

## \* Correlation

- Correlation measures the degree to which two variables move in relation to each other.
- Correlation shows the direction of relationship as well as the strength of it, and represent it numerically as correlation coefficient.
- Correlation coefficient value range in -1 to +1.
- Positive correlation coefficient means that when one variable is increasing other tends to follow the same.
- Closer the coefficient to 1, stronger the directly proportionality relation.
- A negative correlation coefficient means that both variable tends to follow inversely proportional relation.
- Closer the coefficient to -1, stronger the inversely proportional relationship.
- It solves the major drawback of covariance by limiting the value of correlation coefficient within a range of -1 to 1, which help in comparing strength of relationships in for accurate way.
- There are different type of correlation coefficients available,

## \* Pearson's product-moment correlation coefficient

- We tweak the formula of covariance to limit its range between -1 to 1.

$$\rho(x,y) = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y}, \sigma_x \sigma_y > 0$$

- if  $\rho(x,y) = 1$ , perfect direct increasing or linear relationship.
- if  $\rho(x,y) = -1$ , perfect inverse linear relationship.
- all other cases within (-1,1)
- if  $\rho(x,y) = 0$ , no relation.

## \* Spearman's Rank Correlation.

- We tweak the formula of Pearson's coefficient to get better and more precise values.

$$\text{Spearman rank} = \frac{\text{cov}[R(x), R(y)]}{\sigma_{R(x)} \sigma_{R(y)}}$$

- to get the Rank of a variable ( $R(x)$ ), we simply allot a value within (1 to  $\infty$ ) starting from smallest value to the largest.
- if two values are same they will get same rank.

ex -

X	Y	R(x)	R(y)
10	20	4	1
8	30	3	2
7	40	2	3
6	50	1	4
6	10	1	5

then use formula accordingly.