Summarizing Data through Numbers.

* Measure of Central Tendency — Mean of the distribution
  ↓
  Mean is at exactly 50% area in the histogram.
* Dispersion → Variance
* Skewness & kurtosis.
* Minimum Value
* Maximum Value.

Measure of Central tendency.

* Mean
* Median
* Mode.

Ex $X = [3, 4, 3, 1, 2, 3, 9, 5, 6, 7, 4, 8]$

$$\text{Mean} = \frac{\sum_{i=1}^{n} x_i}{n} = 4.583$$

Median

arrange the data in accending order.

Ex

1, 2, 3, 3, 3, 4, 4, 5, 6, 7, 8, 9.

here $n = 12$ Median = 4.

hence Median is mean of the central pair

$$\text{Median} = \frac{x_{n/2} + x_{n/2+1}}{2} \quad \text{if } n \text{ is even}$$

$$= x_{\lceil n/2 \rceil} \quad \text{if } n \text{ is odd}$$

Mode: which particular outcome have highest probability (frequency).

Mode $[x] = 3$

Where do we want to use Mean, Median, Mode

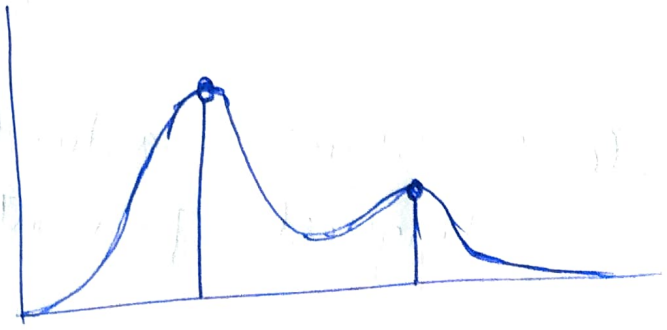
* Choosing between mean & Median.
  * If the data have a Outlier and it is not contributing in the problem then Median is more robust. (error in the data). Ex Salary data/Income
  * If the outlier is the part of the story. then mean is more appropriate as outlier is contributing in the inference (transient behavior of the system)
    ex ~~level of water in the Dam reservoir~~
    Recruiting number of employes for a Bank Branch. & No. of arrival of customer per day or time.

Mode

Multi Modal distribution: Level of water in the Reservoir (DAM)

decision for height & Strength of the dam (Reservoir)

Mean & Median will not help in taking the decision



Skewness: is a measure of the asymmetry of the probability distribution of a random variable about its mean.

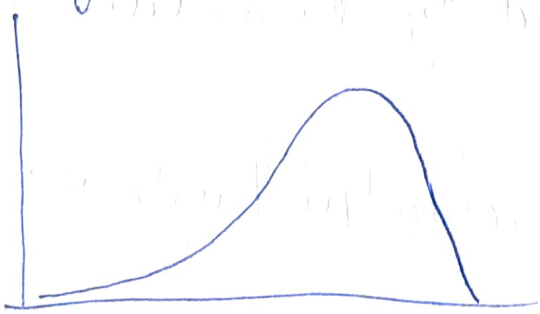
Skewness can have values positive, negative zero & undefined.

for unimodal distribution

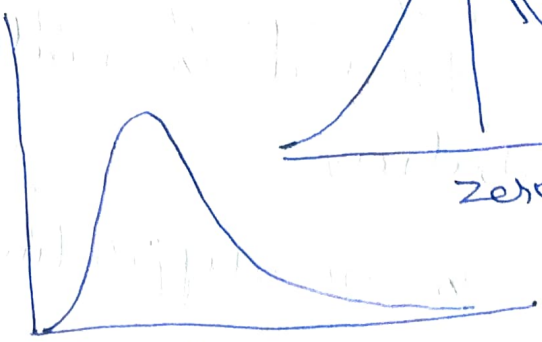
-ive Skewness means the tail is on the left side of the distribution.

+ive Skewness means the tail is on the right side of the distribution.

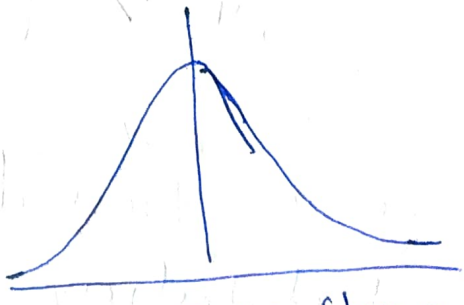
Zero Skewness means tails on the both side balance each other. it is also called Symmetric distribution.



-ive Skew



+ive Skew



zero Skew.

Skewness of a random variable is defined as third Standardized Moment $\hat{\mu}_3$

$$\hat{\mu}_3 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3}$$

μ_3 : third Central Moment.

Scaled version of the (Proportional)

Kurtosis: ^ fourth Central moment of the distribution.
(by Karl Pearson)

* Kurtosis of Univariate Normal distribution is 3.

* Distribution with kurtosis less than 3 are said to be platykurtic, it means that the distribution produces very less extreme outliers than the normal.

* ~~Distribution with kurtosis great~~
Example of Platykurtic distribution is ~~Not~~ Uniform distribution.

Distribution with kurtosis greater than 3 are said to be leptokurtic distribution

Ex Laplace distribution, which has tails that asymptotically approach zero more slowly than a Gaussian. & hence produce more outliers than the Normal distribution

$$\text{Kurtosis}[X] = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$$

$$= \frac{E[(X-\mu)^4]}{\left(E[(X-\mu)^2]\right)^2} = \frac{\mu_4}{\sigma^4}$$

where μ_4 is fourth central moment.

Relation between Skewness & kurtosis

Kurtosis is bounded ~~below~~ by Skewness

$$\frac{\mu_4}{\sigma^4} \geq \left(\frac{\mu_3}{\sigma^3}\right)^2 + 1$$

Lower bound is realized by Bernoulli distribution

There is no upper limit to the kurtosis