A [hypothesis test](https://statisticsbyjim.com/glossary/hypothesis-tests/) evaluates two mutually exclusive statements about a [population](https://statisticsbyjim.com/glossary/population/) to determine which statement is best supported by the sample data. These two statements are called the null hypothesis and the [alternative hypothesis](https://statisticsbyjim.com/glossary/alternative-hypothesis/).

Test statistics assess how consistent your sample data are with the null hypothesis. As a test statistic becomes more extreme, it indicates a larger difference between your sample data and the null hypothesi

Hypothesis tests are not 100% accurate because they use a [random sample](https://statisticsbyjim.com/glossary/sample/) to draw conclusions about entire populations. When you perform a hypothesis test, there are two types of errors related to drawing an incorrect conclusion.

* [Type I error](https://statisticsbyjim.com/glossary/type-i-error/): The rejects a null hypothesis that is true. You can think of this as a false positive.
* [Type II error](https://statisticsbyjim.com/glossary/type-ii-error/): The test fails to reject a null hypothesis that is false. You can think of this as a false negative.

A test result is statistically significant when the sample statistic is unusual enough relative to the null hypothesis that you can reject the null hypothesis for the entire population. “Unusual enough” in a hypothesis test is defined by how unlikely the [effect](https://statisticsbyjim.com/glossary/effect/) observed in your sample is if the null hypothesis is true.

If your sample data provide sufficient evidence, you can reject the null hypothesis for the entire population. Your data favor the alternative hypothesis.

**Hypothesis testing** is a formal procedure for investigating our ideas about the world using [statistics](https://www.scribbr.com/?cat_ID=34372). It is most often used by scientists to test specific predictions, called hypotheses, that arise from theories.

There are 5 main steps in hypothesis testing:

1. State your research hypothesis as a null hypothesis and alternate hypothesis (Ho) and (Ha or H1).
2. [Collect data](https://www.scribbr.com/methodology/data-collection/) in a way designed to test the hypothesis.
3. Perform an appropriate [statistical test](https://www.scribbr.com/statistics/statistical-tests/).
4. Decide whether to reject or fail to reject your null hypothesis.
5. Present the findings in your [results](https://www.scribbr.com/dissertation/results/) and [discussion](https://www.scribbr.com/dissertation/discussion/) section.

Though the specific details might vary, the procedure you will use when testing a hypothesis will always follow some version of these steps.

1. [Step 1: State your null and alternate hypothesis](https://www.scribbr.com/statistics/hypothesis-testing/#step-1)
2. [Step 2: Collect data](https://www.scribbr.com/statistics/hypothesis-testing/#step-2)
3. [Step 3: Perform a statistical test](https://www.scribbr.com/statistics/hypothesis-testing/#step-3)
4. [Step 4: Decide whether to reject or fail to reject your null hypothesis](https://www.scribbr.com/statistics/hypothesis-testing/#step-4)
5. [Step 5: Present your findings](https://www.scribbr.com/statistics/hypothesis-testing/#step-5)

## What is a Test Statistic?

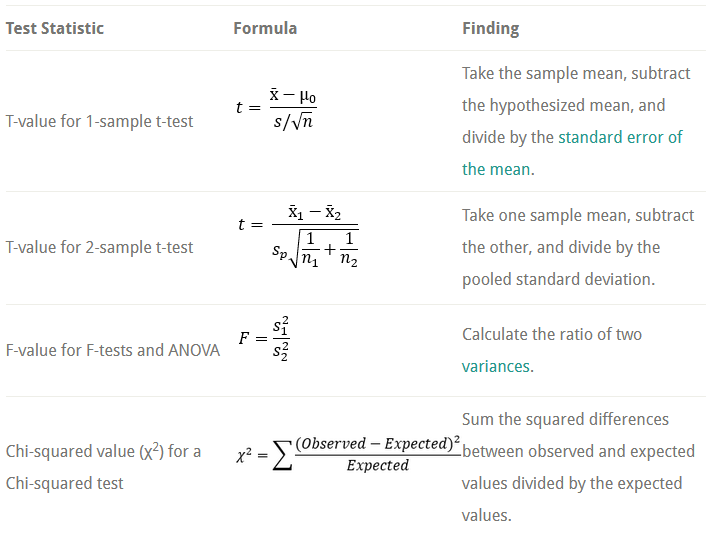
A test statistic assesses how consistent your [sample](https://statisticsbyjim.com/glossary/sample/) data are with the null hypothesis in a [hypothesis test](https://statisticsbyjim.com/glossary/hypothesis-tests/). Test statistic calculations take your sample data and boil them down to a single number that quantifies how much your sample diverges from the null hypothesis. As a test statistic value becomes more extreme, it indicates larger differences between your sample data and the null hypothesis.

When your test statistic indicates a sufficiently large incompatibility with the null hypothesis, you can reject the null and state that your results are statistically significant—your data support the notion that the sample [effect](https://statisticsbyjim.com/glossary/effect/) exists in the [population](https://statisticsbyjim.com/glossary/population/). To use a test statistic to evaluate statistical significance, you either compare it to a critical value or use it to calculate the [p-value](https://statisticsbyjim.com/glossary/p-value/).

[Statisticians](https://statisticsbyjim.com/glossary/statistics/) named the hypothesis tests after the test statistics because they’re the quantity that the tests actually evaluate. For example, t-tests assess t-values, F-tests evaluate F-values, and chi-square tests use chi-square values.

## How to Find Test Statistics

Each test statistic has its own formula. I present several common test statistics examples below. To see worked examples for each one, click the links to my more detailed articles.



### T-Tests, Null = 0

When a t-value equals 0, it indicates that your sample data match the null hypothesis exactly.

For a 1-sample t-test, when the sample mean equals the hypothesized mean, the numerator is zero, which causes the entire t-value ratio to equal zero. As the sample mean moves away from the hypothesized mean in either the positive or negative direction, the test statistic moves away from zero in the same direction.

A similar case exists for 2-sample t-tests. When the two sample means are equal, the numerator is zero, and the entire test statistic ratio is zero. As the two sample means become increasingly different, the absolute value of the numerator increases, and the t-value becomes more positive or negative.

### F-tests including ANOVA, Null = 1

When an F-value equals 1, it indicates that the two variances in the numerator and denominator are equal, matching the null hypothesis.

As the numerator and denominator become less and less similar, the F-value moves away from one in either direction.

### Chi-squared Tests, Null = 0

When a chi-squared value equals 0, it indicates that the observed values always match the expected values. This condition causes the numerator to equal zero, making the chi-squared value equal zero.

As the observed values progressively fail to match the observed values, the numerator increases, causing the test statistic to rise from zero.

You’ll never see a test statistic that equals the null value precisely in practice. However, trivial differences been sample values and the null value are not uncommon.

### Formulas for Test Statistics

## What Are P values?

P values are the probability that a sample will have an effect at least as extreme as the effect observed in your sample if the null hypothesis is correct.

**General Process for How to Find the P value**

To find the p value for your [sample](https://statisticsbyjim.com/glossary/sample/), do the following:

1. Identify the correct test statistic.
2. Calculate the test statistic using the relevant properties of your sample.
3. Specify the characteristics of the test statistic’s sampling distribution.
4. Place your test statistic in the sampling distribution to find the p value.

### Key Takeaways

* A p-value is a statistical measurement used to validate a hypothesis against observed data.
* A p-value measures the probability of obtaining the observed results, assuming that the null hypothesis is true.
* The lower the p-value, the greater the statistical significance of the observed difference.
* A p-value of 0.05 or lower is generally considered statistically significant.1
* P-value can serve as an alternative to—or in addition to—preselected confidence levels for hypothesis testing.

## How Is P-Value Calculated?

P-values are usually found using p-value tables or spreadsheets/statistical software. These calculations are based on the assumed or known [probability distribution](https://www.investopedia.com/terms/p/probabilitydistribution.asp) of the specific statistic tested. P-values are calculated from the deviation between the observed value and a chosen reference value, given the probability distribution of the statistic, with a greater difference between the two values corresponding to a lower p-value.

Mathematically, the p-value is calculated using integral calculus from the area under the probability distribution curve for all values of statistics that are at least as far from the reference value as the observed value is, relative to the total area under the probability distribution curve.

The calculation for a p-value varies based on the type of test performed. The three test types describe the location on the probability distribution curve: lower-tailed test, upper-tailed test, or [two-tailed test](https://www.investopedia.com/terms/t/two-tailed-test.asp).

In a nutshell, the greater the difference between two observed values, the less likely it is that the difference is due to simple random chance, and this is reflected by a lower p-value.

## The P-Value Approach to Hypothesis Testing

The p-value approach to hypothesis testing uses the calculated probability to determine whether there is evidence to reject the null hypothesis. The null hypothesis, also known as the conjecture, is the initial claim about a population (or data-generating process). The alternative hypothesis states whether the population parameter differs from the value of the population parameter stated in the conjecture.

In practice, the significance level is stated in advance to determine how small the p-value must be to reject the null hypothesis. Because different researchers use different levels of significance when examining a question, a reader may sometimes have difficulty comparing results from two different tests. P-values provide a solution to this problem.

For example, suppose a study comparing returns from two particular [assets](https://www.investopedia.com/terms/a/asset.asp) was undertaken by different researchers who used the same data but different significance levels. The researchers might come to opposite conclusions regarding whether the assets differ.

If one researcher used a confidence level of 90% and the other required a confidence level of 95% to reject the null hypothesis, and if the p-value of the observed difference between the two returns was 0.08 (corresponding to a confidence level of 92%), then the first researcher would find that the two assets have a difference that is [statistically significant](https://www.investopedia.com/terms/s/statistically_significant.asp), while the second would find no statistically significant difference between the returns.

To avoid this problem, the researchers could report the p-value of the hypothesis test and allow readers to interpret the statistical significance themselves. This is called a p-value approach to hypothesis testing. Independent observers could note the p-value and decide for themselves whether that represents a statistically significant difference or not.

## Example of P-Value

An investor claims that their investment [portfolio](https://www.investopedia.com/terms/p/portfolio.asp)’s performance is equivalent to that of the [Standard & Poor’s (S&P) 500 Index](https://www.investopedia.com/terms/s/sp500.asp). To determine this, the investor conducts a two-tailed test.

The null hypothesis states that the portfolio’s returns are equivalent to the S&P 500’s returns over a specified period, while the alternative hypothesis states that the portfolio’s returns and the S&P 500’s returns are not equivalent—if the investor conducted a [one-tailed test](https://www.investopedia.com/terms/o/one-tailed-test.asp), the alternative hypothesis would state that the portfolio’s returns are either less than or greater than the S&P 500’s returns.

The p-value hypothesis test does not necessarily make use of a preselected confidence level at which the investor should reset the null hypothesis that the returns are equivalent. Instead, it provides a measure of how much evidence there is to reject the null hypothesis. The smaller the p-value, the greater the evidence against the null hypothesis.

Thus, if the investor finds that the p-value is 0.001, there is strong evidence against the null hypothesis, and the investor can confidently conclude that the portfolio’s returns and the S&P 500’s returns are not equivalent.

Although this does not provide an exact threshold as to when the investor should accept or reject the null hypothesis, it does have another very practical advantage. P-value hypothesis testing offers a direct way to compare the relative confidence that the investor can have when choosing among multiple different types of [investments](https://www.investopedia.com/terms/i/investment.asp#:~:text=An%20investment%20can%20refer%20to%20any%20mechanism%20used%20for%20generating,can%20be%20considered%20an%20investment.) or portfolios relative to a [benchmark](https://www.investopedia.com/terms/b/benchmark.asp) such as the S&P 500.

For example, for two portfolios, A and B, whose performance differs from the S&P 500 with p-values of 0.10 and 0.01, respectively, the investor can be much more confident that portfolio B, with a lower p-value, will actually show consistently different results.

## Is a 0.05 P-value Significant?

A p-value less than 0.05 is typically considered to be statistically significant, in which case the null hypothesis should be rejected. A p-value greater than 0.05 means that deviation from the null hypothesis is not statistically significant, and the null hypothesis is not rejected.1

## What Does a P-value of 0.001 Mean?

A p-value of 0.001 indicates that if the null hypothesis tested were indeed true, then there would be a one-in-1,000 chance of observing results at least as extreme. This leads the observer to reject the null hypothesis because either a highly rare data result has been observed or the null hypothesis is incorrect.

## How Can You Use P-value to Compare Two Different Results of a Hypothesis Test?

If you have two different results, one with a p-value of 0.04 and one with a p-value of 0.06, the result with a p-value of 0.04 will be considered more statistically significant than the p-value of 0.06. Beyond this simplified example, you could compare a 0.04 p-value to a 0.001 p-value. Both are statistically significant, but the 0.001 example provides an even stronger case against the null hypothesis than the 0.04.

## The Bottom Line

The p-value is used to measure the significance of observational data. When researchers identify an apparent relationship between two variables, there is always a possibility that this correlation might be a coincidence. A p-value calculation helps determine if the observed relationship could arise as a result of chance.

**Reverences :**

<https://statisticsbyjim.com/glossary/hypothesis-tests/>

<https://www.scribbr.com/statistics/hypothesis-testing/>

<https://www.investopedia.com/terms/p/p-value.asp>