Correlation

Correlation is a measure of association between two variables .
The variables are not designated as dependent or independent.
The value of a correlation coefficient can vary from minus one to plus one (-1 to +1), where the calculated value of the correlation coefficient indicates the strength of the relationship while the negative or positive signal indicates the direction of the relationship (direct or negative correlation).
A minus one (-1) indicates a perfect negative correlation, while a plus one (+1) indicates a perfect positive correlation. A correlation of zero means there is no relationship between the two variables.
When there is a negative correlation between two variables, as the value of one variable increases, the value of the other variable decreases, and vise versa.
In other words, for a negative correlation, the variables work opposite each other. When there is a positive correlation between two variables, as the value of one variable increases, the value of the other variable also increases. The variables move together.

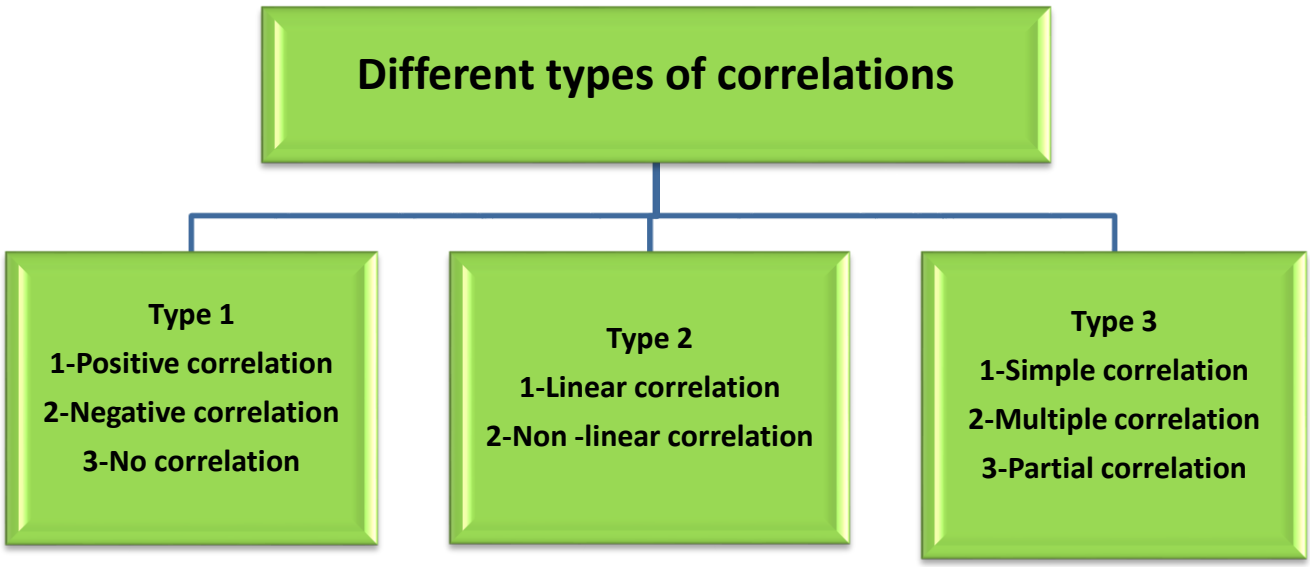
Note: In general,

- the relationship can be considered weak if the correlation coefficient value is less than 0.30.
- the relationship can be considered as medium if the correlation coefficient value ranges from 0.30 to 0.70.
- if the correlation coefficient value is more than 0.70 the strong relationship between the two variables.

- **Note**: we can use scatter diagramed {*the value of the first variable on the x-axis and the value of the second variable on the y-axis*}to give a quick idea of the strength and direction of the correlation between two variables.

Different types of correlations

There are three ways to classify the correlation:



Type1

• **Positive correlation**: If two related variables are such that when one increases (decreases), the other also increases (decreases)

•**Negative correlation**: If two variables are such that when one increases (decreases), the other decreases increases)

• **No correlation**: If both the variables are independent.

Type 2

• **Linear correlation**: When plotted on a graph it tends to be a perfect line.

• **Non-Linear correlation**: When plotted on a graph it is not a straight line.

Type 3

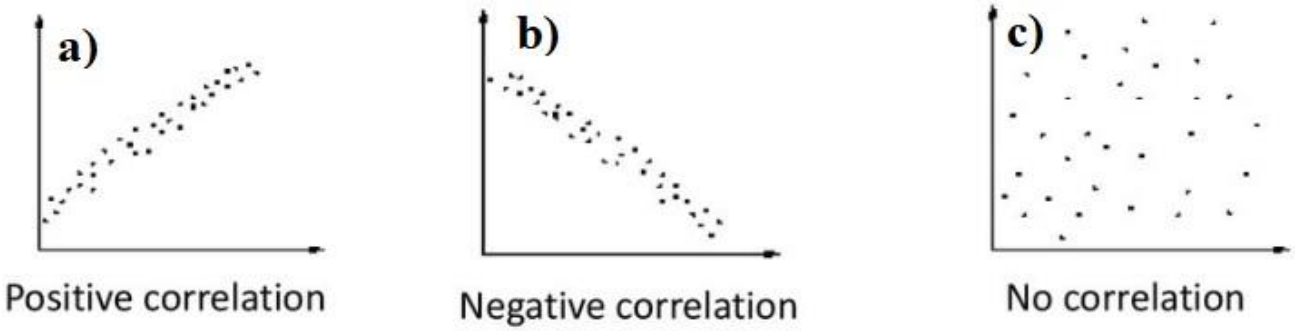
Simple correlation: In this only two variables are studied.

Multiple correlation: In this three or more variables are studied simultaneously.

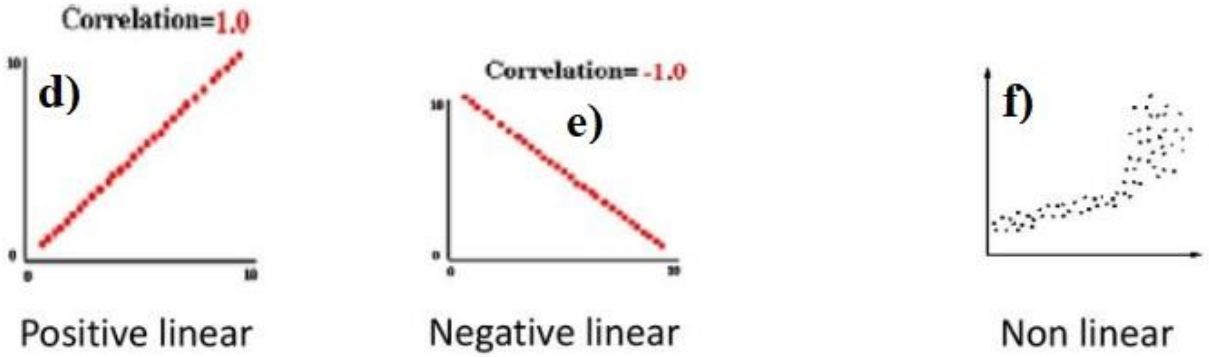
Partial correlation: we recognize more than two variables but consider only two variables to be influencing each other and effect of other influencing variables being kept constant.

Graphical representation of type1 and type 2 correlation

Type 1



Type 2



Interpret a Correlation Coefficient

Correlation Coefficient = 0	No linear relationship
Correlation Coefficient = ± (0.01 – 0.49)	A weak linear relationship
Correlation Coefficient = ± (0.50 – 0.69)	A moderate relationship
Correlation Coefficient = ± (0.70 – 0.90)	A strong linear relationship
Correlation Coefficient = Exactly ±1.	A perfect linear relationship

Types of correlation coefficient formulas

Usually, in statistics, we measure four types of correlations:

1) Pearson correlation
2) Kendall rank correlation
3) Spearman correlation
4) Point-Biserial correlation.

1) Pearson Correlation(r)

A Pearson correlation is a statistical formula that measures linear correlation between two variables X and Y. It has a value between (+1 and −1), where 1 is total positive linear correlation, 0 is no linear correlation, and −1 is total negative linear correlation.

Pearson correlation is widely used in the sciences.

Pearson Correlation (r) – Formula

A Pearson correlation between variables X and Y is calculated by

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

Where,

- r = Pearson Coefficient
- n= number of the pairs of the stock
- ∑xy = sum of products of the paired stocks
- ∑x = sum of the x scores
- ∑y= sum of the y scores
- $\sum x^2$ = sum of the squared x scores
- $\sum y^2$ = sum of the squared y scores

Example 1

Find the Pearson Coefficient (r) for the following table:

No	(x)	(y)
1	40	78
2	21	70
3	25	60
4	31	55
5	38	80
6	47	66

Solution:

For the Calculation of the Pearson Correlation Coefficient, we will first calculate the following values,

Sr. No	(x)	(y)	xy	x^2	y^2
1	40	78	3120	1600	6084
2	21	70	1470	441	4900
3	25	60	1500	625	3600
4	31	55	1705	961	3025
5	38	80	3040	1444	6400
6	47	66	3102	2209	4356
Total (Σ)	202	409	13937	7280	28365

Here the total number of variables are 6 so, n=6

Now the calculation of the Pearson (r) is as follows,

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

$$r = \frac{6 * (13937) - (202)(409)}{\sqrt{[6 * 7280 - (202)^2] * [6 * 28365 - (409)^2]}}$$

$$r = 0.35$$

Thus the value of the Pearson correlation coefficient is 0.35

(A weak linear relationship)