

Hypothesis Testing
Normal Distribution
Statistical Significance

What is P-Value:-

Trick:-*If p value is LOW null_hypothesis is GO:- i,e,,(rejected) or (Statistically_significant) or (alternative hypothesis) or (Independent variable affect the dependent variable, then it shows the alternative hypothesis condition.)

*If p value is HIGH null_hypothesis is FLY:- i,e,,(accepted) or (No_Statistically_significant) or (null_hypothesis) or (Independent variable doesn't affect the dependent variable, then it shows the null_hypothesis condition.)

In Statistical hypothesis testing, the P-value or sometimes called probability value, is used to observe the test results or more extreme results by assuming that the null hypothesis (H_0) is true. In data science, there are lots of concepts that are borrowed from different disciplines, and the p-value is one of them. The concept of p-value comes from statistics and widely used in machine learning and data science

*P-value is also used as an alternative to determine the point of rejection in order to provide the smallest significance level at which the null hypothesis is least or rejected.

*It is expressed as the level of significance that lies between 0 and 1, and if there is smaller p-value, then there would be strong evidence to reject the null hypothesis. If the value of p-value is very small, then it means the observed output is feasible but doesn't lie under the null hypothesis conditions (H_0).

*The p-value of 0.05 is known as the level of significance (α). Usually, it is considered using two suggestions, which are given below:

*If $p\text{-value} > 0.05$: The large p-value shows that the null hypothesis needs to be accepted.

*If $p\text{-value} < 0.05$: The small p-value shows that the null hypothesis needs to be rejected, and the result is declared as statically significant.

*In Statistics, our main goal is to determine the statistical significance of our result, and this statistical significance is made on below three concepts:

- a)Hypothesis Testing
- b)Normal Distribution
- c)Statistical Significance

Hypothesis

Testing:-

Hypothesis testing can be defined between two terms; Null hypothesis and Alternative Hypothesis.

It is used to check the validity of the null hypothesis or claim made using the sample data. Here, the null hypothesis (H_0) is defined as the hypothesis with no statistical significance between two variables, while an alternative hypothesis is defined as the hypothesis with a statistical significance between the two variables. No significant relationship between the two variables tells that one variable will not affect the other variable. Thus, the Null hypothesis tells that what you are going to prove doesn't actually happen. If the independent variable affects the dependent variable, then it shows the alternative hypothesis condition.

*In simple terms Hypothesis is an assumption. World is dynamic assumption at time is may be dangerous.

*When we test the assumption, there are two possibilities.

1) The assumption continues to hold good. i.e., Null Hypothesis

2) We have a reason to believe that the assumption doesn't hold good anymore. i.e., Alternate hypothesis

see that video (<https://www.youtube.com/watch?v=Q1yu6TQZ79w>) for complete understanding (https://youtu.be/d0eVIUyt_Uc)

statistical significance:- means that the result that was produced has a reason behind it, it was not produced randomly, or by chance.

*hypothesis is a claim about a population parameter.

*A hypothesis test is a formal procedure to check if a hypothesis is true or not. For example, we want to check the following claim:

"The average height of people in Denmark is equal to 170 cm."

Null Hypothesis:-

it says that changes in assumptions and reality are by chance and there is no relationship b/w them.

Ex:- The average height of people in Denmark is equal to 170 cm, suppose if the randomly selected samples have avg height of 160 (not equal to 170/assumed value). We consider it happens only by selecting people who have 160 in the whole population. It is called as null hypothesis. There is no relationship b/w them.

Alternate Hypothesis:-

it says that changes in assumptions and reality are not by chance and there is no relationship b/w them or it may be 2 different populations.

See the video for complete understanding about null and alternative hypothesis (https://youtu.be/d0eVIUyt_Uc)

In a simple way, we can say that in hypothesis testing, first, we make a claim that is assumed as a null hypothesis using the sample data. If this claim is found invalid, then the alternative hypothesis is selected. This assumption or claim is validated using the p-value to see if it is statistically significant or not using

the evidence. If the evidence supports the alternative hypothesis, then the null hypothesis is rejected.

Steps for Hypothesis testing:-

Below are the steps to perform an experiment for hypothesis testing:

Claim or state a Null hypothesis for the experiment.
State the alternative hypothesis, which is opposite to the null hypothesis.
Set the value of alpha to be used in the experiment.
Determine the z-score using the normal distribution.
Compare the P-value to validate the statistical significance.

Normal

Distribution:-

The normal distribution, which is also known as Gaussian distribution, is the Probability distribution function. It is symmetric about the mean, and use to see the distribution of data using a graph plot. It shows that data near the mean is more frequent to occur as compared to data which is far from the mean, and it looks like a bell-shaped curve. The two main terms of the normal distribution are mean(μ) and standard deviation(σ). For a normal distribution, the mean is zero, and the standard deviation is 1.

In hypothesis testing, we need to calculate z-score. Z-score is the number of standard deviations from the mean of data-point.

P-Value

Here, the z-score inform us that where the data lies compared to the average population.

Statistical

significance:-

To determine the statistical significance of the hypothesis test is the goal of calculating the p-value. To do this, first, we need to set a threshold, which is said to be alpha. We should always set the value of alpha before the experiment, and it is set to be either 0.05 or 0.01(depending on the type of problem).

The result is concluded as a significant result if the observed p-value is lower than alpha.

Errors in P-value

Two types of errors are defined for the p-value; these errors are given below:

Type I error

Type II error

Type I Error:

It is defined as the incorrect or false rejection of the Null hypothesis. For this error, the maximum probability is α , and it is set in advance. The error is not affected by the sample size of the dataset. The type I error increases as we increase the number of tests or endpoints.

Type II error

Type II error is defined as the wrong acceptance of the Null hypothesis. The probability of type II error is β , and the β depends upon the sample size and value of α . The β cannot be determined as the function of the true population effect. The value of β is inversely proportional to the sample size, and it means β decreases as the sample size increases.

The value of β also decreases when we increase the number of tests or endpoints.

We can understand the relationship between hypothesis testing and decision on the basis of the below table: