

Pearson Correlation Coefficient

Pearson correlation coefficient measures strength between the different variables and their relationships. Pearson's correlation coefficient returns a value between -1 and 1. The interpretation of the correlation coefficient is as under:

- If the correlation coefficient is -1, it indicates a strong negative relationship. It implies a perfect negative relationship between the variables.
- If the correlation coefficient is 0, it indicates no relationship.
- If the correlation coefficient is 1, it indicates a strong positive relationship. It implies a perfect positive relationship between the variables.

A higher absolute value of the correlation coefficient indicates a stronger relationship between variables. Thus, a correlation coefficient of 0.78 indicates a stronger **positive correlation** as compared to a value of say 0.36. Similarly, a correlation coefficient of -0.87 indicates a stronger **negative correlation** as compared to a correlation coefficient of say -0.40.

Pearson's Correlation Coefficient formula is as follows,

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

Where,

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

Example question: Find the value of the correlation coefficient from the following table:

SUBJECT	AGE X	GLUCOSE LEVEL Y
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

Solution:

SUBJECT	AGE X	GLUCOSE LEVEL Y	XY	X ²	Y ²
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
Σ	247	486	20485	11409	40022

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

$$r = 6(20,485) - (247 \times 486) / [\sqrt{[6(11,409) - (247^2)] \times [6(40,022) - 486^2]}]$$

$$= 0.5298$$

The range of the correlation coefficient is from -1 to 1. Our result is 0.5298 or 52.98%, which means the variables have a moderate positive correlation.