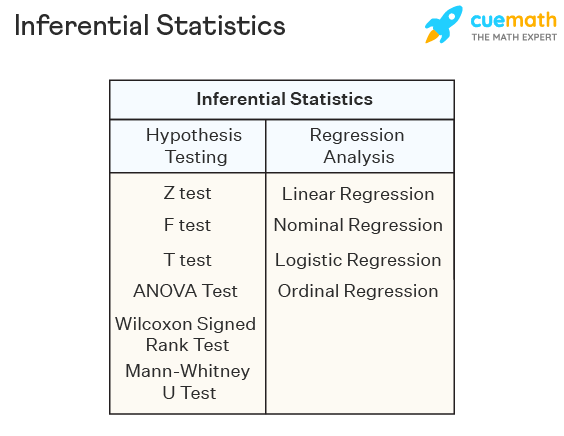
**Types of Inferential Statistics:**

1. Hypothesis Testing:
2. Regression Analysis:



**Hypothesis Testing:**

Hypothesis testing is a type of inferential statistics that is used to test assumptions and draw conclusions about the population from the available sample data.

It involves setting up a **null hypothesis** and an **alternative hypothesis** followed by conducting a **statistical test of significance**. A conclusion is drawn based on the value of the test statistic, the critical value, and the confidence intervals.

**Null Value Testing/Null Hypothesis:**

This is the default or initial assumption. It often represents a statement of no effect, no difference, or no change. It is denoted as H0.

Null hypothesis testing is a formal approach to decide whether a statistical relationship in a sample reflects a real relationship in the population or is just due to chance.

**Alternative hypothesis:**

This is the statement you want to test or the claim you're investigating. It represents the opposite of the null hypothesis and is what you hope to find evidence for. It is denoted as Ha or sometimes as H1.

The alternative hypothesis is a statement used in statistical inference experiments. **It is contradictory to the null hypothesis.**

Given below are certain important hypothesis tests that are used in inferential statistics.

Some technical terms of Hypothesis Testing

1. Confidence Interval
2. Significance Value
3. P-Value

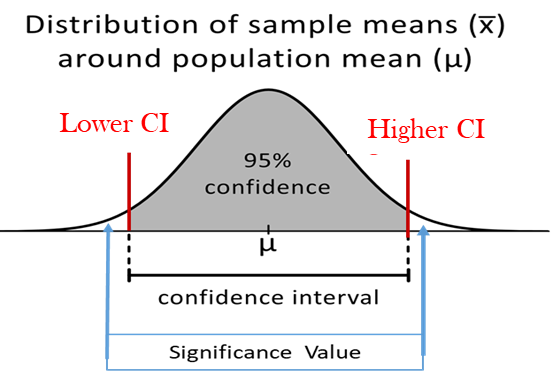
**Confidence Interval(CI):**

This is the range of values you expect your estimate to fall between, which is usually decided by the Domain Expert

**Significance Value(SV):**

The significance level is the probability of **rejecting** the null hypothesis when it is true

SV = 1-CI  
SV = 1-0.95  
SV = 0.5



**Points to be noted for Hypothesis Testing:**

If the output value **is within** the Confidence Interval then we can say the **Null hypothesis is True and Accepted**. In this scenario, we can say we fail to reject the Alternate Hypothesis.

If the output value **is not within** the Confidence Interval then we can say **Null HyPothis is Rejcted** and **Alternate Hypothesis gets Accepted.**

The above steps are called conclusions.

**How to find out Lower CI and higher CI:**

To find out Lower CI and higher CI, we need to 1st understand the Point Estimate

**Point Estimate:**

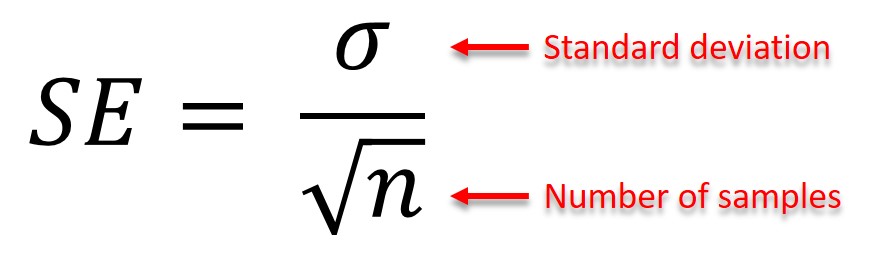
In statistics, a point estimate is a single value that is used to estimate an unknown population parameter based on sample data. When estimating a population parameter, such as the mean, proportion, or standard deviation, a point estimate provides the best guess or approximation of that parameter.

For hypothesis testing we will take **mean** as a Point Estimate.

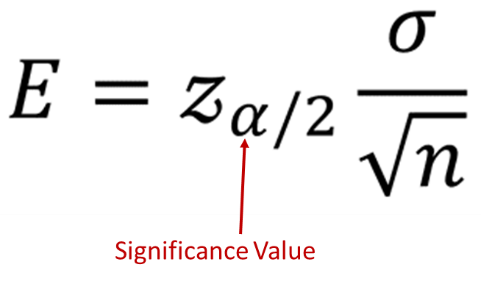
Lower CI: Point estimate – Margin of Error

Higher CI: Point Estimate + Margin of Error

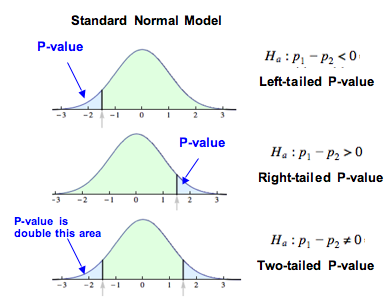
**Formula for Standard Error(SE) :**



**Formula for Margin of Error(E) :**



A hypothesis test can be **left-tailed, right-tailed, or two-tailed.**



**P-Value:**

The p-value is a statistical measure used to determine the significance of a model's results or the effect of a specific feature on the target variable. P-Value is often used to draw conclusions from sample data and make inferences about population parameters.

In hypothesis testing related to model evaluation or feature significance, researchers typically set up null and alternative hypotheses.

1. The null hypothesis (H0) represents the assumption of no effect or no difference,
2. while the alternative hypothesis (H1) represents the opposite.

In the context of machine learning, the p-value is particularly relevant in the following scenarios:

1. **Model Evaluation:**
2. **Feature Selection:**

**Model Evaluation:**

When comparing the performance of different machine learning models, researchers may use statistical tests to assess whether one model significantly outperforms the other. The p-value is used to determine the probability of observing the performance difference (or a more extreme difference) between the models, assuming that there is no real difference in their performance. If the p-value is below a chosen significance level (often 0.05 or 0.01), it is considered statistically significant, indicating that there is evidence to support the claim that one model performs better than the other.

**Feature Selection:**

In some cases, researchers may want to assess the significance of individual features on the target variable. The p-value can be used to evaluate whether the relationship between a particular feature and the target variable is statistically significant. A low p-value suggests that the feature is likely to have a significant effect on the target variable, while a high p-value indicates that the feature is less likely to be important for prediction. This information can guide feature selection and help focus on the most relevant features.

It's important to note that the p-value has limitations, especially when dealing with large datasets or multiple hypothesis tests. Researchers should interpret p-values cautiously and consider other techniques, such as cross-validation, effect size calculations, and practical significance, to make more informed decisions about model features and performance.

**Errors in Hypothesis Testing**

**Type 1 error :**

it happens when the null hypothesis of an experiment is true but rejected often called a false positive.

**type 2 error:**

it occurs when the null hypothesis is false but still not rejected, also known as a false negative.

**Which error is considered to be more dangerous?**

Type 1 error is considered to be worse or more dangerous than type 2 because to reject what is true is more harmful than keeping the data that is not true.

**Different hypothesis tests are used in Hypothesis Testing.**

* Z Test:
* T-Test:
* F Test:
* Annona Test: -
* Chi-Square Test: