CORRELATION ANALYSIS

What is Correlation?

It is a statistical method or a statistical technique that measures the degree of association between two or more variables.

DEFINITION

- According to L.R.Connor,''When two or more quantities vary in sympathy so that movements in the one tend to be accompanied by corresponding movements in the others ,then they are said to be correlated.''
- According to Ya Lun Chow ,'' Correlation analysis attempts to determine the degree of relationship between variables.''.
- According to W.L.King,''Correlation means that between two series or groups of data ,there exists some casual connection,''

IMPORTANCE OF CORRELATION

- The study of Correlation shows the direction and degree of relationship between the variables .This has helped the formation of different laws and concept in economic theory.
- It is very helpful in understanding economic behaviour .This is helpful in studying factors by which economic events are affected.
- Study of correlation reduces the range of uncertainties in matter of prediction.
- Helpful in investigation and research.
- It is also helpful in policy formulation.

KINDS OF CORRELATION

- Correlation can be:
- Positive and Negative Correlation
- Linear and Non- Linear Correlation
- Simple ,Multiple and Partial Correlation

Kinds of Correlation
Positive and Negative Correlation.

- **Positive correlation** - When two variables X and Y move in the same direction,i.e.,when one increases the other also increases and when one decreases the other also decreases, the correlation between the two is positive .For example, Price and supply of a commodity.

Positive Correlation	
X	Y
8	4
16	6
24	10

Positive Correlation	
X	Y
15	35
10	25
5	20

Kinds of Correlation
Positive and Negative Correlation

- **Negative correlation:** When two variables X and Y move in the opposite direction, the correlation is negative .For example ,reverse relationship between price and demand of a commodity.

Negative Correlation	
X	Y
8	32
16	24
24	8

Negative Correlation	
X	Y
30	45
20	55
10	70

Kinds of Correlation
Linear and Non-Linear Correlation

- **Linear Correlation:If the ratio of change between two variables is uniform ,it is called Linear Correlation.If the changes are plotted on a graph paper ,their relationship will be indicated by a straight line**
- **Example:.**

Linear Correlation	
X	Y
2	5
4	10
6	15

Kinds of Correlation
Linear and Non-Linear Correlation

- Non-Linear Correlation :If the ratio of change between two variables is not uniform,It is called Non-Linear Correlation.If these changes are plotted on a graph paper ,they will not form a straight line but a curve.
- Example:

Non- Linear Correlation	
X	Y
2	5
4	8
6	12

Kinds of Correlation :Simple,Multiple and Partial Correlation

- **Simple Correlation**: Relationship between two variables is known as Simple Correlation.For example ,relationship between price and demand of a commodity.
- **Multiple Correlation:**When the relationship among three or more than three variables is studied simultaneously, it is called Multiple Correlation.For example, agricultural production depends on rainfall, amount of mannures,seeds etc. This will be called Multiple Correlation.
- **Partial Correlation**: Relationship between two variables is established keeping other variables constant. For example, If we study the relationship between degree of rainfall and agricultural production assuming amount of fertilizers, quality of seeds as constant ,it will be known as Partial Correlation.

DEGREE OF CORRELATION

- Perfect Correlation
- Absence of Correlation or No Correlation
- Limited degree of Correlation

(a)High degree of Correlation

(b)Moderate degree of Correlation

(c)Low degree of Correlation

DEGREE OF CORRELATION:PERFECT CORRELATION

- Degree of Correlation refers to the Coefficient of Correlation. There may be following types of Positive and Negative Correlation:
- 1.Perfect Correlation: When two variables change in the same proportion. It may be of two kinds:
- (a)Perfect Positive :When proportional change in two variables is in the same direction, it is called Perfect Positive Correlation .In this case coefficient of correlation(r)=+1
- (b)Perfect Negative :When proportional change in two variables is in opposite direction ,it is called perfect negative correlation. In this case, coefficient of correlation(r)= -1

DEGREE OF CORRELATION:ABSENCE OF CORRELATION

- 2.Absence of Correlation: If there is no relation between two variables ,it is called no correlation or absence of correlation. In this case ,coefficient of correlation(r)=0.

DEGREE OF CORRELATION:LIMITED CORRELATION

- Limited Correlation : Between perfect correlation and no correlation there is a situation of limited degree of correlation .In this case the coefficient of correlation (r) is more than zero and less than one ,i.e. r>0 but <1.
- There are three types of limited degree of correlation:
- (a)High Correlation :When Correlation between two series is close to one, it is called high degree of correlation. In this case value of r lies between±0.75and±1.
- (b)Moderate Correlation :When Correlation between two series is neither large nor small, it is called Moderate degree of Correlation .In this case value of r lies between±0.25and±0.75.
- (c)Low Correlation :When the Correlation coefficient between two series is very small, it is called Low degree Correlation. In this case value of r lies between 0and±0.25.

DEGREE OF CORRELATION

DEGREE	POSITIVE	NEGATIVE
Perfect	+1	-1
High	Between +0.75and+1	Between -0.75and -1
Moderate	Between +0.25and+0.75	Between -0.25 and -0.75
Low	Between 0and+0.25	Between 0and -0.25
Zero	0	0

METHODS OF STUDYING CORRELATION:VARIOUS METHODS

- SCATEER DIAGRAM
- KARL PEARSON'S COEFFICIENT OF CORRELATION
- COEFFICIENT OF CORRELATION BY RANK DIFFERENCES
- COEFFICIENT OF CONCURRENT DEVIATION
- REGRESSION LINES AND REGRESSION COEFFICIENT

METHODS OF STUDYING CORRELATION:SCATTER DIAGRAM METHOD

- 1.Scatter Diagram Method :The existence of Correlation between variables can be shown graphically by means of a Scatter diagram. It is obtained by plotting value on a graph paper .The chart is prepared by measuring X-variable on horizontal axis and the Y-variable on vertical axis and all the observations are plotted on a graph . The cluster points ,so obtained on graph paper is called the Scatter diagram or dot diagram.

METHODS OF STUDYING CORRELATION:SCATTER DIAGRAM

- By observing the points we can know the degree and direction of Correlation.
- If the trend of the dotted points is Upward, rising from left bottom and going up towards the right top, Correlation is positive. On the other hand ,If the dotted point show a downward trend from the left top to the right bottom ,correlation is negative.
- If the plotted point do not show any trend ,the two variables are not correlated.

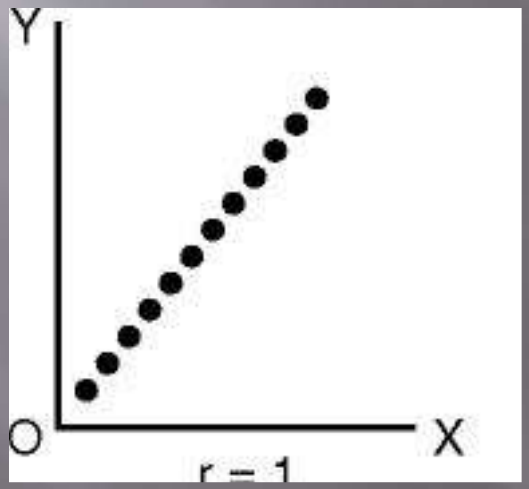
- Closeness of dots towards each other in a particular direction indicating higher degree of correlation.

METHODS OF STUDYING CORRELATION:SCATTER DIAGRAM

PERFECT POSITIVE CORRELATION

- If all the points lies on a straight line rising from the lower left hand corner to the upper right hand corner. In this case ,r=+1

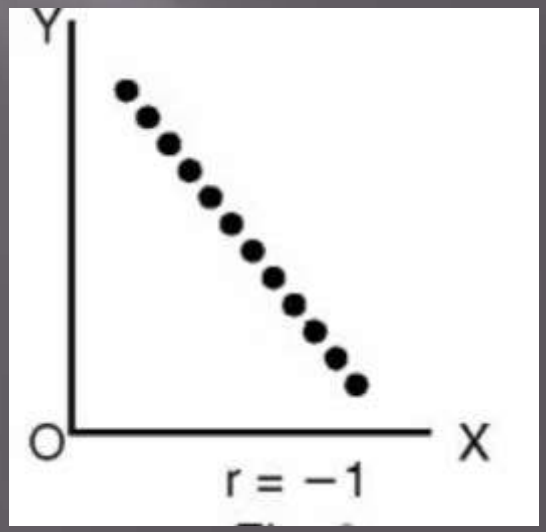
- r=+1



PERFECT NEGATIVE CORRELATION

If all the points lie on a straight line falling from the upper left corner to the lower right hand corner of the diagram. In this case,r=-1

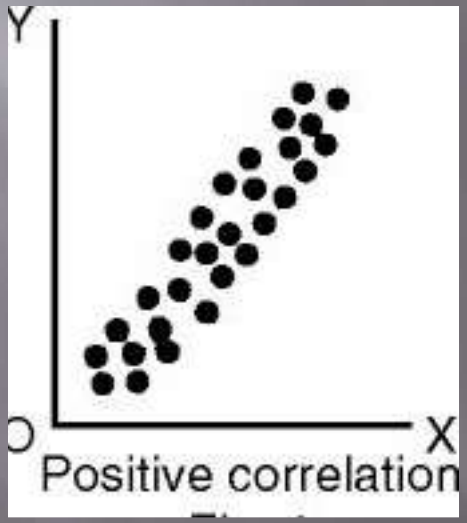
r =-1



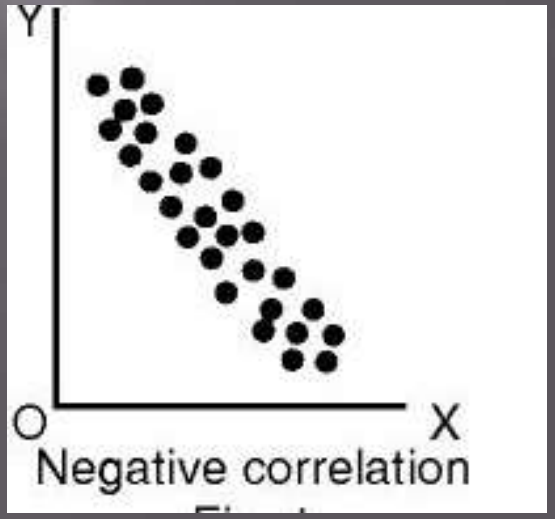
METHODS OF STUDYING CORRELATION:SCATTER DIAGRAM

- IF THE PLOTTED POINTS ARE VERY CLOSE TO EACH OTHER IT SHOWS HIGH DEGREE CORRELATION

High degree of positive correlation



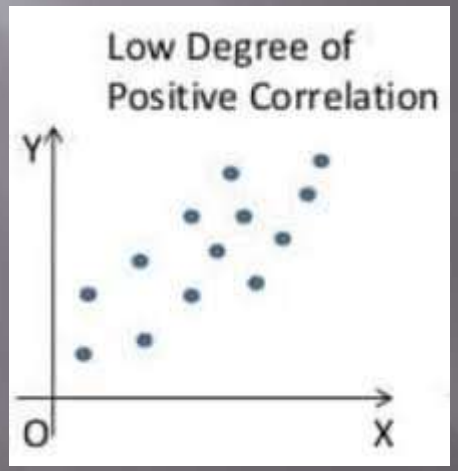
- High degree of negative correlation



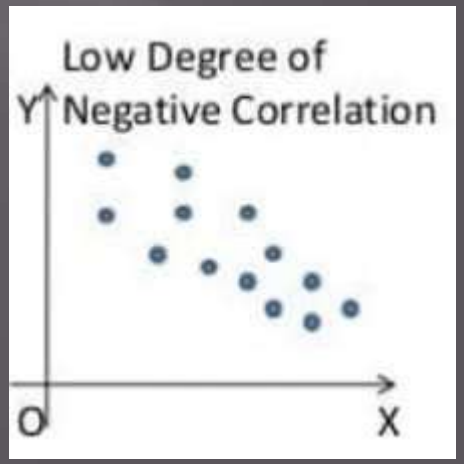
METHODS OF STUDYING CORRELATION:SCATTER DIAGRAM

If the plotted points are not very close to each other but show some upward or downward bend, there is a low degree of correlation

LOW DEGREE POSITIVE CORRELATION



LOW DEGREE NEGATIVE CORRELATION

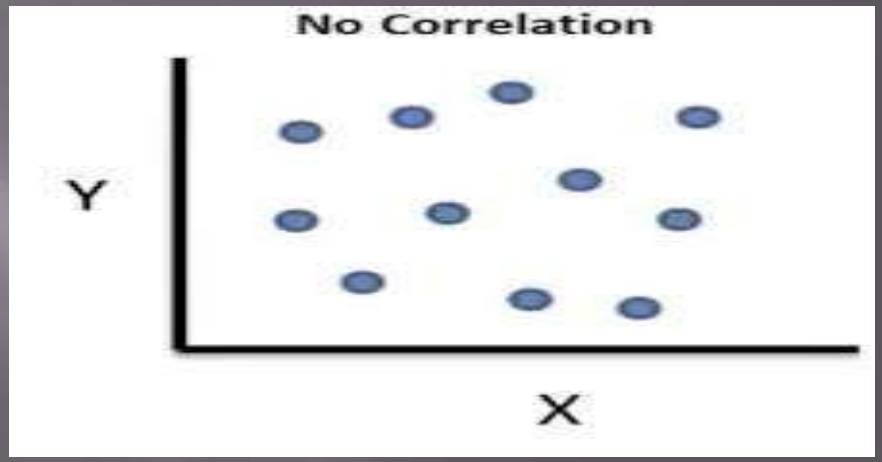


METHODS OF STUDYING CORRELATION:SCATTER DIAGRAM

ABSENCE OF CORRELATION OR NO CORRELATION

- If scatter diagram does not have any trend line and all the points in the diagram are highly scattered, it indicates No Correlation.

NO CORRELATION



MERITS AND DEMERITS OF SCATTER DIAGRAM

- **MERITS OF SCATTER DIAGRAM:**
- It is a simple and non mathematical method of knowing correlation.
- It gives a rough idea at a glance whether there is positive correlation, negative correlation or absence of correlation between variables.
- It is not affected by extreme values.
- **DEMERITS OF SCATTER DIAGRAM:**
- The exact degree of correlation can not be obtained from it.
- It gives only an approximate idea of relationship.

KARL PEARSON'S COEFFICIENT OF CORRELATION

- A mathematical method for measuring the linear relationship between the variable X and Y was suggested by the great biologist and statistician Karl Pearson.
- This method is also called Product Moment Method.
- According to Karl Pearson ,coefficient of correlation(written as r)of two variables is determined by dividing the total of the products by the corresponding deviation of the various items of two series from their respective means by the product of their standard deviation and the number of pairs of observation.

CALCULATION OF CORRELATION COEFFICIENT:DIRECT METHOD

- Formula
- According to Karl Pearson's method ,the coefficient of correlation is measured as

- $$r=\frac{\sum xy}{N\sigma x\sigma Y}$$
- Here, r=Coefficient of Correlation.
- X=$(X-\bar{X})$ =Deviation of value X from mean
- Y=$(Y-\bar{Y})$= Deviation of value Y from mean
- σx=Standard deviation of X series.
- σy=Standard deviation of Y series.
- N= No. of observations.

CALCULATION OF CORRELATION COEFFICIENT:DIRECT METHOD

- **A modified version of Karl Pearson's formula**

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}}$$

- **Here, r=Coefficient of Correlation.**
- **X=(X-$\bar{X}$) =Deviation of value X from mean**
- **Y=(Y-$\bar{Y}$)= Deviation of value Y from mean**
- **Xy=(X- $\overline{X}$) (Y-$\overline{Y}$)**
- **x^2=(X−$\bar{X}$)^2**
- **y^2=(Y−$\bar{Y}$) ^{2}**

DIRECT METHOD

- STEPS:Calculate arithmetic mean of X and Y series
- Calculate deviation of observation in X-series from $\bar{X}$ and denote it by x.
- Calculate deviation of observation in Y-series from $\bar{Y}$ and denote it by y.
- Calculate square of deviation in both series and obtain their aggregate.i.e. $\sum x^2$ and $\sum y^2$.
- Multiply the corresponding deviation of the X and Y series to obtain $\sum xy$.
- Apply the formula: $\mathbf{r=} \dfrac{\sum \mathbf{xy}}{\sqrt{\sum \boldsymbol{x^2} \times \sum \boldsymbol{y^2}}}$

INDIRECT METHOD OR ASSUMED MEAN METHOD

$$r = \frac{\Sigma dxdy - \frac{(\Sigma dx) \times (\Sigma dy)}{N}}{\sqrt{\Sigma dx^2 - \frac{(\Sigma dx)^2}{N}} \times \sqrt{\Sigma dy^2 - \frac{(\Sigma dy)^2}{N}}}$$

Here,

dx = Deviation of X series from the assumed mean = $(X - A)$

dy = Deviation of Y series from the assumed mean = $(Y - A)$

$\Sigma dxdy$ = Sum of the multiple of dx and dy

Σdx^2 = Sum of square of dx

Σdy^2 = Sum of square of dy

Σdx = Sum of deviation of X series

Σdy = Sum of deviation of Y series

N = Total number of items.

INDIRECT METHOD OR ASSUMED MEAN METHOD

STEPS:

1.Any convenient value in X and Y series is taken as assumed mean Ax and Ay.

2.With the help of assumed mean of both the series ,deviation of the values of individual variable,i.e., dx(X-Ax)and dy (Y-Ay) are calculated.

3.$\sum dx$ and $\sum dy$ are found by adding the deviations.

4.Deviations of the two series are multiplied ,as dx.dy,and the multiples added up to obtain $\sum dxdy$.

INDIRECT METHOD OR ASSUMED MEAN METHOD

- **5. Squares of the deviations dx^2 and dy^2 are added up to find out $\sum dx^2$ and $\sum dy^2$**

6. Finally ,Coefficient of Correlation is calculated using the formula

$$r = \frac{\Sigma dxdy - \frac{(\Sigma dx) \times (\Sigma dy)}{N}}{\sqrt{\Sigma dx^2 - \frac{(\Sigma dx)^2}{N}} \times \sqrt{\Sigma dy^2 - \frac{(\Sigma dy)^2}{N}}}$$

STEP- DEVIATION METHOD

- **FORMULA:**

$$r = \frac{\Sigma dx'dy' - \frac{(\Sigma dx') \times (\Sigma dy')}{N}}{\sqrt{\Sigma dx'^2 - \frac{(\Sigma dx')^2}{N}} \times \sqrt{\Sigma dy'^2 - \frac{(\Sigma dy')^2}{N}}}$$

STEP DEVIATION METHOD

STEPS:

1.Any convenient value in X and Y series is taken as assumed mean Ax and Ay.

2.With the help of assumed mean of both the series ,deviation of the values of individual variable,i.e., dx(X-Ax)and dy(Y-Ay) are calculated.

3. Now divide dx and dy by some common factor as $dx' = \frac{dx}{C}, dy' = \frac{dy}{C}$ **; here C is common factor for series X and series Y .And dx′ and dy′ are step deviations.**

4.Deviations of the two series are ,as dx′× dy′ ,and the multiples added up to obtain $\sum dx'dy'$.

STEP DEVIATION METHOD

5.Squares of the deviations dx'^2 and dy'^2 are added up to find out $\sum dx'^2$ and $\sum dy'^2$

6.finally,apply the formula

$$r = \frac{\Sigma dx'dy' - \frac{(\Sigma dx') \times (\Sigma dy')}{N}}{\sqrt{\Sigma dx'^2 - \frac{(\Sigma dx')^2}{N}} \times \sqrt{\Sigma dy'^2 - \frac{(\Sigma dy')^2}{N}}}$$

PROPERTIES OF COEFFICIENT OF CORRELATION

- Karl Pearson correlation coefficient lies between -1 and +1,e.i.,$-1 \leq r \leq +1$.
- r is a pure number and it is independent of the units of measurement.
- -ve value of r indicates an inverse relationship between variables and if r is+ve,the two variables move in the same direction .
- If r=0 ,there is no correlation between variables.
- If r=+1,The correlation is perfect positive
- If r=-1,The correlation is perfect negative.
- The coefficient of correlation is not affected by change of scale or origin.

PROBABLE ERROR OF THE COEFFICIENT OF CORRELATION AND ITS INTERPRETATION.

- Probable error is used to test the reliability of Karl Pearson's correlation coefficient.
- Probable error (P.E.)=$0.6745 \times \frac{1-r^2}{\sqrt{N}}$
- Here, r stands for the coefficient of correlation and n for the no. of pairs of observation.
- Probable Error is used to interpret the value of the correlation coefficient.

*If the value of r is less than the P.E. there is no evidence of correlation

*If the value of r is more than six times of P.E.,it is significant correlation.

*By adding and subtracting the value of P.E. from the coefficient of correlation we get respectively the upper and lower limit within which coefficient of correlation in the population can be expected to lie.

*P.E. as a measure of interpreting coefficient of correlation should be used only when the n is large.

* P.E. as a measure of interpreting coefficient of correlation should be used only when a sample study is being made and the sample is unbiased and representative.

MERITS AND DEMERITS OF KARL PEARSON'S COEFFICIENT OF CORRELATION

- MERITS
- Practical and popular method.
- Meaningful conclusion.
- Measurement of degree and direction simultaneously.
- DEMERITS:
- Greater influence of extreme values.
- Calculation process is long and time consuming.
- Possibility of wrong interpretation.
- Assumption of Linear relationship between the variables.

METHODS OF STUDYING CORRELATION:COEFFICIENT OF CORRELATION BY RANK DIFFERENCES

- This method was propounded by the British psychologist, Charles Edward Spearman, in the year 1904 to calculate coefficient of correlation of qualitative variables whose quantitative measurement is not possible. For example, we can not measure honesty ,intelligence,beauty,leadership etc. quantitatively but these variables can be assigned ranks. These ranks are used for the calculation of coefficient of Correlation under Rank Correlation method. If ranks of X series are denoted as R1 ,ranks of Y series is denoted by R2 and the difference between R1 and R2 is denoted by D then we can find out Rank Correlation with the help of formula

$$R=1-\frac{6\sum D^2}{N(N^2-1)}$$

- Here, R=Rank Coefficient of Correlation ,$\sum D^2$=The total of squares of differences of corresponding ranks.N= Number of pairs of observation.
- As in case of r, $-1 \leq R \leq +1$.
- $\sum D$ Or the sum of the differences between R1 and R2 is always equal to zero.

METHODS OF STUDYING CORRELATION:COEFFICIENT OF CORRELATION BY RANK DIFFERENCES

- WHEN ACTUAL RANKS ARE GIVEN :
- Steps:
- i.Compute the difference of ranks (R1-R2) and denote them by D
- ii.Compute $D^2$and total them to get $\sum D^2$.
- iii.Apply the formula:
- $$R=1-\frac{6\sum D^2}{N(N^2-1)}$$

METHODS OF STUDYING CORRELATION:COEFFICIENT OF CORRELATION BY RANK DIFFERENCES

- When ranks are not given:
- Steps:
- I. First of all ranks are allotted either in ascending or in descending order to all the values of the two series.
- Steps:i.Compute the difference of ranks (R1-R2) and denote them by D
- ii.Compute $D^2$and total them to get $\sum D^2$.
- iii.Apply the formula:
- $$R=1-\frac{6\sum D^2}{N(N^2-1)}$$

METHODS OF STUDYING CORRELATION:COEFFICIENT OF CORRELATION BY RANK DIFFERENCES

- When ranks are equal:
- If there is two or more items have equal values in a series, then the problem of determining the rank arises. In this situation a common rank (Average of the ranks)are assigned to each equal value. In order to avoid the possibility of error, Formula,
- $$R=1-\frac{6[\ \sum D^2+\frac{1}{12}(m1^3-\mathrm{m1})+\frac{1}{12}(m2^3-\mathrm{m2})+\ldots.]}{N^3-\mathrm{N}},$$
- Is used for the calculation of rank correlation in such situation. The correction factor $\frac{m^3-\mathrm{m}}{12}$ is added to the value of $\sum D^2$.Here m =no. of items which have common ranks. $\frac{m^3-\mathrm{m}}{12}$ is added as many times as the number of such groups having equal ranks.

MERITS AND DEMERITS OF COEFFICIENT OF CORRELATION BY RANK DIFFERENCES

- MERITS:
- Its calculation is easier as compared to Karl Pearson's Method.
- This method can be used as a measure of degree of association between qualitative variables.

- DEMERITS:
- This method is not suitable for calculating coefficient of correlation of grouped frequency distribution.
- If the no. of items are large , this method becomes difficult and unsuitable.

STAY HOME STAY SAFE