Measure of Dispersion

Measures of dispersion describe the spread or variability of a dataset. They indicate how much the values in a dataset differ from the central tendency

Common Measures of Dispersion

$X = \{1, 2, 3, 4, 5, 6, 7, 8\}$

1) Range
2) Variance
3) Standard Deviation
4) Interquartile Range (IQR) → Percentiles

① Range

Defn : Range is the difference between the maximum and minimum value in a dataset.

Range = Maximum Value − Minimum Value

Eg : Ages $\{14, 13, 10, 20, 25, 75, 15\}$.

Range = 75 − 10 = 65

Characteristics

1) Simple to calculate
2) Sensitive to outliers
3) Rough measure of dispersion

outlier

Weight = $\{35, 40, 45, 39, 30, 70\}$

$\left.\begin{array}{l} \text{Range} = 40 - 30 = 10 \\ \text{Range} = 70 - 30 = 40 \end{array}\right\}$

② Variance

Defn : Variance measures the average squared deviation of each value from the mean. It provides a sense of how much the values in a dataset vary.

Population Variance

$$\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N}$$

$x_i \rightarrow$ DATA Points

$\mu \rightarrow$ population mean

$N \rightarrow$ population size

Sample Variance

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

$x_i \rightarrow$ DATA points

$\bar{x} \rightarrow$ Sample Mean

$n \rightarrow$ Sample Size.

Example $\rightarrow$ Size of a flower petals

$\{5, 8, 12, 15, 20\} \Rightarrow$ Variance of this distribution.

$N = 5$

$$\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N}$$

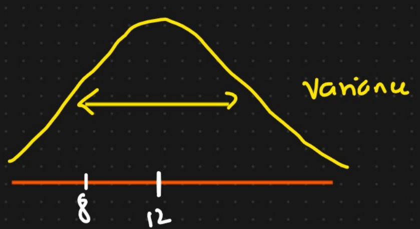
$$\mu = \frac{5+8+12+15+20}{5} = 12$$

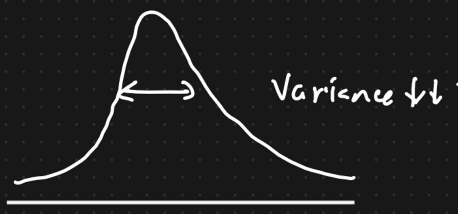
$$\text{Variance} = \frac{(5-12)^2 + (8-12)^2 + (12-12)^2 + (15-12)^2 + (20-12)^2}{5}$$

$$\text{Variance} = 27.6$$

Characteristics

1. Provide a precise measure of variability
2. Units are squared of the original data units.
3. More Sensitive to outliers than the Range.





③ Standard Deviation :

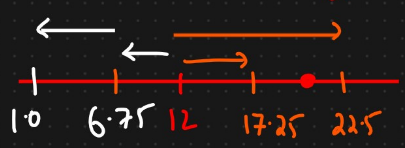
Defn: The standard deviation is the square root of the variance

$\sigma = \sqrt{27.6} \approx 5.25$ $\{5, 8, 12, 15, 20\}$

$$\begin{array}{r} 12.00 \\ 5.25 \\ \hline 6.75 \end{array}$$

$\boxed{\text{Z-score}}$ ←

↳ Standard Normal Distribution



Characteristics

① Provides a clear measure of spread in the same units as the data

② Sensitive to outliers

Key Differences and Similarities

Relationship:

Standard deviation is the square root of variance. If you have the **variance**, you can find the standard deviation by taking the square root of the variance.
Conversely, if you have the standard deviation, you can find the variance by squaring the standard deviation.

Units:

Variance: The units of variance are the square of the units of the original data. For example, if the data is in meters, the variance will be in square meters.

Standard Deviation: The units of standard deviation are the same as the units of the original data. If the data is in meters, the standard deviation will also be in meters.
Interpretation:

Variance: Provides a measure of the dispersion of data points in squared units, which can be difficult to interpret directly.

Standard Deviation: Provides a measure of the dispersion in the same units as the original data, making it easier to interpret and understand.