



**Cavalier Institute** - <https://cavalierinstitutions.com>

1

---

Date	Sep-28-2024	Session No	3
------	-------------	------------	---

Topic : Statistics in BI and Data Analytics
---

---

#### Mean Calculation:

##### 1. Mean (Arithmetic Average):

- Formula:

$$\text{Mean} = \frac{\sum_{i=1}^n x_i}{n}$$

- Where  $x_i$  are the elements in the list, and  $n$  is the number of elements.

- Write a function that takes a list of numbers as input and returns the mean of the list.

#### Median Calculation:

## 2. Median:

- **Formula:**

- If the list is sorted and has **odd** number of elements:

$$\text{Median} = x_{\left(\frac{n+1}{2}\right)}$$

- If the list has **even** number of elements:

$$\text{Median} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}$$

- Where  $x_i$  are the elements sorted in ascending order and  $n$  is the number of elements.

- Write a function to find the median of a list of numbers. Handle both even and odd length lists.

### Mode Calculation:

## 3. Mode (Most Frequent Value):

- **Definition:** The mode is the value(s) that appear most frequently in the list.
- If the list has multiple values that occur with the same highest frequency, all of them are considered modes.

- Create a function that returns the mode(s) of a list of numbers. If multiple modes exist, return them all.

### Variance Calculation:

#### 4. Variance:

- Formula:

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

- Where  $x_i$  are the elements in the list,  $\mu$  is the mean of the list, and  $n$  is the number of elements.

- Write a function to calculate the variance of a list of numbers.

#### Standard Deviation Calculation:

#### 5. Standard Deviation:

- Formula:

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

- The standard deviation is simply the square root of the variance.

- Write a function to calculate the standard deviation of a list of numbers using the variance.

#### Combining Mean, Median, Mode:

- Write a function that calculates and returns the mean, median, and mode of a list in a single output.

#### Range of Numbers:

## 6. Range:

- Formula:

$$\text{Range} = \text{Max}(x) - \text{Min}(x)$$

- Where  $\text{Max}(x)$  is the maximum value in the list and  $\text{Min}(x)$  is the minimum value.

- Write a function to calculate the range (difference between the maximum and minimum) of a list of numbers.

## Weighted Mean:

## 7. Weighted Mean:

- Formula:

$$\text{Weighted Mean} = \frac{\sum_{i=1}^n w_i \cdot x_i}{\sum_{i=1}^n w_i}$$

- Where  $x_i$  are the data points,  $w_i$  are the corresponding weights, and  $n$  is the number of elements.

- Write a function that calculates the weighted mean of a list of numbers, given their corresponding weights.

## Compare Mean and Median:

- Given a list of numbers, write a function to compare the mean and the median and state whether the list is positively skewed, negatively skewed, or symmetrical.

## Outlier Detection:

## 8. Outlier Detection (Using Standard Deviation Method):

- An outlier is any value that lies more than **2 or 3 standard deviations** from the mean.
- **Formula:**

$$\text{Outlier} = x_i \text{ if } |x_i - \mu| > 2\sigma \text{ or } 3\sigma$$

- Write a function that identifies outliers in a list of numbers. Use the standard deviation method to flag values that are more than 2 standard deviations from the mean.

### Rolling Mean (Window Size 3):

- Write a function that calculates the rolling mean of a list with a window size of 3.

### Remove Outliers:

- Write a function that removes any outliers from a list based on the standard deviation threshold.

### Frequency Distribution:

- Write a function that returns the frequency distribution (count of each unique number) from a list.

### Z-Scores:

- Write a function to calculate the z-scores (standard scores) for a list of numbers.

### Grouped Mean:

- Given a list of lists, where each inner list contains numbers from a group, write a function to calculate the mean of each group and return them as a list.

### Find Percentiles:

## 10. Percentile:

- **Formula** (for the  $k$ -th percentile):

$$P_k = x_{\left(\frac{k}{100} \times n\right)}$$

- Where  $k\%$  is the desired percentile and  $n$  is the number of elements in the sorted list.

- Write a function that returns the percentile value for a given percentile rank from a list of numbers.

### Quartiles:

## 11. Quartiles:

- **Q1 (First Quartile):** 25th percentile.
- **Q3 (Third Quartile):** 75th percentile.
- **Formula:**
  - $Q1$  is the 25th percentile of the sorted list.
  - $Q3$  is the 75th percentile of the sorted list.

- Write a function to compute the first (Q1) and third (Q3) quartiles of a list of numbers.

### Coefficient of Variation:

- Write a function that calculates the coefficient of variation (ratio of standard deviation to the mean) of a list of numbers.

### Find Interquartile Range (IQR):

## 12. Interquartile Range (IQR):

- Formula:

$$\text{IQR} = Q3 - Q1$$

- Where  $Q1$  is the first quartile, and  $Q3$  is the third quartile.

- Write a function that calculates the interquartile range ( $Q3 - Q1$ ) of a list of numbers.

### Mean Absolute Deviation:

## 13. Mean Absolute Deviation:

- Formula:

$$\text{MAD} = \frac{\sum_{i=1}^n |x_i - \mu|}{n}$$

- Where  $\mu$  is the mean of the list and  $x_i$  are the elements of the list.

- Write a function to calculate the mean absolute deviation from the mean for a list of numbers.

## 14. Coefficient of Variation (CV):

- Formula:

$$\text{CV} = \frac{\sigma}{\mu}$$

- Where  $\sigma$  is the standard deviation and  $\mu$  is the mean of the list.

---

END