Question

Calculate the correlation co-efficient for the following heights (in inches) of fathers X their sons Y. [A.U. N/D 2004]

X :	65	66	67	67	68	69	70	72
Y :	67	68	65	68	72	72	69	71

Solution :

X	Y	XY	X^2	Y^2
65	67	4355	4225	4489
66	68	4488	4356	4624
67	65	4355	4489	4225
67	68	4556	4489	4624
68	72	4896	4624	5184
69	72	4968	4761	5184
70	69	4830	4900	4761
72	71	5112	5184	5041
544	552	37560	37028	38132

$$\overline{X} = \frac{\Sigma X}{n} = \frac{544}{8} = 68 \; ; \; \overline{Y} = \frac{\Sigma Y}{n} = \frac{552}{8} = 69$$

$$\overline{X}\,\overline{Y} = 68 \times 69 = 4692$$

$$\sigma_X = \sqrt{\frac{1}{n}\Sigma X^2 - (\overline{X})^2} = \sqrt{\frac{1}{8}(37028) - 68^2} = \sqrt{4628.5 - 4624} = 2.121$$

$$\sigma_Y = \sqrt{\frac{1}{n}\Sigma Y^2 - (\overline{Y})^2} = \sqrt{\frac{1}{8}(38132) - 69^2} = \sqrt{4766.5 - 4761} = 2.345$$

$$\text{Cov}(X, Y) = \frac{1}{n}\Sigma XY - \overline{X}\,\overline{Y} = \frac{1}{8}(37560) - 68 \times 69$$

$$= 4695 - 4692 = 3$$

The correlation co-efficient of X and Y is given by,

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{3}{(2.121)(2.345)} = \frac{3}{4.973} = 0.6032$$

Spearman Rank Correlation

3. RANK CORRELATION

Let us suppose that a group of 'n' individuals is arranged in order of merit or proficiency in possession of two characteristics A and B. These ranks in the two characteristics will, in general, be different. For example, if we consider the relation between intelligence and beauty, it is not necessary that a beautiful individual is intelligent also.

If (X_i, Y_i), $i = 1, 2, \dots n$ are the ranks of the individuals in two characteristics A and B respectively, then the rank correlation co-efficient is given by,

$$r = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} d_i^2$$

$$d_i = x_i - y_i \; ; \qquad n = \text{no. of items}$$

where d_i is the different between the ranks. This formula is called Karl Pearson's formula for the rank correlation co-efficient.

Find the rank correlation co-efficient from the following data :

Rank in X	1	2	3	4	5	6	7
Rank in Y	4	3	1	2	6	5	7

Solution :

X	Y	$d_i = x_i - y_i$	d_i^2
1	4	−3	9
2	3	−1	1
3	1	2	4
4	2	2	4
5	6	−1	1
6	5	1	1
7	7	0	0
		0	20

Here, $n = 7$

∴ Rank correlation co-efficient

$$r(X, Y) = 1 - \frac{6 \Sigma d_i^2}{n(n^2-1)} = 1 - \frac{6 \times 20}{7(49-1)} = 0.6429$$

Question

Ten participants were ranked according to their performance in a mustical test by the 3 Judges in the following data.

	1	2	3	4	5	6	7	8	9	10
Rank by X	1	6	5	10	3	2	4	9	7	8
Rank by Y	3	5	8	4	7	10	2	1	6	9
Rank by Z	6	4	9	8	1	2	3	10	5	7

Using rank correlation method, discuss which pair of judges has the nearest approach to common likings of music.

Solution :

x_i	y_i	z_i	$d_1 = x_i - y_i$	$d_2 = y_i - z_i$	$d_3 = x_i - z_i$	d_1^2	d_2^2	d_3^2
1	3	6	−2	−3	−5	4	9	25
6	5	4	1	1	2	1	1	4
5	8	9	3	−1	−4	9	1	16
10	4	8	6	−4	2	36	16	4
3	7	1	−4	6	2	16	36	4
2	10	2	−8	8	0	64	69	0
4	2	3	2	−1	1	4	1	1
9	1	10	8	−9	−1	64	81	1
7	6	5	1	1	2	1	1	4
8	9	7	−1	2	1	1	4	1

The rank correlation co-efficient between X and Y is given by

$$r(X, Y) = 1 - \frac{6 \Sigma d_1^2}{n(n^2 - 1)} \qquad \text{Here, } n = 10$$

$$= 1 - \frac{6 \times 200}{10(100 - 1)} = -0.212$$

The rank correlation co-efficient between Y and Z is given by

$$r(Y, Z) = 1 - \frac{6 \Sigma d_2^2}{n(n^2 - 1)} = 1 - \frac{6 \times 214}{10(100 - 1)} = -0.296$$

The rank correlation co-efficient between X and Z is given by

$$r(X, Z) = 1 - \frac{6 \Sigma d_3^2}{n(n^2 - 1)} = 1 - \frac{6 \times 60}{10(100 - 1)} = 0.636$$

Since the rank correlation coefficient between X and Z is positive and maximum, we conclude that the pair of judges X and Z has the nearest approach to common likings in music.

Repeated Rank Correlation

4. REPEATED RANKS

If any two or more individuals are equal in any classification with respect to characteristic A or B, or if there is more than one item with the same value in the series then Spearman's formula for calculating the rank correlation coefficients breaks down. In this case common ranks are given to the repeated ranks. This common rank is the average of the ranks which these items would have assumed if they are slightly different from each other and the next item will get the rank next to the ranks already assumed. As a result of this, following adjustment or correction is made in the correlation formula.

In the correlation formula, we add the factor $\frac{m(m^2-1)}{12}$ to Σd^2 where m is the number of times an item is repeated. This correction factor is to be added for each repeated value.

Question

Obtain the rank correlation coefficient for the following data :

X	68	64	75	50	64	80	75	40	55	64
Y	62	58	68	45	81	60	68	48	50	70

X	Y	Rank of X	Rank of Y	$d_i = x_i - y_i$	d_i^2
68	62	4	5	1	1
5) 64	58	6	7	1	1
2) 75	68 3)	2.5	~~3.5~~ 3.5	1	1
50	45	9	10	1	1
6) 64	81	6	1	5	25
80	60	1	6	5	25
3) 75	68 4)	2.5	3.5	1	1
40	48	10	9	1	1
55	50	8	8	0	0
7) 64	70	6	2	4	16

2) 75

$\frac{2+3}{2} \Rightarrow 5/2 \Rightarrow 2.5$

3) 64

$\frac{5+6+7}{3} \Rightarrow \frac{18}{3} \Rightarrow 6$

2) 68

$\frac{3+4}{2} \Rightarrow 7/2 \Rightarrow 3.5$

$\Sigma d_i^2 \Rightarrow 72$

$\boxed{n=10}$

$$r_{xy} \Rightarrow 1 - \frac{6\left[\Sigma d_i^2 + \frac{1}{12} m_1(m_1^2-1) + \frac{1}{12} m_2(m_2^2-1) + \cdots\right]}{n(n^2-1)}$$

75 → 2 times   64 → 3 times   68 → 2 times

$$\Rightarrow 1 - \frac{6\left[72 + \frac{1}{12}(2)(4-1) + \frac{1}{12}3(6-1) + \frac{1}{12}2(4-1) + \right]}{10(99)}$$

$$\Rightarrow 1 - \frac{6\left[72 + \frac{1}{12}(2)(3) + \frac{1}{12}(3)(5) + \frac{1}{12}(2)(3)\right]}{990}$$

$$\Rightarrow 1 - \frac{6[75]}{990} \Rightarrow 0.5454$$

$\boxed{r_{xy} \Rightarrow 0.5454}$

Question

Example:

Following is the data on heights and weights of ten students in a class:

Heights (in cm)	140	142	140	160	150	155	160	157	140	170
Weights (in cm)	43	45	42	50	45	52	57	48	49	53

Calculate rank correlation coefficient between heights and weights of students.

Solution:

Let height be a variable X and weight be a variable Y. Since, the data contains tied observations, we associate average ranks to the tied observations. The spearman's rank correlation coefficient is given by

X	Y	R_x	R_y	$d = R_x - R_y$	d^2
140	43	9	9	0	0
142	45	7	7.5	-0.5	0.25
140	42	9	10	-1	1
160	50	2.5	4	-1.5	2.25
150	45	6	7.5	-1.5	2.25
155	52	5	3	2	4
160	57	2.5	1	1.5	2.25
157	48	4	6	-2	4
140	49	9	5	4	16
170	53	1	2	-1	1
				TOT	33

From the table, we have,

$$n = 10, \sum d_2 = 33, S_1 = 3, S_2 = 2, S_3 = 33$$

Thus,

$$Adj \sum d^2 = \sum d^2 + \left[\frac{S_1^{\ 3} - S_1}{12}\right] + \left[\frac{S_2^{\ 3} - S_2}{12}\right] + \left[\frac{S_3^{\ 3} - S_3}{12}\right] + \ldots\ldots$$

$$Adj \sum d^2 = 33 + \left[\frac{3^3 - 3}{12}\right] + \left[\frac{2^3 - 2}{12}\right] + \left[\frac{2^3 - 2}{12}\right] + \ldots\ldots$$

$$= 33+2+0.5+0.5$$

$$= 36$$

$$\rho = 1 - \left[\frac{6(36)}{10(100-1)}\right]$$

$$\rho = 1 - 0.2182$$

$$= 0.7818$$

We say that there is high degree of positive rank correlation between heights and weights of students.

Question **(Important)**

The joint probability mass function of X and Y is given below.

$y \backslash x$	-1	$+1$
0	$\frac{1}{8}$	$\frac{3}{8}$
1	$\frac{2}{8}$	$\frac{2}{8}$

Solution :

$y \backslash x$	-1	1	$p(y) = p_{*j}$
0	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{4}{8}$
1	$\frac{2}{8}$	$\frac{2}{8}$	$\frac{4}{8}$
$p_{i\bullet} = p(x)$	$\frac{3}{8}$	$\frac{5}{8}$	1

(i) $E[X] = \Sigma x_i\, p(x_i) = (-1)\left(\frac{3}{8}\right) + (1)\left(\frac{5}{8}\right) = \frac{2}{8}$

(ii) $E[X^2] = \Sigma x_i^2\, p(x_i) = (1)\,\frac{3}{8} + (1)\left(\frac{5}{8}\right) = 1$

(iii) $E[Y] = \Sigma y_j\, p(y_j) = (0)\left(\frac{4}{8}\right) + (1)\left(\frac{4}{8}\right) = \frac{4}{8} = \frac{1}{2}$

(iv) $E[Y^2] = \Sigma y_j^2\, p(y_j) = (0)\left(\frac{4}{8}\right) + (1)\left(\frac{4}{8}\right) = \frac{4}{8} = \frac{1}{2}$

(v) $E[XY] = \sum_i \sum_j x_i y_j \; p(x_i y_j)$

$$= (0)\,(-1)\left(\frac{1}{8}\right) + (0)\,(1)\left(\frac{3}{8}\right) + (1)\,(-1)\left(\frac{2}{8}\right) + (1)(1)\left(\frac{2}{8}\right)$$

$$= 0 + 0 - \frac{2}{8} + \frac{2}{8} = 0$$

(vi) $\sigma_x{}^2 = E[X^2] - [E(X)]^2 = 1 - \left(\frac{1}{4}\right)^2 = 1 - \frac{1}{16} = \frac{15}{16}$

$$\sigma_x = \frac{\sqrt{15}}{4}$$

(vii) $\sigma_y^2 = E[Y^2] - [E[Y]]^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$

$$\sigma_y = \frac{1}{2}$$

(viii) $r_{XY} = \frac{\text{Cov}(X,\ Y)}{\sigma_X \sigma_Y} = \frac{E[XY] - E[X]E[Y]}{\sigma_X \ \sigma_Y}$

$$= \frac{0 - \left(\frac{1}{4}\right)\left(\frac{1}{2}\right)}{\left(\frac{\sqrt{15}}{4}\right)\left(\frac{1}{2}\right)} = \frac{-\left(\frac{1}{8}\right)}{\left(\frac{\sqrt{15}}{8}\right)} = \frac{-1}{\sqrt{15}} = -0.258$$

Question **(Important)**

Let X and Y be discrete RVs with probability function

$$f(x, y) = \frac{x+y}{21}, \; x = 1, 2, 3 \; ; \; y = 1, 2.$$

Find (i) Mean and Variance of X and Y. (ii) Cov (X, Y) (iii) Correlation of X and Y. [A.U CBT A/M 2011]

[A.U CBT A/M 2011]

Y	X	1	2	3	$f(y)$
	1	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{9}{21}$
	2	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{12}{21}$
	$f(x)$	$\frac{5}{21}$	$\frac{7}{21}$	$\frac{9}{21}$	1

$$E(X) = \Sigma x f(x) = 1 \times \frac{5}{21} + 2 \times \frac{7}{21} + 3 \times \frac{9}{21}$$

$$= \frac{5}{21} + \frac{14}{21} + \frac{27}{21} = \frac{46}{21}$$

$$E(Y) = \Sigma y f(y) = 1 \times \frac{9}{21} + 2 \times \frac{12}{21} = \frac{9}{21} + \frac{24}{21} = \frac{33}{21}$$

$$E(X^2) = \Sigma x^2 f(x) = 1^2 \times \frac{5}{21} + 2^2 \times \frac{7}{21} + 3^2 \times \frac{9}{21}$$

$$= \frac{5}{21} + \frac{28}{21} + \frac{81}{21} = \frac{114}{21}$$

$$E(Y^2) = \Sigma y^2 f(y) = 1^2 \times \frac{9}{21} + 2^2 \times \frac{12}{21} = \frac{9}{21} + \frac{48}{21} = \frac{57}{21}$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{114}{21} - \left(\frac{46}{21}\right)^2 = \frac{278}{441}$$

$$V(Y) = E(Y^2) - [E(Y)]^2 = \frac{57}{21} - \left(\frac{33}{21}\right)^2 = \frac{108}{441}$$

$$E(XY) = 1.1.\frac{2}{21} + 1.2.\frac{3}{21} + 1.3.\frac{4}{21} + 2.1.\frac{3}{21} + 2.2.\frac{4}{21} + 2.3.\frac{5}{21}$$

$$= \frac{2}{21} + \frac{6}{21} + \frac{12}{21} + \frac{6}{21} + \frac{16}{21} + \frac{30}{21} = \frac{72}{21}$$

$$\text{Cov}(X, Y) = \frac{72}{21} - \frac{46}{21} \cdot \frac{33}{21} = \frac{1512 - 1518}{441} = \frac{-6}{441} = -0.0136$$

$$\therefore r_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{\frac{-6}{441}}{\frac{\sqrt{278}}{21} \cdot \frac{\sqrt{108}}{21}} = \frac{-6}{(16.673)(10.392)}$$

Question (Important)

Two random variables X and Y have the joint density

$$f(x, y) = \begin{cases} 2 - x - y, & 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Show that $\text{cov}(X, Y) = \frac{-1}{144}$.

[AU, N/D. 2004, M/J 2006, N/D 2010, Tvli A/M 2009, M/J 2010]
[A.U. CBT M/J 2010] [A.U N/D 2011]

Solution : The marginal density function of X is,

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^1 (2 - x - y)\, dy$$

$$= \left[(2 - x)\, y - \frac{y^2}{2} \right]_0^1 = 2 - x - \frac{1}{2} = \frac{3}{2} - x, \; 0 < x < 1.$$

Similarly, the marginal density function of Y is,

$$f(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^1 (2 - x - y)\, dx$$

$$= \left[(2 - y)\, x - \frac{x^2}{2} \right]_0^1 = 2 - y - \frac{1}{2} = \frac{3}{2} - y, \; 0 < y < 1$$

Now, $E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^1 x \left(\frac{3}{2} - x \right) dx = \int_0^1 \left(\frac{3}{2} x - x^2 \right) dx$

$$= \left[\frac{3x^2}{4} - \frac{x^3}{3} \right]_0^1 = \frac{3}{4} - \frac{1}{3} = \frac{9 - 4}{12} = \frac{5}{12}$$

$$E(Y) = \int_{-\infty}^{\infty} y f(y)\, dy = \int_0^1 y \left(\frac{3}{2} - y \right) dy = \int_0^1 \left(\frac{3}{2} y - y^2 \right) dx$$

$$= \left[\frac{3y^2}{4} - \frac{y^3}{3} \right]_0^1 = \frac{3}{4} - \frac{1}{3} = \frac{9 - 4}{12} = \frac{5}{12}$$

$$E(XY) = \int_0^1 \int_0^1 xy f(x, y)\, dx\, dy = \int_0^1 \int_0^1 xy (2 - x - y)\, dx\, dy$$

$$= \int_0^1 \int_0^1 (2xy - x^2 y - xy^2)\, dx\, dy$$

$$= \int_0^1 \left[x^2 y - \frac{x^3 y}{3} - \frac{x^2 y^2}{2} \right]_0^1 dy$$

$$= \int_0^1 \left(y - \frac{y}{3} - \frac{y^2}{2} \right) dy = \int_0^1 \left(\frac{2y}{3} - \frac{y^2}{2} \right) dy$$

$$= \left[\frac{y^2}{3} - \frac{y^3}{6} \right]_0^1 = \frac{1}{3} - \frac{1}{6} = \frac{2 - 1}{6} = \frac{1}{6}$$

$\therefore$ Cov (X, Y) = E(XY) − E(X) . E(Y)

$$= \frac{1}{6} - \frac{5}{12} \cdot \frac{5}{12} = \frac{1}{6} - \frac{25}{144} = \frac{24 - 25}{144} = \frac{-1}{144}$$

Suppose that the 2D RVs (X, Y) has the joint p.d.f.

$$f(x, y) = \begin{cases} x+y, & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Obtain the correlation co-efficient between X and Y. Check whether X and Y are independent.

[AU, N/D, 2003, 2004] [A.U Tvli M/J 2010] [A.U A/M 2010] [A.U CBT N/D 2011]

Solution : The marginal density function of X is given by,

$$f(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^1 (x + y)\, dy = \left[xy + \frac{y^2}{2}\right]_0^1$$

$$= x + \frac{1}{2},\ 0 < x < 1$$

The marginal density function of Y is given by,

$$f(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^1 (x + y)\, dx = \left[\frac{x^2}{2} + xy\right]_0^1 = y + \frac{1}{2},\ 0 < y < 1$$

$$E(X) = \int_0^1 x f(x)\, dx = \int_0^1 x\left(x + \frac{1}{2}\right) dx = \int_0^1 \left(x^2 + \frac{x}{2}\right) dx$$

$$= \left[\frac{x^3}{3} + \frac{x^2}{4}\right]_0^1 = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$$

$$E(Y) = \int_0^1 y f(y)\, dy = \int_0^1 y\left(y + \frac{1}{2}\right) dy = \left[\frac{y^3}{3} + \frac{y^2}{4}\right]_0^1 = \frac{7}{12}$$

$$E(X^2) = \int_0^1 x^2 f(x)\, dx = \int_0^1 x^2\left(x + \frac{1}{2}\right) dx = \int_0^1 \left(x^3 + \frac{x^2}{2}\right) dx$$

$$= \left[\frac{x^4}{4} + \frac{x^3}{6}\right]_0^1 = \frac{1}{4} + \frac{1}{6} = \frac{3 + 2}{12} = \frac{5}{12}.$$

Similarly $E(Y^2) = \frac{5}{12}$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{5}{12} - \frac{49}{144} = \frac{11}{144}$$

$$\therefore \sigma_X = \frac{\sqrt{11}}{12} \text{ and } |||\text{ly } \sigma_Y = \frac{\sqrt{11}}{12}$$

$$\therefore r_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{E(XY) - E(X)\,E(Y)}{\sigma_X \cdot \sigma_Y}$$

$$E(XY) = \int_0^1 \int_0^1 (xy)(x+y)\,dxdy = \int_0^1 \int_0^1 (x^2 y + xy^2)\,dx\,dy$$

$$= \int_0^1 \left[\frac{x^3 y}{3} + \frac{x^2 y^2}{2}\right]_0^1 dy = \int_0^1 \left(\frac{y}{3} + \frac{y^2}{2}\right) dy$$

$$= \left[\frac{y^2}{6} + \frac{y^3}{6}\right]_0^1 = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$$

$\therefore$ $\text{Cov}(X, Y) = E(XY) - E(X)\,E(Y)$

$$= \frac{1}{3} - \frac{7}{12} \cdot \frac{7}{12} = \frac{1}{3} - \frac{49}{144} = \frac{48-49}{144} = \frac{-1}{144}$$

$\text{Cov}(X, Y) \neq 0$

$\therefore$ X and Y are not independent.

$$\therefore r_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y} = \frac{\frac{-1}{144}}{\frac{\sqrt{11}}{12} \cdot \frac{\sqrt{11}}{12}} = \frac{-1}{11} = 0.0909 \ (+\text{ve})$$

Question (Important)

Example

Let X be a random variable with p.d.f $f(x) = \frac{1}{2}$, $-1 \le x \le 1$ and let $Y = X^2$. Prove that, the correlation co-efficient between X and Y is zero.

Solution :

$$E(X) = \int_{-1}^{1} x f(x)\, dx = \frac{1}{2}\int_{-1}^{1} x\, dx = \frac{1}{2}\left[\frac{x^2}{2}\right]_{-1}^{1} = \frac{1}{2}\left[\frac{1}{2} + \frac{1}{2}\right] = 0$$

$$E(XY) = E(X^3) = \int_{-1}^{1} x^3 f(x)\, dx = \int_{-1}^{1} x^3 \frac{1}{2}\, dx$$

$$= \frac{1}{2}\int_{-1}^{1} x^3\, dx = 0 \quad [\because \text{ for odd fn.}]$$

$$E(Y) = E(X^2) = \int_{-1}^{1} x^2 f(x)\, dx = \frac{1}{2}\left[\frac{x^3}{3}\right]_{-1}^{1} = \frac{1}{2}\left[\frac{1}{3} + \frac{1}{3}\right] = \frac{1}{3}$$

$$\text{Cov}(X, Y) = E(XY) - E(X)\, E(Y) = 0 \Rightarrow r_{XY} = 0$$

Question (Important)

Two independent random variables X and Y are defined by,

$f(x) = 4ax,\ 0 \le x \le 1$ $\qquad$ $f(y) = 4by,\ 0 \le y \le 1$

$= 0$, otherwise $\qquad$ $= 0$, otherwise

Show that $U = X + Y$ and $V = X - Y$ are uncorrelated.

[AU A/M 2003, N/D 2012]

Solution :

Covariance $(U, V) = E(UV) - E(U)E(V) \neq 0$

$E(UV) = E((X+Y)(X-Y))$

$E(UV) = E(X^2 - Y^2) \Rightarrow \boxed{E(X^2) - E(Y^2)}$

$E(U) = E(X+Y) \Rightarrow \boxed{E(X) + E(Y)}$

$E(V) = E(X-Y) \Rightarrow \boxed{E(X) - E(Y)}$

(i) $f(x) = 4ax,\ 0 \le x \le 1$ $= 0$, otherwise $f(x)$ is the density function of X $\int_0^1 f(x)\,dx = 1$ $\int_0^1 4ax\,dx = 1$ $4a\left[\frac{x^2}{2}\right]_0^1 = 1$ $2a = 1;\ a = \frac{1}{2}$	$f(y) = 4by,\ 0 \le y \le 1$ $= 0$, otherwise $f(y)$ is the density function of Y $\int_0^1 f(y)\,dy = 1$ $\int_0^1 4by\,dy = 1$ $4b\left[\frac{y^2}{2}\right]_0^1 = 1$ $2b = 1;\ b = \frac{1}{2}$

To prove $U = X + Y$ and $V = X - Y$ are uncorrelated

i.e., to prove $\text{Cov}(U, V) = 0$

$$\text{Cov}(U, V) = E(UV) - E(U)E(V)$$

$$E(U) = E[X+Y] = E(X) + E(Y)$$

$$E(V) = E[X-Y] = E(X) - E(Y)$$

$$E(UV) = E[X^2 - Y^2]$$

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^1 x\,(4ax)\, dx = 4a \int_0^1 x^2\, dx = 4a \left[\frac{x^3}{3}\right]_0^1 = \frac{4a}{3} = \frac{2}{3}$$

$$E(Y) = \int_{-\infty}^{\infty} y f(y)\, dy = \int_0^1 y\,(4by)\, dy = 4b \int_0^1 y^2\, dy = 4b \left[\frac{y^3}{3}\right]_0^1 = \frac{4b}{3} = \frac{2}{3}$$

$$E[XY] = E[X]\,E[Y] \qquad [\because \text{X and Y are independent}]$$

$$= \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$$

$$E[U] = E[X+Y] = E[X] + E[Y] = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$$

$$E[V] = E[X-Y] = E[X] - E[Y] = \frac{2}{3} - \frac{2}{3} = 0$$

$$E[UV] = E[X^2 - Y^2] = E[X^2] - E[Y^2] = \frac{1}{2} - \frac{1}{2} = 0$$

$$E[X^2] = \int_0^1 x^2 f(x)\, dx = \int_0^1 x^2 (2x)\, dx = \int_0^1 2x^3\, dx = 2\left[\frac{x^4}{4}\right]_0^1$$

$$= \frac{1}{2}(1-0) = \frac{1}{2}$$

$$E[Y^2] = \int_0^1 y^2 f(y)\, dy = \int_0^1 y^2 (2y)\, dy = \int_0^1 2y^3\, dy = 2\left[\frac{y^4}{4}\right]_0^1$$

$$= \frac{1}{2}(1-0) = \frac{1}{2}$$

$$\mathrm{Cov}(U,V) = E[UV] - E[U]\,E[V] = 0 - \frac{4}{3}(0) = 0$$

$\therefore$ U and V are uncorrelated.

Partial Correlation

Partial & Multiple correlation

Partial correlation

X, Y, Z

Partial correlation coefficient of X & Y, controlling Z

$$r_{xy.z} \Rightarrow \frac{r_{xy} - r_{xy}\, r_{yz}}{\sqrt{1-r_{xz}^2}\,\sqrt{1-r_{yz}^2}} \quad \text{(or)} \quad r_{12.3} \Rightarrow \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{1-r_{12}^2}\,\sqrt{1-r_{23}^2}}$$

Multiple correlation

$$R_{1(23)} \Rightarrow \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}\, r_{23}\, r_{31}}{1-r_{23}^2}}$$

$X_1 \Rightarrow$ dependent

$X_2, X_3 \Rightarrow$ independent

$$R_{2(31)} \Rightarrow \sqrt{\frac{r_{23}^2 + r_{21}^2 - 2r_{23}\, r_{21}\, r_{31}}{1-r_{31}^2}}$$

$X_2 \Rightarrow$ dependent

$X_3, X_1 \Rightarrow$ independent

$b_{xy} = b_{yx}$

regression coefficient } Not symmetric

correlation coefficient } Symmetric

$r_{21} = r_{12}$

Partial and Multiple Correlation

Let us consider the example of yield of rice in a firm. It may be affected by the type of soil, temperature, amount of rainfall, usage of fertilizers etc. It will be useful to determine how yield of rice is influenced by one factor or how yield of rice is affected by several other factors. This is done with the help of partial and multiple correlation analysis.

The basic distinction between multiple and partial correlation analysis is that in the former, the degree of relationship between the variable Y and all the other variables $X_1, X_2, ..., X_n$ taken together is measured, whereas, in the later, the degree of relationship between Y and one of the variables $X_1, X_2, ..., X_n$ is measured by removing the effect of all the other variables.

Partial correlation

Partial correlation coefficient provides a measure of the relationship between the dependent variable and other variable, with the effect of the rest of the variables eliminated. If there are three variables X_1, X_2 and X_3, there will be three coefficients of partial correlation, each studying the relationship between two variables when the third is held constant. If we denote by $r_{12.3}$, that is, the coefficient of partial correlation X_1 and X_2 keeping X_3 constant, it is calculated as

Relationship between 1 and 2

Relationship between 1 and 3

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}, \qquad r_{13.2} = \frac{r_{13} - r_{12} r_{23}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{23}^2}},$$

$$r_{23.1} = \frac{r_{23} - r_{12} r_{13}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{13}^2}}$$

3 as constant

2 as constant

1 as constant

Relationship between 2 and 3

1. In a trivariate distribution, it is found that $r_{12} = 0.7$, $r_{13} = 0.61$ and $r_{23} = 0.4$. Find the partial correlation coefficients.

Solution:

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}} = \frac{0.7 - (0.61 \times 0.4)}{\sqrt{1 - (0.61)^2}\sqrt{1 - (0.4)^2}} = 0.628$$

$$r_{13.2} = \frac{r_{13} - r_{12} r_{23}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{23}^2}} = \frac{0.61 - (0.7 \times 0.4)}{\sqrt{1 - (0.7)^2}\sqrt{1 - (0.4)^2}} = 0.504$$

$$r_{23.1} = \frac{r_{23} - r_{12} r_{13}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{13}^2}} = \frac{0.4 - (0.7 \times 0.61)}{\sqrt{1 - (0.7)^2}\sqrt{1 - (0.61)^2}} = -0.048$$

Multiple Correlation

Multiple Correlation

In multiple correlation, we are trying to make estimates of the value of one of the variable based on the values of all the others. The variable whose value we are trying to estimate is called the dependent variable and the other variables on which our estimates are based are known as independent variables.

The coefficient of multiple correlation with three variables X_1, X_2 *and* X_3 are $R_{1.23,}$ $R_{2.13}$ and $R_{3.21}$. $R_{1.23,}$ is the coefficient of multiple correlation related to X_1 as a dependent variable and X_2, X_3 as two independent variables and it can be expressed in terms of $r_{12,}$ r_{23} and r_{13} as

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{23}r_{13}}{1 - r_{23}^2}},$$

$$R_{2.13} = \sqrt{\frac{r_{12}^2 + r_{23}^2 - 2r_{12}r_{23}r_{13}}{1 - r_{13}^2}},$$

$$R_{3.12} = \sqrt{\frac{r_{13}^2 + r_{23}^2 - 2r_{12}r_{23}r_{13}}{1 - r_{12}^2}}$$

Dependent variable

Independent variable

PROPERTIES OF MULTIPLE CORRELATION COEFFICIENT

The following are some of the properties of multiple correlation coefficients:

1. Multiple correlation coefficient is the degree of association between observed value of the dependent variable and its estimate obtained by multiple regression,

2. Multiple Correlation coefficient lies between 0 and 1.

3. If multiple correlation coefficient is 1, then association is perfect and multiple regression equation may said to be perfect prediction formula.

4. If multiple correlation coefficient is 0, dependent variable is uncorrelated with other independent variables. From this, it can be concluded that multiple regression equation fails to predict the value of dependent variable when values of independent variables are known.

5. Multiple correlation coefficient is always greater or equal than any total correlation coefficient. If $R_{1.23}$ is the multiple correlation coefficient than $R_{1.23} \geq r_{12}$ or r_{13} or r_{23} and

6. Multiple correlation coefficient obtained by method of least squares would always be greater than the multiple correlation coefficient obtained by any other method.

Question

2. From the following data, obtain $R_{1.23}$, $R_{2.13}$ and $R_{3.12}$

X_1	2	5	7	11
X_2	3	6	10	12
X_3	1	3	6	10