ASSIGNMENT TITLE:

Skewness, Variance, Deviation

زهر جلال السقاف

21-16-0028

Skewness:

- **Definition:**

  Skewness is a statistical measure that describes the asymmetry of a probability distribution. It quantifies the extent to which a dataset deviates from a symmetric distribution.

- **Uses:**
  1. Data analysis: Skewness helps in understanding the shape of a distribution. Positive skewness indicates a longer right tail, negative skewness indicates a longer left tail, and zero skewness suggests a symmetric distribution.
  2. Risk assessment: Skewness is used in finance and investment to assess the risk associated with an investment. Skewed distributions may indicate non-normality and can influence investment decisions.
  3. Feature engineering: Skewness is considered when transforming variables in machine learning to improve model performance and meet assumptions.

- **Formula:**

$$\frac{Sk_1 = \frac{X - Mo}{s}}{Sk_2 = \frac{3(X - Md)}{s}}$$

where:

$Sk_1 =$ Pearson's first coefficient of skewness and Sk_2 the second

$s =$ The standard deviation for the sample

$\bar{X} =$ Is the mean value

$Mo =$ The modal (mode) value

$Md =$ Is the median value

Variance:

- **Definition:**

  Variance is a statistical measure that quantifies the spread or dispersion of a dataset. It measures how much the data points deviate from the mean.

- **Uses:**
  1. Descriptive statistics: Variance provides a measure of the variability or spread of data. A higher variance indicates greater dispersion, while a lower variance suggests less variability.
  2. Assessing model performance: Variance is used in regression analysis to evaluate the accuracy of a model. It helps determine how well the model fits the data by measuring the dispersion of residuals.
  3. Quality control: Variance is employed in manufacturing and quality control processes to assess the consistency and variability of product measurements.

- **Formula:**

$$\sigma^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2}{N}$$

where:

x_i = Each value in the data set

$\overline{x}$ = Mean of all values in the data set

N = Number of values in the data set

Deviation:

- **Definition:**

  Deviation, in statistics, refers to the difference between a data point and a reference point, such as the mean or median.

- **Uses:**
  1. Descriptive statistics: Deviation provides insights into the distance of each data point from the central tendency, such as the mean or median.
  2. Outlier detection: Deviation can be used to identify outliers in a dataset by comparing the distance of data points from the mean or median.
  3. Data transformation: Deviation is considered when transforming variables in statistical analyses or machine learning to normalize or standardize the data.

- **Formula:**

$$\text{Standard Deviation} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

where:

$x_i = \text{Value of the } i^{th} \text{ point in the data set}$

$\overline{x} = \text{The mean value of the data set}$

$n = \text{The number of data points in the data set}$