

# Hypothesis Testing

A statistical hypothesis test is a method of statistical inference used to decide whether the data at hand sufficiently support a particular hypothesis. Hypothesis testing allows us to make probabilistic statements about population parameters.

## Null and Alternate Hypothesis

In simple terms, the null hypothesis is a statement that assumes there is no significant effect or relationship between the variables being studied. It serves as the starting point for hypothesis testing and represents the **status quo** or the assumption of no effect until proven otherwise. The purpose of hypothesis testing is to gather evidence (data) to either reject or fail to reject the null hypothesis in favour of the alternative hypothesis, which claims there is a significant effect or relationship.

## Alternative hypothesis ( $H_1$ or $H_a$ ):

The alternative hypothesis, is a statement that contradicts the null hypothesis and claims there is a significant effect or relationship between the variables being studied. It represents the **research hypothesis** or the claim that the researcher wants to support through statistical analysis.

### Important Points

- How to decide what will be Null hypothesis and what will be Alternate Hypothesis (Typically the Null hypothesis says nothing new is happening)
- We try to gather evidence to reject the null hypothesis
- It's important to note that failing to reject the null hypothesis doesn't necessarily mean that the null hypothesis is true; it just means that there isn't enough evidence to support the alternative hypothesis.

Hypothesis tests are similar to jury trials, in a sense. In a jury trial,  $H_0$  is similar to the not-guilty verdict, and  $H_a$  is the guilty verdict. You assume in a jury trial that the defendant isn't guilty unless the prosecution can show beyond a reasonable doubt that he or she is guilty. If the jury says the evidence is beyond a reasonable doubt, they reject  $H_0$ , not guilty, in favour of  $H_a$ , guilty.

## Steps involved in Hypothesis Testing

### Rejection Region Approach

1. Formulate a Null and Alternate hypothesis
2. Select a significance level(This is the probability of rejecting the null hypothesis when it is actually true, usually set at 0.05 or 0.01)
3. Check assumptions (example distribution)
4. Decide which test is appropriate(Z-test, T-test, Chi-square test, ANOVA)
5. State the relevant test statistic
6. Conduct the test
7. Reject or not reject the Null Hypothesis.
8. Interpret the result

### Performing a Z test Example 1

Suppose a company is evaluating the impact of a new training program on the productivity of its employees. The company has data on the average productivity of its employees before implementing the training program. The average productivity was 50 units per day with a known population standard deviation of 5 units. After implementing the training program, the company measures the productivity of a random sample of 30 employees. The sample has an average productivity of 53 units per day. The company wants to know if the new training program has significantly increased productivity.

### Example 2

Suppose a snack food company claims that their Lays wafer packets contain an average weight of 50 grams per packet. To verify this claim, a consumer watchdog organization decides to test a random sample of Lays wafer packets. The organization wants to determine whether the actual average weight differs significantly from the claimed 50 grams. The organization collects a random sample of 40 Lays wafer packets and measures their weights. They find that the sample has an average weight of 49 grams, with a known population standard deviation of 4 grams.

### Rejection Region

**Significance level**-denoted as  $\alpha$  (alpha), is a predetermined threshold used in hypothesis testing to determine whether the null hypothesis should be rejected or not. It represents the probability of rejecting the null hypothesis when it is actually true, also known as Type 1 error.

**The rejection region** is the region of values that corresponds to the rejection of the null hypothesis at some chosen probability level.

## Type 1 vs Type 2 Error

In hypothesis testing, there are two types of errors that can occur when making a decision about the null hypothesis: Type I error and Type II error.

**Type-I (False Positive) error** occurs when the sample results, lead to the rejection of the null hypothesis when it is in fact true.

In other words, it's the mistake of finding a significant effect or relationship when there is none. The probability of committing a Type I error is denoted by  $\alpha$  (alpha), which is also known as the significance level. By choosing a significance level, researchers can control the risk of making a Type I error.

**Type-II (False Negative) error** occurs when based on the sample results, the null hypothesis is not rejected when it is in fact false.

This means that the researcher fails to detect a significant effect or relationship when one actually exists. The probability of committing a Type II error is denoted by  $\beta$  (beta).

Trade-off between Type 1 and Type 2 errors

## One sided vs two sided test

**One-sided (one-tailed) test:** A one-sided test is used when the researcher is interested in testing the effect in a specific direction (either greater than or less than the value specified in the null hypothesis). The alternative hypothesis in a one-sided test contains an inequality (either ">" or "<").

Example: A researcher wants to test whether a new medication increases the average recovery rate compared to the existing medication.

**Two-sided (two-tailed) test:** A two-sided test is used when the researcher is interested in testing the effect in both directions (i.e., whether the value specified in the null hypothesis is different, either greater or lesser). The alternative hypothesis in a two-sided test contains a "not equal to" sign ( $\neq$ ).

Example: A researcher wants to test whether a new medication has a different average recovery rate compared to the existing medication.

The main difference between them lies in the directionality of the alternative hypothesis and how the significance level is distributed in the critical regions.

### Advantages and Disadvantages?

#### Two-tailed test (two-sided):

##### Advantages:

- 1. Detects effects in both directions:** Two-tailed tests can detect effects in both directions, which makes them suitable for situations where the direction of the effect is uncertain or when researchers want to test for any difference between the groups or variables.
- 2. More conservative:** Two-tailed tests are more conservative because the significance level ( $\alpha$ ) is split between both tails of the distribution. This reduces the risk of Type I errors in cases where the direction of the effect is uncertain.

##### Disadvantages:

- 1. Less powerful:** Two-tailed tests are generally less powerful than one-tailed tests because the significance level ( $\alpha$ ) is divided between both tails of the distribution. This means the test requires a larger effect size to reject the null hypothesis, which could lead to a higher risk of Type II errors (failing to reject the null hypothesis when it is false).
- 2. Not appropriate for directional hypotheses:** Two-tailed tests are not ideal for cases where the research question or hypothesis is directional, as they test for differences in both directions, which may not be of interest or relevance.

#### One-tailed test (one-sided):

##### Advantages:

- 1. More powerful:** One-tailed tests are generally more powerful than two-tailed tests, as the entire significance level ( $\alpha$ ) is allocated to one tail of the distribution. This means that the test is more likely to detect an effect in the specified direction, assuming the effect exists.
- 2. Directional hypothesis:** One-tailed tests are appropriate when there is a strong theoretical or practical reason to test for an effect in a specific direction.

##### Disadvantages:

- 1. Missed effects:** One-tailed tests can miss effects in the opposite direction of the specified alternative hypothesis. If an effect exists in the opposite direction, the test will not be able to detect it, which could lead to incorrect conclusions.
- 2. Increased risk of Type I error:** One-tailed tests can be more prone to Type I errors if the effect is actually in the opposite direction than the one specified in the alternative hypothesis.

## Where can be Hypothesis Testing Applied?

1. **Testing the effectiveness of interventions or treatments:** Hypothesis testing can be used to determine whether a new drug, therapy, or educational intervention has a significant effect compared to a control group or an existing treatment.
2. **Comparing means or proportions:** Hypothesis testing can be used to compare means or proportions between two or more groups to determine if there's a significant difference. This can be applied to compare average customer satisfaction scores, conversion rates, or employee performance across different groups.
3. **Analysing relationships between variables:** Hypothesis testing can be used to evaluate the association between variables, such as the correlation between age and income or the relationship between advertising spend and sales.
4. **Evaluating the goodness of fit:** Hypothesis testing can help assess if a particular theoretical distribution (e.g., normal, binomial, or Poisson) is a good fit for the observed data.
5. **Testing the independence of categorical variables:** Hypothesis testing can be used to determine if two categorical variables are independent or if there's a significant association between them. For example, it can be used to test if there's a relationship between the type of product and the likelihood of it being returned by a customer.
6. **A/B testing:** In marketing, product development, and website design, hypothesis testing is often used to compare the performance of two different versions (A and B) to determine which one is more effective in terms of conversion rates, user engagement, or other metrics.

# Hypothesis Testing ML Applications

**1. Model comparison:** Hypothesis testing can be used to compare the performance of different machine learning models or algorithms on a given dataset. For example, you can use a paired t-test to compare the accuracy or error rate of two models on multiple cross-validation folds to determine if one model performs significantly better than the other.

**2. Feature selection:** Hypothesis testing can help identify which features are significantly related to the target variable or contribute meaningfully to the model's performance. For example, you can use a t-test, chi-square test, or ANOVA to test the relationship between individual features and the target variable. Features with significant relationships can be selected for building the model, while non-significant features may be excluded.

**3. Hyperparameter tuning:** Hypothesis testing can be used to evaluate the performance of a model trained with different hyperparameter settings. By comparing the performance of models with different hyperparameters, you can determine if one set of hyperparameters leads to significantly better performance.

**4. Assessing model assumptions:** In some cases, machine learning models rely on certain statistical assumptions, such as linearity or normality of residuals in linear regression. Hypothesis testing can help assess whether these assumptions are met, allowing you to determine if the model is appropriate for the data.