**Standard Deviation:**

Standard deviation is a measure of the amount of variation or dispersion in a set of values. It quantifies the amount of uncertainty or deviation from the mean (average) of a dataset.

Here's how you calculate the standard deviation for a set of numbers:

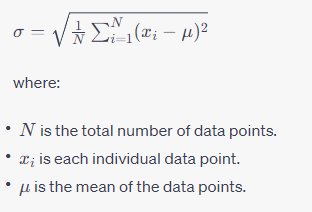
1. Calculate the mean (average): Add up all the numbers and divide by the total count of numbers.

2. Calculate the squared differences from the mean: Subtract the mean from each number and square the result.

3. Calculate the variance: Find the average of the squared differences.

4. Calculate the standard deviation: Take the square root of the variance.

The formula for the standard deviation (σ) is:



The standard deviation is a valuable tool in statistics for understanding the spread of data and making comparisons between different data sets. A higher standard deviation indicates greater variability, while a lower standard deviation indicates that the data points are closer to the mean.

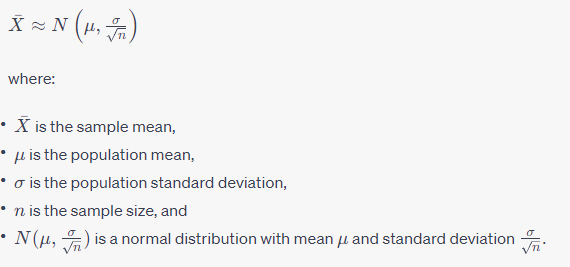
**Central limit theorem**:

The Central Limit Theorem (CLT) is a fundamental principle in probability theory and statistics. It describes the behavior of the sum or average of a large number of independent, identically distributed random variables, regardless of the shape of their original distribution. The theorem states that, under certain conditions, the distribution of the sum or average will be approximately normal, regardless of the underlying distribution of the individual variables.

More formally, the Central Limit Theorem can be stated as follows:

Suppose are independent and identically distributed random variables with mean and standard deviation Then, as n (the sample size) becomes large, the distribution of the sample mean approaches a normal distribution with mean and standard deviation .

Mathematically, this can be expressed as:



The Central Limit Theorem is significant in statistics because it allows us to use the properties of the normal distribution to make inferences about the population, even if the original population is not normally distributed. It is the basis for many statistical methods and hypothesis testing procedures.

**Bias and variance-**

Bias and variance are two fundamental concepts in machine learning and statistics that are related to the performance and generalization ability of models.

1. Bias:

- Bias refers to the error due to overly simplistic assumptions in the learning algorithm.

- A high bias means the model is too simple and may miss important patterns in the data, leading to underfitting.

- Underfitting occurs when the model cannot capture the underlying trend of the data and performs poorly on both the training and unseen data.

2. Variance:

- Variance refers to the error due to too much complexity in the learning algorithm.

- A high variance means the model captures noise in the training data and does not generalize well to unseen data, leading to overfitting.

- Overfitting occurs when the model learns the detail and noise in the training data to the extent that it negatively impacts its performance on unseen data.

**Here's a visual representation of bias-variance tradeoff:**

**1- High Bias, Low Variance:**

- The model is overly simplistic, making strong assumptions about the data.

- It tends to miss important patterns and performs consistently poorly across different datasets.

- Example: Linear regression on a non-linear dataset.

**2- High Variance, Low Bias:**

- The model is overly complex, capturing noise in the training data.

- It fits the training data well but does not generalize well to unseen data.

- Example: Very deep neural networks with no regularization on a small dataset.

**3- Optimal Tradeoff:**

- The goal is to find the right level of model complexity to achieve a good balance between bias and variance.

- This leads to a model that generalizes well to unseen data.

- Example: A well-tuned random forest model.

To build a robust and accurate model, it's important to manage bias and variance effectively. Techniques like cross-validation, regularization, ensemble methods, and appropriate model selection help strike a good bias-variance balance and improve model performance.