Starting with Statistics

Measures of Central Tendency

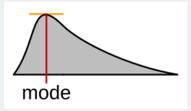
Mean

The mean is the arithmetic average of a set of numbers. It is calculated by adding up all the values in the dataset and then dividing the sum by the number of values.

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Mode

The mode is the value that appears most frequently in a dataset. A dataset can have one mode, more than one mode, or no mode at all if no value repeats.



Median

The median is the middle value in a dataset when the values are arranged in ascending or descending order. If there is an odd number of values, the median is the value exactly in the middle. If there is an even number of values, the median is the average of the two middle values.

$$\text{Med}(X) = \begin{cases} X\left[\frac{n+1}{2}\right] & \text{if n is odd} \\ \frac{X\left[\frac{n}{2}\right] + X\left[\frac{n}{2}+1\right]}{2} & \text{if n is even} \end{cases}$$

Measures of Variability

Range

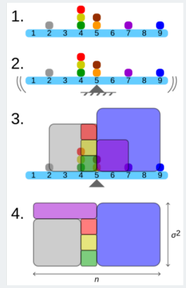
The difference between the highest and lowest values in the dataset.

Range = Max - Min

Variance

The average of the squared differences from the mean. It shows how spread out the numbers are in relation to the mean.

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$



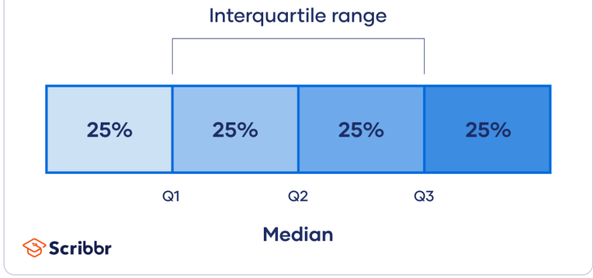
Standard Deviation

The square root of the variance. It gives a measure of how spread out the numbers are in the same unit as the data, making it easier to interpret.

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

Interquartile Range (IQR)

The IQR **describes the middle 50% of values when ordered from lowest to highest.**



Outliers Formula

Lower Outlier = Q1 – (1.5 x IQR)

Higher Outlier = Q3 + (1.5 x IQR)