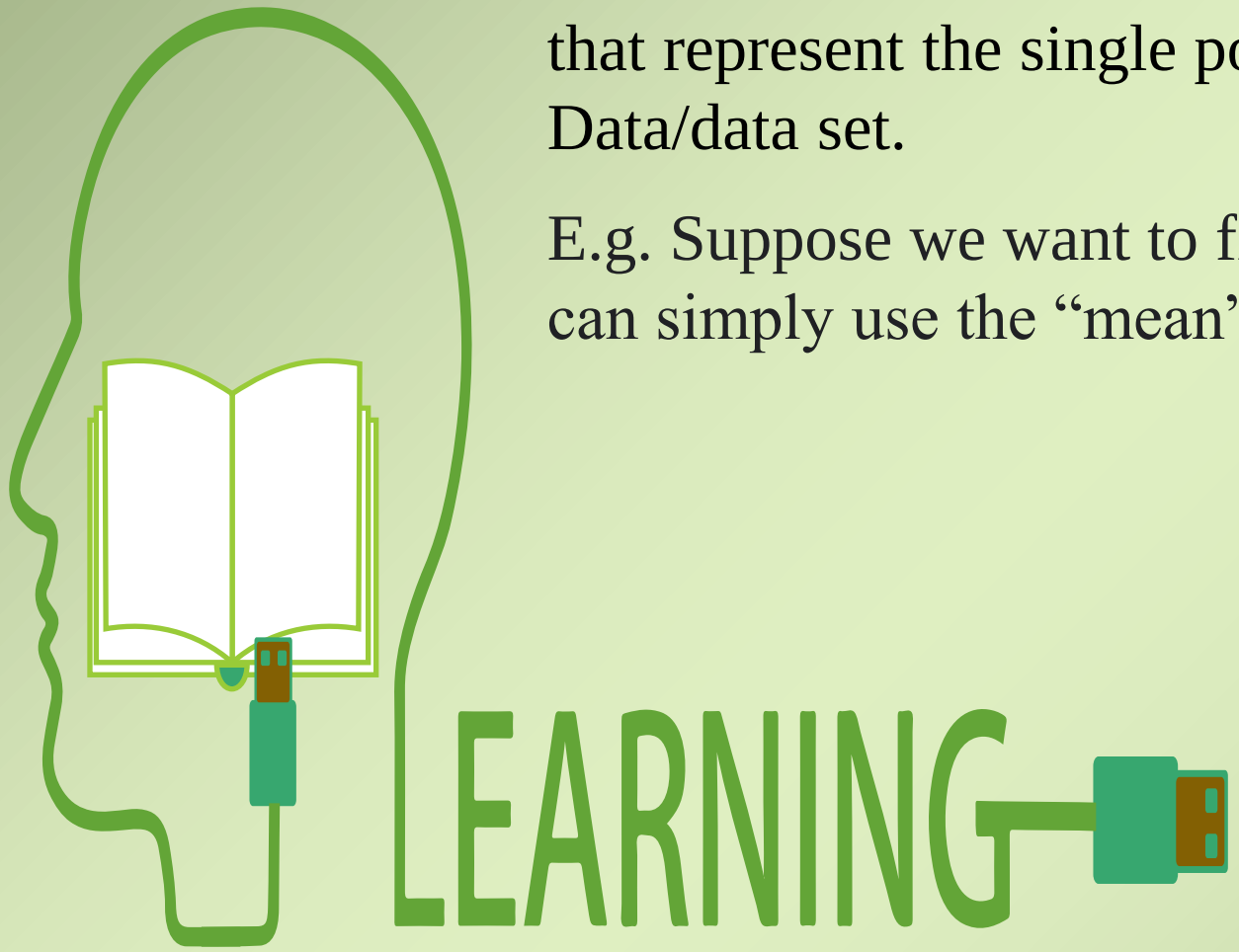
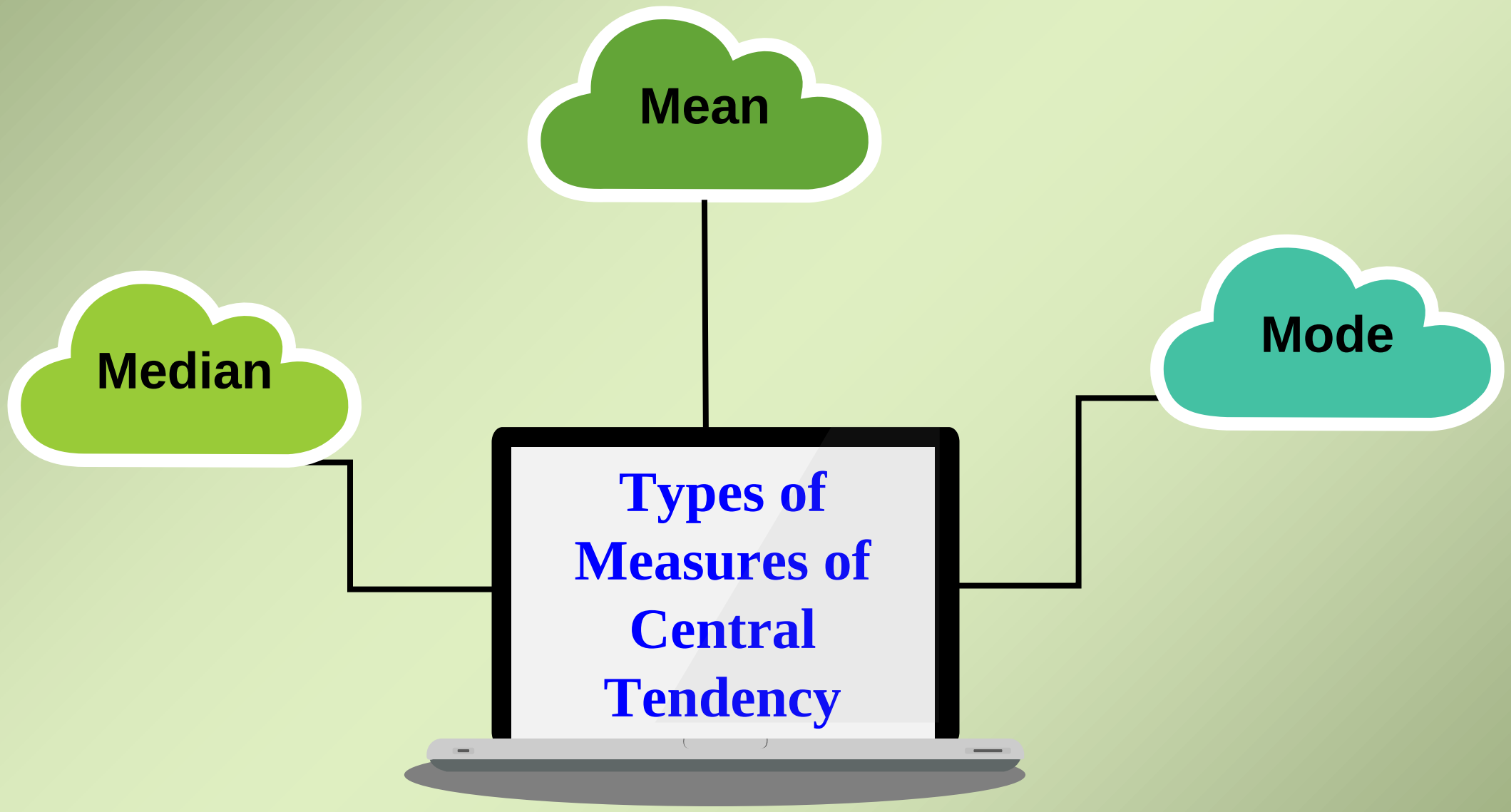
Measures Of Central Tendency



A measure of central tendency is a summary of the statistics that represent the single point as the representative of the Data/data set.

E.g. Suppose we want to find out the height of 500 students, we can simply use the “mean”.


Mean
Median
Mode
Types of Measures of Central Tendency

Mean

VisionNLP

The arithmetic mean is defined as the sum of the value of the given observation divided by the no of observation.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{N} X_i .$$

Merits and demerits:-

Merits:-

1. It's easy to calculate and understand, that's why it's popular measure of central tendency.

Demerits:-

1. It can’t be located graphically.
2. It’s extremely affected by outliers since it's based on all the observations.

Median

VisionNLP

The median is the middle number of the data after sorting it into either Ascending or descending order.

Median divides the data into two equal parts.The median is the response that divides the population into the top and bottom 50%.

The formula is,

For odd numbers:-

$$Median = Size\ of\ \frac{(n+1)}{2}\ th\ Observation$$

For even numbers:-

$$median = \frac{\left(\frac{n}{2}\right) + \frac{(n+1)}{2}}{2}\ th\ Observation$$

1, 3, 3, **6**, 7, 8, 9

Median = 6

1, 2, 3, **4**, **5**, 6, 8, 9

Median = (4 + 5) ÷ 2

= 4.5

E.g.. .Median age divides the population into two parts, with one half aged above the 50 and the other half below the 50 age.

Median

VisionNLP

Merits and demerits:-

Merits:-

1. It is easily understood and is easy to calculate.
2.It’s not affected by extreme Values like outliers.

Demerits:-

1.It is not based on all the observations of the data.
2.For calculation of median data must be arranged in ascending or descending order.

Mode

The mode is defined as the observation having maximum frequency or the most repeated value in the data Set called mode.
The mode is calculated using the formula,
Mode = Observation having maximum frequency

E.g.
1. In election a party with a high vote is selected as the winner.

Merits:-
1.Mode can be calculated for both qualitative and Quantitative Data.
2.It's not affected by extreme Values like outliers.

Demerits:-
1 .It is not based on all the observations of the data.

Quartiles

VisionNLP

Quartiles :-

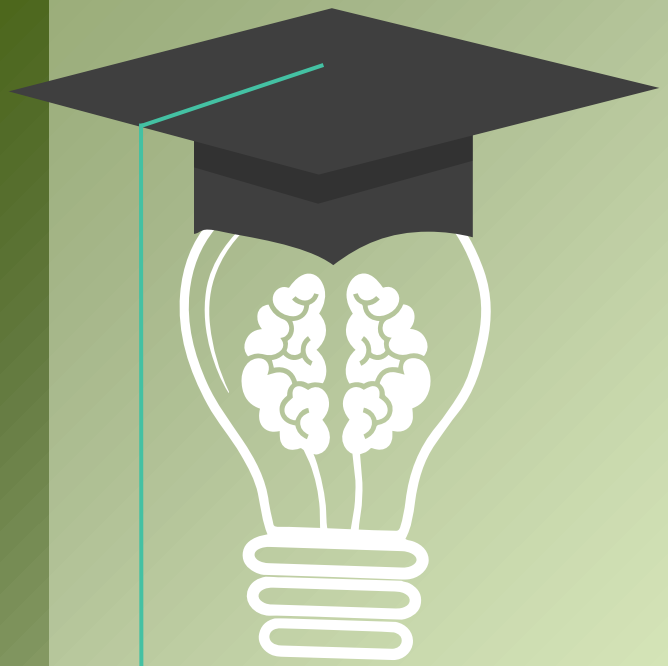
The values which divide the data into four equal parts are called Quartiles.
There are three quartiles namely Q1 ,Q2 ,Q3.

We called these points as — a lower quartile or first quartile (Q1) , median or second quartile (Q2) , and third quartile or upper quartile(Q3)

Formula:-

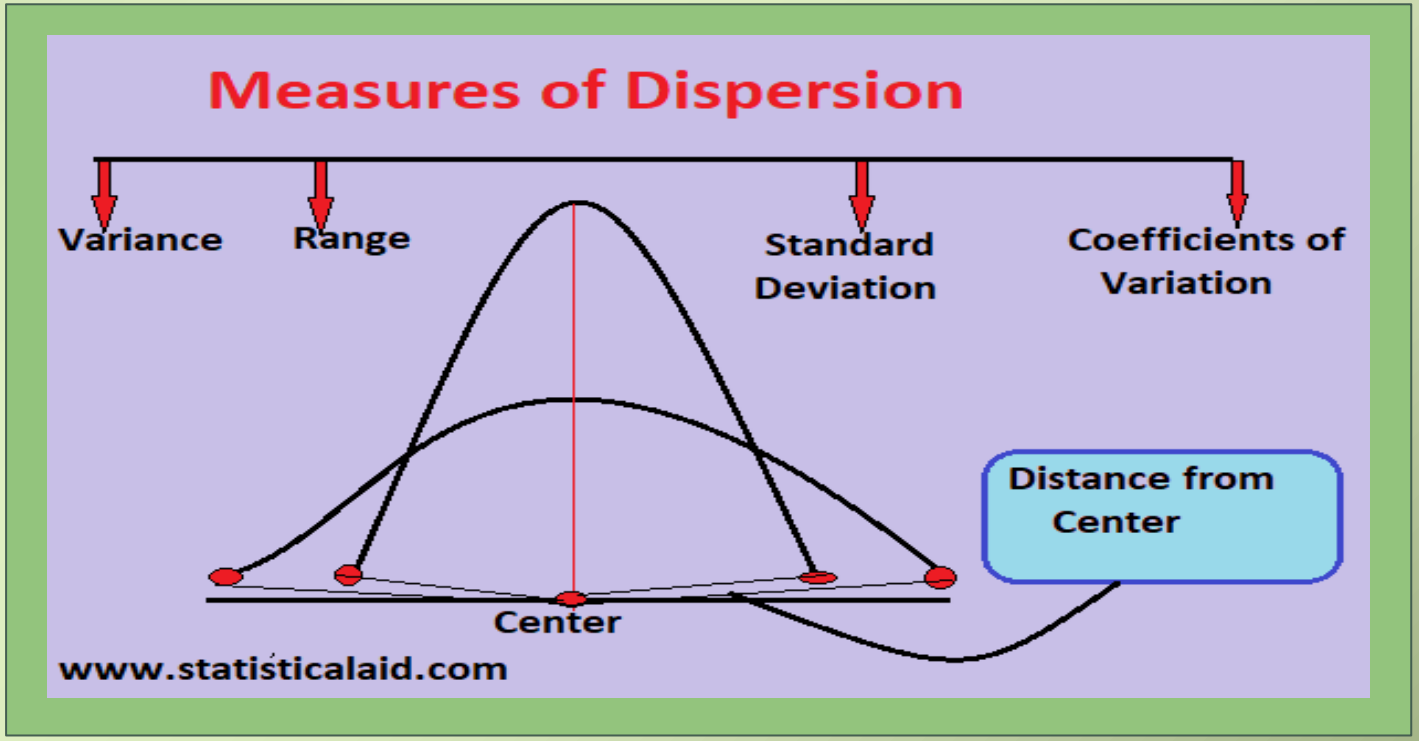
Q_k = k(N+1) / 4 th observation ,Where k = 1,2,3

Quartile which is useful in determining the shape of the distribution.



Measures of Dispersion

Dispersion is a statistical measure of how much spread in data or how the data scattered around the mean.



Range

VisionNLP

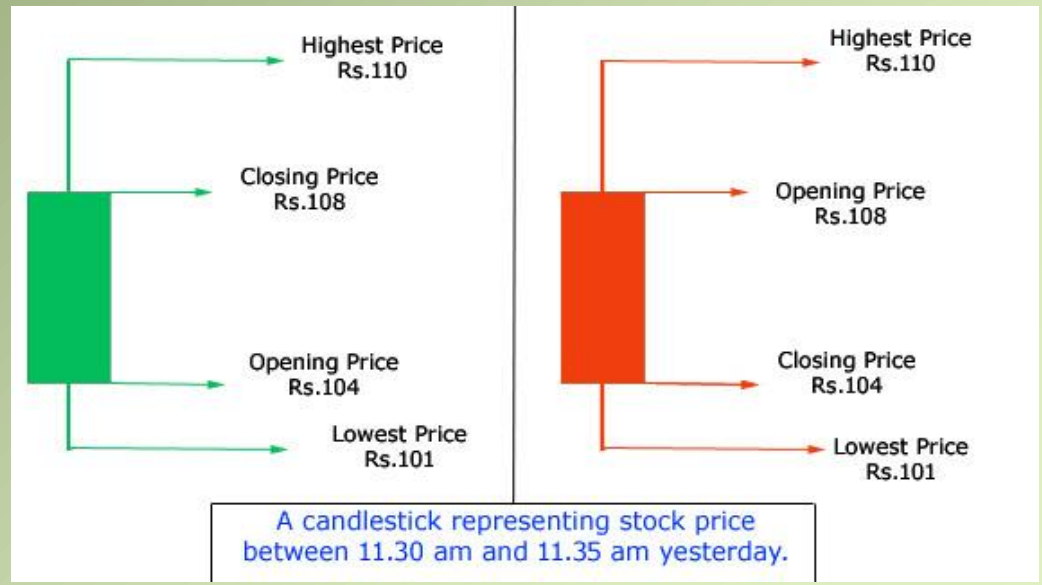
Range is the simplest method to find the dispersion.

Range is defined as largest value minus smallest value and is given by,

Range = L-S Where , L =Largest Value and S = Smallest Value

Note: It's suitable for small data with less variation

Range



It's used in share market the lowest and highest value are recorded

In the medical field it is used to record the , oxygen level of patients.

Variance

VisionNLP

Variance is the measure of how the data point differs from the mean OR Spread around the mean.

The variance is denoted by V(X) and is given by,

$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \qquad \mu = \frac{1}{n}\sum_{i=1}^{n} x_i\,.$$

E.g. Take an example of a patients. The people have different heights. There is an average height, and most people's heights deviate from the average (are either higher or lower, by some amount).

Properties:

The greater the value of variance the more scatteredness/Spread in the data.

Variance is always positive or zero, because squares are positive.

Variance of constant (single value) is zero.

μ

Variance

VisionNLP

Example: Rolling a single die 6 times,, outcomes will be 1 to 6 values each with equal probability ⅙.

X = 1, 2, 3, 4, 5, 6

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{N} X_i.$$

Mean = (1+2+3+4+5+6)/6
= 7/2

Now, variance of x is,

$$\operatorname{Var}(X) = \sum_{i=1}^{6} \frac{1}{6}\left(i - \frac{7}{2}\right)^2$$

$$= \frac{1}{6}\left((-5/2)^2 + (-3/2)^2 + (-1/2)^2 + (1/2)^2 + (3/2)^2 + (5/2)^2\right)$$

$$= \frac{35}{12} \approx 2.92.$$

S.D.

VisionNLP

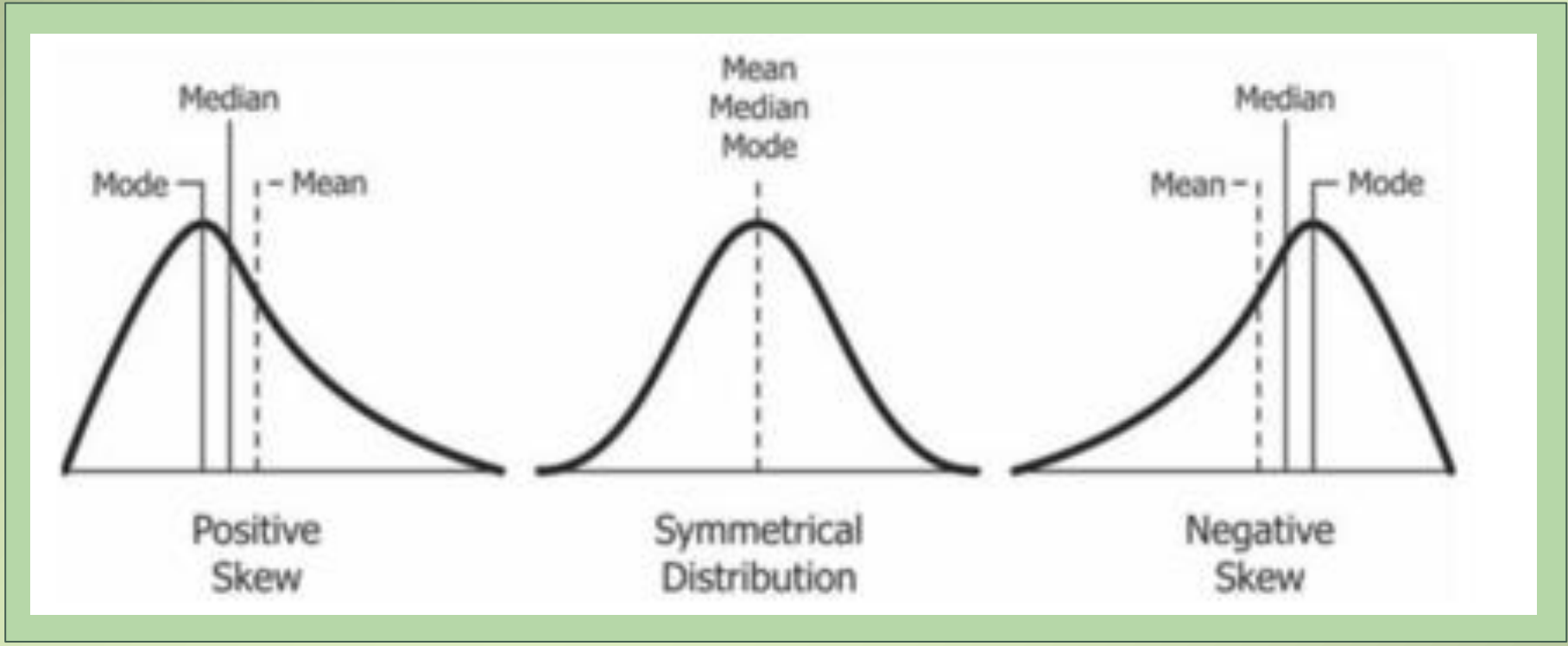
Standard Deviation:- It is the positive square root of variance and is given below

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}, \text{ where } \mu = \frac{1}{N}\sum_{i=1}^{N} x_i .$$

Skewness and its types

VisionNLP

Skewness is a measure of asymmetry of the probability distributions.



Skewness and its types

VisionNLP

1.Symmetric Distribution:-

The distribution of values in which the right half is a mirror image of the left half is said to be symmetrical

Mean= Median = Mode

2.Positive Skewness:- The right tail is longer. The mass of distribution is concentrated on the left side.

Mean > Median > Mode

3.Negative Skewness:- The left tail is longer. The mass of distribution is concentrated on the right side.

Mean <Median <Mode.

Kurtosis

VisionNLP

Kurtosis is a measure of tailedness in probability distributions.

- The peakedness of frequency curve as compared to the normal peaked curve.
- The measure of the respective sharpness of the curve, in the frequency distribution.

Types of Kurtosis:

Leptokurtic: The curve is more peaked than the Normal curve. Leptokurtic distribution has fatter tails.

Platykurtic: The curve is less peaked than the Normal curve. Platykurtic distribution has thinner tails.

Mesokurtic: The curve which is neither flat or peaked is called Normal curve or Mesokurtic. It looks like bell shaped curve.