

# Learn all About Hypothesis Testing!

[BEGINNER](#)[MATHS](#)[STATISTICS](#)

This article was published as a part of the [Data Science Blogathon](#).

## Table of Contents

- 1) Introduction
- 2) Types of Errors
- 3) Types of Hypothesis Tests
- 4) All about Parametric and Non-Parametric Tests
- 5) Parametric vs Non-Parametric Tests
- 6) Hypothesis Tests of the Mean and Median
- 7) Reasons to use Parametric Tests
- 8) Reasons to use Non-Parametric Tests
- 9) Conclusion

## Hypothesis Testing

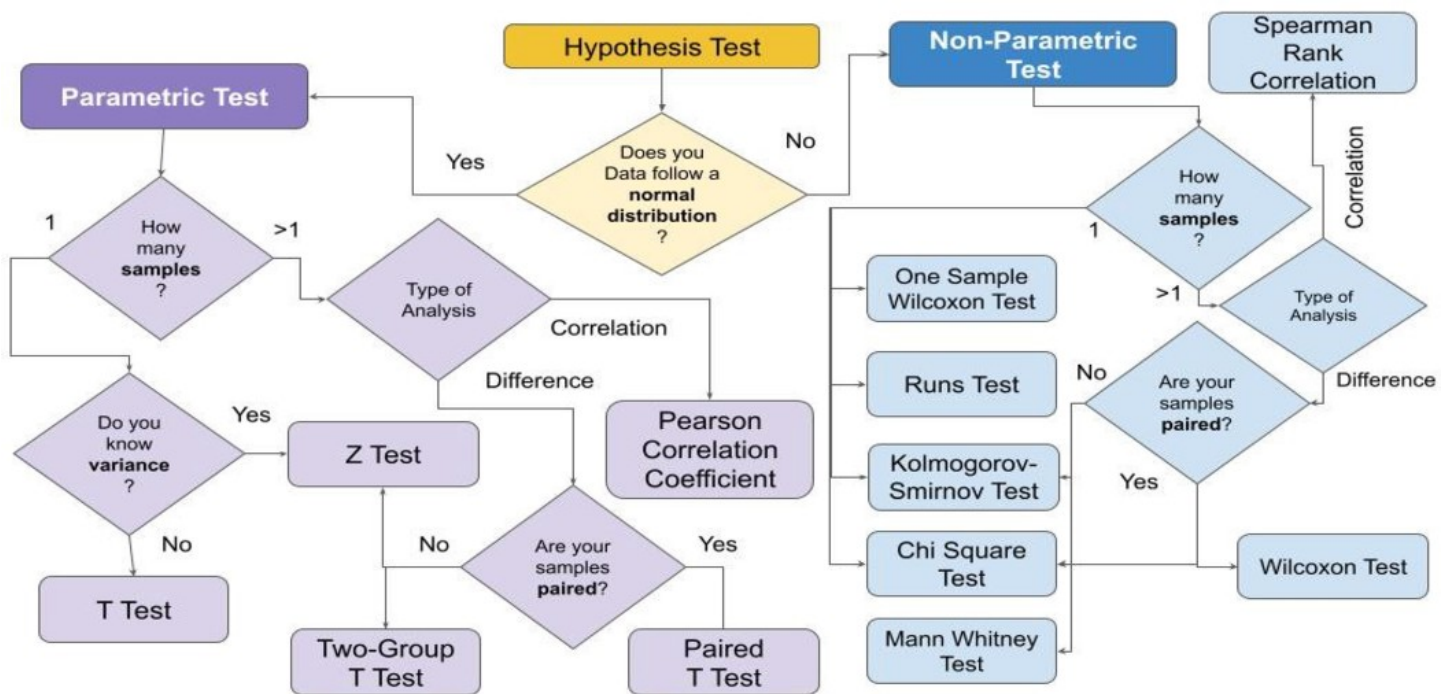
Hypothesis testing is a sort of inferential statistics that allows us to make assumptions about a full population based on a representative sample. In most cases, it is just impossible to observe the entire population in order to comprehend its properties. The only option is to select a random sample and use statistics to analyse it.

Prior to conducting Hypothesis Testing, you must first establish a hypothesis. Hypotheses include things like “there is a difference between two groups in a population” or “there is a correlation between weight and gender in a population.”

The thesis to be shown is usually referred to as the Alternative Hypothesis ( $H_A$ ), while the Null Hypothesis is its polar opposite ( $H_0$ ). The Null Hypothesis states that nothing new is happening in the population in practice.

Null Hypotheses could be written as follows in the previous examples: there is no link between weight and gender in the population, and there is no difference between the two groups.

The purpose of Hypothesis Testing is to see if the Null Hypothesis can be rejected or not. In most cases, rejecting the Null Hypothesis does not imply that the alternative hypothesis is true. Rejecting the Null Hypothesis, on the other hand, can lead to acceptance of the Alternative Hypothesis in some cases.



**There are two types of errors that can occur when executing a Hypothesis Test:**

- **Type-I Error:** Rejecting the Null Hypothesis when it is actually true is a Type-I Error.
- **Type-II Error:** Accepting the Null Hypothesis when it is false is a Type-II Error.

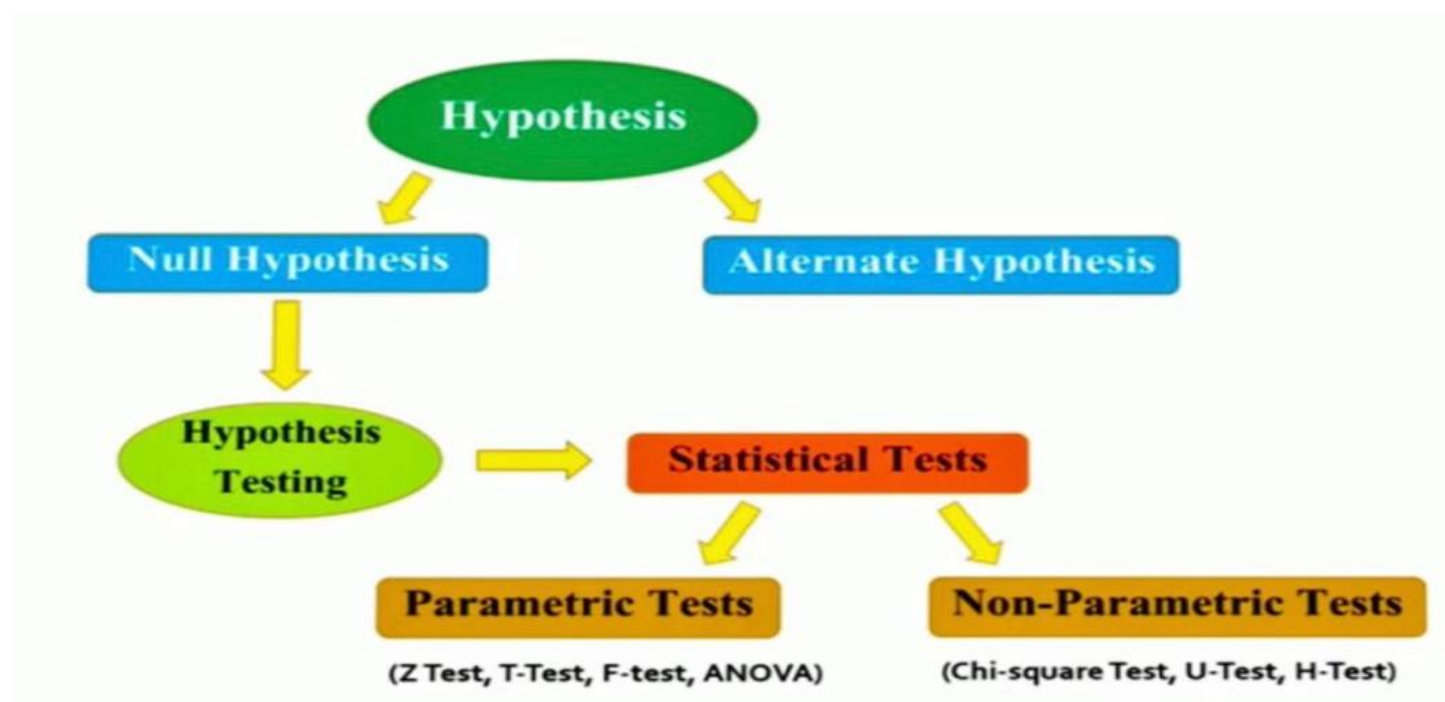
		Truth	
Hypothesis Testing		The Null Hypothesis Is True	The Alternative Hypothesis Is True
Research	The Null Hypothesis Is True	Accurate	Type II Error
	The Alternative Hypothesis Is True	Type I Error	Accurate

## Types of Hypothesis Testing

- **Hypothesis tests are divided into two categories:**

1) **Parametric tests** – are used when the samples have a normal distribution. In general, samples with a mean of 0 and a variance of 1 follow a normal distribution.

2) **Non-Parametric tests** – If the samples do not follow a normal distribution, non-parametric tests are used.



• **Two types of Hypothesis Testing can be created depending on the number of samples to be compared:**

• **One Sample** – If there is only one sample that must be compared to a specific value, it is called a single sample.

• **Two Samples** – if you're comparing two or more samples. Correlation and sample difference are two tests that could be used in this situation. Samples can be paired or not in both circumstances. Dependent samples are sometimes known as paired samples, while independent samples are known as unpaired samples. Natural or matched couplings occur in paired samples.

## Parametric and Non-Parametric Tests

A parametric test is one in which the parameters are predetermined and the population distribution is always known. A mean value is used to calculate the central tendency. These tests are common, which makes conducting research relatively simple and time-consuming. The Non-parametric test makes no assumptions and measures using the median value. Kruskal-Wallis, Mann-Whitney, and other non-parametric tests are examples. This article will teach you what parametric and non-parametric tests are, the benefits and drawbacks of parametric and non-parametric tests, parametric and non-parametric statistics, and the distinction between parametric and non-parametric tests.

### What is a Parametric Test?

In statistics, the parametric test provides generalisations for generating data on the original population's mean. This test is also classified as a hypothesis test. A t-test is utilised, which is based on a student's t-

test and is often used in this value. This is known as a parametric test.

The t-measurement test assumes an ordinary distribution for a variable. In this scenario, the mean value is known, or it is assumed or taken to be known. The population variance is used to find the sample from the population. The population is estimated using an interval scale, and the variables of concern are hypothesised.

**The table below lists some of the most common parametric tests as well as what they measure.**

**What is a Non-Parametric Test?**

There is no necessity for population distribution in the non-parametric test. In addition, the non-parametric test is a hypothesis test that is independent of any underlying hypothesis. The median value is used in the non-parametric test. Distribution-free testing is another name for this method of testing. Test values are determined at the ordinal or nominal level. The parametric test is typically employed when the independent variables are non-metric. A non-parametric test is what this is called.

The table below lists some of the most common non-parametric tests as well as what they measure.

# Parametric vs. Non-Parametric: What’s the Difference?

Properties	Parametric	Non-Parametric
1) Assumptions	Yes, assumptions are made before using parametric tests	No, assumptions are not made before using non-parametric tests
2) Value for Central Tendency	Mean is the CT for parametric tests	Median is the CT for non-parametric tests
3) Correlation	Pearson Correlation	Spearman Correlation
4) Probabilistic Distribution	Normal pd	Arbitrary pd
5) Population Knowledge	Required	Not required
6) Uses	Used for finding interval data	Used for finding nominal data
7) Applications	Applicable to variables	Applicable to variables and attributes
8) Examples	T-test, Z-test	Friedman test etc

In this article, I’ll explain when each test should be used:

- To test group means, use parametric analysis.
- Group medians were tested using nonparametric analysis.

## Hypothesis Testing of the Mean and Median

Nonparametric tests are like a parallel universe to parametric tests. The table shows related pairs of hypothesis tests.

Parametric tests (means)	Nonparametric tests (medians)
1-sample t-test	1-sample Sign, 1-sample Wilcoxon
2-sample t-test	Mann-Whitney test
One-Way ANOVA	Kruskal-Wallis, Mood’s median test
Factorial DOE with one factor and one blocking variable	Friedman test

# Reasons to use Parametric Tests

**Reason 1: Parametric tests can handle skewed and non-normal distributions well.**

It may come as a surprise, but parametric tests can perform well with non-normal continuous data if the sample size criteria in the table below are followed.

Parametric analyses	Sample size guidelines for non-normal data
1-sample t-test	Greater than 20
2-sample t-test	Each group should be greater than 15
One-Way ANOVA	<ul style="list-style-type: none"><li>• If you have 2-9 groups, each group should be greater than 15.</li><li>• If you have 10-12 groups, each group should be greater than 20.</li></ul>

**Reason 2: When the dispersion of each group is different, parametric tests can perform well.**

Nonparametric tests don't require that your data follow a normal distribution, but they do have other requirements that can be difficult to achieve. Whereas nonparametric tests compare data of the groups that must have the same dispersion. Nonparametric tests may not yield accurate results if your groups have a different distribution.

**Reason3: Because of the Statistical Power of Parametric Tests**

By comparison, it can be seen that Nonparametric tests have less statistical power than parametric tests.

# Reasons to use Non-Parametric Tests

**Reason 1: The median represents your field of study better.**

The fact that you can run a parametric test on non-normal data doesn't mean you should.

The median, for example, can better measure the centre of a skewed distribution, such as income, when 50% are above and 50% are below the median. Even if the average person's income remains the same, when a few billionaires are included in a sample, the mathematical mean climbs dramatically.

When your distribution is sufficiently skewed, changes far out in the tail have a large impact on the mean, yet the median continues to reflect the distribution's centre. A random sampling of 100 people from each distribution yields statistically different means, but not significantly different medians for these two distributions.

**Reason 2: Your sample size is quite small.**

If you don't meet the sample size requirements for parametric testing and aren't confident that your data is normally distributed, you should use a nonparametric test. When your sample size is limited, you may not be able to identify the distribution of your data because distribution tests lack the power to provide meaningful results.

You're in a tight spot with no feasible solutions in this situation. Nonparametric tests have lower power, to begin with, and when you combine that with a small sample size, you have a double whammy!

**Reason 3: You can't get rid of your ordinal or ranking data or outliers.**

Outliers can dramatically influence the outcomes of typical parametric tests, which can only examine continuous data. Some nonparametric tests, on the other hand, can handle ordinal and ranking data while avoiding outliers. Check the nonparametric test assumptions because each one has different data requirements.

## Conclusion

The requirement to choose between a parametric and nonparametric test is usually assumed to arise when your data fails to meet a parametric test assumption. When you have a limited sample size and non-normal data, this can happen. However, while parametric tests can typically manage non-normal data, other factors often come into play. Nonparametric tests, on the other hand, have rigid assumptions that you can't ignore.

Whether you choose the mean or the median to depict the centre of your data's distribution is typically a deciding factor.

- Consider a parametric test if the mean adequately represents the centre of your distribution and your sample size is large enough. They are more powerful.
- Even if you have a big sample, consider the nonparametric test if the median better represents the centre of your distribution.

Finally, if your sample size is exceedingly tiny, you may be forced to use a nonparametric test. If at all feasible, please collect more data next time! As you can see, the sample size requirements aren't that stringent. When you have a limited sample size and must utilise a less efficient nonparametric test, your chances of identifying a significant effect when one exists can be slim!

## EndNotes

Thank you for following up with me all the way to the end. In this article, I have described the concept of Hypothesis Testing, as well as the most popular tests and when they can be used.

For those who still have difficulty understanding hypothesis testing, there is a Python library, called "easy-ht", which runs the main hypothesis tests without any knowledge of statistics. A tutorial on how to use "easy-ht" is available on YouTube.

I hope you found this article useful. Please feel free to distribute it to your peers.

### Author

Hello, I'm **Gunjan Agarwal** from Gurugram, and I earned a master's degree in Data Science from Amity University in Gurgaon. I enthusiastically participate in Data Science hackathons, blogathons, and workshops.

I'd like to connect with you on [LinkedIn](#). Mail me [here](#) for any queries.

### Image Sources

- [https://miro.medium.com/max/960/0\\*oJWBEpXPplcZZX-K.jpg](https://miro.medium.com/max/960/0*oJWBEpXPplcZZX-K.jpg)
- <https://media.cheggcdn.com/study/41b/41b2c31c-895f-4fc0-8007-0dced3df21bf/stat-1-img-1.png>
- <https://editor.analyticsvidhya.com/uploads/49317hp.jpeg>
- [https://miro.medium.com/max/1400/0\\*rn4ekgAiHvy4lprV.jpg](https://miro.medium.com/max/1400/0*rn4ekgAiHvy4lprV.jpg)
- [https://miro.medium.com/max/875/0\\*-WxQpdTKIU95zncU.jpg](https://miro.medium.com/max/875/0*-WxQpdTKIU95zncU.jpg)

The media shown in this article is not owned by Analytics Vidhya and are used at the Author's discretion.



## **Gunjan Agarwal**

I am Data Science Fresher. And I'm open to work.