SKEWNESS ,VARIANCE AND DEVIATION,

أتراب أحمد الدبعي

Skewness:

is a statistical measure that describes the asymmetry and shape of a probability distribution. It quantifies the degree of distortion from a symmetrical bell curve (normal distribution) in a data set. In simpler terms, skewness measures the extent to which the data values are concentrated on one side of the mean compared to the other side.

Skewness can take on different values, indicating different types of skewness:

1. Positive Skewness: Also known as right skewness, it occurs when the tail of the distribution extends towards the right side. In a positively skewed distribution, the majority of the data points are concentrated on the left side, while the right side has a few extreme values. The mean is typically greater than the median in a positively skewed distribution.

2. Negative Skewness: Also known as left skewness, it occurs when the tail of the distribution extends towards the left side. In a negatively skewed distribution, the majority of the data points are concentrated on the right side, while the left side has a few extreme values. The mean is typically less than the median in a negatively skewed distribution.

3. Zero Skewness: A distribution is considered to have zero skewness (symmetrical) if it has equal probabilities of values on both sides of the mean. In a symmetrical distribution, the mean and median are equal.

Skewness is an important measure because it provides insights into the shape of the data distribution and can impact various statistical analyses and modeling techniques. It helps to identify departures from normality and understand the nature of the data.

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 individual data points deviate from the average value.

Mathematically, variance is calculated as the average of the squared differences between each data point and the mean of the data set. It is represented by the symbol σ^2 (sigma squared) for a population variance or s^2 (squared) for a sample variance.

For a population variance (σ^2), the formula is:

$\sigma^2 = \Sigma(x - \mu)^2 / N$

where:

- Σ denotes the sum of the values
- x represents each individual data point in the population
- μ is the mean of the population
- N is the total number of data points in the population

For a sample variance (s^2), the formula is slightly different to account for the degrees of freedom:

$s^2 = \Sigma(x - \bar{x})^2 / (n - 1)$

where:

- Σ denotes the sum of the values
- x represents each individual data point in the sample
- $\bar{x}$ is the mean of the sample
- n is the total number of data points in the sample

The variance is always a non-negative value. A larger variance indicates a greater dispersion of data points from the mean, while a smaller variance indicates a more concentrated distribution around the mean.

Variance is commonly used in statistics and data analysis to understand the variability of a data set, compare different data sets, assess the performance of statistical models, and make inferences about the population based on sample data.

Deviation:

in statistics, refers to the difference between a data point and a reference point, such as the mean or median of a dataset. It measures how much a given value varies from the average or central tendency of the data.

There are two commonly used types of deviation:

1. Deviation from the mean: This type of deviation is calculated by subtracting the mean of the dataset from each individual data point. The formula for deviation from the mean (d) is:

$d = x - \mu$

where:

- x is a data point
- μ is the mean of the dataset

The deviation from the mean can be positive or negative, depending on whether the data point is greater or smaller than the mean, respectively.

2. Deviation from the median: This type of deviation is calculated by subtracting the median of the dataset from each individual data point. The formula for deviation from the median (d) is:

d = x - M

where:

- x is a data point
- M is the median of the dataset

Similar to deviation from the mean, the deviation from the median can also be positive or negative.

Deviation is a fundamental concept in statistics and is used to understand the spread or dispersion of data points around a central value. It plays a crucial role in various statistical calculations, such as calculating variance, standard deviation, and other measures of variability.