know, because the sampling distribution of the sample mean allows us to use the rules of statistics and probability to define the probability of getting less or equal to a value, we can define the probability of making a Type I error and then find the number of standard errors in the probability distribution that correspond to that probability.

The key idea of hypothesis testing is defining my level of “tolerance” to making a type I error, and then finding the number of standard errors that correspond to that probability. If the sample mean is further away from that number of standard errors, it means that the likelihood of making a type 1 error is smaller than the likelihood of the null being actually false.

A type II error is the opposite of a type I error. It is the probability of not rejecting the null hypothesis when it is false. This is often denoted as β , and named the *type II error rate* or the power of the test. We normally do not explicitly define the type II error rate, but we can use it to find the power of the test. The power of the test is the probability of rejecting the null hypothesis when it is false.

Because we normally do not control for the possibility of making a type II error, we make our wording of the rejection of the null hypothesis in a way that we do not explicitly mention the type II error rate. For example, if we find evidence against the existence of the alternative hypothesis, we say that we “fail to reject the null hypothesis”. This is because we do not want to say that we “accept the null hypothesis”, as we know that there is a chance that the null hypothesis is false and that we are making a type II error.

Knowing the concept of the hypothesis test, we can now review the types of hypothesis tests that we can perform and see how to use the sampling distribution of the sample mean to perform hypothesis tests.

The z-test (one sample)

The z-test is a hypothesis test that is used to test whether the population mean is equal to a certain value. It is a hypothesis test that is used when we know the population standard deviation, just like the case of the confidence interval when we know the population standard

deviation. Though this is not a very common case, it is important to know how to perform this test as it is possible one can know the population standard deviation through historical data. Further, knowing the z-test will help us understand the t-test, which is a hypothesis test that is used when we do not know the population standard deviation, and is one of the most important tests in statistics.

There is an additional caveat to this test: it assumes that we either have a population distributed normally, or that we have a large sample size. This is because the sampling distribution of the sample mean is normally distributed when the population is normally distributed, or when the sample size is large.

Within the z-test, we have three cases: a left-tailed test, a right-tailed test, and a two-tailed test.

Left-tailed test

In a left-tailed test, the null hypothesis is that the population mean is greater or equal to a certain value, and the alternative hypothesis is that the population mean is less than that value.

$$H_0 : \mu \geq \mu_0$$

$$H_1 : \mu < \mu_0$$

The test statistic is the sample mean, and the test statistic follows a normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. This standard deviation of the population is considered to be known by us.

Consider an example. We have a population of students with population mean μ and standard deviation $\sigma = 0.3$. We want to test whether the population mean is greater than or equal to 3.0. We take a sample of 100 students and find that the sample mean is 2.8. We then write the hypothesis test as follows:

$$H_0 : \mu \geq 3.0$$

$$H_1 : \mu < 3.0$$

Carefully understand what are we trying to do in the test. We are trying to test whether the population mean is greater than or equal to 3.0. We've acquired a sample, which has given us a value of the sample mean that is 2.8. Of course, 2.8 is less than 3.0, but we want to make statistical inference to know that, having obtained this value of the sample mean, it is still possible that the population (which is unknown to us) has a mean that is greater than or equal to 3.0.

How will we do that? We can use the sampling distribution of the sample mean to try and guess how far is 2.8 from the hypothesized value of the population mean (which is the expected value of the sampling distribution). We know that the sampling distribution of the sample mean is normally distributed with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. We know that the population standard deviation is 0.3, and that the sample size is 100. We can then calculate the standard deviation of the sampling distribution (standard error) of the sample mean as $\frac{0.3}{\sqrt{100}} = 0.03$.

We can then calculate the number of standard errors that 2.8 is away from the expected value of the sampling distribution, which is 3.0. We can do this by calculating the z-score of 2.8, which is $\frac{2.8-3.0}{0.03} = -6.67$. This means that 2.8 is 6.67 standard errors away from the expected value of the sampling distribution. In hypothesis testing language, we call this the *test statistic*. For all z-tests, the test statistic is the z-score of the sample mean.

What to do then? There are two ways to proceed.

1. Calculate the *p*-value (or prob value) of the test statistic and compare to the significance level α . If the *p*-value is less than α , we reject the null hypothesis. If the *p*-value is greater than α , we fail to reject the null hypothesis.
2. Calculate a *critical value* of the significance level and compare it to the test statistic. If the test statistic is less than or equal to the critical value, we reject the null hypothesis. If the test statistic is greater than the critical value, we fail to reject the null hypothesis.

How to calculate the *p*-value? It depends on the type of test. For a left-tailed test like this one, the *p* value is the probability of getting a value *less than or equal* to the test statistic. So, we use our standard normal table and look for the “larger portion”, “left tail” or “less than” area. That will be the *p*-value of the test statistic.

In this case, the *p*-value is the probability of getting a value less than or equal to -6.67. The *p*-value is `pnorm(-6.67) = 1.56e-11`. This is a very small number, which means that the probability of getting a value less than or equal to -6.67 is very small, so there is evidence against the null hypothesis.

We then need to define a significance level α . Let’s say that we define $\alpha = 0.05$. This means that we will reject the null hypothesis if the *p*-value is less than 0.05. In this case, the *p*-value is less than 0.05, so we reject the null hypothesis.

We can say that there is enough statistical evidence to reject the null hypothesis and say that the population mean is less than 3.0.

Using the critical value approach also depends on the type of test that we are using. For a left-tailed test, the critical value is the number of standard errors that have an area of α , the significance level, to the left of it. In this case, the critical value is the number of standard errors that have an area of 0.05 to the left of it. We can use the standard normal table to find this value. The critical value is -1.645.

Since the test statistic is less than the critical value, we reject the null hypothesis. We can say that there is enough statistical evidence to reject the null hypothesis and say that the population mean is less than 3.0, with 95% confidence ($1-\alpha$ is the confidence level). In both using the p -value approach and the critical value approach, we have the same conclusion, and that always has to happen.

Why are there two ways to do a hypothesis test? The p -value approach is more common, but the critical value approach is useful when we don't have a computer that can calculate the p -value for us. Looking for the p -value in the standard normal table is not easy because sometimes test statistics have many decimals, so we can use the critical value approach instead.

Right-tailed test

In a right-tailed test, the null hypothesis is that the population mean is less than or equal to a certain value, and the alternative hypothesis is that the population mean is greater than that value.

$$H_0 : \mu \leq \mu_0$$

$$H_1 : \mu > \mu_0$$

The test statistic is the sample mean, and the test statistic follows a normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. This standard deviation of the population is considered to be known by us.

Consider an example. We have a population of students with population mean μ and standard deviation $\sigma = 0.3$. We want to test whether the population mean is less than or equal to 3.0. We take a sample of 100 students and find that the sample mean is 3.1. We then write the hypothesis test as follows:

$$H_0 : \mu \leq 3.0$$

$$H_1 : \mu > 3.0$$

All tests, no matter if they're left-tailed, right-tailed or two-tailed, have the same steps. The only difference is the p -value or critical value calculation. So, produce the test statistic:

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{3.1 - 3.0}{\frac{0.3}{\sqrt{100}}} = 3.33$$

The p -value is the probability of getting a value *greater than or equal* to the test statistic. So, we use our standard normal table and look for the “smaller portion”, “right tail” or “greater than” area. That will be the p -value of the test statistic.

In this case, the p -value is the probability of getting a value greater than or equal to 6.67. The p -value is 6.8713794×10^{-4} . This is a very small number, which means that the probability of getting a value greater than or equal to 3.2 is very small, so there is evidence against the null hypothesis.

We then need to define a significance level α . Let's say that we define $\alpha = 0.05$. This means that we will reject the null hypothesis if the p -value is less than 0.05. In this case, the p -value is less than 0.05, so we reject the null hypothesis with 95% confidence.

We can say that there is enough statistical evidence to reject the null hypothesis and say that the population mean is greater than 3.0.

Using the critical value approach also depends on the type of test that we are using. For a right-tailed test, the critical value is the number of standard errors that have an area of α , the significance level, to the right of it. In this case, the critical value is the number of standard errors that have an area of 0.05 to the right of it. We can use the standard normal table to find this value. The critical value is 1.645.

In the left tailed test, if we saw that the test statistic is smaller than the critical value, we rejected the null hypothesis. In the right tailed test, if we see that the test statistic is **greater** than the critical value, we reject the null hypothesis. We can say that there is enough statistical evidence to reject the null hypothesis and say that the population mean is greater than 3.0. In both using the p -value approach and the critical value approach, we have the same conclusion, and that always has to happen.

Two-tailed test

A two-tailed test is a test where the null hypothesis is that the population mean is equal to a certain value, and the alternative hypothesis is that the population mean is not equal to that value.

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

The test statistic is the sample mean, and the test statistic follows a normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. This standard deviation of the population is considered to be known by us.

Consider an example. We have a population of students with population mean μ and standard deviation $\sigma = 0.3$. We want to test whether the population mean is equal to 3.0. We take a sample of 100 students and find that the sample mean is 3.1 We then write the hypothesis test as follows:

$$H_0 : \mu = 3.0$$

$$H_1 : \mu \neq 3.0$$

All tests, no matter if they're left-tailed, right-tailed or two-tailed, have the same steps. The only difference is the p -value or critical value calculation. So, produce the test statistic:

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{3.1 - 3.0}{\frac{0.3}{\sqrt{100}}} = 3.33$$

The p -value is the probability of getting a value *greater than or equal* to the test statistic multiplied by two. So, we use our standard normal table and look for the “smaller portion”, “right tail” or “greater than” area. That, multiplied by two, will be the p -value of the test statistic.

In this case, the p -value is the probability of getting a value greater than or equal to 6.67. The p -value is 0.0013743. This is a very small number, which means that the probability of getting a value greater than or equal to 3.2 is very small, so there is evidence against the null hypothesis.

We then need to define a significance level α . Let's say that we define $\alpha = 0.05$. This means that we will reject the null hypothesis if the p -value is less than 0.05. In this case, the p -value is less than 0.05, so we reject the null hypothesis with 95% confidence.

We can say that there is enough statistical evidence to reject the null hypothesis and say that the population mean is not equal to 3.0.

With a critical value approach, we must again adjust our calculation to reflect a double-tailed test. There will be two critical values for a two tailed test. The first one is the critical value that has an area of $\frac{\alpha}{2}$ to the right of it. The second one is the critical value that has an area of $\frac{\alpha}{2}$ to the left of it. In this case, the critical values are 1.96 and -1.96. It will always be true that we have a positive and negative critical values.

To compare, if we see that the test statistic is smaller than the negative critical value or greater than the positive critical value, we reject the null hypothesis. Because the test statistic is greater than the positive critical value, we reject the null hypothesis. We can say that there is enough statistical evidence to reject the null hypothesis and say that the population mean is not equal to 3.0. In both using the p -value approach and the critical value approach, we have the same conclusion, and that always has to happen.

Summary

- Hypothesis testing is a statistical method that allows us to make conclusions about a population parameter based on a sample statistic.
- The null hypothesis is the hypothesis that we want to test. The alternative hypothesis is the hypothesis that we want to accept if we reject the null hypothesis.
- The test statistic is the statistic that we use to test the null hypothesis. The test statistic follows a certain distribution, which depends on the type of test that we are performing.
- The p -value is the probability of getting a value equal to or more extreme than the test statistic. The p -value is compared to the significance level α to make a conclusion.
- The critical value is the value that has an area of α to the right of it. The critical value is compared to the test statistic to make a conclusion.
- When performing a z -test, we have to follow the following steps:
 1. State the null and alternative hypothesis, and determine the type of test (left, right or two-tailed)
 2. Define a significance level α
 3. Calculate the test statistic, which is the z -score
 4. Calculate the p -value or critical value
 5. Compare the p -value or critical value to the significance level α or the test statistic to the critical value
 6. Make a conclusion, using the right language.
- The table below summarises the calculation of a p -value and a critical value for a z -test.

Type of test	p -value	Critical value
Left-tailed	Left tail probability of z	Std. errors which produces a left area of α
Right-tailed	Right tail probability of z	Std. errors which produces a right area of α
Two-tailed	Double the right tail probability of z	Std errors which produce a right and a left area of $\alpha/2$

- The table below summarises the comparison rules of the p -value and the critical value.

Type of test	p -value	Critical value
Left-tailed	Reject H_0 if $p\text{-value} < \alpha$	Reject H_0 if $z < \text{critical value}$

Type of test	p -value	Critical value
Right-tailed	Reject H_0 if p -value $< \alpha$	Reject H_0 if $z >$ critical value
Two-tailed	Reject H_0 if p -value $< \alpha$	Reject H_0 if $z <$ -critical value or $z >$ critical value

- The next table gives you the most common critical values which correspond to the significance levels of α which are common in practice, for each kind of test (left, right or two-tailed).

Type of test	α	Critical value
Left-tailed	0.05	-1.645
	0.01	-2.33
	0.1	-1.28
Right-tailed	0.05	1.645
	0.01	2.33
	0.1	1.28
Two-tailed	0.05	+/- 1.96
	0.01	+/- 2.58
	0.1	+/- 1.64

Don't ever forget that when you make a conclusion which does support the null hypothesis, you can never say that the null hypothesis is true. You can only say that there is not enough statistical evidence to reject the null hypothesis. Further, remember that based on your significance level, you can define your confidence level, which is $1 - \alpha$. So, if you have a significance level of 0.05, you have a confidence level of 95%. If you have a significance level of 0.01, you have a confidence level of 99%. If you have a significance level of 0.1, you have a confidence level of 90%. 95% is the most common confidence level, but you can use any confidence level you want.