**Learn all about Hypothesis Testing**

## Introduction

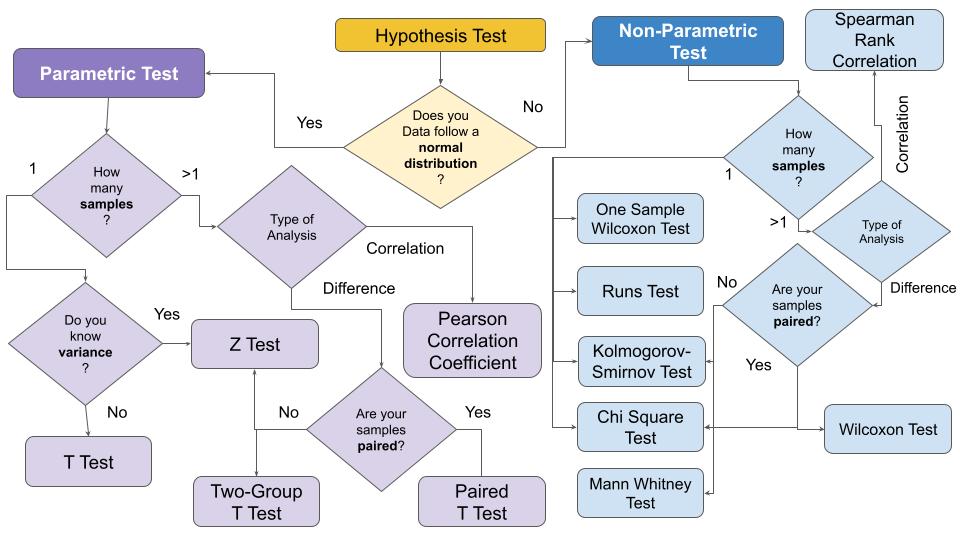
According to Jim Frost, hypothesis testing is a sort of inferential statistics that allows us to make assumptions about a full population based on a representative sample. In most cases, it is just impossible to observe the entire population in order to comprehend its properties. The only option is to 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 (HA), while the Null Hypothesis is its polar opposite (H0). The Null Hypothesis states that nothing new is happening in the population in practise.

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 goal of Hypothesis Testing is to determine whether or not the Null Hypothesis can be rejected. In general, rejecting the Null Hypothesis does not imply acceptance of the alternative hypothesis. In other situations, however, rejecting the Null Hypothesis can lead to acceptance of the Alternative Hypothesis.

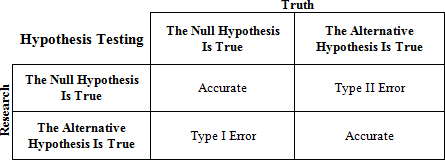


[https://miro.medium.com/max/960/0\*oJWBEpXPpIcZZX-K.jpg](https://miro.medium.com/max/960/0*oJWBEpXPpIcZZX-K.jpg)

**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.



<https://media.cheggcdn.com/study/41b/41b2c31c-895f-4fc0-8007-0dced3df21bf/stat-1-img-1.png>

## Types of Hypothesis Tests

**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 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.

<https://editor.analyticsvidhya.com/uploads/49317hp.jpeg>

**Two types of Hypothesis Tests 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.

## All about 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 exactly is a Parametric Test?**

The parametric test in statistics provides generalisations for creating data about tlhe mean of the original population. This test is a type of hypothesis test as well. A t-test is utilised, which is based on the t-test of students, which is commonly used in this value. A parametric test is what this is called.

The t-measurement test assumes that a variable has an ordinary distribution. The mean value is known, or it is assumed or taken to be known, in this case. To find the sample from the population, the population variance is calculated. An interval scale is used to estimate the population, and the variables of concern are hypothesised.

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

Table

Description automatically generated

[https://miro.medium.com/max/1400/0\*rn4ekgAiHvy4IprV.jpg](https://miro.medium.com/max/1400/0*rn4ekgAiHvy4IprV.jpg)

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

In the non-parametric test, there is no requirement for population distribution. Furthermore, the non-parametric test is a hypothesis test that does not rely on any underlying hypothesis. The non-parametric test is based on the value of the median. This testing technique is also known as distribution-free testing. The ordinal or nominal level is used to determine test values. When the independent variables are non-metric, the parametric test is usually used. This is referred to as a non-parametric test.

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

Table

Description automatically generated

[https://miro.medium.com/max/875/0\*-WxQpdTKlU95zncU.jpg](https://miro.medium.com/max/875/0*-WxQpdTKlU95zncU.jpg)

**Parametric vs. Non-Parametric Statistics: What's the Difference?**

* The parametric tests are based on assumptions made using data from the normal distribution that was used in the investigation. It is critical to have a good understanding of the parameters. Non-parametric tests, on the other hand, are independent of any parameters.
* It isn't based on preconceptions.
* In the case of a parametric test, the analysis produced from the data is known as probabilistic distribution. Non-parametric probability distributions, on the other hand, are arbitrary and their central tendency cannot be measured.
* Data is measured using the mean value in parametric tests and the median value in non- parametric tests.
* In both circumstances, the information used is different. Because of the parameters associated to population in parametric, prior knowledge of the population is required, but in non-parametric, no prior knowledge of the population is required.
* A parametric test uses defined parameters, prior information, and assumptions, whereas non-parametric statistics uses independent variables freely.
* In the case of parametric statistics, the result is based on a distribution, whereas in the case of non-parametric statistics, no distribution is used.
* In parametric analysis, the variable of interest is assessed on an interval or ratio level, whereas in non-parametric analysis, the variable of interest is measured on a nominal or ordinal scale.
* Non-parametric tests, unlike parametric tests, do not require population information.
* Non-parametric tests are relevant to both variables and attributes, whereas parameter tests are only applicable to variables.
* In parametric testing, Person's coefficient or co-relation is used to evaluate the degree of association between two variables, whereas in non-parametric tests, Spearman's rank correlation is employed.

**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 Tests 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 | * If you have 2-9 groups, each group should be greater than 15. * If you have 10-12 groups, each group should be greater than 20. |

**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, 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 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

In simplest terms, if information about the population is accessible in the form of parameters, parametric statistics can be employed to conduct the research.

The test that is run using this statistical technique is known as a parametric test.

Non-parametric statistics are employed when a population hypothesis needs to be evaluated without the availability of population knowledge. Non-parametric test is the test that is used to analyse non-parametric statistical reports.

### End Notes:

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

For those who still have difficulty understanding hypothesis tests, 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 at 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](https://www.linkedin.com/in/gunjan-agarwal-b0a69418b?lipi=urn%3Ali%3Apage%3Ad_flagship3_profile_view_base_contact_details%3BEIcPSCIQRROdk0KyqfaKaw%3D%3D). Mail me [here](mailto:gunjan.ag9729@gmail.com)for any queries.