**Measures of Association OR Covariance**

**Correlation Coefficient:**

**Covariance:**

A measure of how much two variables change together. It indicates the degree to which two variables vary together.

1. Positive covariance
2. Negative Covariance

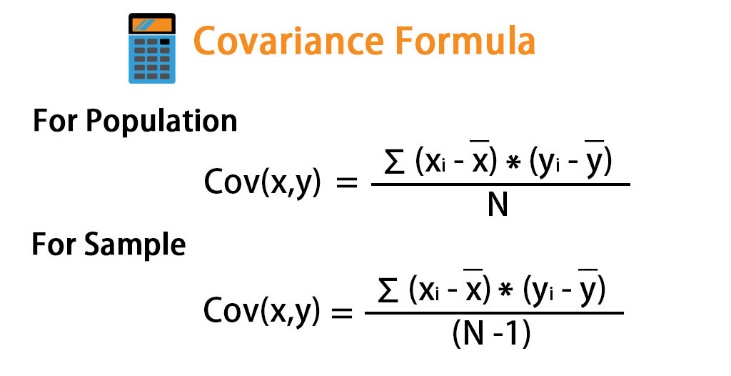
**Positive covariance**

If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values, the covariance is positive.

**Negative Covariance**

In the opposite case, when the greater values of one variable mainly correspond to the lesser values of the other, the covariance is negative.

The sign of the covariance, therefore, shows the tendency in the linear relationship between the variables. The magnitude of the covariance is not easy to interpret because it is not normalized and hence depends on the magnitudes of the variables. The normalized version of the covariance, the correlation coefficient, however, shows by its magnitude the strength of the linear relation.



**+ Ve Covariance**

**X is Increasing and Y is Increasing  
X is Decreasing and Y is Decreasing**

**-Ve Covariance**

**X is Increasing and Y is Decreasing  
X is Decreasing and Y is Increasing**

**Covariance 0:**

**No relation between X and Y**

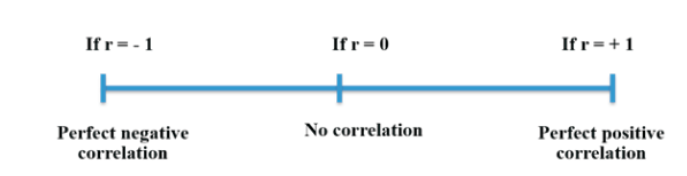
**Correlation Coefficient:**

Correlation estimates the depth of the relationship between variables.

The correlation coefficient is a measure of the strength and direction of the linear relationship between two variables. It quantifies the degree to which changes in one variable are associated with changes in another variable. There are several types of correlation coefficients, including

1. Pearson correlation coefficient,
2. Spearman rank correlation coefficient, and

The relationship between the correlation coefficient and covariance is given by;

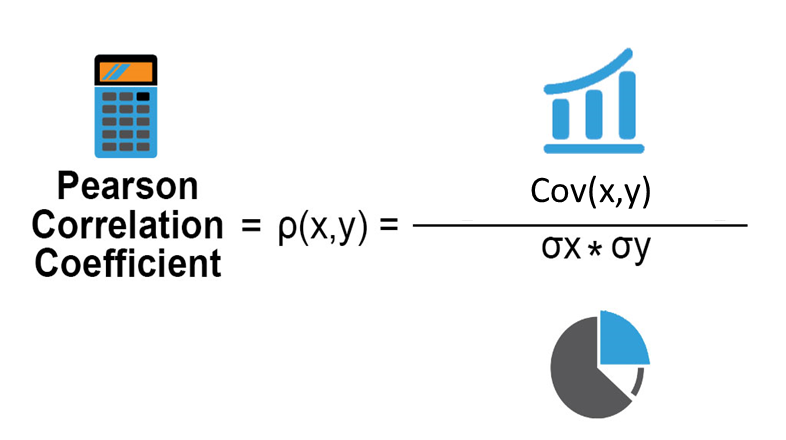


**Correlation,ρ(X,Y) = Cov(X,Y)/σXσy**

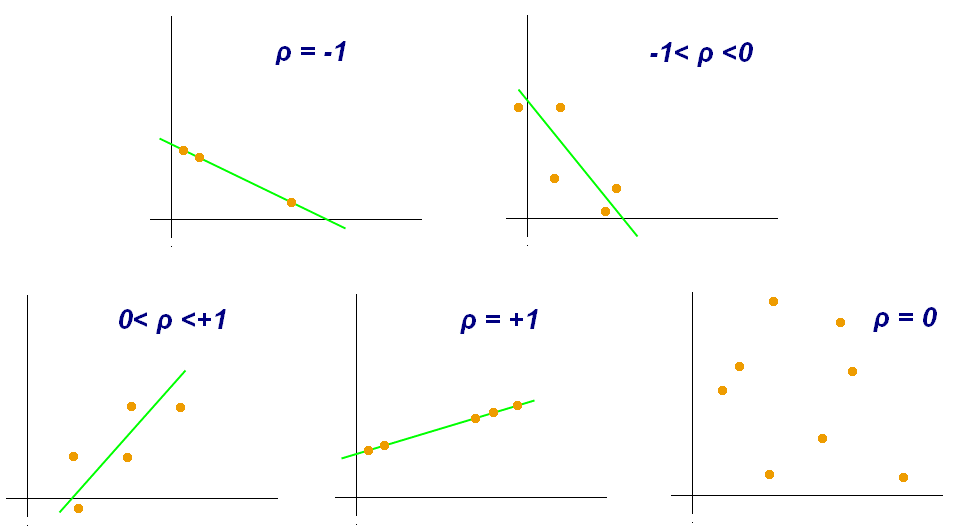
**Pearson correlation coefficient:**

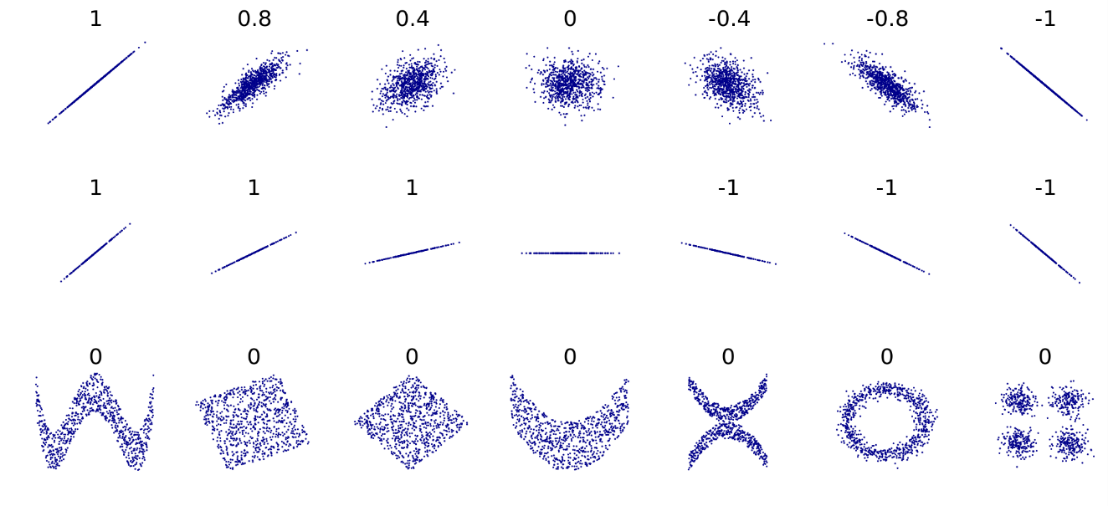
Restricting the values between(-1 to +1)

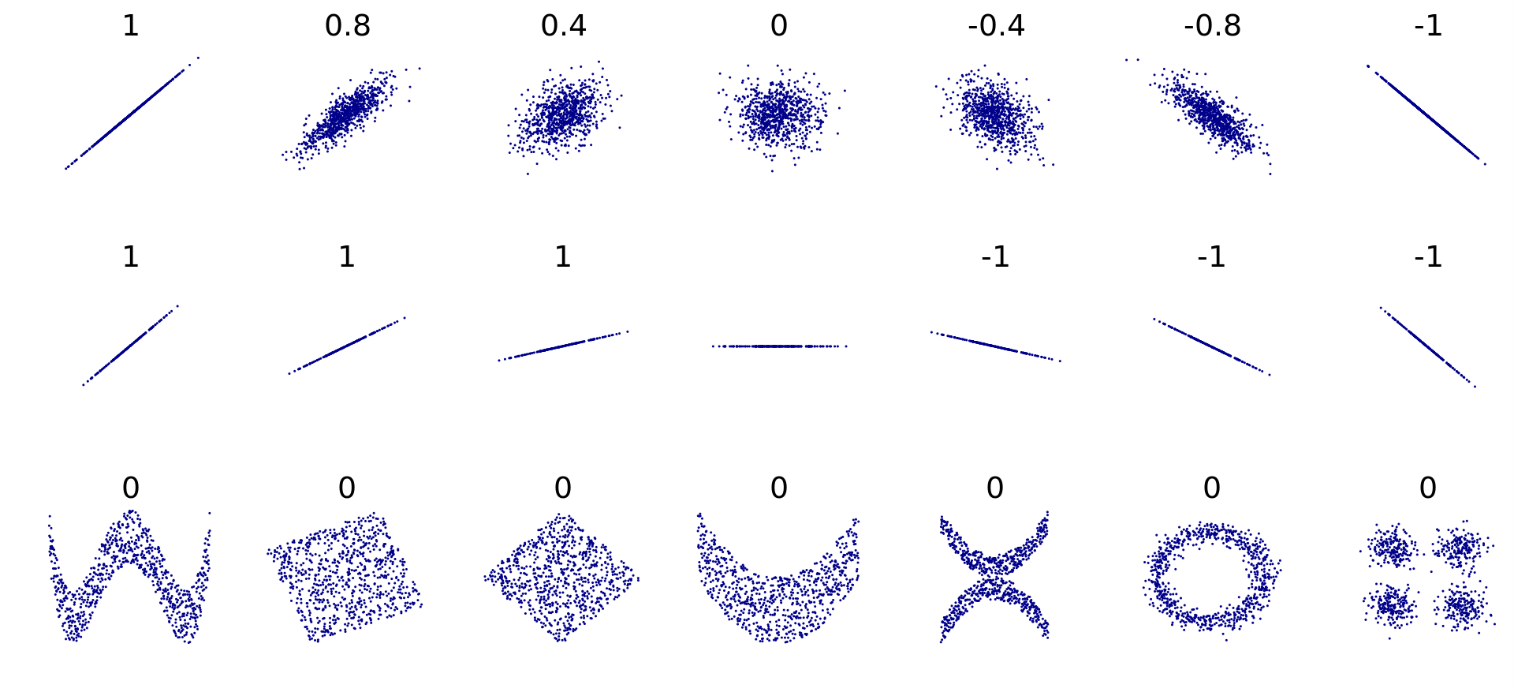
The more the values toward the +1 more positively correlated it is and vice-versa.

****

**Different types of correlation are shown in the below figures.**





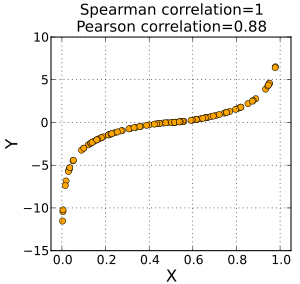


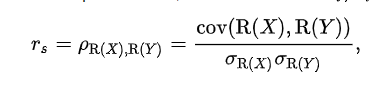
**Spearman correlation:**

It is used only for non-linear data.

It assigns ranks

Spearman's rank correlation measures the strength and direction of association between two ranked variables. It basically gives the measure of monotonicity of the relation between two variables i.e. how well the relationship between two variables could be represented using a monotonic function.



****

**Covariance vs correlation**

In summary, covariance measures the directional relationship between variables but does not give a sense of the strength of the relationship. The correlation coefficient, on the other hand, standardizes covariance, providing a clear indication of both the direction and strength of the linear relationship between two variables.

Each type of correlation coefficient has its own strengths and weaknesses and is appropriate for different types of data and research questions. Choosing the appropriate correlation coefficient depends on the nature of the variables being studied and the assumptions underlying the analysis.

|  |  |
| --- | --- |
| Pearson Correlation Coefficient (r): | Spearman Rank Correlation Coefficient (ρ): |
|  |  |
| The Pearson strength and direction of a linear relationship between two continuous variables, | The Spearman correlation coefficient measures the strength and direction of the monotonic relationship between two non linear variables. |
| It ranges from -1 to +1, where: | Spearman's rho ranges from -1 to +1, similar to Pearson's correlation coefficient. |
| It is based on actual data values. | It is based on the ranks of the data rather than the actual values. |
| +1 indicates a perfect positive linear relationship (both variables increase together). | +1 indicates a perfect monotonic positive relationship, meaning that higher ranks of one variable are associated with higher ranks of the other variable. However, this doesn't necessarily imply a linear relationship in the traditional sense. |
| -1 indicates a perfect negative linear relationship (one variable increases as the other decreases). | -1 indicates a perfect monotonic negative relationship, where higher ranks of one variable are associated with lower ranks of the other variable. Again, this doesn't imply a linear relationship in the conventional sense, but rather a consistent trend in the rankings of the data. |
| 0 indicates no linear relationship. | 0 indicates that there is no monotonic association between the variables, but it doesn't rule out the presence of other types of relationships or associations between the variables. |
| Example: In a study examining the relationship between hours of study and exam scores, a high positive Pearson correlation coefficient suggests that students who study more tend to achieve higher exam scores. | Ranking of Student Performance: Suppose you want to assess the relationship between students' ranks in a class based on their performance in two different subjects, such as math and English. You collect data on the ranks of students in each subject, where 1 indicates the highest performer and n indicates the lowest performer. By calculating Spearman's ρ, you can determine the strength and direction of the relationship between the rankings in math and English. |
|  | Consumer Preferences: Imagine you conduct a survey to evaluate the preferences of consumers for two competing brands of smartphones. Each respondent ranks the two brands based on factors like price, features, and customer service. By analyzing the rankings using Spearman's ρ, you can assess if there is a consistent trend in the preferences of consumers between the two brands. |
|  | Comparison of Job Satisfaction: Consider a study investigating the job satisfaction of employees in different departments of a company. Employees in each department rank their level of satisfaction on a scale, and Spearman's ρ is used to analyze the relationship between department rankings and job satisfaction rankings. This helps identify if certain departments have higher or lower levels of job satisfaction compared to others. |
|  | Evaluating Sports Rankings: In sports analytics, Spearman's ρ can be used to assess the consistency of rankings between different performance metrics for athletes or teams. For example, you might analyze the rankings of tennis players based on their performance in various tournaments to determine if there is a consistent relationship between their rankings in different tournaments. |