MEASURES OF CORRELATION

INTERPRETATION OF CORRELATION

A correlation shows how two variables in a series are related to one another, meaning that changes in one variable's value will also affect changes in the other variable's value.

CORRELATION AND CAUSATION

1. A third variable influencing both variables: A high level of correlation between the two variables could be the result of a third variable that was left out of the analysis.

2. Mutual Dependence (Cause and Effect): Two variables that exhibit a high degree of correlation may be reacting to one another, making it challenging to determine which is the cause and which is the effect.

3. Pure Chance : It's possible that the correlation between the two variables was the result of random chance or pure coincidence. We refer to this kind of association as spurious. It is crucial to determine whether there is a chance of a relationship between the variables under investigation in order to properly understand the correlation coefficient.

IMPORTANCE OR SIGNIFICANCE OF CORRELATION

The following factors make correlation research extremely important:

1. The degree of link between two variables in a single figure can be determined with the use of the correlation coefficient.
2. Correlation analysis makes it easier to comprehend how the economy behaves and aids in identifying the crucial variables that other factors depend on.
3. One variable's value can be estimated based on the value of another when there is a correlation between the two. The use of regression equations is used for this.
4. In the business realm, correlation makes decision-making easier. Because correlation-based forecasts are more likely to be accurate and close to reality, it lowers the range of uncertainty.

TYPES OF CORRELATION

1. Positive Correlation: A relationship is said to be positive when two variables move in the same direction, that is, when one rises, the other rises as well, and when one falls, the other falls.

2. Negative Correlation: This type of relationship occurs when two variables move in the opposite directions, that is, when one increases, the other lowers, and vice versa.

LINEAR AND NON-LINEAR (CURVILINEAR) CORRELATION

1. Linear Correlation: A relationship between two variables is considered to be linear if there is a constant ratio between the changes in the two variables. When variables with this kind of relationship are plotted, the result is a straight line.
2. Non-Linear (Curvilinear) Correlation: In this type of correlation, there is no continuous relationship between the change in one variable and the change in the curvilinear other associated variable.

SIMPLE, MULTIPLE AND PARTIAL CORRELATION

1. Simple Correlation: A simple correlation problem is one in which just two variables are examined.
2. Multiple Correlation: The term "multiple correlation" refers to a relationship that is examined concurrently between three or more variables.
3. Partial Correlation: In partial correlation, two variables are compared while holding the other variables constant.

DEGREE OF CORRELATION

Perfect Correlation

A perfect correlation exists when there is an equal proportion of change in the values of the two variables as a result of their association.

Zero correlation

Zero correlation occurs when there is no relationship at all between the two variables.

METHODS OF MEASUREMENT CORRELATION

The correlation between two variables can be measured using a variety of techniques. Among the first are the scatter diagrams.
2. The Correlation Coefficient of Karl Pearson
3. The correlation coefficient of Spearman's Rank

SCATTER DIAGRAM

A scatter diagram is a straightforward and visually appealing way to diagrammatically display a bivariate distribution and ascertain the type of correlation that exists between the variables.

- It is the most straightforward technique for examining the relationship between two variables without the need to compute any numbers.
- Drawing a Scatter Diagram: How to Do It Plot the variables X and Y's values along the corresponding X- and Y-axes. On the graph, indicate the values of two variables with dots. Every dot stands for a set of two values.

- An illustration of the type of relationship between the two variables can be seen in a scatter diagram.

MERITS AND DEMERITS OF SCATTER DIAGRAM

MERITS OF SCATTER DIAGRAM

1. Simplicity: The technique for examining correlation between two variables is straightforward and non-mathematical.
2. Easy to understand: It is simple to comprehend and interpret. It makes it possible for us to quickly determine whether correlation is there in the diagram.
3. Not impacted by extreme items: Unlike most mathematical techniques, it is unaffected by the magnitude of extreme values.
4. First stage: Examining the relationship between two variables is the first stage in the process.

DEMERITS OF SCATTER DIAGRAM

1. Non-mathematical technique: Unlike other mathematical ways of correlation, this method is unable to provide the precise numerical value of correlation.

2. Rough Measure: This just provides a general and approximative understanding of the strength and type of connection between two variables. As such, it is not a quantitative expression; rather, it is simply a qualitative one.

3. Unsuitable for extensive observations: When there are more than two variables, a scatter diagram cannot be drawn on graph paper.

KARL PEARSON'S COEFICIENT OF CORRELATION

Karl Pearson states that one can calculate the coefficient of correlation by dividing the total of the products of the deviations from each mean by the number of pairs and the standard deviations.

i.e. $r = \frac{\Sigma xy}{N \times \sigma_x \times \sigma_y}$

PROPERTIES OF COEFFICIENT OF CORRELATION

1. The coefficient of correlation is between -1 and + 1: This characteristic of r will be helpful in ensuring that our computations are accurate. There is a calculation mistake if the calculated value of r is outside of these bounds.
2. The correlation coefficient remains unaffected by changes in the origin and scale of measurements: We shall note that the correlation coefficient is unaffected by changes in the origin and scale of both variables.
3. The linear relationship is measured by the coefficient of correlation (r): When two variables rise or fall together, the relationship is positive (r is positive); when one variable's values rise as

the other's values fall and vice versa, the relationship is negative.
4. The correlation coefficient between two independent variables: X and Y, will be zero.

MERITS AND DEMERITS OF COEFICIENT OF CORRELATION

Merits

The advantages of the Karl Pearson approach are:

1. Popular Method: When examining the correlation between two variables, this is the most well-liked and frequently applied mathematical method.
2. Degree and direction of correlation: The correlation coefficient condenses both the degree and the direction of correlation, or whether the correlation is positive or negative, into a single figure.

Demerits

The following are Karl Pearson's method's primary limitations:
1. Affected by extreme values: The value of extreme items has an excessive impact on the correlation values.
2. Time Consuming procedure: This procedure takes longer than the others.

3. Assumption of linear relationship: Whether or not this assumption is true, the correlation coefficient consistently uses it to infer a linear relationship.

4. Potential for incorrect interpretation: The coefficient of correlation is frequently understood incorrectly, thus one must exercise extreme caution when evaluating its value.

SPEARMAN'S RANK CORRELATION

Using this method, different things are ranked based on their attributes, and a correlation between these ranks is calculated.

COMPUTATION OF RANK CORRELATION

1. When Ranks are given.

2. When Ranks are not given.

Apply the following formula : $r_k = 1 - \frac{6\Sigma D^2}{N^3 - N}$

3. When Ranks are Equal or Repeated

$$r_k = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \cdots..\right]}{N^3 - N}$$

MERITS AND DEMERITS OF RANK CORRELATION

MERITS

1. Compared to Karl Pearson's method, this one is simpler to compute and comprehend. The computation time is reduced.
2. When dealing with qualitative data, such as honesty, beauty, intelligence, voice quality, etc., the rank approach is quite helpful. In these situations, several factors under consideration are given ranks.
3. When we are provided the ranks but not the actual data, this is the only approach that can be applied.
4. This method can be used to acquire a general impression of the degree of association when actual values are provided (instead of ranks).

DEMERITS

1. In a bivariate (grouped) frequency distribution, correlation cannot be determined using the rank approach.
2. Determining the ranks and their differences becomes challenging when there are a lot of values. Because of this, it is best to apply this strategy only in cases when there are fewer than thirty observations.
3. Compared to Karl Pearson's method, this approach is less precise. Ranks are used in place of the original values.

FORMULAE AT A GLANCE

Karl Pearson's Coefficient of Correlation

1.	Actual mean method	r = $\frac{\Sigma xy}{N\times\sigma_x\times\sigma_y} = \frac{\Sigma xy}{N\times\sqrt{\frac{\Sigma X^2}{N}\times\sqrt{\frac{\Sigma y^2}{N}}}} = \frac{\Sigma xy}{\sqrt{\Sigma x^2\times\Sigma y^2}}$
2.	Direct Method	r = $\frac{N\Sigma XY-\Sigma X.\Sigma Y}{\sqrt{N\Sigma X^2-\Sigma X^2\sqrt{N\Sigma Y^2}}-\Sigma Y^2}$
3.	Short-cut Method	r = $\frac{N\Sigma dxdy - \Sigma X\times\Sigma Y}{\sqrt{N\Sigma dx'^2-\Sigma dx'^2\times\sqrt{N\Sigma dY^2}}-\Sigma dY^2}$
4.	Step Deviation Method	r = $\frac{N\Sigma dx'dy'- \Sigma dx'\times\Sigma dy'}{\sqrt{N\Sigma dx'^2-\Sigma dx'^2\times\sqrt{N\Sigma dy'^2}}-\Sigma dy'^2}$
Spearman's Rank Correlation Coefficient		
When Ranks are not equal		$r_k = 1 - \frac{6\Sigma D^2}{N^3 - N}$

When Ranks are equal	$r_k = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \cdots ..\right]}{N^3 - N}$