Module - 3

Correlation & Regression :

* Correlation & Regression
* Partial & Multiple correlation
* Multiple regression

Correlation

* If the change in one variable affects the change in another variable then the variables are said to be correlated.

+ve Correlation

If the two variables deviate in the same direction then the variables are +vely correlated.

-ve correlation

If the two variables constantly deviate in the opposite direction then the variables are -vely correlated.

Examples

* The income & Expenditure is +vely correlated
* Volume & Pressure of a perfect gas is -vely correlated.

Scanned by CamScanner

Perfect Correlation:

If the deviation in one variable is followed by a corresponding & proportional deviation in the other.

Note

A measure of itensity (or) degree of linear relationship between two variables.

Karl pearson's coefficient of correlation (or) correlation coefficient:

$$r(x,y) = r = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

Where $cov(x,y) = E\{(X-E(X))(Y-E(Y))\}$

$$= \frac{1}{n} \sum (x_i - \bar{x})(y - \bar{y})$$

$$= \frac{1}{n} \sum x_i y_i - \bar{x}\bar{y}$$

$$\sigma_x = \frac{1}{n} \sum x_i^2 - \left(\frac{1}{n} \sum x_i\right)^2 = \frac{1}{n} \sum x_i^2 - \bar{x}^2$$

$$\sigma_y = \frac{1}{n} \sum y_i^2 - \left(\frac{1}{n} \sum y_i\right)^2 = \frac{1}{n} \sum y_i^2 - \bar{y}^2$$

Note

$$cov(x,y) = \sigma_{xy} = M_{11}$$

which is also known as Product moment correlation coefficient

Scanned by CamScanner

Properties

(1) $-1 \leq r \leq 1$ (Limits for Correlation Coefficient)

(2) If $r = 1$, The correlation is perfect & +ve
If $r = -1$, The Correlation is perfect & -ve

(3) Correlation Coefficient is independent of change of origin & scale.

(4) If x & y are random variables & a, b, c, d are any no/s provided only that $a \neq 0$, $c \neq 0$ then
$$r(ax+b, cy+d) = \frac{ac}{|ac|} r(x, y)$$
$$\& \quad r(ax, cy) = \frac{ac}{|ac|} r(x, y)$$

(5) Two independent variables are uncorrelated.
i.e., If x & y are uncorrelated then
$$cov(x, y) = 0$$
$$\Rightarrow r(x, y) = 0$$
but the converse is not true.

(8) Sums

Calculate the correlation coefficient for the following heights (in inches) of fathers (x) & their sons (y):

Scanned by CamScanner

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

Sol

X	Y	$U = X - 68$	$V = Y - 69$	U^2	V^2	UV
65	67	-3	-2	9	4	6
66	68	-2	-1	4	1	2
67	65	-1	-4	1	16	4
67	68	-1	-1	1	1	1
68	72	0	3	0	9	0
69	72	1	3	1	9	3
70	69	2	0	4	0	0
72	71	4	2	16	4	8
Total		$\Sigma U = 0$	$\Sigma V = 0$	$\Sigma U^2 = 36$	$\Sigma V^2 = 44$	$\Sigma UV = 24$

$n = 8$

$$\bar{U} = \frac{\Sigma U}{n} = 0/8 = 0$$

$$\bar{V} = \Sigma V/n = 0/8 = 0$$

$$\text{Cov}(U,V) = \frac{1}{n}\Sigma UV - \bar{U}\bar{V} = \frac{1}{8} \times 24 = 3$$

$$\sigma_U^2 = \frac{1}{n}\Sigma U^2 - (\bar{U})^2 = \frac{1}{8} \times 36 = 4.5$$

$$\sigma_V^2 = \frac{1}{n}\Sigma V^2 - (\bar{V})^2 = \frac{1}{8} \times 44 = 5.5$$

$$\therefore \gamma = \frac{\text{Cov}(U,V)}{\sigma_U \sigma_Y} = \frac{3}{\sqrt{4.5}\sqrt{5.5}} = 0.603$$

$$= \gamma(U,V)$$

$$= \gamma(X,Y)$$

Scanned by CamScanner

2. A Computer while calculating correlation coefficient between the variables x & y from 25 pairs of observations obtained ~~from~~ the following results.

$n=25$, $\sum x=125$, $\sum x^2=650$, $\sum y=100$, $\sum y^2=460$

$\sum xy=508$. It was, however, later discovered at the time of checking that he had copied down two pairs as

x	y
6	14
8	6

while the correct values were

x	y
8	12
6	8

Obtain the correct value of correlation coefficient.

Solution

Corrected $\sum x = 125 - 6 - 8 + 8 + 6 = 125$

Corrected $\sum y = 100 - 14 - 6 + 12 + 8 = 100$

Corrected $\sum x^2 = 650 - 6^2 - 8^2 + 8^2 + 6^2 = 650$

Corrected $\sum y^2 = 460 - 14^2 - 6^2 + 12^2 + 8^2 = 436$

Corrected $\sum xy = 508 - 6\times14 - 8\times6 + 8\times12 + 6\times8$
$= 520$

$\therefore \operatorname{Cov}(x,y) = \frac{1}{n}\sum xy - \bar{x}\bar{y} = \frac{1}{25}(520) - 5\times4 = 4/5$

$\sigma_x^2 = \frac{1}{n}\sum x^2 - (\bar{x})^2 = \frac{1}{25}\times650 - (5)^2 = 1$

$\sigma_y^2 = \frac{1}{n}\sum y^2 - (\bar{y})^2 = \frac{1}{25}\times436 - 16 = 36/25$

Scanned by CamScanner

Corrected $r(x,y) = \dfrac{cov(x,y)}{\sigma_x \sigma_y} = \dfrac{4/5}{1 \times 6/5} = 2/3 = 0.67$

Note

① If $z = ax + by$ & r is the correlation coefficient between x & y then $\sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2abr\sigma_x\sigma_y$

②

Correlation Coefficient when $a=1, b=-1$ & $z = x-y$

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2\sigma_x\sigma_y}$$

$\sigma_x, \sigma_y, \sigma_{x-y}$ are S.D's of x, y & $x-y$ respectively.

③ $V(ax+by) = a^2 v(x) + b^2 v(y) + 2ab\, Cov(x,y)$

$V(ax-by) = a^2 v(x) + b^2 v(y) - 2ab\, cov(x,y)$

Sum

①. The independent variables x & y are defined as

$$f(x) = \begin{cases} 4ax, & 0 \le x \le r \\ 0 & \text{otherwise} \end{cases} \quad \& \quad f(y) = \begin{cases} 4by, & 0 \le y \le s \\ 0 & \text{otherwise} \end{cases}$$

Find the correlation between U & V where

$U = X + Y$ & $V = X - Y$

Ans: $r(U,V) = \dfrac{b-a}{b+a}$ where $a = \dfrac{1}{2r^2}$ & $b = \dfrac{1}{2s^2}$

Scanned by CamScanner

Sol

Since X & Y are independent & total probability is 1

$$\therefore \int_0^r f(x)\,dx = 4a\int_0^r x\,dx = 1$$

$$\Rightarrow \cancel{4}^{2}a\left(\frac{x^2}{\cancel{2}}\right)_0^r = 1$$

$$2ar^2 = 1$$

$$\therefore r^2 = \frac{1}{2a}$$

(or) $a = \frac{1}{2r^2}$

$$\int_0^s f(y)\,dy = 1$$

$$\Rightarrow 4b\int_0^s y\,dy = 1$$

$$2\cancel{4b}\left(\frac{y^2}{\cancel{2}}\right)_0^s = 1$$

$$2b(s^2) = 1$$

$$s^2 = \frac{1}{2b}$$

(or) $b = \frac{1}{2s^2}$

$$\therefore f(x) = 4ax = \frac{2x}{r^2}, \quad 0 \le x \le r$$

$$\& \; f(y) = 4by = \frac{2y}{s^2}, \quad 0 \le y \le s$$

Since X & Y are independent variates,

$$r(X,Y) = 0 \Rightarrow \operatorname{cov}(X,Y) = 0$$

$$\therefore \operatorname{cov}(U,V) = \operatorname{cov}(X+Y,\, X-Y)$$

$$= \operatorname{cov}(X,X) - \operatorname{cov}(X,Y) + \operatorname{cov}(Y,X) - \operatorname{cov}(Y,Y)$$

$$= \operatorname{cov}(X,X) - \operatorname{cov}(Y,Y) = \sigma_X^2 - \sigma_Y^2$$

Scanned by CamScanner

$$V(U) = V(X+Y) = V(X) + V(Y) + 2\,cov(X,Y)$$
$$= \sigma_X^2 + \sigma_Y^2$$

$$V(V) = V(X-Y) = \sigma_X^2 - \sigma_Y^2$$

$$\therefore \quad r(U,V) = \frac{cov(U,V)}{\sigma_U \sigma_V} = \frac{\sigma_X^2 - \sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}$$

$$\therefore \quad E(X) = \int_0^r x f(x)\,dx = \frac{2}{r^2}\int_0^r x^2\,dx = \frac{2r}{3}$$

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

$$= \frac{1}{2r^2}(r^4) = \frac{r^2}{2}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{r^2}{2} - \frac{4r^2}{9} = \frac{r^2}{18}$$
$$= \frac{1}{36a}$$

$|||^{ly}$ $E(Y) = \frac{2s}{3}$, $\cancel{E(Y^2)}$

i.e., $E(Y) = \int_0^s y f(y)\,dy = \frac{4}{2s^2}\int_0^s y^2\,dy$

$$= \frac{2}{s^2}\left(\frac{y^3}{3}\right)_0^s = \frac{2s^3}{3s^2} = \frac{2s}{3}$$

Scanned by CamScanner

$$E(y^2) = \frac{2}{s^2}\int_0^s y^3\,dy = \frac{2}{s^2}\left(\frac{y^4}{4}\right)_0^s =$$

$$= \frac{1}{2s^2}\left(s^4\right) = \frac{s^2}{2}$$

$$V(y) = E(y^2) - (E(y))^2 = \frac{s^2}{2} - \frac{4s^2}{9}$$

$$= \frac{9s^2 - 8s^2}{18} = \frac{s^2}{18} = \frac{1}{36b}$$

$$\therefore \quad r(u,v) = \frac{\frac{1}{36a} - \frac{1}{36b}}{\frac{1}{36a} + \frac{1}{36b}} = \frac{36b - 36a}{36b + 36a}$$

$$= \frac{b-a}{b+a}$$

<u>Rank correlation</u> (<u>Spearman's Rank Correlation Coefficient</u>)

* Rank Correlation Coefficient $= \rho = 1 - \dfrac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$

Where $d_i = x_i - y_i$ (difference between ranks of x & y)

n — no. of observation.

<u>Note</u>

<u>In case of repeated ranks</u>

In the formula, with $\sum d^2$ we add the correltion factor $\frac{m(m^2-1)}{12}$ where m is the no. of times an item is repeated.

* This correction factor is to be added for each repeated value in both the x series & y series.

Note

Limit for the rank correlation coefficient

$$-1 \leq \rho \leq 1.$$

Note

* For Tie ranks

If two or more observations have the same rank we assign to them the mean rank.

In this case,

$$\text{correlation coefficient} = \rho = 1 - \frac{6\left(\sum d^2 + \frac{\sum m(m^2-1)}{12}\right)}{n(n^2-1)}$$

where m — number of times a rank is repeated.

* Rank correlation also lies between -1 to 1

Scanned by CamScanner

Sums

Ten competitors in a beauty contest are ranked by three judges in the following order:

First judge	1	4	6	3	2	9	7	8	10	5
Second judge	2	6	5	4	7	10	9	3	8	1
Third judge	3	7	4	5	10	8	9	2	6	1

Use the method of rank correlation coefficient to determine which pair of judges have the nearest approach to common taste in beauty?

Sol

Let x, y, z denotes the ranks by 1^{st}, 2^{nd} & 3^{rd} judges respectively.

x	y	z	Rank x ~ y d_1	Rank of y ~ z d_2	Rank of x & z d_3	d_1^2	d_2^2	d_3^2
1	2	3	−1	−1	−2	1	1	4
4	6	7	−2	−1	−3	4	1	9
6	5	4	1	1	2	1	1	4
3	4	5	−1	−1	−2	1	1	4
2	7	10	−5	−3	−8	25	9	64
9	10	8	−1	2	1	1	4	1
7	9	9	−2	0	−2	4	0	4
8	3	2	5	1	6	25	1	36
10	8	6	2	2	4	4	4	16
5	1	1	4	0	4	16	0	16
Total						82	22	158

Scanned by CamScanner

$\because n = 10$

$$\rho_{xy} = 1 - \frac{6\sum d_1^2}{n(n^2-1)} = 1 - \frac{6\times 82}{10\times 99} = 0.503$$

$$\rho_{yz} = 1 - \frac{6\sum d_2^2}{n(n^2-1)} = 1 - \frac{6\times 22}{10\times 99} = 0.867$$

$$\rho_{zx} = 1 - \frac{6\sum d_3^2}{n(n^2-1)} = 1 - \frac{6\times 158}{10\times 99} = 0.04$$

* The rank correlation coefficient between y & z is +ve & highest among all the three.

$\therefore$ Judges y & z have the nearest approach for common taste in beauty.

② 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

Scanned by CamScanner

Sol

X	Y	Rank of X	Rank of Y	d = x − y	d^2
68	62	4	5	−1	1
64	58	6	7	−1	1
75	68	2.5	3.5	−1	1
50	45	9	10	−1	1
64	81	6	1	5	25
80	60	1	6	−5	25
75	68	2.5	3.5	−1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2	4	16
Total				$\Sigma d = 0$	$\Sigma d^2 = 72$

In X Series,

75 Occurs 2 times

68 Occurs 3 times

In Y series,

68 occurs twice

$m_3 = 2$

$\therefore\ m_1 = 2,\ m_2 = 3$

$\therefore$ Correction factors

$$CF = \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12} + \frac{m_3(m_3^2-1)}{12}$$

$$= \frac{2(4-1)}{12} + \frac{3(9-1)}{12} + \frac{2(4-1)}{12} = \frac{1}{2} + 2 + \frac{1}{2} = 3$$

Scanned by CamScanner

∴ rank correlation coefficient

$$= \rho = 1 - \frac{6\left(\sum d^2 + \frac{m_1(m_1^2-1)}{12} + \frac{m_2(m_2^2-1)}{12} + \frac{m_3(m_3^2-1)}{12}\right)}{n(n^2-1)}$$

$$\rho = 1 - \frac{6(72+2)}{10 \times 99} = 0.545$$

Multiple & partial correlation

The value of one varible influenced by many others. is known as multiple correlation.

If x_1 is the dependent variable & $x_2, x_3 \ldots x_n$ are independent variables then we denote the multiple correlation by $R_{1.23\cdots n}$

For example

* The yield of crop may depend upon the rainfall, fertilizer, the average temperature, the period between sowing & harvesting

Scanned by CamScanner

Partial correlation

The correlation between only two variates eliminating the linear effect of other variates in them is called as partial correlation.

i.e., A partial correlation coefficient measures the relationship between any two variables when the other variables connected with those variables are kept constant.

Example

$\star$. Correlation between the heights & weights of the boys of the same age.

~~If the variable X_1 & X_2 are par~~

If we want to study the effect of one variable X_2 on another variable X_1 after eliminating the effects of all other variables the measure of relationship between X_1 & X_2 is called partial correlation & it ~~is~~ denoted by

$r_{12 \cdot 3}$ where X_3 kept as constant & X_1 & X_2 are partially correlated.

Scanned by CamScanner

Formulae

* $$W = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix} = 1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2 r_{12} r_{13} r_{23}$$

& W_{11} = The minor of 1 $= \begin{vmatrix} 1 & r_{23} \\ r_{32} & 1 \end{vmatrix} = 1 - r_{23}^2$

W_{12} = The minor of r_{12} $= - \begin{vmatrix} r_{21} & r_{23} \\ r_{21} & 1 \end{vmatrix} = -r_{21} + r_{23} r_{31}$

W_{13} = The minor of r_{13} $= \begin{vmatrix} r_{21} & 1 \\ r_{31} & r_{32} \end{vmatrix} = r_{21} r_{32} - r_{31}$

and so on.

* Multiple Correlation

If x_1 is correlated with x_2 & x_3 then the multiple correlation is

$$R_{1.23}^2 = 1 - \frac{W}{W_{11}} = \frac{r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23}}{1 - r_{23}^2}$$

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

Scanned by CamScanner

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

Properties of Multiple Correlation Coefficient

* $0 \le R_{1.23} \le 1$
* If $R_{1.23} = 0$ then all the total & partial correlations involving x_1 are zero.
* $R_{1.23}$ is not less than any total correlation coefficient.

$$R_{1.23} \ge r_{12}, r_{13}, r_{23}.$$

Partial correlation

The correlation coefficients between x_1 & x_2 after the linear effect of x_3 on each of them has been eliminated is called the partial correlation coefficient between x_1 & x_2 & it is denoted by

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

|||ly

$$r_{13.2} = \frac{r_{13} - r_{12} r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{13}^2)}}$$

Scanned by CamScanner

Problems

From the data relating to the yield of dry bark (x_1), height (x_2) & girth (x_3) for 18 cinchona plants, the following correlation coefficient were obtained:

$r_{12} = 0.77$, $r_{13} = 0.72$ & $r_{23} = 0.52$

Find the partial correlation coefficients $r_{12.3}$, $r_{13.2}$, $r_{23.1}$ & Multiple correlation coefficient $R_{1.23}$.

Sol

i) $$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = \frac{0.77 - 0.72 \times 0.52}{\sqrt{(1-(0.72)^2)(1-(0.52)^2)}}$$

$$= 0.62$$

ii) $$r_{13.2} = \frac{r_{13} - r_{12} r_{23}}{\sqrt{(1-r_{12}^2)(1-r_{23}^2)}} = \frac{0.72 - (0.77)(0.52)}{\sqrt{(1-(0.77)^2)(1-(0.52)^2)}}$$

$$= \frac{0.72 - 0.4004}{\sqrt{0.4071 \times 0.7296}} = \frac{0.3196}{0.4655}$$

$$= 0.6866$$

Scanned by CamScanner

iii) $$r_{23.1} = \frac{r_{23} - r_{12} r_{13}}{\sqrt{(1-r_{12}^2)(1-r_{13}^2)}} = \frac{0.52 - 0.77 \times 0.72}{\sqrt{(1-(0.77)^2)(1-(0.72)^2)}}$$

$$= \frac{-0.0344}{0.3073} = -0.1119$$

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

$$= \frac{(0.77)^2 + (0.72)^2 - 2(0.77)(0.72)(0.52)}{1-(0.52)^2}$$

$$= 0.7334$$

$R_{1.23}^2 = 0.7334$

$R_{1.23} = 0.8564$ (Since multiple correlation coefficient is non -ve)

Formulae

$$b_{12.3} = r_{12.3} \frac{\sigma_{1.3}}{\sigma_{2.3}}$$ = regression coefficient of $X_{1.3}$ on $X_{2.3}$

where $$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

Scanned by CamScanner

$$\sigma_{1.3} = \sigma_1 \sqrt{(1 - r_{13}^2)}$$

$$\sigma_{2.3} = \sigma_2 \sqrt{(1 - r_{23}^2)}$$

III^{ly} $b_{13.2} = r_{13.2} \dfrac{\sigma_{1.2}}{\sigma_{3.2}}$ = coefficient of regression of $X_{2.3}$ on $X_{1.3}$

Where

$r_{13.2}$ – Partial correlation Coefficient of X_1 & X_3 keeping X_2 as constant

$$= \frac{r_{13} - r_{12} r_{32}}{\sqrt{(1 - r_{12}^2)(1 - r_{32}^2)}}$$

$$\sigma_{1.2} = \sigma_1 \sqrt{(1 - r_{12}^2)}$$

$$\sigma_{3.2} = \sigma_3 \sqrt{(1 - r_{32}^2)}$$

* $\sigma_{1.23} = \sigma_1 \sqrt{w / w_{11}}$

Sums

① In a trivariate distribution $\sigma_1 = 2, \sigma_2 = \sigma_3 = 3$, $r_{12} = 0.7$, $r_{23} = r_{31} = 0.5$ then find i) $r_{23.1}$ ii) $R_{1.23}$, iii) $b_{12.3}, b_{13.2}$ & iv) $\sigma_{1.23}$

Scanned by CamScanner

Sol

i) $r_{23.1} = \dfrac{r_{23} - r_{12}r_{13}}{\sqrt{(1-r_{12}^2)(1-r_{13}^2)}} = \dfrac{0.5 - (0.7)(0.5)}{\sqrt{(1-0.49)(1-0.25)}}$

$= 0.2425$

ii) $R_{1.23}^2 = \dfrac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1-r_{23}^2}$

$= \dfrac{0.49 + 0.25 - 2(0.7)(0.5)(0.5)}{1-0.25}$

$= 0.52$

$R_{1.23} = 0.7211$

iii) $b_{12.3} = r_{12.3}\dfrac{\sigma_{1.3}}{\sigma_{2.3}}$ & $b_{13.2} = r_{13.2}\dfrac{\sigma_{1.2}}{\sigma_{3.2}}$

$r_{12.3} = \dfrac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} = 0.6$

$r_{13.2} = \dfrac{r_{13} - r_{12}r_{32}}{\sqrt{(1-r_{12}^2)(1-r_{32}^2)}} = \dfrac{0.5 - (0.7)(0.5)}{\sqrt{(1-(0.7)^2)(1-(0.5)^2)}}$

$= 0.2425$

$\sigma_{1.3} = \sigma_1\sqrt{(1-r_{13}^2)} = 2\sqrt{1-0.25} = 1.7320$

Scanned by CamScanner

$$\sigma_{2.3} = \sigma_2\sqrt{(1-r_{23}^2)} = 3\sqrt{(1-0.25)} = 2.5980$$

$$\sigma_{1.2} = \sigma_1\sqrt{(1-r_{12}^2)} = 2\sqrt{(1-0.49)} = 1.4282$$

$$\sigma_{3.2} = \sigma_3\sqrt{(1-r_{32}^2)} = 3\sqrt{(1-0.25)} = 2.5980$$

$$b_{12.3} = 0.6 \times \frac{1.7320}{2.5980} = 0.4$$

$$b_{13.2} = 0.2425 \times \frac{1.4282}{2.5980} = 0.1333$$

iv) $$\sigma_{1.23} = \sigma_1\sqrt{w/w_{11}}$$

where $$w = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix} = 1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{23}r_{13}$$

$$= 1 - 0.49 - 0.25 - 0.25 + 2(0.7)(0.5)(0.5)$$

$$= 0.36$$

& $$w_{11} = \text{Minor of } 1 = \begin{vmatrix} 1 & r_{23} \\ r_{23} & 1 \end{vmatrix} = 1 - r_{23}^2$$

$$= 1 - 0.25 = 0.75$$

$$\therefore \sigma_{1.23} = 2 \times \sqrt{\frac{0.36}{0.75}} = 2 \times \sqrt{0.48}$$

$$= 2(0.6928) = 1.3856$$

Scanned by CamScanner

Relation between partial & Multiple Correlation Coefficient

* $1 - R^2_{1.23} = (1 - r^2_{12})(1 - r^2_{13.2})$

* If $R_{1.23} = 0$ then x_1 is uncorrelated with x_2 & x_3

* $R_{1.23} \geq r_{12}$

Linear regression

* If the Variable in a bivariate distributions are related we will find the point in the Scatter diagram will cluster round some curve called the curve of regression.

* If the curve is a straight line then it is called the line of regression

Formulae

Equation of the line of regression of y on x

$$y - \bar{y} = r \, \sigma_y / \sigma_x \, (x - \bar{x})$$

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

Scanned by CamScanner

where b_{yx} = regression coefficient of y on x

$$= r\,\sigma_y/\sigma_x$$

r – Correlation coefficient

σ_y – S.D of y, σ_x – SD of x.

Equation of the line of regression of x on y:

$$(x-\bar{x}) = r.\frac{\sigma_x}{\sigma_y}(y-\bar{y})$$

(or)

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

where b_{xy} = regression coefficient of x on y

$$= r.\frac{\sigma_x}{\sigma_y}$$

Properties of regression coefficients:

i) Correlation coefficient is the geometric mean between the regression coefficients.

i.e., $r = \pm\sqrt{b_{xy}\,b_{yx}}$

Note:

1. If both b_{xy} & b_{yx} are +ve then r is +ve.
2. If both b_{xy} & b_{yx} are –ve then r is –ve.

Scanned by CamScanner

(ii) If one of the regression coefficients is greater than unity then the other must be less than unity.

i.e., If $b_{yx} > 1$ then $b_{xy} < 1$ vice versa

(iii)
$$\left| \frac{1}{2}(b_{xy} + b_{yx}) \right| > |r|$$

iv) Regression coefficients are independent of change of origin but not scale.

Angle between two lines of regression:

$$\theta = \tan^{-1}\left\{ \frac{1-r^2}{|r|} \left(\frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \right) \right\}$$

Note

① If $r = 0$ then $\theta = \pi/2$

i.e., The lines of regression become $\perp^r$ to each other

② If $r = \pm 1$ then $\theta = 0$ (or) π

i.e., The two lines are $\parallel^l$ or coincide.

But both lines of regression are passing through $(\bar{x}, \bar{y})$. They cannot be $\parallel^l$.

$\therefore$ They coincide.

Scanned by CamScanner

Sums

① In a partially destroyed laboratory, record of an analysis of correlation data, the following results are legible:

Variance of $x = 9$, Regression equations:

$8x - 10y + 66 = 0$; $40x - 18y = 214$.

What are i) the mean values of x & y

ii) The correlation coefficient between x & y

& iii) The standard deviation of y?

Sol

Since both lines of regression passes through $(\bar{x}, \bar{y})$

$$\therefore \quad 8\bar{x} - 10\bar{y} + 66 = 0 \quad \text{--- (1)}$$

$$40\bar{x} - 18\bar{y} = 214 \quad \text{--- (2)}$$

$$(1) \times 5 \Rightarrow \quad \cancel{40\bar{x}} - 50\bar{y} = -330$$

$$(2) \Rightarrow \quad \cancel{40\bar{x}} - 18\bar{y} = 214$$

$$\underline{(-) \qquad (+) \qquad (-)}$$

$$\cancel{-}32\bar{y} = \cancel{-}544$$

$$\bar{y} = \frac{544}{32} = 17$$

Put $\bar{y} = 17$ in (1)

$$8\bar{x} = -66 + 170$$

$$\bar{x} = \frac{104}{8} = 13$$

Scanned by CamScanner

$\therefore (\bar{x}, \bar{y}) = (13, 17)$

ii) Let

Regression of y on x

$$8x - 10y + 66 = 0$$

$$y = 8/10\, x + 66/10$$

Regression of x on y

$$40x - 18y = 214$$

$$x = 18/40\, y + \frac{214}{40}$$

$\therefore b_{yx}$ = regression coefficient of y on x

$$= 8/10 = 4/5$$

b_{xy} = regression coefficient of x on y

$$= 18/40 = 9/20$$

$$\therefore r^2 = b_{xy} \cdot b_{yx} = 4/5 \cdot 9/20$$

$$= 9/25$$

$$r = \pm 3/5 = \pm 0.6$$

Since b_{xy} & b_{yx} are +ve

$$\therefore r = 0.6$$

Scanned by CamScanner

iii) we know that

$$b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$$

$$\cancel{4}/\cancel{8} = \cancel{3}/\cancel{8} \times \frac{\sigma_y}{\cancel{3}}$$

$$\sigma_y = 4$$

Scanned by CamScanner