Unit - 5

DATA ANALYSIS

Correlation Analysis :

Correlation is a statistical method to study the strength of relationship between two variables (or) data sets.

Karl Pearson coefficient of correlation :

It is classified into two types.

1. Arithmetic Mean Method
2. Assumed Mean Method

Arithmetic Mean Method :

Formula :

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

where,

$$x = X - \bar{X}$$
$$y = Y - \bar{Y}$$

Assumed Mean Method :

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

where,

dx (or) dy = deviation from assumed mean

$dx = X - A$, $dy = Y - A$

Note :

when $\bar{x}$ and $\bar{y}$ are whole number, it is arithmetic mean method.

where,

$$\bar{x} = \frac{\Sigma x}{N} \quad , \quad \bar{y} = \frac{\Sigma y}{N}$$

Example :

Calculate the karl pearson's coefficient of correlation from the following data relating to the age of employees and the number of days they were reported sick in the month.

Age (x)	30	32	35	40	48	50	52	55	57	61
Sick days (Y)	1	0	2	5	2	4	6	5	7	8

Solution :

$$\bar{x} = \frac{\Sigma x}{N} = \frac{460}{10} = 46$$

$$\bar{y} = \frac{\Sigma y}{N} = \frac{40}{10} = 4$$

x	y	x^2	y^2	xy
30	1	900	1	30
32	0	1024	0	0
35	2	1225	4	70
40	5	1600	25	200
48	2	2304	4	96
50	4	2500	16	200
52	6	2704	36	312
55	5	3025	25	275
57	7	3249	49	399
61	8	3721	64	488
$\Sigma x =$ 460	$\Sigma y =$ 40	$\Sigma x^2 =$ 22,252	$\Sigma y^2 =$ 224	$\Sigma xy =$ 2070

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

$$= \frac{2070}{\sqrt{22,252 \times 224}}$$

$$= \frac{2070}{\sqrt{4984448}}$$

$$= \frac{2070}{2232.58} = 0.927$$

$$\boxed{r = 0.927}$$