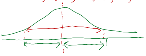
Measue of dispersion

Measure of dispersion describe the spread or variability of a dataset. They indicate how much the value in a data set differ from the central tendency.

Common measures of dispersion

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

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



1) Range

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

$$\text{Range} = \text{Maximum Value} - \text{Minimum Value}$$

Example

Age = $\{14, 13, 10, 20, 25, 75, 15\}$

$\text{Range} = 75 - 13 = 62$

Characteristics:

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

outlier

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

without outlier

$\text{Range} = 45 - 30 = 15$

with outlier

$\text{Range} = 70 - 30 = 40$

2) VARIANCE

Definition measures the average squared deviation of each value from the mean. It provides a sences of how much the values in a dataset vary.

Population variance

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

$X_i \rightarrow$ Data points

$\mu \rightarrow$ population mean

$N \rightarrow$ population size

Sample variance

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

$X_i \rightarrow$ data points

$\bar{X} \rightarrow$ Sample mean

$n \rightarrow$ Sample size

Example :

Size of a flower petals {5,8,12,15,20} => Variance of this distribution

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

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

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

$$\text{Variance} = 27.5$$

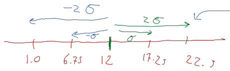
Characteristics:

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 desviation:

Definition: The standard deviation is the square not of the variance.

$$\sigma = \sqrt{\sigma^2} = \sqrt{27.6} \simeq 5.25 \quad \{5, 8, 12, 15, 20\}$$



Z-Score

$\rightarrow$ Standard Normal Distribution

Characteristies

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

2) Sensitive to outliers

3) Commonly