Rank correlation.

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

→ error.

$D = R_x - R_y$ & $m \rightarrow$ no. of times a particular rank gets repeated.

↓ Quality of two observations.

$x \rightarrow$ ht of the student.

$y \rightarrow$ marks scored by the students.

Ten Competitors in a beauty Contest were ranked by 3 judges.

R_x:	4	2	8	6	1	5	3	9	10	7
R_y:	2	5	9	3	6	7	1	10	8	4
R_z:	4	3	6	9	2	8	7	5	1	10

which pair of judges has the nearest approach to common taste?

Soln

R_x	R_y	R_z	$D_1=R_x-R_y$	D_1^2	$D_2=R_x-R_z$	D_2^2	D_3	D_3^2
4	2	4	2	4	0	0	-2	4
2	5	3	-3	9	-1	1	2	4
8	9	6	-1	1	2	4	3	9
6	3	9	3	9	-3	9	-6	36
1	6	2	-5	25	-1	1	4	16
5	7	8	-2	4	-3	9	-1	1
3	1	7	2	4	-4	16	-6	36
9	10	5	-1	1	4	16	5	25
10	8	1	2	4	9	81	7	49
7	4	10	-3	9	-3	9	-6	36

$\sum D_1^2 = 70$

$\sum D_2^2 = 146$

$\sum D_3^2 = 216.$

$$\rho_{xy} = 1 - \frac{6\sum D_1^2}{N(N^2-1)}$$

$$= 1 - \frac{6(70)}{10(99)} = \underline{0.542}$$

$\rho_{xy} = 0.542$

$\downarrow$

Judges X & Y has the Common taste.

$$\rho_{xz} = 1 - \frac{6\sum D_2^2}{N(N^2-1)}$$

$$= 1 - \frac{6(146)}{10(99)} = \underline{0.115}$$

$$\rho_{yz} = 1 - \frac{6\sum D_3^2}{N(N^2-1)} = 1 - \frac{6(216)}{10(99)} = \underline{-0.309}$$

→ Find the Rank correlation coeff, given

(x) Marks in Maths:	48	57	82	65	39	78	26	47	50	85	48	57	75
(y) Marks in stat:	67	43	80	45	79	50	46	72	65	71	32	59	66

R_x	R_y	$D = R_x - R_y$	D^2
8.5	7	1.5	2.25
5.5	11	-5.5	30.25
2	1	1	1.00
4	10	-6	36
11	2	9	81
3	8	-5	25
12	9	3	9
10	3	7	49
7	5	2	4
1	4	-3	9
8.5	12	-3.5	12.25
5.5	6	-0.5	0.25

$$\sum D^2 = 259$$

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

R(5.5), R(8.5)

$$= 1 - \frac{6\left[259 + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right]}{12(143)}$$

$$= 1 - \frac{6(259+1)}{12(143)} = 1 - \frac{130}{143} = \frac{13}{143}$$

Regression lines

Regression line of X on Y.

(X depends on Y)

$$(x - \bar{x}) = b_{xy}(y - \bar{y})$$

Regression coefficient $b_{xy} = \dfrac{N\Sigma xy - \Sigma x \Sigma y}{N\Sigma y^2 - (\Sigma y)^2} = r\dfrac{\sigma_x}{\sigma_y}$

Regression line of Y on X

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$b_{yx} = \dfrac{N\Sigma xy - \Sigma x \Sigma y}{N\Sigma x^2 - (\Sigma x)^2} = r\dfrac{\sigma_y}{\sigma_x}$$

X, Y & Z

→ $r_{12.3}$

Partial correlation (keep one variable const)

Multiple correlation. (effect of one variable over other variables)

$R_{1.23}$

$y = f(x)$.

(i) $r^2 = b_{xy} b_{yx}$.

Note:- If both b_{xy} & b_{yx} are -ve then

r = -ve.

eg:- $b_{xy} = -0.5$ & $b_{yx} = \frac{-1}{0.5}$

$$r^2 = (-0.5)\left(\frac{-1}{0.5}\right) = 1.$$

$$\Rightarrow \boxed{r = -1}$$

Calculate regression lines for the following:

X:	1	2	3	4	5
Y:	10	12	7	3	8

Sol^n

X	Y	X^2	Y^2	XY
1	10	1	100	10
2	12	4	144	24
3	7	9	49	21
4	3	16	9	12
5	8	25	64	40
Total 15	40	55	366	107

$$b_{YX} = \frac{N\sum XY - \sum X \sum Y}{N\sum X^2 - (\sum X)^2} = \frac{5(107) - (15)(40)}{5(55) - (15)^2}$$

$$= -1.3$$

$$b_{XY} = \frac{N\sum XY - \sum X \sum Y}{N\sum Y^2 - (\sum Y)^2} = \frac{5(107) - (15)(40)}{5(366) - 40^2}$$

$$= -0.283.$$

$$\bar{X} = \frac{\sum X}{N} = \frac{15}{5} = 3; \quad \frac{\sum Y}{N} = \bar{Y} \Rightarrow \bar{Y} = \frac{40}{5} = 8$$

x on y (x depends on y)

$x - \bar{x} = b_{xy}(y - \bar{y})$

$x - 3 = -0.283(y - 8)$

$x = 3 - 0.2834 + 8(0.283) = 5.264 - 0.2834$

$x = 5.264 - 0.2834y$

y on x (y depends on x)

$y - \bar{y} = b_{yx}(x - \bar{x})$

$y - 8 = -1.3(x - 3) \Rightarrow y = 8 - 1.3x + 3.9$

$y = 11.9 - 1.3x$

X on Y $\Rightarrow$ $X = 5.264 - 0.283Y$.

Y on X $\Rightarrow$ $Y = 11.9 - 1.3X$.

given X, to find Y, use Y on X.

given Y, to find X, use X on Y.

find Y, given $X = 6$.

Y on X is $Y = 11.9 - 1.3X \Rightarrow y(6) = 11.9 - 1.3(6)$

$= 11.9 - 7.8 = 4.1$

find X, given $Y = 6$.

X on Y is $X = 5.264 - 0.283Y \Rightarrow X(6) = 5.264 - 0.283(6)$

$= 3.566$.

Y on X in $y = 11.9 - 1.3x$.

$X = 1$

$y = 11.9 - 1.3(1)$

$= 11.9 - 1.3$

$= 10.6$