The following table gives the marks obtained by 100 students. Find the coefficient of correlation.

$(-1 \le r \le 1)$

Marks \ Age	18	19	20	21	Total
10-20	4	2	2	—	8
20-30	5	4	6	4	19
30-40	6	8	10	11	35
40-50	4	4	6	8	22
50-60	—	2	4	4	10
60-70	—	2	3	1	6
Total	19	22	31	28	100

X	Y	r
$\uparrow$	$\downarrow$	$-$
$\downarrow$	$\uparrow$	
$\uparrow$	$\uparrow$	$+$
$\downarrow$	$\downarrow$	

$$r = \frac{\overbrace{N\sum xy - \sum x \sum y}^{\text{Cov}(x,y)}}{\underbrace{\sqrt{N\sum x^2 - (\sum x)^2}}_{S.D(x)}\ \underbrace{\sqrt{N\sum y^2 - (\sum y)^2}}_{S.D(Y)}}$$

Properties:-

(i) $-1 \le r \le 1$.

(ii) Coeff of correlation is unaffected change of scale ~~scale~~ & origin.

$$r = \frac{N\sum' dx\,dy - \sum' dx \sum' dy}{\sqrt{N\sum' dx^2 - \left(\sum' dx\right)^2}\sqrt{N\sum' dy^2 - \left(\sum' dy\right)^2}}$$

<u>no</u> of observations

$N = \cancel{\sum' f}$

$dx = \frac{x - A}{h}$

$dy = \frac{y - B}{k}$

$$r = \frac{N\sum' f\,dx\,dy - \sum' f\,dx \sum' f\,dy}{\sqrt{N\sum' f\,dx^2 - \left(\sum' f\,dx\right)^2}\sqrt{N\sum' f\,dy^2 - \left(\sum' f\,dy\right)^2}}$$

$N = \sum' f$

$dx = \frac{x - A}{h}$

$dy = \frac{y - B}{k}$

y	dy \ x, dx	18 (−2)	19 (−1)	20 (0)	21 (1)	f	fdy	fdy^2	fdxdy
15 (10–20)	−3	4 · 24	2 · 6	2 · 0	0 · 0	8	−24	72	30
25 (20–30)	−2	5 · 20	4 · 8	6 · 0	4 · −8	19	−38	76	20
35 (30–40)	−1	6 · 12	8 · 8	10 · 0	11 · −11	35	−35	35	9
45 (40–50)	0	4 · 0	4 · 0	6 · 0	8 · 0	22	0	0	0
55 (50–60)	1	0 · 0	2 · −2	4 · 0	4 · 4	10	10	10	2
65 (60–70)	2	0 · 0	2 · −4	3 · 0	1 · 2	6	12	24	−2
f	—	19	22	31	28	100	*	*	59
	fdx	−38	−22	0	28				
	fdx^2	76	22	0	28				
	fdxdy	56	16	0	−13	59			

$dx = x - 20$

$dy = \frac{y-45}{10}$

$\sum' f\,dx\,dy = 59$

$\sum' f\,dx = -32$

$\sum' f\,dy = -75$

$\sum' f\,dx^2 = 126$

$\sum' f\,dy^2 = 217$

$N = \sum' f = 100.$

$69 + 148 = 217$

$$r = \frac{N\sum f dx dy - \sum f dx \sum f dy}{\sqrt{N\sum f dx^2 - \left(\sum f dx\right)^2}\sqrt{N\sum f dy^2 - \left(\sum f dy\right)^2}}$$

$$= \frac{100(59) - (-32)(-75)}{\sqrt{100(126) - (-32)^2}\sqrt{100(217) - (-75)^2}}$$

$$= 0.2566.$$

→ A computer while calculating coeff of correlation b/w X & Y from 25 pairs of observations obtained the following:

$n = 25$; $\sum X = 125$; $\sum X^2 = 650$; $\sum Y = 100$; $\sum Y^2 = 460$ & $\sum XY = 508$

It was later discovered that at the time of checking he had copied two pairs

X	Y
6	14
9	6

wrongly & whereas correct values are

X	8	6
Y	6	8

. obtain correct correlation coefficient.

Soln

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

Corrected $\sum x = 125 - \cancel{6} - 9 + 8 + \cancel{6} = 124$

Corrected $\sum y = 100 - 14 - \cancel{6} + \cancel{6} + 8 = 94.$

Corrected $\sum x^2 = 650 - \cancel{36} - 81 + 64 + \cancel{36} = 633$

Corrected $\sum y^2 = 460 - 196 - \cancel{36} + \cancel{36} + 64 = 328$

Corrected $\sum xy = 508 - 6\times 14 - 9\times 6 + 6\times 8 + 8\times 6 = 466.$

$$\begin{array}{r} 508 \\ 48 \\ 48 \\ \hline 604 \end{array} \qquad \begin{array}{r} 196 \\ 64 \\ \hline 132 \end{array} \qquad \begin{array}{r} 84 \\ 54 \\ \hline 138 \end{array} \qquad \begin{array}{r} 460 \\ 132 \\ \hline 328 \end{array} \qquad \begin{array}{r} 604 \\ -138 \\ \hline 466 \end{array}$$

$$\text{Corrected } r = \frac{25(466) - (124)(94)}{\sqrt{25(633) - (124)^2}\ \sqrt{25(328) - (94)^2}}$$

$=$

$\rightarrow$ $\underline{(-ve)}$

Rank correlation. (Spearman's correlation coefficient).

Data is given according to two different characteristics.
If we require coeff of correlation, we use rank correlation

X : $\longrightarrow$ Data arranged according to ht of the student

Y : $\longrightarrow$ Data arranged according to Marks.

We first assign ranks & find the correlation.

$$\rho = 1 - \left[\frac{6 \sum D^2}{N(N^2-1)} \right]$$

$D = R_x - R_y$.

$N \rightarrow$ No of observations.

Repeated ranks.

$$\rho = 1 - \frac{6 \left\{ \sum D^2 + \frac{1}{12}(m^3 - m) + \cdots \right\}}{N(N^2-1)}$$

$m \rightarrow$ no of times a particular rank gets repeated.

Ex:- Find the Rank correlation Coeff for the following:

X: 65 67 63 62 65

Y: 19 18 23 18 17

Soln

X	Y	R_x	R_y	$D=R_x-R_y$	D^2
65	19	2.5	2	−0.5	0.25
67	18	1	3.5	−2.5	6.25
63	23	4	1	3	9.00
62	18	5	3.5	1.5	2.25
65	17	2.5	5	−2.5	6.25
					24.00

$$\rho = 1 - \frac{6\left\{\sum D^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m)\right\}}{N(N^2-1)}$$

$$= 1 - \frac{6\left\{24 + \frac{1}{12}(8-2) + \frac{1}{12}(8-2)\right\}}{5(25-1)}$$

$$= 1 - \frac{6\left\{\cancel{24 + \frac{1}{2} + \frac{1}{2}}\right\}^{5}}{\cancel{5} \times 24}$$

$$= 1 - \frac{\cancel{30}\,5}{\cancel{24}\,4} = \frac{-1}{4} = -0.25$$