

Module - 04

Co-relation and Analysis

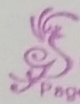
* Meaning of Correlation:-

Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate together. A positive correlation indicates the extent to which these variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases.

* Types of Correlation:-

Types of Correlation

On the basis of Direction	On the Basis of number of sets	On the Basis of change
Same (positive/direct)	Only 2 sets (Simple)	Linear (Straight line)
opp. (Negative or indirect)	More than 2 sets (Multiple)	Non-linear (Curve type)
partial (One depending on other)		



I On the Basis of Direction

1. Positive Correlation

The correlation is said to be positive correlation if the values of 2 variables change with same direction.

2. Negative Correlation

The correlation is said to be negative correlation when the values of variables change with opposite direction.

II On the Basis of Number of Sets

1. Simple Correlation

when only 2 variables are studied it is a case of simple correlation.

2. Multiple Correlation

when more than these variables are studied it is known as multiple correlation.

III On the Basis of change

1. Linear Correlation

Correlation is said to be linear when the amt of change in one variable tends to bear a constant ratio to the amt of change in the other. The graph of the variables having a linear relationship will form a straight line.

2. Non linear correlation:-

The correlation would be non linear if the amt of change in one variable does not bear a constant ratio to the amt of change in the other variable.

→ Karl Pearson's Co-efficient of correlation

Karl Pearson's Co-efficient of correlation denoted by ' r ' ($-1 \leq r \leq +1$). The coefficient of correlation ' r ' measures the degree of linear relationship between two variables say x & y .

Degree of correlation is expressed by a value of coefficient. Direction of change is indicated by sign (-ve or +ve).

When deviation taken from actual mean: $r = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}}$

→ Merits:-

This method indicates the presence or absence of correlation between 2 variables and gives the exact degree of their correlation.

To ascertain the direction of the correlation - positive or negative.

This method has many algebraic properties for which the calculation of Co-efficient of correlation



and other related factors are made easy

→ Demerits

- It is more difficult to calculate than other methods of calculations.

- It is much affected by the value of the extreme items.

- It is very much likely to be misinterpreted in case of homogeneous data.

→ problems:-

1. The following data relate to age of employees and the number of days they were reported sick in a month.

Age (x)	Sick days (y)	$x - \bar{x}$ (x)	x^2	$y - \bar{y}$ (y)	y^2	xy
30	1	-16	256	-3	9	48
33	0	-14	196	-4	16	56
35	2	-11	121	-2	4	22
40	5	-6	36	1	1	-6
48	2	2	4	-2	4	-4
50	4	4	16	0	0	0
52	6	6	36	2	4	12
55	5	9	81	1	1	9
57	7	11	121	3	9	33
61	8	15	225	4	16	60
460	40		1092		64	230

$$\bar{x} = \frac{\sum x}{n} = \frac{460}{10} = 46$$

$$\bar{y} = \frac{\sum y}{n} = \frac{40}{10} = 4$$

$$\frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}} = \frac{230}{\sqrt{1092 \times 64}} = \frac{230}{\sqrt{69888}}$$

$$= \frac{230}{264.36}$$

$$= 0.87$$

2. Compute Karl Pearson's Co-efficient of correlation and probable error from the following data.

Marks in Accountancy (x)	Marks in English (y)	$x - \bar{x}$ (x)	x^2	$y - \bar{y}$ (y)	y^2	xy
77	35	28	784	-10	100	-280
54	58	5	25	13	169	65
27	60	-22	484	15	225	-330
52	40	3	9	-5	25	-15
14	50	-35	1225	5	25	-175
35	40	-14	196	-5	25	70
90	35	41	1681	-10	100	-410
25	56	-24	576	11	121	-264
56	34	7	49	-11	121	-77
60	42	11	121	-3	9	-33
490	450		5150		920	-1449

$$\bar{x} = \frac{\sum x}{n} = \frac{490}{10} = 49$$

$$\bar{y} = \frac{\sum y}{n} = \frac{450}{10} = 45$$



Date : _____

Page No : _____

$$r = \frac{S_{xy}}{\sqrt{S_x^2 \times S_y^2}}$$

$$= \frac{-1449}{\sqrt{5150 \times 920}}$$

$$= \frac{-1449}{\sqrt{4738000}}$$

$$= \frac{-1449}{2176.69}$$

$$= -0.66$$

$$\text{probable error} = 0.6745 \times \frac{1 - r^2}{\sqrt{n}}$$

$$= 0.6745 \times \frac{1 - (-0.66)^2}{\sqrt{10}}$$

$$= 0.6745 \times \frac{1 - 0.4356}{3.16}$$

$$= 0.6745 \times \frac{0.5644}{3.16}$$

$$= 0.6745 \times 0.1786$$

$$= 0.12$$

~~2~~
14/11