

What is Hypothesis Testing?

Hypothesis testing can be defined as the statistical framework which can be used to test whether the claim made about anything is true or otherwise. Take a look at the examples of the following claim. This will require hypothesis testing.

- Students of class X studying for more than 6 hours a day on average secures more than 75% marks.
- Anyone walking more than 4 km for 30 consecutive days will lose more than 2 kg weight.

Hypothesis testing requires the formulation of the **null hypothesis** and the **alternate hypothesis**. The null hypothesis represents the default state of belief in the real world. For example, the coin is fair. Or, the dice is fair. The alternate hypothesis represents something different and unexpected. The following represent key steps in hypothesis testing. Note the usage of p-value:

- Formulate the null and alternative hypotheses.
- Define the test statistic (such as z-value, or, t-value) that summarizes the strength of evidence against the null hypothesis.
- Determine the appropriate level of significance (0.1, 0.05, or 0.01)
- **Compute the p-value** that quantifies the probability of having obtained a comparable or more extreme value of the test statistic given that the null hypothesis is true.
- **Based on the p-value and level of significance**, decide whether the test outcome is statistically significant and hence reject the null hypothesis or otherwise.

What is P-VALUE?

In hypothesis testing, once the test statistics are determined to evaluate the null hypothesis, the next step is to compute the probability of observing a test statistic as extreme, under the assumption that the null hypothesis H_0 is true. This probability is called the P-value. If the value of p-value is smaller than the level of significance, it provides evidence against the null hypothesis. For example, let's say we are testing the claim that the students studying more than 6 hours on average get more than 75% marks. Here are the steps followed:

- The null hypothesis is that there is no relationship between students for more than 6 hours a day on average and them getting more than 75% marks.
- The data sample for 30 students is gathered.
- Test statistics is **t-value** and the level of significance is set as 0.05.
- The mean of 30 students is found to be 79% with a standard deviation of 4%
- T-value comes out to be +5.48.
- P-value comes out to be less than 0.00001.
- As the p-value is less than the significance level of 0.05, the test outcome is statistically significant.
- The null hypothesis can thus be rejected. Thus, based on the evidence, the claim can be accepted that students studying more than 6 hours a day on average score more than 75%.

The p-value can be defined as the probability of obtaining test statistics more extreme than the ones observed if we repeated the experiment many many times, provided the null hypothesis holds good.

It is measured using techniques such as determining the test statistics such as Z, T, or chi-square and calculating P-value using the related distribution tables such as z-distribution, t-distribution, or chi-square distribution respectively. The distribution of the test statistic for null hypothesis testing will depend on the details of what type of null hypothesis is being tested, and what type of test statistic is used. In general, most commonly-used test statistics follow a well-known statistical distribution under the null hypothesis — such as a normal distribution, at-distribution, a χ^2 -distribution, or an F-distribution.

P-VALUE explained with Examples

The examples below do not take into consideration the test statistics and p-value for explaining whether the null hypothesis can be rejected or otherwise. These examples however intend to explain the concepts regarding the p-value and the notion regarding rejecting or failing to reject the null hypothesis.

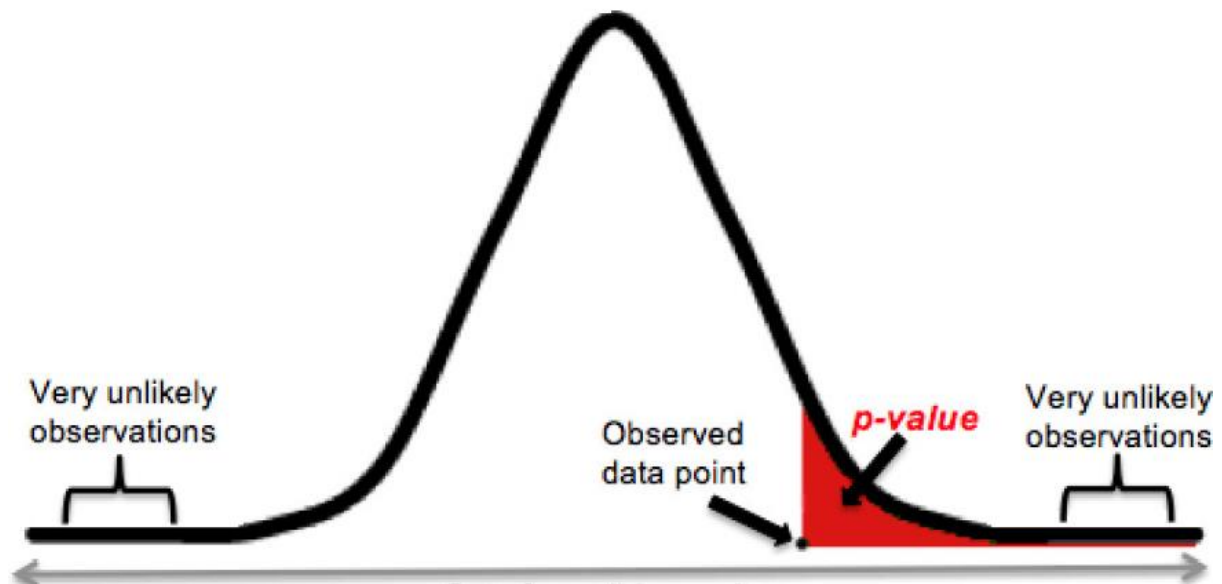
Let's take a quick example to understand the concept of P-value. Given the co-ed school consisting of both boys and girls students, let's **test the hypothesis that the boys on average always score a greater percentage of marks than girls in the school**. In order to test the hypothesis, as a first step, we will need to formulate the **null and alternate hypotheses**.

- Null hypothesis: The null hypothesis is the fact that it is not true that boys on average always score a greater percentage of marks than girls.
- Alternate hypothesis: The alternate hypothesis is that the boys on average always score a greater percentage of marks than girls. As part of the test, several random samples of 30 students are taken comprising of both boys.

Out of 10 different samples of 30 students taken comprising of both boys and girls, in 6 samples, the boys on average were found to score a greater percentage of marks than girls. Do we have enough evidence to reject the null hypothesis? It does not look like that. In case, out of 10 samples, in 9 samples, the boys would have been found to score the greater marks than girls, this would not have been a likely result given the null hypothesis is true. Thus, we would have rejected the null hypothesis in favor of the alternate hypothesis – the claim we made in the beginning. However, given that only 6 samples are found where the boys on average scored greater marks than girls, we do have some evidence but not enough to reject the null hypothesis. In this scenario, we will fail to reject the null hypothesis.

The P-VALUE is used to represent whether the outcome of a hypothesis test is statistically significant enough to be able to reject the null hypothesis. It lies between 0 and 1.

The threshold value below which the P-VALUE becomes statistically significant is usually set to be 0.05. The threshold value is called the **level of significance** and is a function of confidence level. One could choose to set different threshold values (such as 0.025 or 0.01) based on the confidence level based on which one could choose to reject the null hypothesis. A detailed explanation on significance level is provided in one of my related posts titled Level of significance and hypothesis testing. The following diagram represents the p-value of the test statistics as the area of the shaded region (with red).



What is P-Value?

P Value is a probability score that is used in statistical tests to establish the statistical significance of an observed effect.

“a p-value is the probability under a specified statistical model that a statistical summary of the data (e.g., the sample mean difference between two compared groups) would be equal to or more extreme than its observed value.”

When and how is p-value used?

p-values are often reported whenever you perform a statistical significance test (like t-test, chi-square test etc). These tests typically return a computed test statistic and the associated p-value. This reported value is used to establish the statistical significance of the relationships being tested.

The p-value reported is used to make a decision on whether the null hypothesis being tested can be rejected or not.

how to frame a Null hypothesis in general?

While the null hypothesis itself changes with every statistical test, there is a general principle to frame it:

how would the alternate hypothesis would look like?

The **alternate hypothesis (HA)** is always framed to negate the null hypothesis. The corresponding HA for above tests are as follows:

- **Welch Two Sample t-Test:** The true difference in means of two samples is NOT equal to 0
- **Linear Regression:** The beta coefficient(slope) of the X variable is NOT zero
- **Chi Square test:** The difference between expected frequencies and observed frequencies is NOT zero.