

1) Measure of Central Tendency

Center Identify the central / typical value of a dataset.

Mean — Avg value

Median — Middle value when data are in ascending order.

Mode — Most frequent occurring value.

$$\text{Mean} = (\bar{x}) = \frac{\sum x_i}{n}$$

Measures of dispersion

describes spread of the data.

Range — difference b/w highest & lowest

Variance — Avg of squared differences from mean.

Standard deviation — Root square of variance

Interquartile range — d.iff b/w 1st quartile (Q_1) and 3rd " (Q_3) shows middle 50% of data.

Range = (Max - Min) value

$$\text{Variance} = (\sigma^2) = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$\text{Std dev } (\sigma) = \sqrt{\text{Variance}}$$

$$\text{IQR} = Q_3 - Q_1$$

2) Skewness is a measure of asymmetry of the probability distribution of a real valued variable about its mean. Indicates whether data points are skewed to left or right.

- Positive (Right Skew) \Rightarrow Tail on right side is longer
- \Rightarrow most data concentrated on left.
- \Rightarrow mean is greater than median.

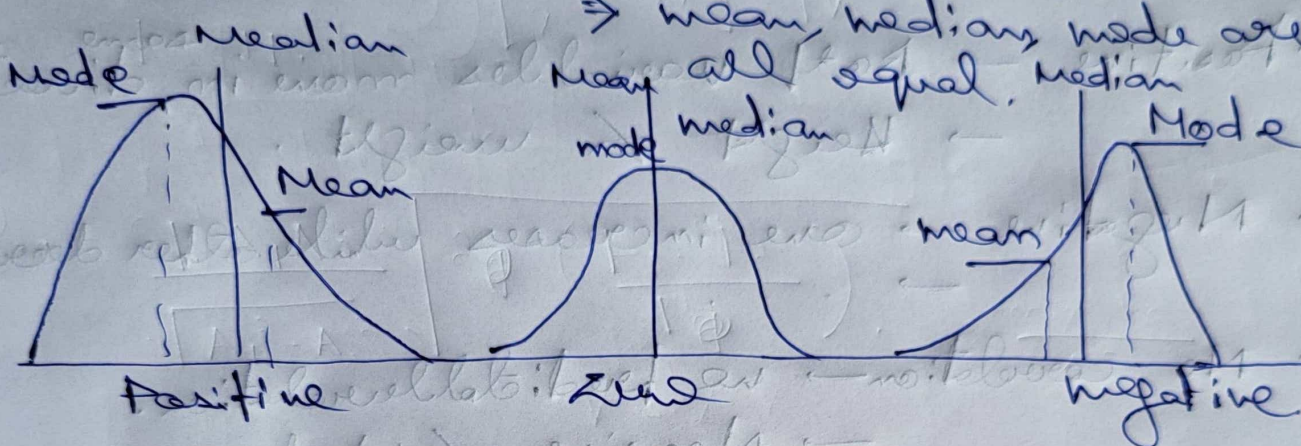
→ negative (left skew) \Rightarrow Tail on left side longer or fatter than right.

\Rightarrow data points conc on right.

\Rightarrow mean < median.

→ Zero skew (symmetrical) \Rightarrow left and right side of distribution are mirror images

\Rightarrow mean, median, mode are all equal.



22) Probability mass func (PMF) Probability den func (PDF)

\Rightarrow Used for discrete random variables.

\Rightarrow Used for continuous random variable.

\Rightarrow Discrete random variable is exactly equal to some value.

\Rightarrow Describes likelihood of a variable to take a particular value. It gives the density of probability.

$\Rightarrow [P(X=x) = f(x)]$

$\Rightarrow [f(x)]$

\Rightarrow Sum of all prob is equal to 1.

\Rightarrow Integral over entire range is 1.

23) Correlation \rightarrow measures strength & direction of linear relationship b/w 2 variables.
It ranges from -1 to $+1$.

-1 \rightarrow perfect negative linear relationship.

$+1$ \rightarrow " " " " " "

0 \rightarrow no relationship " " " "

\Rightarrow Positive \rightarrow Both variables move in same direction
 \rightarrow Height & weight.

\Rightarrow Negative \rightarrow one increases while other decreases.

\Rightarrow No correlation \rightarrow no predictable relation!
 \rightarrow shoe size & intelligence.

① Pearson's $\rightarrow r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$

② Spearman's rank \rightarrow measures strength and direction of association b/w 2 ranked variables.

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$
 where, $d_i =$ diff b/w ranks of corresponding variable

③ Kendall's Tau $\rightarrow \tau = \frac{C - D}{\frac{1}{2}n(n-1)}$

where, $C =$ no of concordant pairs
 $D =$ " " " " discordant "

24) What to do cannot understand ??