Statistical Methods

Module – 1 Correlation Analysis:-

- **Meaning**
- **Types of correlation**
- **Methods of correlation**
  - * **Karl Pearson's Coefficient of correlation**
  - * **Spearman's Correlation (Rank)**
- **Probable error**

Meaning & Definition

- **Meaning:**

  Correlation is a statistical technique which measures the degree or extent to which two or more variables fluctuate with reference to one another.

  A statistical tool with the help of which we can find the degree of relationship that exists between two or more variables is called correlation.

  Correlation refers to the technique used in measuring the closeness of the relationship between the variables.

DEFINITION OF CORRELATION

Coxton & Cowden:-

when the relationship is of quantitative matter, the appropriate statistical tool for discovering and measuring the relationship and expressing it in brief formula is known as correlation.

A.M. Tuttle:-

Correlation is an analysis of the co variation between two or more variables.

Ya Lun Chou:-

Correlation analysis attempts to determine the degree of relationship between variables.

IMPORTANCE OF CORRELATION

Correlation used in Business & Economics:

- To study the relationship between related variables. such as price & demand, price & supply, income & expenditure etc.
- Facilitate estimation – once the relationship is known through correlation, the estimation can be done through regression.
- To study the economic behaviour- aids in locating the critically important variable on which another variable is depended and it also suggests the paths through which stabilising forces may become effective .
- In business- It helps to estimate sales, purchases, cost, expenditure, income etc
- Helps to reduce uncertainty- The prediction based on correlation is more variable and new to reality.

TYPES OF CORRELATION

Three broad categories of correlation:

- Positive, Negative & Zero correlation
- Linear & Non Linear correlation
- Simple & Multiple Correlation
  (Multiple correlation- Partial & Total)

Positive, Negative & Zero correlation:-

If the correlated variables move in the same direction, the correlation is said to be positive.

For ex- X	10	15	17	20	22	
	Y	2	4	5	7	9 or
	X	27	23	22	19	18
	Y	8	5	4	2	0

If the interrelated variables trend to move in the opposite direction, the correlation is said to be negative.

For ex- X	11	15	18	19	20
	Y	17	13	9	6	5

If there is no correlation between the variables taken for study , the correlation is said to be zero correlation.

For ex- Marks in accountancy & Temperature.

TYPES OF CORRELATION

- **Linear & Non Linear correlation**

(It is based on the ratio of change)

If the amount of change in one variable tends to bear constant ratio to the change of another variable over the entire range, the correlation between them is said to be linear.

For ex:	X	10	15	20	25	30
	Y	20	30	40	50	60

The ratio between to variables is 1:2

In non linear correlation, the amount of change in one variable does not bear the constant ratio to the change in another variable.

For ex:	A	8	9	11	13	17
	B	7	9	12	15	18

TYPES OF CORRELATION

- **Simple & Multiple Correlation:**

(On the basis of variables)

The correlation designed between only two variables, then the correlation is called **simple correlation.**

For ex: Amount of rainfall & Yield of wheat- Simple correlation

If the correlation if designed between more than two variables it is called **multiple correlation.**

For ex: Production of Wheat, Amount of rainfall, quantity of fertilizers, Pesticides

Multiple correlation is subdivided into Partial & Total

METHODS OF CORRELATION

CORRELATION

Graphic method

Scattered diagrams

Graphs

Algebraic method

Karl pearson's correlation

Rank Correlation

Concurrent deviation

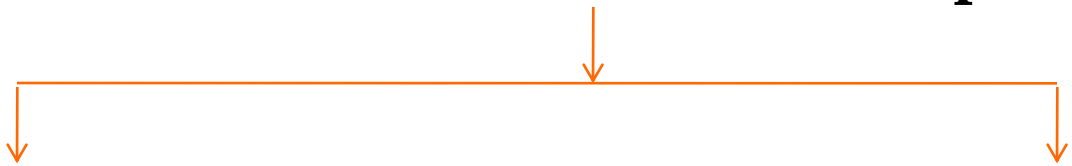
KARL PEARSON'S COEFFICIENT OF CORRELATION

It is based on the covariance of the concerned variables

Characteristics:-

- Based on mean and standard deviation.
- Determines the direction of relationship.
- Establishes the size of relationship.
- Ideal measure- covariance a standard statistical tool.

Two methods for computing



Direct method

Indirect/ Short Cut method

DIRECT METHOD

This method is used when deviation of the observations are taken from actual arithmetic mean.

Formula for computing 'r'

$$r = \frac{\sum xy}{N \sigma_x \sigma_y}$$ **(Original formula)**

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

X= Deviations taken from mean of x series.

Y= Deviations taken from mean of y series

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}}$$ **(Simplified formula)**

INDIRECT / SHORTCUT METHOD

Under this method, deviations are taken from assumed mean. Usually it is used when mean if in fraction.

For ungrouped data:-

$$r = \frac{N\sum d_x d_y - \sum d_x \times \sum d_y}{\sqrt{N\sum d_x^2 - (\sum d_x)^2}\ \sqrt{N\sum d_y^2 - (\sum d_y)^2}}$$

For grouped data:-

$$r = \frac{N\sum f d_x d_y - \sum f d_x \times \sum f d_y}{\sqrt{N\sum f d_x^2 - (\sum f d_x)^2}\ \sqrt{N\sum f d_y^2 - (\sum f d_y)^2}}$$

Direct Method

Under this method , the correlation coefficient if calculated without taking any deviations from wither arithmetic mean nor assumed mean.

Formula for calculation:

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{N \sum X^2 - (\sum X)^2}\ \sqrt{N \sum Y^2 - (\sum Y)^2}}$$

Where

$\sum XY$ = Sum of the product of values in X and Y series.

$\sum X$ = Sum of values in X series.

$\sum Y$ = Sum of values in Y series.

$\sum X^2$ =Sum of the square of the values of X series.

$\sum Y^2$ =Sum of the square of the values of Y series.

Degree of Correlation

The degree or the intensity of the correlation between two variables is ascertained on the basis of its coefficient (**coefficient lies between +1 & -1**)

Degree	Positive	Negative
Perfect	+1	-1
Very high degree	+.9 or more	-9 or more
Fairly high	+.75 to .9	-.75 to -.9
Moderate degree	+.5 to .75	-.5to -.75
Low degree	+.25 to .5	-.25 to -.5
Very little	Less than +.25	Less than -.25
Zero	0	0

PROBABLE ERROR

Meaning:

Probable error is an instrument which measures the reliability and dependability of the value of 'r'. As the coefficient of correlation is generally computed from samples which are subject to error of sampling.

Definitions:

Weldon: "Probable error defines the limits above and below the size of the coefficient determined within which there is an equal chance that coefficient of correlation similarly calculated from other samples will fall."

Horace Secrist: " The probable error of 'r' is an amount which if added to and subtracted from the average correlation coefficient produces amounts within which the chance are even that a coefficient of correlation from a series selected at random will fall."

$$P.E = .6745 \frac{1-r^2}{\sqrt{N}}$$

PROBABLE ERROR

Uses/ Functions of PE:

- **Determination of limits:**

  They are used to measure the limits within which the population coefficient correlation is expected to lie.

  $$r \pm PE = \text{Limits}$$

- **Interpretation of correlation:**

  a. r is less than PE = r is not significant.

  b. r is greater than 6 times of PE = r is more significant.

ILLUSTRATION:

- Find the coefficient of correlation between X & Y

$\bar{X} = \frac{\sum x}{N} = \frac{60}{10} = 6$ $\quad \bar{Y} = \frac{\sum Y}{N} = \frac{40}{10} = 4$ $\quad r = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}} = 0.576$

X	Y	x (X-6)	y (Y-4)	x^2	y^2	xy
5	1	-1	-3	1	9	3
10	6	4	2	16	4	58
5	2	-1	-2	1	4	2
11	8	5	4	25	16	20
12	5	6	1	36	1	6
4	1	-2	-3	4	9	6
3	4	-3	0	9	0	0
2	6	-4	2	16	4	-8
7	5	1	1	1	1	1
1	2	-5	-2	25	4	10
60	**40**	**0**	**0**	**134**	**52**	**48**

Rank correlation- Spearman correlation

This correlation is formulated by s British psychologist Edward Spearman in 1904. this method is especially used when qualitative measures for certain factors are to be evaluated.

Under this method, for each observation ranks will be assigned, While assigning the ranks the largest or the smallest, any one may be considered as 1 rank.

Next apply the following formula

$$r_s = 1 - \frac{6 \sum d^2}{N^3 - N}$$

r_s = Rank correlation

d = difference in ranks

N = No of paired observations

The above formula holds good when ranks are not repeated

When the ranks are repeated:-

$$r_s = 1 - \frac{6 \{ \sum d^2 + 1/12(m^3 - m) + 1/12(m^3 - m) \ldots\ldots\ldots}{N^3 - N}$$

N = No of times the value is repeated in the given series