Interpretation of Coefficient of Correlation

What is the Correlation Coefficient?

In statistics, the correlation coefficient indicates the strength of the relationship between two variables. When we say that two variables are correlated, it means that there exists a definable relationship to each other. If there is a positive correlation between them, it means that when the value of one variable goes up, the value of the other goes up as well and conversely when the value of one variable goes down, the value of the other variable goes down as well. If there is a negative correlation between the two variables, this means that as the value of one variable goes up, the value of the other variable falls and conversely when the value of one variable falls, the value of the other variable goes up.

An example of positive correlation in simple words - is the relationship between the Hot during the summer versus the AC usage. When the hot will be more the more consumption of AC will come into picture. And the vice versa. Similarly for the negative correlation is the relationship between the supply and demand of any products. Hence if the supply of the product rises in the market, its demand decreases and the vice versa.

The correlation coefficient is used to measure the strength of the linear relationship between two variables on a graph. It also plots the direction of their relationship. The correlation coefficient is calculated by the following formula:

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{\left[n\Sigma x^2 - (\Sigma x)^2\right]\left[n\Sigma y^2 - (\Sigma y)^2\right]}}$$

r: This is the correlation coefficient

n: n specifies the number of values we're looking at. If we had five instances we were calculating the correlation coefficient for, the value of n would be 5.

x: This is the first data variable

y: This is the second data variable.

Σ: The Sigma symbol (Greek) is used to calculate the sum of anything placed next to it.

Let's break it down to understand in a better way. If we were calculating the relationship between weight loss and exercise, for example, weight loss would be variable x and exercise would be variable y. If we were doing it for 10 people, the value of n would be 10. After calculating the results, we would get the value of r.

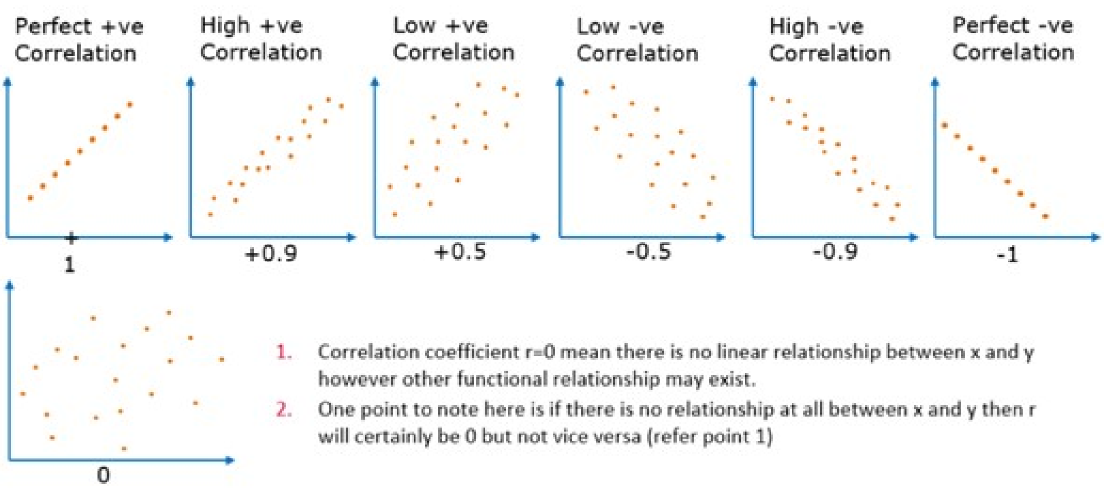
Interpreting the Correlation Coefficient

Let's continue using the same example as above to help us to interpret the correlation of coefficient. You're probably thinking that the more you exercise, the more the weight you lose right? That's true in some cases, of course, but not all the time. Some people gain weight at the gym, others lose it. This jumbles the relationship between the two variables we're studying (exercise and weight loss).

It's very rare to get a perfect positive (+1) or perfect negative (-1) relationship between two variables in the real world. Because of this, the value of the correlation coefficient will usually drifts between 1 and -1, depending on the strength of the relationship between the two variables.

Other examples of a positive correlation are:

- The more education years you complete the higher degree you will gain.
- The less time you spend doing business marketing, the fewer new clients you get.
- The more time you invest money, the more compound interest it gains
- The more money you save, the more secure you feel financially.



Other examples of a negative correlation are:

- If a train increases speed, the length of time to reach the destination decreases.
- If it is darker outside, more light is needed inside.
- If a car tire has more air/ pressure, the car may use less gas per mile.

Understanding Data Points

What are these data variables we keep mentioning? They can be anything, like the height of a person, a person's age or goods like candy, or a basketball. Why are these data variables relevant? It's obvious that a child will consume more candy than a grown up. However, a child may not play as much basketball as a teenager or an adult. There is an inverse (negative) relationship between the data variables the age of the person and the candy consumed- as a person grows older, he consumes less candy. However, there is a direct (positive) relationship between the age of a person and time spent playing basketball- a person plays more basketball as he grows older.

If you were in a company who is making candy or basketballs, you will want to know the relationship between these data variables, so that you can target your product toward a particular spectrum of the population.

Example: Let's calculate the correlation of coefficient for a set of data to help you understand the formula better. Let's take ages of children as our x variable and the candy consumed as y variable. Let's assume that after research we found that as kids got older, they ate less candy. The following are the value of the data variables we received for three children (or a single child at three stages of his life.

x (Age of Child)	Y (Candy Consumed)
6	10
7	9
8	8

Step 1: Find all the values we need

Here, n (number of variables in x and y) would be 3.

x	y	xy	x2	y2
6	10	60	36	100
7	9	63	49	81
8	8	64	64	64

The other values we need are:

Σx =6 + 7 + 8 = 21

Σy = 10 + 9 + 8 = 27

Σxy = 60 + 63 + 64 = 187

Σx2 = 36 + 49 + 64 = 149

Σy2 = 100 + 81 + 64 = 245

Step 2: Input Values into the Formula

(r) =[nΣxy – (Σx)(Σy) / Sqrt([nΣx2 – (Σx)2][nΣy2 – (Σy)2])]

r = [3(187) – (21)(27) / Sqrt ([3(149) – (21) 2][3(245)-(27) 2])]

r= [561-567/ Sqrt ([447-441][735-729])]

r= [-6/ Sqrt ([6] [6])]

r=[-6/ Sqrt (36)]

r= -6/ 6

r= -1

Observations -

- The co relational coefficient we obtained is a perfect minus 1.
- This shows that there is a perfect negative (inverse) match between the two variables.
- As the value of one variable increases, the value of the other variable goes down. It is obvious that in reality, there would be some kids who didn't eat candy when they were young, kids who didn't eat much candy when they were young and kids who ate different amounts of candy at different ages.