Descriptive Statistics

Descriptive statistics involve methods for summarizing and describing data, providing insights into its main features, such as central tendency, variability, and distribution.

Inferential Statistics

Inferential statistics, on the other hand, involves making inferences or predictions about populations based on sample data, testing hypotheses, and drawing conclusions.

Mean

Calculated by summing all values in the dataset and dividing by the number of values. It represents the average value and is sensitive to extreme values.

Mean

$$= \frac{\sum x}{n}$$

$$= \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Median

The middle value when the data is ordered from least to greatest. It is less affected by extreme values and provides a measure of central tendency.

Median

$$= x_{(n+1)/2}$$

$$= \frac{x_{(n/2)} + x_{((n/2)+1)}}{2}$$

Mode

The value that appears most frequently in the dataset. It is suitable for both numerical and categorical data.

Mode

Higher Frequency

Range

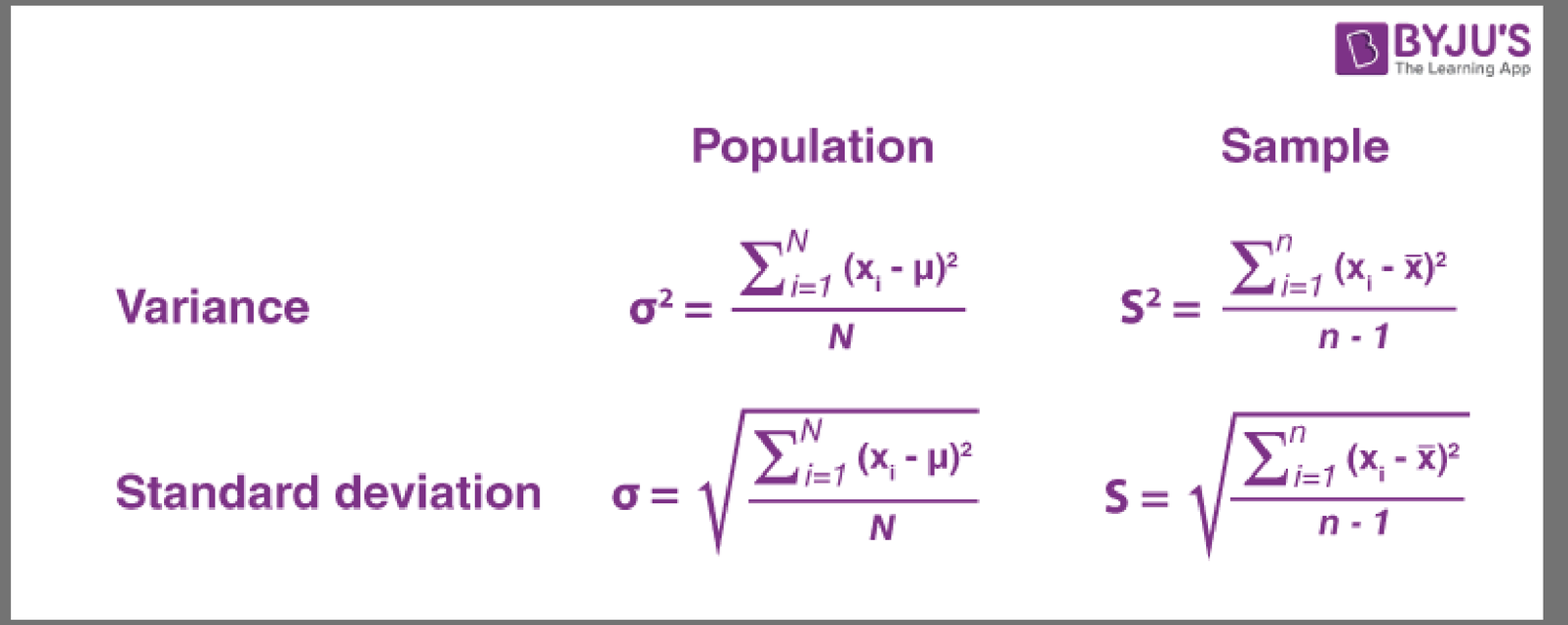
The difference between the maximum and minimum values in the dataset, providing a measure of dispersion.

Range

Highest value

Minus

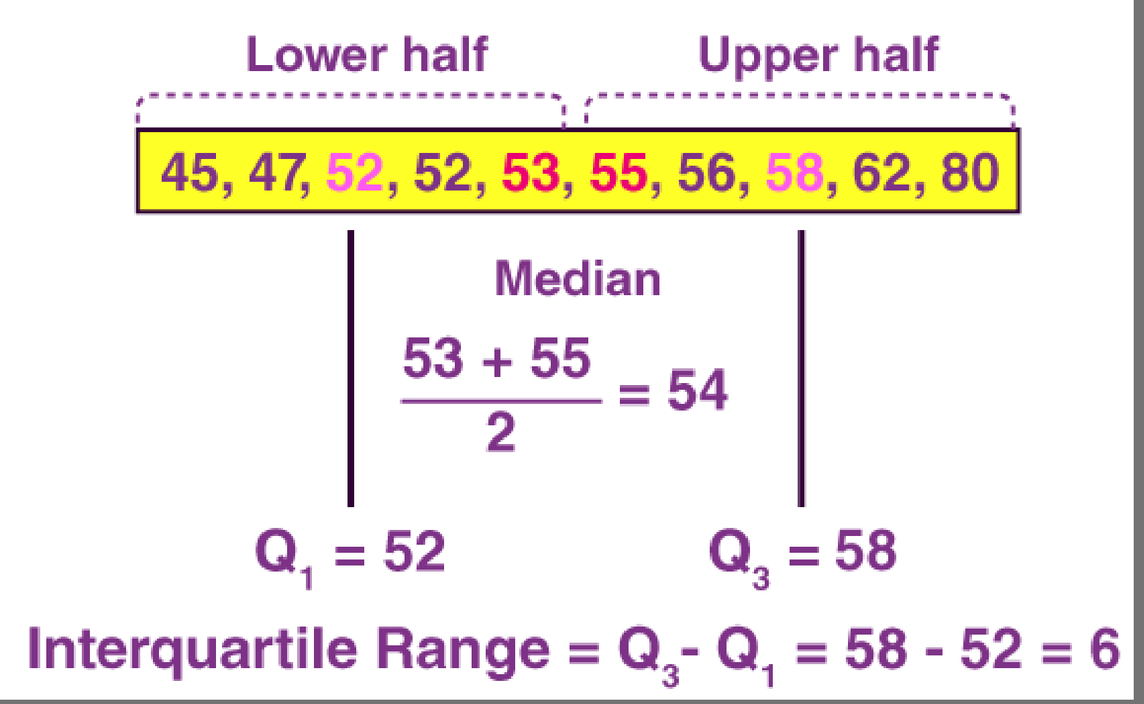
Lowest value

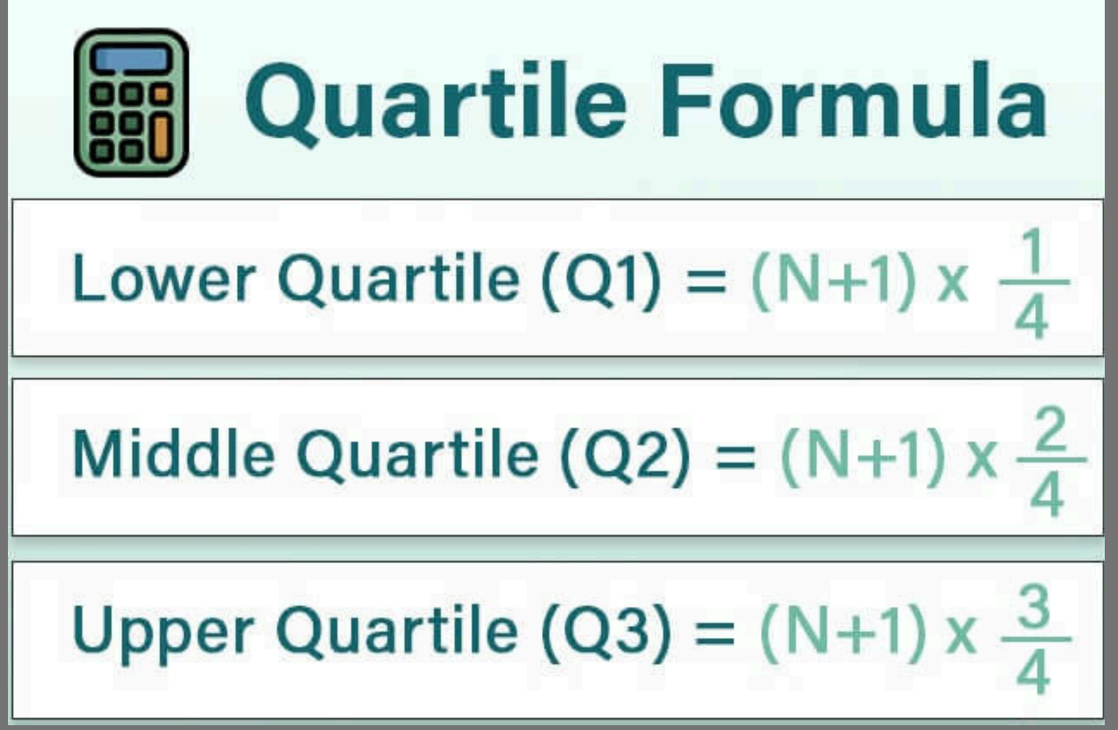


Variance:

Variance measures the dispersion of data points around the mean:

1. Calculate the mean of the dataset.
2. For each data point, find the squared difference between it and the mean.
3. Sum up all the squared differences.
4. Divide the sum by the total number of data points.





Interquartile Range & Any Outliers:

The interquartile range (IQR) measures the spread of the middle 50% of the data:

1. Arrange the dataset in ascending order.
2. Find the median (Q2), which divides the data into two halves.
3. Find the median of the lower half (Q1) and the upper half (Q3).
4. IQR = Q3 - Q1. Outliers can be detected using the IQR method:
5. Calculate the IQR.
6. Define the lower and upper fences as Q1 - 1.5 * IQR and Q3 + 1.5 * IQR, respectively.
7. Any data point outside these fences is considered an outlier.