Correlation:

Correlation is the measure of the strength of the linear relationship between two or more variables.

Example: The amount of rainfall to some extent is accompanied by an in the volume of production.

Different types of correlation:

(i) Positive and negative correlation
(ii) Simple vs Multiple vs Partial correlation
(iii) Linear vs Non-linear correlation

Positive Correlation:

If two variables vary in the same direction, then the correlation between them is called positive correlation.

i.e increase (or decrease) in the value of one variable results increase (or decrease) in the value of other variable, then the two variables are said to have positive correlation.

Example:

Correlation between height and weight.

Negative Correlation:

Two variables are said to have negative correlation if two variables move in the opposite direction.

i.e if one variable increase (or decrease) the second decreases (or increases).

Example:

Correlation between price and demand of a product.

Simple Correlation:

When correlation is concerned with only two variables then it is known as simple correlation.

Example:

The correlation between height and weight of EEE students.

Multiple correlation:

When in a study, we consider more than two variables at a time then the correlation is known as multiple correlation.

Example:

If we consider the correlation between the academic results of a student and his family size, family income, study hour, then this problem can be termed as multiple correlation.

Partial correlation:

When we have more than two variables in the study, but we consider only two variables for influencing each other – where the other variables are considered as constant, then the correlation is termed as partial correlation.

Example:

Let us consider four variables such as the academic results of a student, his family size, family income, study hour. If we consider family size and family income as constant and determine the relationship between academic results and study hour, then it can be termed as partial correlation.

Linear Correlation and non-linear Correlation:

The correlation between two variables is said to be linear when a unit change in one variable results a constant change in the other variable over the entire range of the values.

If corresponding to a unit change in one variable, there is no constant change in other variables, then correlation is said to be non-linear.

Methods of correlation analysis:

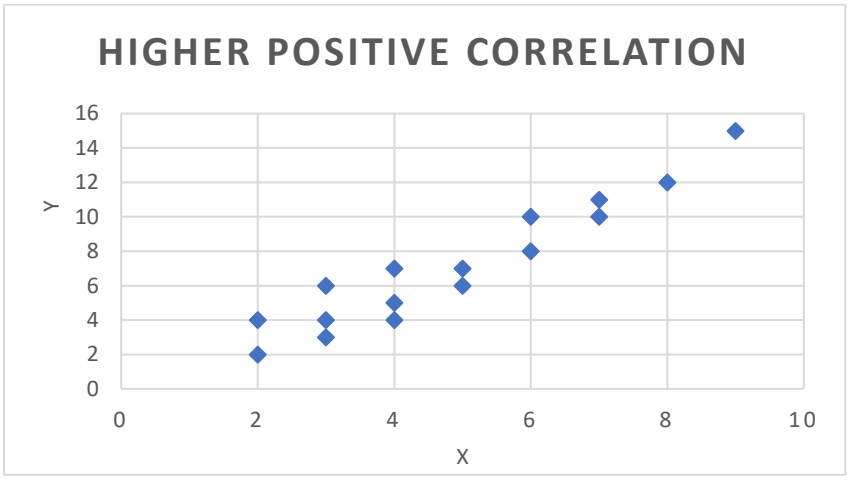
i. Scatter diagram method
ii. Karl Pearson's coefficient of correlation.
iii. Spearman's Rank Correlation Coefficient.

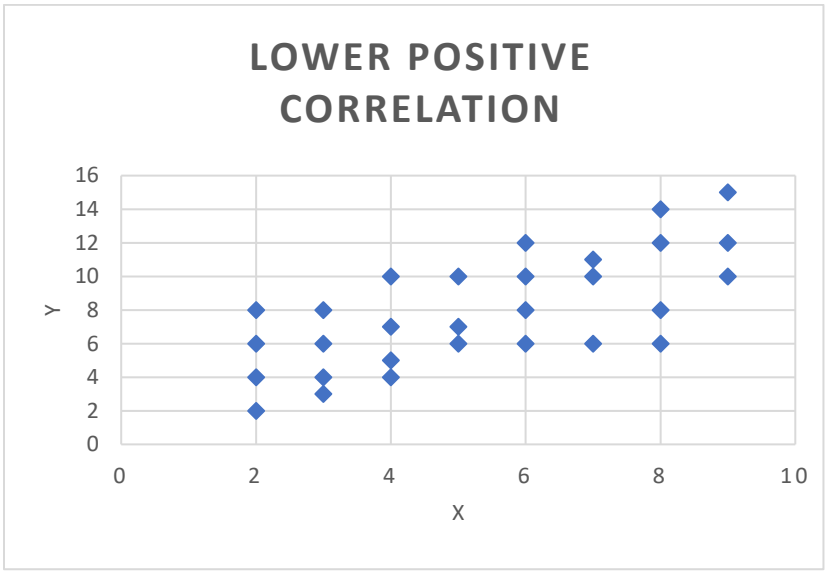
Scatter Diagram Method:

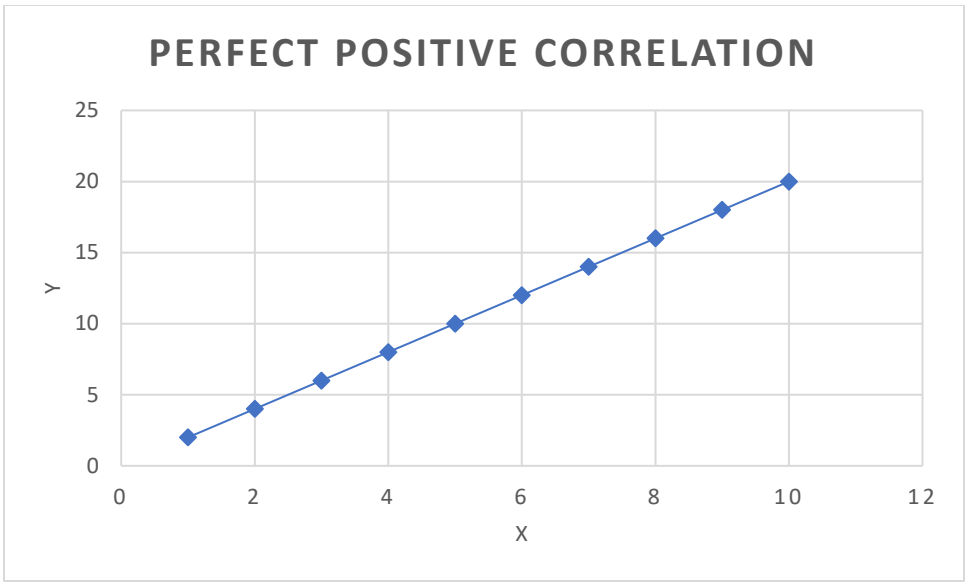
By the scatter diagram method, we mean to plotting a dot (or a point) on the graph sheet for each corresponding pair of values.

Hence by using the scatter of the various points we can realize whether the variables are related or not. The more plotted points are scattered over a chart, the lesser the degree of relationship between the variables.

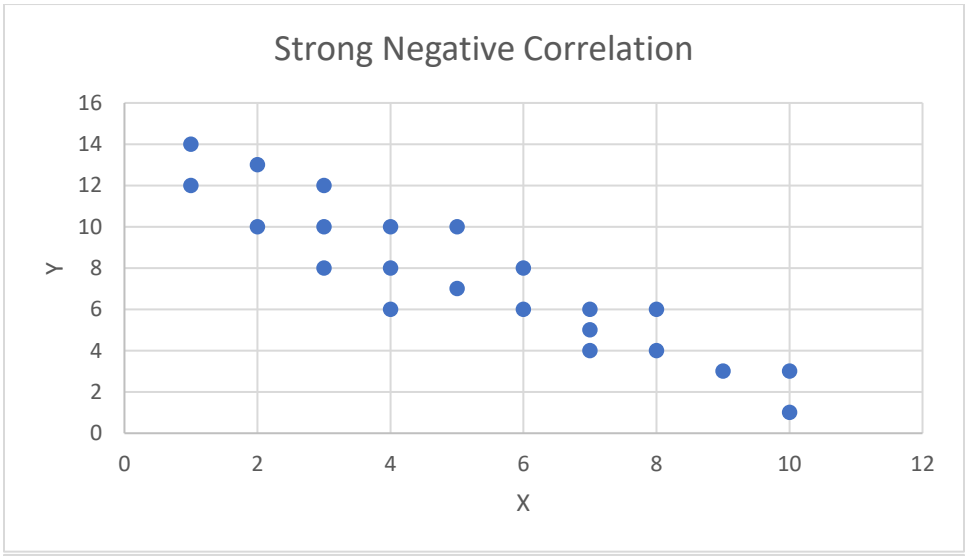
Positive correlation:

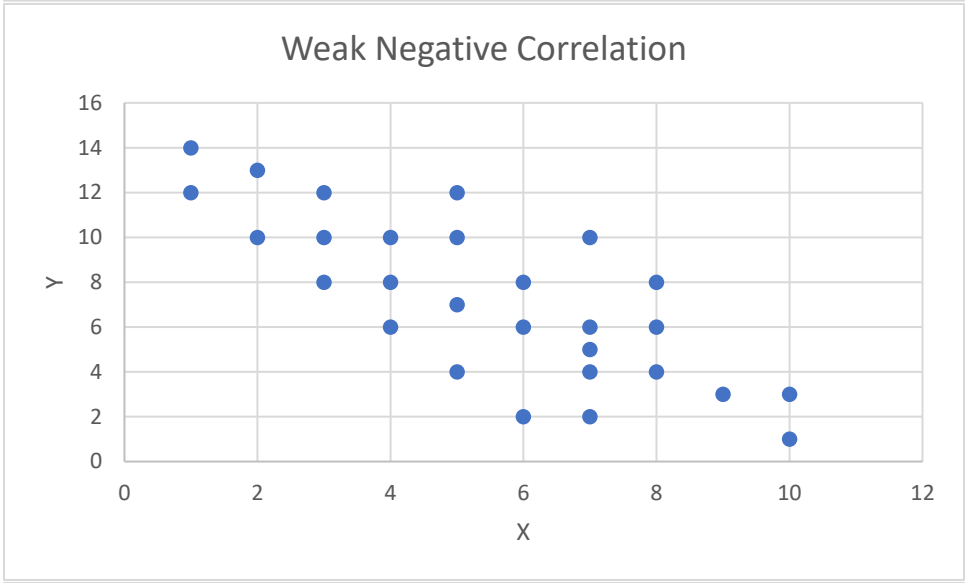


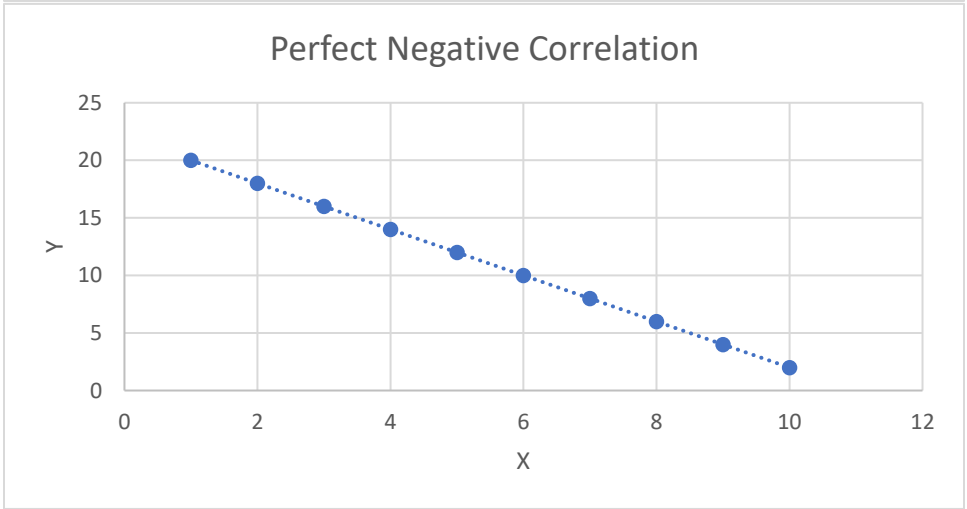


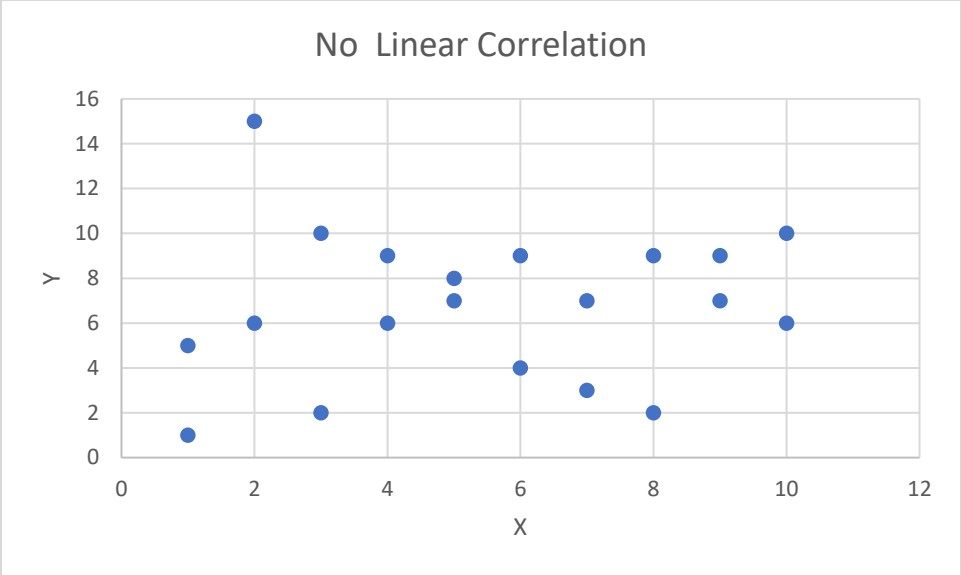

PERFECT POSITIVE CORRELATION
25
20
15
10
5
0
Y
0
2
4
6
8
10
12
X

Negative Correlation:









Karl Pearson's Coefficient of Correlation:

Let, (x_1, y_1), (x_2, y_2), (x_3, y_3), . . . , (x_n, y_n) be n pairs of values of two variables x and y. The correlation coefficient between the variables x and y is denoted by r and is defined by –

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

$$= \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\{\sum {x_i}^2 - \frac{(\sum x_i)^2}{n}\}\{\sum {y_i}^2 - \frac{(\sum y_i)^2}{n}\}}}$$

Assumptions:

i. The range of r is -1 to 1.
ii. The relationship between two variables is linear.

Interpretation:

$Values\ of\ r$	Interpretation
0	No linear correlation
+1	Perfect positive correlation
-1	Perfect negative correlation
$r \leq -0.8$	Higher/Strong negative correlation
$-0.8 < r < -0.2$	Moderate negative correlation

$-.2 \leq r < 0$	Weak negative correlation
$0 \leq r < 0.2$	Weak positive correlation
$0.2 \leq r < 0.8$	Moderate positive correlation
$r \geq 0.8$	Strong positive correlation

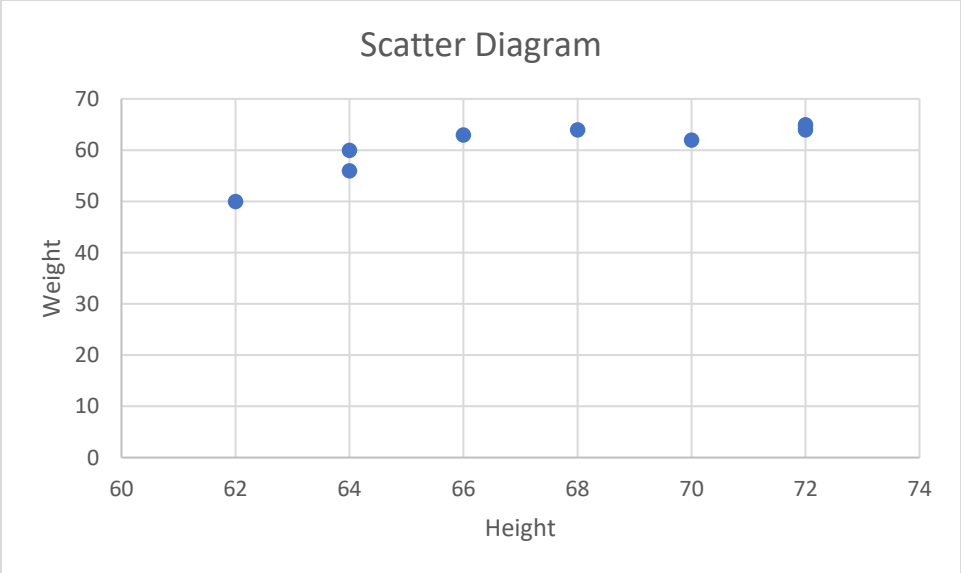
Example 1:

Followings are height and weights of 8 students .

Height (inch):	62	72	68	64	70	66	64	72
Weight (Kg):	50	65	64	56	62	63	60	64

i. Draw a scatter diagram and interpret.
ii. Calculate the coefficient of correlation and interpret.

(i)



It seems from the scatter diagram that there is a strong positive correlation between price and demand of a commodity.

(ii) Let, price and demand be denoted by x and y respectively.

We know, $r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\{\sum {x_i}^2 - \frac{(\sum x_i)^2}{n}\}\{\sum {y_i}^2 - \frac{(\sum y_i)^2}{n}\}}}$

Calculation table

Height (x_i)	Weight (y_i)	${x_i}^2$	${y_i}^2$	$x_i y_i$
62	50	3844	2500	3100
72	65	5184	4225	4680
68	64	4624	4096	4352
64	56	4096	3136	3584
70	62	4900	3844	4340
66	63	4356	3969	4158
64	60	4096	3600	3840
72	64	5184	4096	4608
$\sum x_i$= 538	$\sum y_i$ = 484	$\sum {x_i}^2$= 36284	$\sum {y_i}^2$=29466	$\sum x_i y_i$ =32662

Here, n = 8

Now, r = $\frac{32662 - \frac{538 \times 484}{8}}{\sqrt{(36284 - \frac{538^2}{8})(29466 - \frac{484^2}{8})}}$

= $\frac{113}{138}$

= 0.819

Interpretation:

This result show that the relationship between height and weight is strong and positive.

i.e. weight of a person increase if height increases and vice versa.