

# Hypothesis testing

by Marco Taboga, PhD

Hypothesis testing is a method of making statistical inferences in which:


- we establish an hypothesis, called null hypothesis;
- we use some data to decide whether to reject or not to reject the hypothesis.

This lecture provides a rigorous introduction to the mathematics of hypothesis tests, and it provides several links to other pages where the single steps of a test of hypothesis can be studied in more detail.



## What you need to know to get started

Remember that a [statistical inference](#) is a statement about the probability distribution from which a [sample](#) has been drawn.

In mathematical terms, the sample  $\xi$  can be regarded as a realization of a [random vector](#)  $\Xi$ , whose unknown [joint distribution function](#)  $F_{\Xi}(\xi)$  is assumed to belong to a set of distribution functions , called [statistical model](#).

**Example** We observe the realizations  $x_1, \dots, x_n$  of  $n$  [independently and identically distributed](#) (IID) random variables having a [normal distribution](#). The sample

$$\xi = [x_1 \ \dots \ x_n]$$

can be regarded as a realization of a random vector  $\Xi$  whose entries are all independent of each

other. The statistical model is a set of distribution functions satisfying certain conditions

$$\Phi = \{F_{\Xi} : \text{the marginals of } F_{\Xi} \text{ are IID and normal}\}$$

We will continue this example in the following sections.

---

## Testing restrictions

In hypothesis testing we make a statement about a model restriction involving a subset  $\Phi_R \subset \Phi$  of the original model.

The statement we make is chosen between two possible statements:

1. reject the restriction  $F_{\Xi} \in \Phi_R$ ;
2. do not reject the restriction  $F_{\Xi} \in \Phi_R$ .

Roughly speaking, we start from a large set  $\Phi$  of distributions that might possibly have generated the sample  $\xi$  and we would like to restrict our attention to a smaller set  $\Phi_R$ .

In a test of hypothesis, we use the sample  $\xi$  to decide whether or not to indeed restrict our attention to the smaller set  $\Phi_R$ .

**Example** In the case of our normal sample, we might want to test the restriction that the mean of the distribution is equal to zero. The restriction would be:

$$\Phi = \{F_{\Xi} : \text{the marginals of } F_{\Xi} \text{ are IID, normal and have zero mean}\}$$

---

## Parametric tests

Remember that in a parametric model the set of distribution functions  $\Phi$  is put into correspondence with a set  $\Theta \subseteq \mathbb{R}^p$  of  $p$ -dimensional real vectors called the [parameter space](#).

The elements of  $\Theta$  are called [parameters](#) and the true parameter is denoted by  $\theta_0$ . The true parameter is the parameter associated with the unknown distribution function  $F_{\Xi}(\xi)$  from which the sample  $\xi$  was actually drawn. For simplicity,  $\theta_0$  is assumed to be unique.

In parametric hypothesis testing we have a restriction  $\Theta_R \subset \Theta$  on the parameter space and we choose one of the following two statements about the restriction:

1. reject the restriction  $\theta_0 \in \Theta_R$ ;
2. do not reject the restriction  $\theta_0 \in \Theta_R$ .

For concreteness, we will focus on parametric hypothesis testing in this lecture, but most of the things we will say apply with straightforward modifications to hypothesis testing in general.

**Example** In the above example, a normal distribution is completely described by its mean  $\mu$  and variance  $\sigma^2$ . Thus, each distribution in the set  $\Phi$  is put into correspondence with a parameter vector  $\theta = (\mu, \sigma^2)$ . In this case the parameter space is  $\Theta = \mathbb{R} \times \mathbb{R}_+$ . The restriction to be tested is that the mean of the distribution be equal to zero. Therefore, the parametric restriction is

$$\Theta_R = \{(\mu, \sigma^2) \in \Theta : \mu = 0\}$$

---

## Null hypothesis

The hypothesis that the restriction is true is called null hypothesis and it is usually denoted by  $H_0$ :

$$H_0 : \theta_0 \in \Theta_R$$

Understanding how to formulate a null hypothesis is a fundamental step in hypothesis testing. We suggest to read a thorough discussion of null hypotheses [here](#).

**Example** In our example, the null hypothesis is

$$H_0 : \mu_0 = 0$$

---

## Alternative hypothesis

The restriction  $\theta_0 \in \Theta_R^c$  (where  $\Theta_R^c$  is the complement of  $\Theta_R$ ) is often called [alternative hypothesis](#) and it is denoted by  $H_1$ :

$$H_1 : \theta_0 \in \Theta_R^c$$

Statisticians sometimes take into consideration as an alternative hypothesis a set smaller than  $\Theta_R^c$ . In these cases, the null hypothesis and the alternative hypothesis do not cover all the possibilities contemplated by the parameter space  $\Theta$ .

For some authors, "rejecting the null hypothesis  $H_0$ " and "accepting the alternative hypothesis  $H_1$ " are synonyms. For other authors, however, "rejecting the null hypothesis  $H_0$ " does not necessarily imply "accepting the alternative hypothesis  $H_1$ ".

Although this is mostly a matter of language, it is possible to envision situations in which, after rejecting  $H_0$ , a second test of hypothesis is performed whereby  $H_1$  becomes the new null hypothesis and it is rejected (this may happen for example if the model is [mis-specified](#)).

In these situations, if "rejecting the null hypothesis  $H_0$ " and "accepting the alternative hypothesis  $H_1$ " are treated as synonyms, then some confusion arises, because the first test leads to "accept  $H_1$ " and the second test leads to "reject  $H_1$ ".

**Example** In our example, the alternative hypothesis could be

$$H_1 : \mu_0 \neq 0$$

## Types of errors

When we decide whether to reject a restriction or not to reject it, we can incur in two types of errors:

1. reject the restriction  $\theta_0 \in \Theta_R$  when the restriction is true; this is called an error of the first kind or a [Type I error](#);
2. do not reject the restriction  $\theta_0 \in \Theta_R$  when the restriction is false; this is called an error of the second kind or a [Type II error](#).

**Example** In our example, if we reject the restriction  $\mu_0 = 0$  when it is true, we commit a Type I error.

## Critical region

Remember that the sample  $\xi$  is regarded as a realization of a random vector  $\Xi$  having support  $R_\Xi$ .

A test of hypothesis is usually carried out by explicitly or implicitly subdividing the support  $R_\Xi$  into two disjoint subsets.

One of the two subsets, denoted by  $C_\Xi$  is called the critical region (or rejection region) and it is the set of all values of  $\xi$  for which the null hypothesis is rejected:

$$C_\Xi = \{\xi \in R_\Xi : H_0 \text{ is rejected whenever } \xi \text{ is observed}\}$$

The other subset is the complement of the critical region:

$$C_\Xi^c = \{\xi \in R_\Xi : H_0 \text{ is not rejected whenever } \xi \text{ is observed}\}$$

and it is, of course, such that

$$C_\Xi \cup C_\Xi^c = R_\Xi$$

This mathematical formulation is made more concrete in the next section.

## Test statistic

The critical region is often implicitly defined in terms of a test statistic and a critical region for the test statistic.

A test statistic is a random variable

$$S = s(\Xi)$$

whose realization is a function of the sample  $\xi$ .

A critical region for  $S$  is a subset  $C_S \subset \mathbb{R}$  of the set of real numbers and the test is performed based on the test statistic, as follows:

$$s(\xi) \in C_S \Rightarrow \xi \in C_\Xi \Rightarrow H_0 \text{ is rejected}$$

$$s(\xi) \notin C_S \Rightarrow \xi \notin C_\Xi \Rightarrow H_0 \text{ is not rejected}$$

If the complement of the critical region  $C_S$  is an interval, then its extremes are called critical values of

the test. See [this glossary entry](#) for more details about critical values.

**Example** In our example, where we are testing that the mean of the normal distribution is zero, we could use a test statistic called z-statistic. If you want to read the details, go to the lecture on [hypothesis tests about the mean](#).

---

## Power function

The [power function](#) of a test of hypothesis is the function that associates the probability of rejecting  $H_0$  to each parameter  $\theta \in \Theta$ .

Denote the critical region by  $C_{\Xi}$ .

The power function  $\pi(\theta)$  is defined as follows:

$$\pi(\theta) = \mathbf{P}_{\theta}(\Xi \in C_{\Xi})$$

where the notation  $\mathbf{P}_{\theta}$  is used to indicate the fact that the [probability](#) is calculated using the distribution function  $F_{\Xi}(\xi; \theta)$  associated to the parameter  $\theta$ .

## Size of a test

When  $\theta \in \Theta_R$ , the power function  $\pi(\theta)$  gives us the probability of committing a Type I error, that is, the probability of rejecting the null hypothesis when the null hypothesis is true.

The maximum probability of committing a Type I error is, therefore,

$$\alpha = \sup_{\theta \in \Theta_R} \pi(\theta)$$

This maximum probability is called the [size of the test](#).

The size of the test is also called by some authors the level of significance of the test. However, according to other authors, who assign a slightly different meaning to the term, the level of significance of a test is an upper bound on the size of the test.

In mathematical terms, the level of significance is a constant  $\alpha$  that, to the statistician's knowledge, satisfies

$$\alpha \geq \sup_{\theta \in \Theta_R} \pi(\theta)$$

## Criteria to evaluate tests

Tests of hypothesis are most commonly evaluated based on their size and power.

An ideal test should have:

- size equal to 0 (i.e., the probability of rejecting the null hypothesis when the null hypothesis is true should be 0);
- power equal to 1 when  $\theta_0 \notin \Theta_R$  (i.e. the probability of rejecting the null hypothesis when the null hypothesis is false should be 1).

Of course, such an ideal test is never found in practice, but the best we can hope for is a test with a very small size and a very high probability of rejecting a false hypothesis. Nevertheless, this ideal is routinely used to choose among different tests.

For example:

- if we choose between two tests having the same size, we will always utilize the test that has the higher power when  $\theta_0 \notin \Theta_R$ ;
- if we choose between two tests that have the same power when  $\theta_0 \notin \Theta_R$ , we will always utilize the test that has the smaller size.

Several other criteria, beyond power and size, are used to evaluate tests of hypothesis. We do not discuss them here, but we refer the reader to the very nice exposition in Berger and Casella (2002).

## Examples

Examples of how the mathematics of hypothesis testing works can be found in the following lectures:

1. [Hypothesis tests about the mean](#) (examples of tests of hypothesis about the mean of an unknown distribution);
2. [Hypothesis tests about the variance](#) (examples of tests of hypothesis about the variance of an unknown distribution).

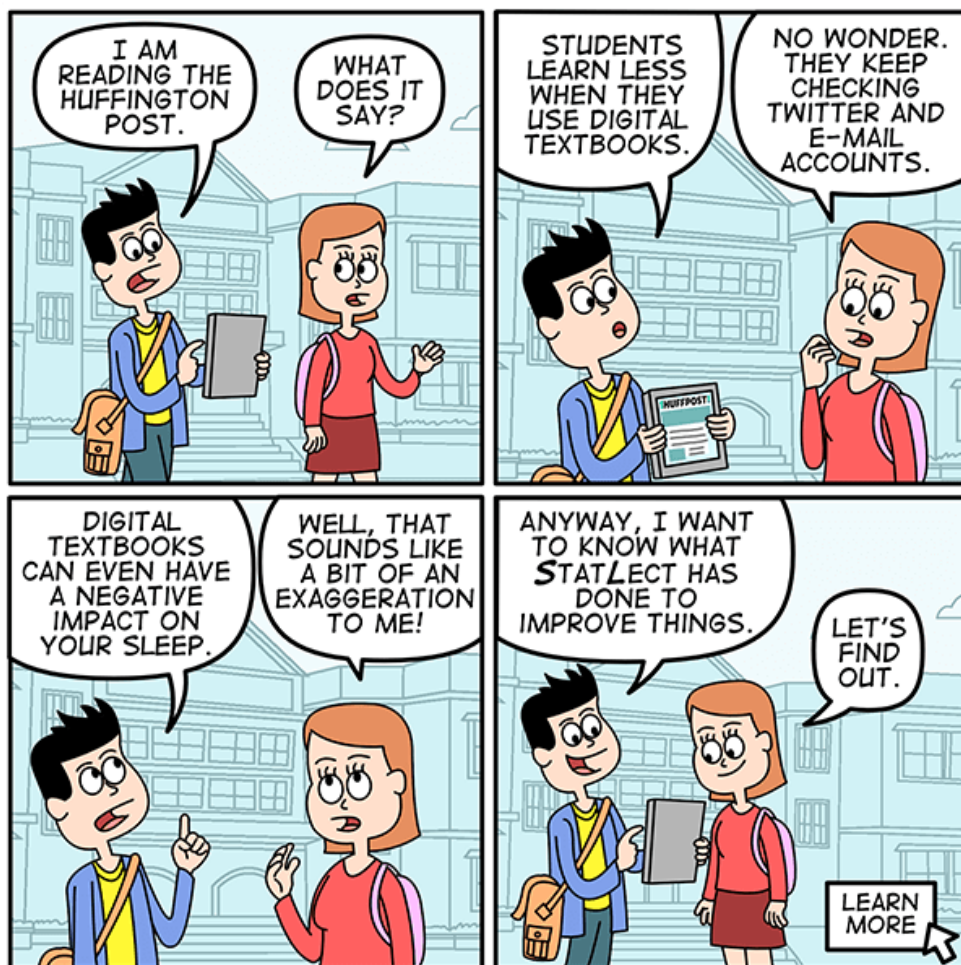
## References

Berger, R. L. and G. Casella (2002) "Statistical inference", Duxbury Advanced Series.

## How to cite

Please cite as:

Taboga, Marco (2021). "Hypothesis testing", Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. <https://www.statlect.com/fundamentals-of-statistics/hypothesis-testing>.





# The books

Most of the learning materials found on this website are now available in a traditional textbook format.

Probability and statistics

Matrix algebra

## Featured pages

Central Limit Theorem  
Beta distribution  
F distribution  
Point estimation  
Score test  
Bernoulli distribution

## Main sections

Mathematical tools  
Fundamentals of probability  
Probability distributions  
Asymptotic theory  
Fundamentals of statistics  
Glossary

## Glossary entries

Critical value  
Almost sure  
Continuous random variable  
Alternative hypothesis  
Probability density function  
Integrable variable

## Explore

Hypothesis testing  
Likelihood ratio test  
Multinomial distribution

## About

About Statlect  
Contacts  
Cookies, privacy and terms of use

## Share

To enhance your privacy,  
we removed the social buttons,  
but **don't forget to share.**