10 - 03 - 2022

Thursday, March 10, 2022 11:55 AM

Unit-6 (Correlation & Regression)

§ Correlation: A quantitative measure of the relationship between two variables, and also an appropriate mathematical or statistical form of the relationship.

$$\begin{cases} X \rightarrow \text{increases} \\ Y \quad \text{increase, decrease} \end{cases}$$

§ Mathematical (statistical) meaning of correlation: In a bivariate distribution, we may be interested to find out if there is any correlation or covariation between the two variables under study.

If the change in one variable affects a change in the other variable, the variables are said to be correlated.

1) If the two variables deviates in the same direction i.e., if the increase (decrease) in one results in a corresponding increase (decrease) in the other, the correlation is direct or positive.

Eg: $X \rightarrow$ time
$Y \rightarrow$ age of a particular person.
there $cor(X, Y)$ is +ve.

2) If the two variables contantly deviate in the opposite direction, i.e. if X increases $\Rightarrow Y$ decreases or if X decreases $\Rightarrow Y$ increases, then the correlation is negative or diverse.

Eg: $X \rightarrow$ volume
$Y \rightarrow$ pressure
-ve correlation.

3) Perfect correlation: Correlation is said to be perfect if the

3) Perfect correlation: Correlation is said to be perfect if the deviation in one variable is followed by a corresponding and proportional deviation in the other.

$X \to k$

$Y \to 2k, 3k, 4k, \ldots, pk$

$\frac{Y}{X} = c$

$Y \propto X$

X	2	3	4	5	6
Y	4	5	6	7	8

Is the corr +ve ?

Yes

Is this perfect corr ?

X	2 $\to$	5 $\to$	8	11	14
Y	3 $\to$	6 $\to$	10	15	21.

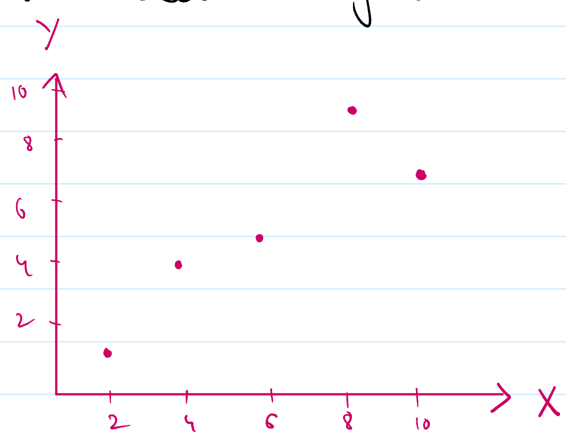
+ve

Yes, perfect..

Not a perfect correlation.

§ Scatter diagram: It is the simplest way of the diagramatic representation of bivariate data.

Thus, for the bivariate distribution (x_i, y_i); $i = 1, 2, \ldots, n$, if the values of X and Y are plotted along X-axis & Y-axis resp, in the x-y plane, then the diagram of dots so obtained is known as scatter diagram.



X	2	4	6	8	10
Y	1	4	5	9	7

If the dots will make a st line, then X & Y are perfectly correlation.

§ If the points are very dense, close to each other, we should expect a fairly good amount of correlation between X & Y.
If the points are scattered, a poor correlation is expected. This works only if n is small.

§ Karl Pearson's co-effecient of correlation :-

$aX + bY = c$

linear correlation.

As a measure of intensity or degree of linear relationship betⁿ 2 variables, Karl Pearson (1867 - 1936), a british biometrician, developed a formula called correlation co-effecients.

Correlation co-effecients betⁿ X & Y, usually denoted by $r(X,Y)$, or r_{XY}, a numerical measure of linear relⁿ. betⁿ them & is defined as -

$$r(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

$aX + bY = c$

if $(X,Y) = (x_i, y_i)$, $i = 1$ to n, is the bivariate distribution, then -

$$Cov(X,Y) = E\left[\{X - E(X)\}\{Y - E(Y)\}\right]$$
$$= \frac{1}{n}\Sigma(x_i - \bar{x})(y_i - \bar{y}) = \mu_{11}$$

$$\sigma_X^2 = E\{X - E(X)\}^2 = \frac{1}{n}\Sigma(x_i - \bar{x})^2$$
$$\sigma_Y^2 = E\{Y - E(Y)\}^2 = \frac{1}{n}\Sigma(y_i - \bar{y})^2.$$

X	2	4	6	8	10
Y	3	9	27	81	243

§ $Cov(X,Y) = \frac{1}{n}\Sigma x_i y_i - \bar{x}\bar{y}$

x_i y_i x_i^2 y_i^2 $x_i y_i$

§ $Cov(X,Y) = \frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}$

$\sigma_x^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2$

$\sigma_y^2 = \frac{1}{n}\sum y_i^2 - \bar{y}^2$

x_i y_i x_i^2 y_i^2 $x_i y_i$

$\bar{x}$

$\bar{x} = \frac{10}{4} = \frac{5}{2}$, $\bar{y} = \frac{20}{4} = 5$.

Eg:

$X(x_i)$	$1 \rightarrow$	$2 \rightarrow$	$3 \rightarrow$	4	
$Y(y_i)$	$2 \rightarrow$	$4 \rightarrow$	$6 \rightarrow$	8	
$x_i y_i$	2	8	18	32	60
x_i^2	1	4	9	16	30
y_i^2	4	16	36	64	120

$Cov(X,Y) = \frac{1}{4}60 - \frac{5}{2}.5$

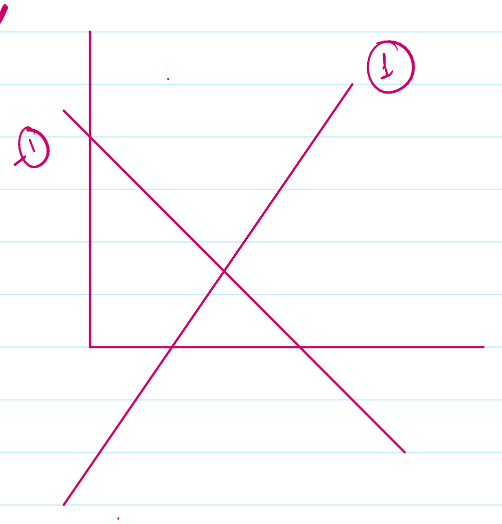
$= 15 - \frac{25}{2} = \frac{5}{2}$.

$\sigma_x^2 = \frac{1}{4}30 - \frac{25}{4} = \frac{5}{4}$

$\sigma_y^2 = \frac{1}{4} \times 120 - 25 = 5$

$r_{xy} = \frac{5/2}{\frac{\sqrt{5}}{2} . \sqrt{5}} = \frac{5}{2} \times \frac{2}{5} = 1$

$r_{xy} \in [-1,1]$



$|r| \leq 1$.

$\frac{Cov(X,Y)}{\sigma_x \sigma_y}$