Variance:

Variance is a measure of the spread or dispersion of data points in a dataset.

It is calculated by finding the average of the squared differences between each data point and the mean.

Variance is always a non-negative value. A variance of zero indicates that all the data points are the same, while a higher variance suggests greater variability.

Variance is sensitive to outliers because it considers the squared differences, which can be heavily influenced by extreme values.

It is denoted by the symbol σ^2

(sigma squared).

Deviation (Standard Deviation):

Standard deviation is a measure of the spread of data points around the mean.

It is the square root of the variance and has the same units as the original data.

Standard deviation provides a more interpretable measure of dispersion compared to variance because it is in the original units.

Like variance, standard deviation is also sensitive to outliers.

It is denoted by the symbol σ (sigma).

Skewness:

Skewness is a measure of the

asymmetry of the distribution of data points.

It quantifies the extent to which a dataset deviates from a symmetric distribution.

A skewness value of zero indicates a perfectly symmetric distribution.

Positive skewness (right-skewed) means the distribution has a longer tail on the right side, with more extreme values on the positive side.

Negative skewness (left-skewed) means the distribution has a longer tail on the left side, with more extreme values on the negative side.

Skewness is calculated by considering the mean, standard deviation, and the differences between each data point and the mean.

Skewness can help identify

departures from normality in a dataset.

example of a dataset of exam scores: [60, 65, 70, 75, 80].

Variance:

To calculate the variance, we follow these steps:

Step 1: Find the mean (average) of the dataset:

(60 + 65 + 70 + 75 + 80) / 5 = 70

Step 2: Subtract the mean from each data point:

(60 - 70) = -10

(65 - 70) = -5

(70 - 70) = 0

(75 - 70) = 5

(80 - 70) = 10

Step 3: Square each difference:

(-10)^2 = 100

(-5)^2 = 25

(0)^2 = 0

(5)^2 = 25

(10)^2 = 100

Step 4: Average the squared differences:

(100 + 25 + 0 + 25 + 100) / 5 = 50

Therefore, the variance for this dataset is 50.

Deviation (Standard Deviation):

To calculate the standard deviation, we take the square root of the variance:

√(50) ≈ 7.07

So, the standard deviation for this dataset is approximately 7.07.

Skewness:

To calculate skewness, we can use a formula involving the mean, standard deviation, and the differences between each data point and the mean:

Step 1: Calculate the mean (already calculated as 70).

Step 2: Calculate the standard deviation (already calculated as approximately 7.07).

Step 3: Calculate the differences between each data point and the mean:

(60 - 70) = -10

(65 - 70) = -5

(70 - 70) = 0

(75 - 70) = 5

(80 - 70) = 10

Step 4: Calculate the cubed differences for each data point:

(-10)^3 = -1000

(-5)^3 = -125

(0)^3 = 0

(5)^3 = 125

(10)^3 = 1000

Step 5: Average the cubed differences:

(-1000 - 125 + 0 + 125 + 1000) / 5 = 0

Therefore, the skewness for this dataset is 0, indicating a symmetrical distribution.