Skewness& Variance& Standard Deviation

Done By:

Manar Qadi

Skewness: Skewness measures the asymmetry of a probability distribution. It indicates the degree to which a distribution deviates from a symmetrical bell curve. Skewness can be positive, negative, or zero. Positive skewness means the tail of the distribution extends towards the right, while negative skewness means the tail extends towards the left.

Type of Skewness:

Positive skewness: Indicates that the distribution has a tail extending towards higher values.
Negative skewness: Indicates that the distribution has a tail extending towards lower values.
Zero skewness: Indicates a perfectly symmetrical distribution.

Example:

Dataset: [160, 165, 170, 168, 172]
To calculate skewness, we need to find the mean, median, and standard deviation of the dataset. Let's assume the mean is 167. The median can be calculated as (165 + 168) / 2 = 166.5.

Skewness is then determined using the following formula:
Skewness = 3 * (mean - median) / standard deviation

Substituting the values, we get:
Skewness = 3 * (167 - 166.5) / standard deviation

If the standard deviation is, for example, 4, then:

Skewness = 3 * (167 - 166.5) / 4 = 0.375

A positive skewness of 0.375 indicates a right skew, meaning the tail of the distribution extends towards higher values.

Variance: Variance is a statistical measure that quantifies the spread or dispersion of a set of data points around their mean. It provides a measure of how much the data points differ from the average. A higher variance indicates greater variability, while a lower variance suggests that the data points are closer to the mean.

Type of Variance:

1. Population Variance: Population variance is calculated using the entire population of data points. It measures the average squared deviation of each data point from the population mean.

2. Sample Variance: Sample variance is calculated using a sample of data points from a larger population. It is an estimate of the population variance and is used when the entire population is not available. Sample variance typically involves dividing by (n-1) instead of n to provide an unbiased estimate.

3. Explained Variance: Explained variance is commonly used in statistical modeling, especially in techniques like regression analysis or principal component analysis (PCA). It quantifies the amount of variance in the dependent variable that is explained by the independent variables or components.

4. Residual Variance: Residual variance, also known as error variance, measures the unexplained or residual variability in a statistical model. It represents the variance that is not accounted for by the explanatory variables or components.

Example:

Dataset: [160, 165, 170, 168, 172]

To calculate the variance, we need to find the mean of the dataset. Assuming the mean is 167, we can calculate the variance using the following formula:

Variance = $\Sigma((x - \text{mean})^2) / n$

Substituting the values, we get:
Variance = $((160 - 167)^2 + (165 - 167)^2 + (170 - 167)^2 + (168 - 167)^2 + (172 - 167)^2) / 5$

Simplifying the equation, we have:
Variance = $(49 + 4 + 9 + 1 + 25) / 5 = 17.6$

Therefore, the variance of the dataset is 17.6.

Standard Deviation: Standard deviation is another measure of the dispersion or spread of a set of data points. It is the square root of the variance. The standard deviation represents the average amount by which individual data points differ from the mean. Like variance, a higher standard deviation indicates greater variability in the data.

Type of Standard Deviation:

1. Population Standard Deviation: This is the standard deviation calculated using the entire population of data points. It provides a measure of the dispersion of the population data.
2. Sample Standard Deviation: When dealing with a sample from a larger population, the sample standard deviation is used. It is an estimate of the population standard deviation and is calculated slightly differently to account for the degrees of freedom.
3. Weighted Standard Deviation: In cases where different data points have different weights or importance, a weighted standard deviation can be calculated. This involves assigning weights to each data point based on their significance and incorporating those weights in the standard deviation calculation.

Example:

Dataset: [160, 165, 170, 168, 172]

The standard deviation is simply the square root of the variance. Using the variance value from the previous calculation:

Standard Deviation = $\sqrt{\text{Variance}} = \sqrt{17.6} \approx 4.19$