Central Tendency in Statistics

1) **Mean:** The mean (or average) is that the most generally used and well-known measure of central tendency. It will be used with both discrete and continuous data, though it's most typically used with continuous data. The mean is adequate the sum of all the values within the data set divided by the number of values within the data set. So, if we have n values in a data set and they have values $x1, x2, \ldots, xn$, the sample mean, usually denoted by "xbar", is:

Population Mean Formula

$$\text{Population Mean} = \frac{\text{Sum of All the Items}}{\text{Number of Items}}$$

$$\text{Sample Mean} = \frac{\text{Sum of All the Items in Sample}}{(\text{Number of Items in Sample} - 1)}$$

★ Bassel's correction $(n-1)$ is adopted to correct for bias in using the sample variance as an estimator of the true variance. The bias in the uncorrected statistic occur because the sample mean is closer to the middle of the observations than the true mean, and so the squared deviations around the sample mean systematically underestimates the squared deviations around the true mean.

1

2) Median:

The median value of a dataset is the middle of the dataset when it is arranged is ascending or descending order. When the dataset has an even number of values, the median value can be calculated by taking the mean of the middle two values.

The following image gives an example for finding the median for odd and even numbers of samples in the dataset.
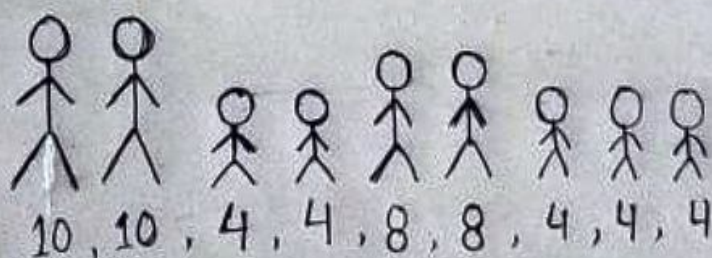
$$1, 3, 3, 6, 7, 8, 9$$
$$\text{Median} = \underline{6}$$

$$1, 2, 3, 4, 5, 6, 8, 9$$
$$\text{Median} = (4 + 5) \div 2$$
$$= \underline{4.5}$$

3) Mode:

The mode is the value that appears the most frequently in your data set. The mode is the highest bar in a bar chart. A multimodal distribution exists when the data contains multiple values that are tied for the most frequently occurring. If no value repeats, the data does not have a mode.

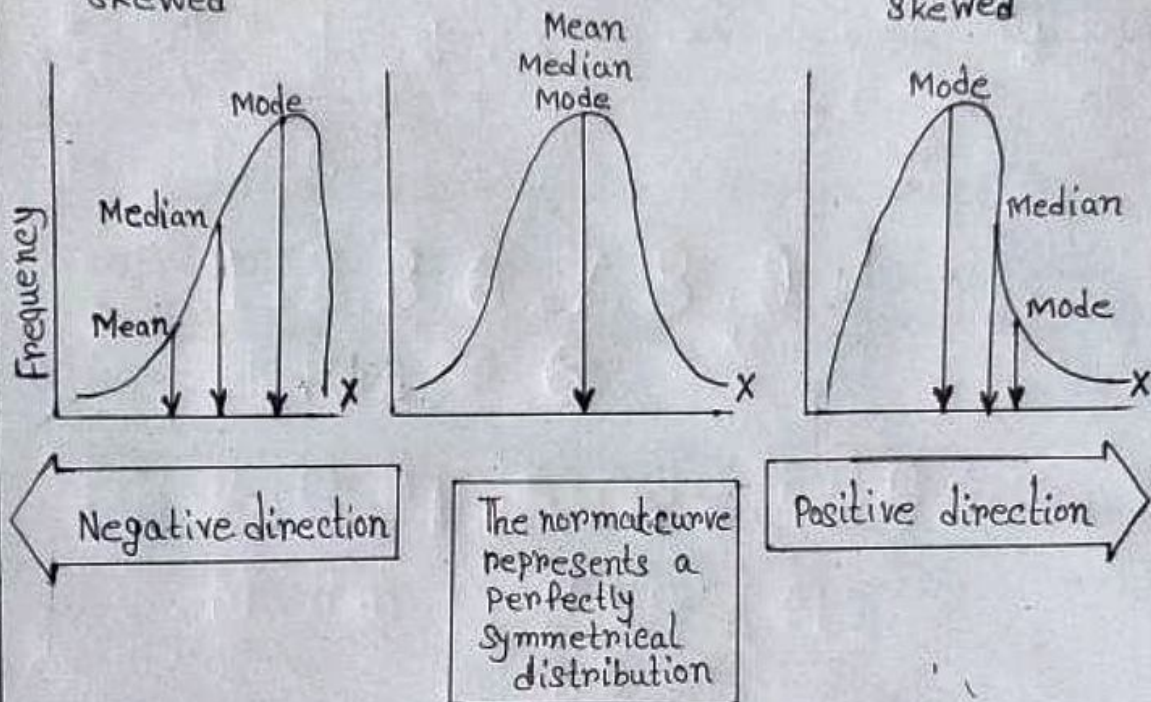$$10, 10, 4, 4, 8, 8, 4, 4, 4$$
$$\text{Mode} = \underline{5}$$

4) Skewness :

Skewness is a metric for symmetry, or more specifically, the lack of it. If a distribution, or data collection, looks the same to the left and right of the center point, it is said to be symmetric.

(a) Negatively skewed

(b) Normal (No skew)

(c) Positively skewed

Negative direction

The normal curve represents a perfectly symmetrical distribution

Positive direction

5) Kurtosis :

Kurtosis is a measure of how heavy-tailed or light-tailed the data are in comparison to a normal distribution. Data sets having a high kurtosis are more likely to contain heavy tails or outliers. Light tails or a lack of outliers are common in data sets with low kurtosis.

Kurtosis

$K > 3$ Leptokurtic

$K = 3$ mesokurtic

$K < 3$ Platykurtic