**Estimation:**

Estimation in inferential statistics involves using sample data to estimate unknown population parameters. The goal is to make educated guesses or predictions about the characteristics of a population based on information gathered from a sample.

**Point Estimation:**

Estimating a population parameter (e.g., mean, proportion) using a single value based on sample data.

Point estimation involves using a single value, typically derived from sample data, to estimate the value of an unknown population parameter. For example, if we want to estimate the population mean, we might use the sample mean as our point estimate. Similarly, if we want to estimate the population proportion, we might use the sample proportion.

Example: If we calculate the mean height of a sample of 100 individuals and find it to be 170 cm, we might use this value (170 cm) as our point estimate for the population's mean height.

**Interval Estimation:**

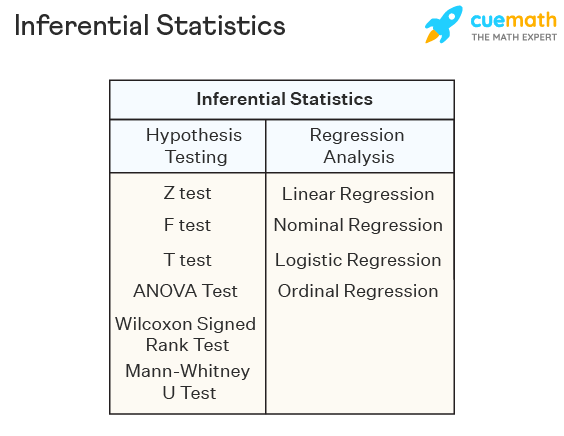
Estimating a range of values within which the population parameter is likely to fall, along with a level of confidence.

Interval estimation involves providing a range of values, known as a confidence interval, within which the true value of the population parameter is likely to lie. Unlike point estimation, which provides a single estimate, interval estimation acknowledges the uncertainty inherent in estimation by providing a range of plausible values.

Example: If we calculate a 95% confidence interval for the population mean height based on our sample data, we might find that the interval ranges from 165 cm to 175 cm. This means that we are 95% confident that the true population mean height falls within this interval.

**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.

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

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

**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.**

Some technical terms of Hypothesis Testing

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

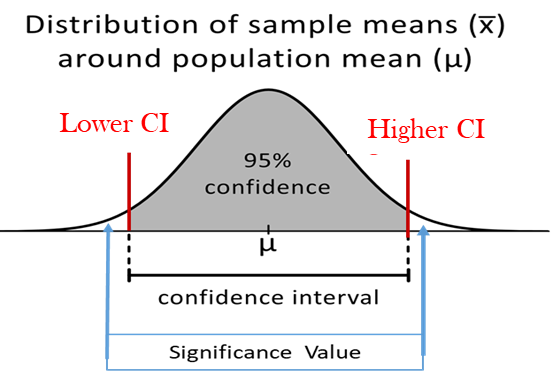
**Confidence Interval(CI):**

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 = 0.05



**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:**

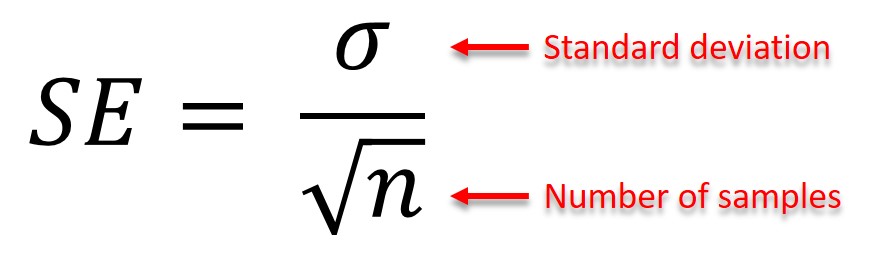
In statistics, a point estimate is a single value that is used to estimate an unknown population parameter based on sample data. When estimating a population parameter, such as the mean, proportion, or standard deviation, a point estimate provides the best guess or approximation of that parameter.

For example, if you want to estimate the mean height of all adults in a country, you might take a sample of individuals and calculate the mean height of that sample. This calculated mean would be the point estimate of the population mean height.

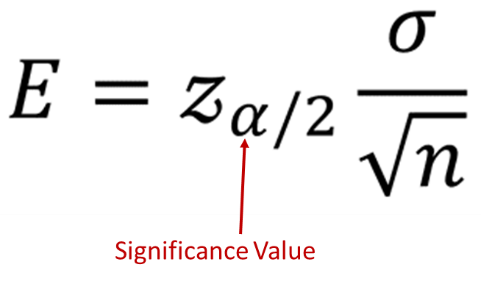
Lower CI: Point estimate – Margin of Error

Higher CI: Point Estimate + Margin of Error

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



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



**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. It is a fundamental concept in statistical hypothesis testing, which 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:

**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.

Given below are certain important hypothesis tests that are used in Hypothesis Testing.

**Z Test:** A [z test](https://www.cuemath.com/data/z-test/) is used on data that follows a [normal distribution](https://www.cuemath.com/normal-distribution-formula/) and has a sample size greater than or equal to 30. It is used to test if the [means](https://www.cuemath.com/data/mean/) of the sample and population are equal when the population variance is known. The right-tailed hypothesis can be set up as follows:

Null Hypothesis: H0: μ=μ0

Alternate Hypothesis: H1 : μ>μ0

Test Statistic: z = ¯¯¯x−μσ√n�¯−���. ¯¯¯x�¯ is the sample mean, μ is the population mean, σ� is the population standard deviation and n is the sample size.

Decision Criteria: If the z statistic > z critical value then reject the null hypothesis.

**T Test:** A[t test](https://www.cuemath.com/t-test-formula/) is used when the data follows a [student t distribution](https://www.cuemath.com/t-distribution-formula/) and the sample size is lesser than 30. It is used to compare the sample and population mean when the population variance is unknown. The hypothesis test for inferential statistics is given as follows:

Null Hypothesis: H0�0 : μ=μ0�=�0

Alternate Hypothesis: H1�1 : μ>μ0�>�0

Test Statistics: t = ¯¯¯x−μs√n�¯−���

Decision Criteria: If the t statistic > t critical value then reject the null hypothesis.

**F Test:** An [f test](https://www.cuemath.com/data/f-test/) is used to check if there is a difference between the [variances](https://www.cuemath.com/data/variance/) of two samples or populations. The right tailed f hypothesis test can be set up as follows:

Null Hypothesis: H0�0 : σ21=σ22�12=�22

Alternate Hypothesis: H1�1 : σ21>σ22�12>�22

Test Statistic: f = σ21σ22�12�22, where σ21�12 is the variance of the first population and σ22�22 is the variance of the second population.

Decision Criteria: If the f test statistic > f test critical value then reject the null hypothesis.

**Type**1**or Type**2**error in statistics.**

* A type 1 error happens when the null hypothesis of an experiment is true but rejected often called a false positive.
* A type 2 error occurs when the null hypothesis is false but still not rejected, also known as a false negative.
* 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.