Unit - I 07/06/2023

Correlation and Regression

If x and y are two variables related in such a way that an increasing one is accompanied by an increase or decrease in other. Such a relationship is called correlation or covariation.

Note: * If x and y increases or decrease together then x and y are positively correlated.

* If x increases as y decreases or vice versa then x and y are negatively correlated.

The correlation is measured by co-efficient of correlation.

The coefficient of correlation 'r' between x and y is given by

$$\boxed{r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n\,\sigma_x\,\sigma_y}}$$

$$r = \frac{\sum X_i Y_i}{n\,\sigma_x\,\sigma_y}$$

where $X_i = x_i - \bar{x}$

$Y_i = y_i - \bar{y}$

$$r = \frac{\sigma_{xy}}{\sigma_x\,\sigma_y}$$

where $\boxed{\sigma_{xy} = \frac{\sum X_i Y_i}{n}}$ is called covariance.

Also, we know that

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

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

therefore r can also be written as

$$r = \frac{\Sigma x_i y_i}{\sqrt{\Sigma x_i^2 \ \Sigma y_i^2}}$$

Note : * $-1 < r < 1$ (correlation lies between -1 and 1)

* If $r = \pm 1$, x and Y are perfectly correlated.
* If $r = 0$, x and Y are not correlated
* Coefficient 'r' is also reffered as Karl-Pearson's coefficient of correlation.

Q. The following table gives the ages of 10 couples. Calculate the covariance and the co-efficient of correlation between these ages.

Age of husband (x)	Age of wife (y)
23	18
27	22
28	23
29	24
30	25
31	26
33	28
35	29
36	30
39	32

$\sigma_{xy} = ?$ $r = ?$

$$\sigma_{xy} = \frac{\Sigma x_i y_i}{n} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$n = 10$

$\bar{x} = \frac{\Sigma x_i}{n} = 31.5$ $\bar{y} = \frac{\Sigma y_i}{n} = 25.7$

x_i	y_i	$X_i = x_i - \bar{x}$	$Y_i = y_i - \bar{y}$	$X_iY_i = (x_i-\bar{x})(y_i-\bar{y})$	X_i^2	Y_i^2
23	18	-8.1	-7.7	62.37	65.61	59.29
27	22	-4.1	-3.7	15.17	16.81	13.69
28	23	-3.1	-2.7	8.37	9.61	7.29
29	24	-2.1	-1.7	3.57	4.41	2.89
30	25	-1.1	-0.7	0.77	1.21	0.49
31	26	-0.1	0.3	-0.03	0.01	0.09
33	28	1.9	2.3	4.37	3.61	5.29
35	29	3.9	3.3	12.87	15.21	10.89
36	30	4.9	4.3	21.07	24.01	18.49
39	32	7.9	6.3	49.77	62.41	39.69
				$\Sigma X_iY_i = 178.30$	$\Sigma X_i^2 = 202.90$	$\Sigma Y_i^2 = 158.1$

Covariance $= \dfrac{\sigma_{xy}}{\sigma_x \sigma_y} = \dfrac{\Sigma X_i Y_i}{n}$

$$= \frac{178.30}{10}$$

$$\boxed{\sigma_{xy} = 17.83}$$

Coefficient of Correlation.

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$\sigma_x^2 = \frac{1}{n} \Sigma X_i^2$ $\quad\quad$ $\sigma_y^2 = \frac{1}{n} \Sigma Y_i^2$

$= \frac{1}{10}(202.9)$ $\quad\quad$ $= \frac{1}{10}(158.1)$

$= 20.29$ $\quad\quad$ $= 15.81$

$\sigma_x = 4.50$ $\quad\quad$ $\sigma_y = 3.976$

x_i^2	y_i^2
65.61	59.29
16.81	13.69
9.61	7.29
4.41	2.89
.21	0.49
.01	0.09
61	5.29
.21	10.89
.01	18.49
41	39.69
Σx_i^2 20290	$\Sigma y_i^2 =$ 158.1

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$= \frac{17.83}{4.5 \times 3.976}$$

$$\boxed{r = 0.996}$$

$\bar{x} = 31.3$

$\bar{y} = 25.7$

$\sigma_x = 4.50$

$\sigma_y = 3.97$

Q. Psycological teast of intalligente and computational ability where applied to 10 adults following is the record showing intelligence ratio I.R and ability ratio A.R. calculate the coefficient of correlation.

I.R	105	104	102	101	100	99	98	96	95	94
A.R	101	103	100	98	95	96	104	97	97	96

$r = 0.5477$ (From calculator). $\bar{x} = 99.4$ $\bar{y} = 98.7$

x	y	$X_i = x_i - \bar{x}$	$Y_i = y_i - \bar{y}$	$X_i Y_i$	X_i^2	Y_i^2
105	101	5.6	2.3	12.88	31.36	5.29
104	103	4.6	4.3	19.78	21.16	18.49
102	100	2.6	1.3	3.38	6.76	1.69
101	98	1.6	-0.7	-1.12	2.56	0.49
100	95	0.6	-3.7	-2.22	0.36	13.69
99	96	-0.4	-2.7	1.08	0.16	7.29
98	104	-1.4	5.3	-7.42	1.96	28.09
96	97	-3.4	-1.7	5.78	11.56	2.89
95	97	-4.4	-1.7	7.48	19.36	2.89
94	96	-5.4	-2.7	14.58	29.16	7.29
				54.2	114.4	88.1

4(var) → reg → r.

Stat mode → press 2 option → AC → shift-1 → 4 (var) → (σx, σ_y, $\bar{x}$, $\bar{y}$)

Q. Find the co-efficient of correlation between industrial production and export using the following data.

Production	55	56	58	59	60	60	62
Export	35	38	38	39	44	43	45

$$r = \frac{\Sigma X_i Y_i}{\sqrt{\Sigma X_i^2 \Sigma Y_i^2}}$$

$\bar{x} = 58.57$

$\bar{y} = 40.28$

$r = 0.931$ (calculator)

x_i	y_i	$X_i(x_i-\bar{x})$	$Y_i = y_i - \bar{y}$	X_i^2	Y_i^2	X_iY_i
55	35	-3.57	-5.28	12.74	27.87	18.8496
56	38	-2.57	-2.28	6.60	5.19	5.8596
58	38	-0.57	-2.28	0.32	5.19	1.2996
59	39	0.43	-1.28	0.18	1.63	-0.5504
60	44	1.43	3.72	2.04	13.83	5.3196
60	43	1.43	2.72	2.04	7.39	3.8896
62	45	3.43	4.72	11.76	22.27	16.1896
				35.68	83.37	$\Sigma X_iY_i = 50.85$

$\Sigma x_i^2 = 35.7$ $\quad \Sigma y_i^2 = 83.4$ $\quad \Sigma x_i y_i = 50.85$

$$r = \frac{\Sigma x_i y_i}{\sqrt{\Sigma x_i^2 \, \Sigma y_i^2}}$$

$$= \frac{50.85}{\sqrt{35.7 \times 83.4}}$$

$$= 0.931$$

for Σx_i^2 and Σy_i^2 stat mode → 2 → insert values of

shift +1 → 3 → $(\Sigma x^2, \Sigma y^2 \ldots)$

If $z = ax+by$ and r is correlation co-efficient between x and y show that $r = \frac{\sigma_z^2 - (a^2\sigma_x^2 + b^2\sigma_y^2)}{2ab\sigma_x\sigma_y}$

Proof: $\sigma_x^2 = \frac{\Sigma x_i^2}{n} = \frac{\Sigma(x_i-\bar{x})^2}{n}$

$$\sigma_y^2 = \frac{\Sigma y_i^2}{n} = \frac{\Sigma(y_i-\bar{y})^2}{n}$$

$$\sigma_z^2 = \frac{\Sigma z_i^2}{n} = \frac{\Sigma(z_i-\bar{z})^2}{n}$$

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

given $z = ax+By$ — 1

$$\bar{z} = a\bar{x} + b\bar{y}$$

$$z_i = ax_i + by_i \quad (z = 1,2,3,4\text{----})$$

Let $(z_i - \bar{z}) = a(x_i-\bar{x}) + b(y_i-\bar{y})$

squaring on both sides

$$(z_i-\bar{z})^2 = a^2(x_i-\bar{x})^2 + b^2(y_i-\bar{y})^2 + 2ab(x_i-\bar{x})(y_i-\bar{y})$$

Summing these terms

$$\Sigma(z_i-\bar{z})^2 = a^2\Sigma(x_i-\bar{x})^2 + b^2\Sigma(y_i-\bar{y})^2 + 2ab\,\Sigma(x_i-\bar{x})(y_i-\bar{y})$$

$$n\sigma_z^2 = a^2n\sigma_x^2 + b^2n\sigma_y^2 + 2ab(nr\sigma_x\sigma_y)$$ (using the Formulae obtained above)

$$n\sigma_z^2 = n(a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab(r\sigma_x\sigma_y))$$

$$\sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab(r\sigma_x\sigma_y)$$

$$\boxed{r = \frac{\sigma_z^2 - (a^2\sigma_x^2 + b^2\sigma_y^2)}{2ab\sigma_x\sigma_y}}$$

Note: $z = ax+by$

(1) If $a=1, b=1$; $r = \frac{\sigma_{x+y}^2 - (\sigma_x^2+\sigma_y^2)}{2\sigma_x\sigma_y}$

(2) If $a=1, b=-1$; $r = \frac{\sigma_{x-y}^2 - (\sigma_x^2+\sigma_y^2)}{-2\sigma_x\sigma_y}$

Q. If standard deviation of x and y are 2 and 3 respectively and if the correlation co-efficient between x and y is 0.4. Find the standard deviation of x+y and x-y.

Sol^n:

$\sigma_x = 2 \quad \sigma_y = 3 \quad r = 0.4$

$\sigma_{x+y} = ? \quad \sigma_{x-y} = ?$

$z = ax + by$

$a = 1, \ b = 1$

$\sigma^2_{x+y} = \sigma_x^2 + \sigma_y^2 + 2r\sigma_x\sigma_y$

$= 4 + 9 + 2(0.4)(2)(3)$

$= 13 + 4.8$

$= 17.8$

$\boxed{\sigma_{x+y} = 4.219}$

$a = 1 \quad b = 1$

$\sigma^2_{x-y} = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y$

$= 4 + 9 - 2(0.4)(2)(3)$

$= 13 - 4.8$

$= 8.2$

$\boxed{\sigma_{x-y} = 2.86}$

Q. If the variables x and y are such that

(i) x+y has variance 15

(ii) x-y has variance 11

(iii) 2x+y has variance 29

Find σ_x, σ_y and r.

$\sigma^2_{x+y} = 15 = \sigma_x^2 + \sigma_y^2 + 2r\sigma_x\sigma_y$ —— (1) $a=1, b=1$

$\sigma^2_{x-y} = 11 = \sigma_x^2 + \sigma_y^2 - 2r\sigma_x\sigma_y$ —— (2) $a=1, b=-1$

$\sigma^2_{2x+y} = 29 = 4\sigma_x^2 + \sigma_y^2 - 4r\sigma_x\sigma_y$ —— (3) $a=2 \ b=1$

(1) + (2) $2(\sigma_x^2 + \sigma_y^2) = 26$

$\Rightarrow \sigma_x^2 + \sigma_y^2 = 13$ —— (4)

respe- x- and x-y.

① - ② $\quad 4 = 4r\sigma_x\sigma_y$

$r\sigma_x\sigma_y = 1$ — ⑥

③-② $\quad 18 = 3\sigma_x^2 + 6r\sigma_x\sigma_y$

$18 = 3\sigma_x^2 + 6$ [from ⑥]

$3\sigma_x^2 = 12$

$\boxed{\sigma_x = 2}$

From ④, $\quad 13 = \sigma_x^2 + \sigma_y^2$

$13 = 4 + \sigma_y^2$

$\boxed{\sigma_y = 3}$

From ⑥, $\quad r = \frac{1}{\sigma_x\sigma_y} = \frac{1}{6} = 0.167$

Regression Analysis.

It is Statistical tool which is employed for the purpose of making estimates.

In this analysis there are 2 types of variables.

Dependent and independent.

The variable whose value is influenced is called dependent variable.

The variable which influence is called independent variable.

Suppose 'n' pairs of values (x_1, y_1) (x_2, y_2) .... $(x_n y_n)$ are given A straight line is fit to this data taking 'x' as independent and 'y' as dependent variable. The straight line obtained is called line of regression of Y on X. It's slope is called 'regression' coefficient of Y on X. If the straight line is fit by taking y as independent & x as dependent then the line obtained is ' line of regression of x on y'. Its the reciprocal of its slope is called regression coefficient of X on Y.

Equations of regression lines

$$(y-\bar{y}) = r\underbrace{\frac{\sigma_y}{\sigma_x}}_{\text{slope}}(x-\bar{x})$$

dependent — $(y-\bar{y})$; independent — $(x-\bar{x})$

This equation is line of regression of y on x. And it's slope $r\frac{\sigma_y}{\sigma_x}$ which is denoted by $b_{yx} = r\frac{\sigma_y}{\sigma_x}$ is called regression coefficient of y on x.

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

This equation is line of regression of x on y. The slope of this equation is $\frac{1}{r\frac{\sigma_y}{\sigma_x}}$ and hence the reciprocal of slope denoted by $b_{xy} = r\frac{\sigma_x}{\sigma_y}$ is regression coefficient of 'x' on 'y'.

Remarks :

1) Since σ_x and σ_y are positive r, σ_x, σ_y have same signs.
2) $|r| = \sqrt{b_{yx} b_{xy}}$
3) Lines of regression always pass through $(\bar{x}, \bar{y})$.

Calculate the coefficient of correlation and obtain lines of regression for the following data. Obtain and estimate for y which corresponds to $x = 6.2$.

x	y	$x_i = x-\bar{x}$	$y_i = y-\bar{y}$	x_i^2	y_i^2	$x_i y_i$
1	9	-4	-3	16	9	12
2	8	-3	-4	9	16	12
3	10	-2	-2	4	4	4
4	12	-1	0	1	0	0
5	11	0	-1	0	1	0
6	13	1	1	1	1	1
7	14	2	2	4	4	4
8	16	3	4	9	16	12
9	15	4	3	16	9	12

$\Sigma x_i^2 = 60$

$\Sigma y_i^2 = 60$

$\Sigma x_i y_i = 57$

$\sigma_x^2 = \frac{\Sigma x_i^2}{n} = \frac{60}{9} = 6.67$

$\sigma_x = \sqrt{6.67} = 2.5819$

$\text{lly} \quad \sigma_y^2 = \frac{\Sigma Y_i^2}{n} = \frac{60}{9} = 6.67$

$\sigma_y = 2.5819$

(i) $r = 0.95$ (calculator)

$\bar{x} = 5$

$\bar{y} = 12$

$\sigma_x = 2.5819$

$\sigma_y = 2.5819$

$r = \frac{\sqrt{\Sigma X_i Y_i}}{\sqrt{\Sigma x_i^2 \Sigma y_i^2}} = \frac{57}{\sqrt{60 \times 60}} = 0.95$

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

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

$(y - 12) = 0.95(x - 5)$

$y = 0.95x + 7.25$

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

$(x - 5) = r \frac{\sigma_x}{\sigma_y}(y - 12)$

$x - 5 = 0.95(y - 12)$

$x - 5 = 0.95y - 11.4$

$x = 0.95y - 6.4$

$y = ?$ when $x = 6.2$

using (ii) $y = 0.95(6.2) + 7.25 = 13.14$

using (iii) $y = 13.26$ (Not necessary)

Q. The following data is found in respect to the prices of a certain consumer item in belagavi and Bengaluru. Average price at Belagavi is 65 Rs and at bengaluru is Rs 67 Standard deviation at belagavi is 2.5 and 3.5 in Bengaluru The coefficient of correlation b/n the prices in two cities is 0.8. Find the most likely price in Bengaluru corresponding to price of 270 in Belagavi.

$\bar{x} = 65$ $\bar{y} = 67$ $r = 0.8$

$\sigma_x = 2.5$ $\sigma_y = 3.5$

From the equation of regression $y - \bar{y} = b_{yx}(x - \bar{x})$

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

$$y - 67 = 0.8 \times \frac{3.5}{2.5}(x - 65)$$

$$y - 67 = 1.12(x - 65)$$

$$y - 67 = 1.12x - 72.8$$

$$y = 1.12x - 5.8$$

for $x = 70$

$$y = 1.12(70) - 5.8$$

$$= 72.6$$

$y = 72.6$ ₹ in Bengaluru.

Q. In a partially destroyed laboratory record of correlation data the following results only are available

(i) variance of x is 9

(ii) Regression equations are $8x - 10y = -66$ and $40x - 18y = 214$.

Find

(a) Mean of x and Y

(b) standard deviation of Y

(c) correlation co-efficient between x and Y

$\sigma_x^2 = 9 \Rightarrow \sigma_x = 3$

$8x - 10y = -66$

$40x - 18\bar{y} = 214$

(a) line of regression always passes through point $(\bar{x}, \bar{y})$.

so

$8\bar{x} - 10\bar{y} = -66$

$40\bar{x} - 18\bar{y} = 214$

$\bar{x} = 13 \quad \bar{y} = 17.$

(b) $r = \sqrt{b_{yx} b_{xy}}$

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

$b_{xy} = r\frac{\sigma_x}{\sigma_y}$

Slope of $8x - 10y = -66$

$y = \frac{8}{10}x + \frac{66}{10}$

$b_{yx} = 8/10$

Slope of $40x - 18y = 214$

$x = \frac{18y}{40} - \frac{9}{20}$

$b_{xy} = \frac{18}{40}$

$r = \sqrt{\frac{8}{10} \times \frac{18}{40}}$

$r = 0.6$

b) $b_{yx} = r \frac{\sigma_y}{\sigma_x}$

$$\sigma_y = \frac{8}{10} \times \frac{8}{0.6}$$

$$\sigma_y = 4$$

Multiple and Partial Correlation.

When the data involves 2 variables, the correlation between the variables is called simple correlation. But when the data involves more than 2 variables then multiple and partial correlations are studied.

Multiple Correlation.

It is used to study the cummulative effect of all independent variables on dependent variable.

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

1- dependent variable
2 & 3- independent variable

$$R_{2,13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2 r_{21} r_{23} r_{13}}{1 - r_{13}^2}}$$

$$R_{3,12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2 r_{31} r_{32} r_{12}}{1 - r_{12}^2}}$$

Note: i) The magnitude of multiple correlation coeffiemt is always between 0 and 1.

ii) If the value of coefficient is 0 then there is an absence of linear relationship. and when it is equal to 1 it indecates perfect correlation.

Partial Correlation

Partial correlation involves

Study of correlation b/n an independent variable and dependent variable holding other independent variable constant statastically.

Here the effect of only one independent variable at a time while others are held constant on a dependent variable is of study.

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

$$r_{13.2} = \frac{r_{13} - r_{12} \times r_{32}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{32}^2}}$$

$$r_{23.1} = \frac{r_{23} - r_{21} \times r_{31}}{\sqrt{1-r_{21}^2}\sqrt{1-r_{31}^2}}$$

Q. Calculate the coefficient of multiple correlation $R_{1,23}$ on the basis of a set of data relating to 3 variables It is found that $r_{12} = -0.7$ $r_{13} = 0.65$ $r_{23} = -0.5$

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

$$= \sqrt{\frac{(-0.7)^2 + (0.65)^2 - 2(-0.7)(0.65)(-0.5)}{1-(-0.5)^2}}$$

$$R_{1,23} = 0.781$$

Q. In a trivariate distribution it is found that

$r_{12} = 0.7$

$r_{13} = 0.61$

$r_{23} = 0.40$ find the values of (i) $r_{12.3}$

(ii) $r_{13.2}$

(iii) $r_{23.1}$

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

$$= 0.627$$

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

$$= 0.504$$

(iii) $$r_{23.1} = \frac{r_{23} - r_{21} r_{31}}{\sqrt{1-r_{21}^2}\sqrt{1-r_{31}^2}}$$

$$= -0.047$$

Q. Obtain the multiple and partial correlation of the following.

x_1	x_2	x_3
28	74	5
33	87	11
21	69	4
24	69	9
38	81	7
46	97	10

$r_{12} = 0.935$ $r_{13} = 0.554$

$r_{23} = 0.668$

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

$R_{2.13} = \cancel{r_{21}^2 +}$

$\bar{x}_1 = 31.66 \quad \bar{x}_2 = 79.5 \quad \bar{x}_3 = 7.66$

x_1	x_2	x_3	$X_1 = x_1 - \bar{x}_1$	$X_2 = X_2 - \bar{X}_2$	$X_3 = X_3 - \bar{X}_3$	X_1X_2	X_1^2	X_2^2
28	74	5	-3.66	-5.5	-2.66	20.13	13.395	30.25
33	87	11	1.34	7.5	3.34	10.05	1.795	56.25
21	69	4	-10.66	-10.5	-3.66	111.93	113.635	110.25
24	69	9	-7.66	-10.5	1.34	80.43	58.675	110.25
38	81	7	6.34	1.5	-0.66	9.51	40.195	2.25
46	97	10	14.34	17.5	2.34	250.95	205.635	306.25
						$\Sigma X_1X_2 = 483$	$\Sigma X_1^2 = 433.33$	$\Sigma X_2^2 = 615.5$

$$r_{12} = \frac{\Sigma X_1X_2}{\sqrt{\Sigma X_1^2 \Sigma X_2^2}} = 0.935$$

X_2X_3	~~x_2x_3~~	~~x_3^2~~	X_3^2	X_1X_3
14.63	~~370~~	~~25~~	7.0756	9.7356
25.05	~~957~~	121	11.1556	4.4756
38.43	~~276~~	16	13.3956	39.0156
-14.07	~~621~~	81	1.7956	-10.2644
-0.99	567	49	0.4356	-4.1844
40.95	970	100	5.4756	33.5556

$\Sigma X_2X_3 = 104 \qquad \Sigma X_3^2 = 39.3336 \qquad \Sigma X_1X_3 = 72.3336$ ~~101.3312~~

$$r_{13} = \frac{\Sigma X_1X_3}{\sqrt{\Sigma X_1^2 \Sigma X_3^2}} = 0.5540$$

$$r_{23} = \frac{\Sigma X_2X_3}{\sqrt{\Sigma X_2^2 \Sigma X_3^2}} = 0.668$$

Multiple Correlation

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

$$= 0.939$$

$$R_{2,13} = \sqrt{\frac{r_{21}^2 + r_{23}^2 - 2r_{21}r_{23}r_{13}}{1 - r_{13}^2}} = 0.952$$

$$R_{3,12} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{31} \times r_{32} \times r_{12}}{1 - r_{12}^2}} = 0.697$$

Partial Correlation

$$r_{12.3} = \frac{r_{12} - r_{13} \times r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}} = 0.911$$

$$r_{13.2} = \frac{r_{13} - r_{12} \times r_{23}}{\sqrt{1 - r_{12}^2}\sqrt{1 - r_{23}^2}} = -0.267$$

$$r_{23.1} = \frac{r_{23} - r_{21} \times r_{31}}{\sqrt{1 - r_{21}^2}\sqrt{1 - r_{31}^2}} = 0.508$$

Q. Obtain the multiple and partial correlation of the following data.

x_1	193	172	113	230	91	125
x_2	1.6	2.2	33	15.5	43	40
x_3	851	1058	1357	816	1201	1115

x_1	x_2	x_3	$X_1 = x_1 - \bar{x}_1$	$X_2 = x_2 - \bar{x}_2$	$X_3 = x_3 - \bar{x}_3$	$X_1 X_2$	X_1^2	X_2^2
193	1.6	851	39	-20.95	-215.33	-817.05	1521	438.9025
172	2.2	1058	+18	-20.35	-8.33	-366.3	324	414.1225
113	33	1357	-41	10.45	290.67	-428.45	1681	~~370.5625~~ 109.2025
230	15.5	816	76	-7.05	-250.33	-535.8	5776	49.7025
91	43	1201	-63	20.45	134.67	-1288.35	3969	418.2025
125	40	1115	-29	17.45	48.67	-506.05	841	304.5025
						~~-1564.3~~ -3538	14112	~~1995.995~~ 1734.635

$\bar{x}_1 = 154$

$\bar{x}_2 = 22.55$

$\bar{x}_3 = 1066.33$

$X_1 X_3$	$X_2 X_3$	X_3^2
-8397.87	4511.1635	46367.0089
-149.94	169.5155	69.3889
-11917.47	~~5595.3975~~ 3037.5015	84489.0489
-19025.08	1764.8265	62665.1089
-8484.21	2754.0015	18136.0089
-1411.43	849.2915	2368.7689
-49386		

$r_{12} = -0.795$

$r_{13} = -0.898$

$r_{23} = 0.679$

$$r_{12} = \frac{\Sigma X_1 X_2}{\sqrt{\Sigma X_1^2 \, \Sigma X_2^2}} = -0.795$$

r

$R_{1,23} = 0.933$ $\quad$ $r_{12.3} = -0.579$

$R_{2,13} = 0.801$ $\quad$ $r_{13.2} = -0.805$

$R_{3,12} = 0.900$ $\quad$ $r_{23.1} = -0.138$

Regression on Multivrate (Analysis)

If x, y, z are 3 variables then the line of regression of z on x and y is given by

$$z = a + bx + cy$$

line of regression of y on x and z

$$y = a + bx + cz$$

line of regression of x on y and z

$$x = a + by + cz.$$

Here the parameters a, b and c can be obtained using the normal equations.

Normal equations (obtained by multiplying the terms other than present one).

$z = a + bx + cy$

$\Sigma z = na + b\Sigma x + c\Sigma y$ (take summation)

$\Sigma xz = a\Sigma x + b\Sigma x^2 + c\Sigma xy$ (multiply by x and summation)

$\Sigma yz = a\Sigma y + b\Sigma xy + c\Sigma y^2$ (multiply by y and summation)

$y = a + bx + cz$

$\Sigma y = na + b\Sigma x + c\Sigma z$

$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma xz$

$\Sigma zy = a\Sigma z + b\Sigma xz + c\Sigma z^2$

$x = a + by + cz$

$\Sigma x = na + b\Sigma y + c\Sigma z$

$\Sigma xy = a\Sigma y + b\Sigma y^2 + c\Sigma zy$

$\Sigma zy = a\Sigma z + b\Sigma zy + c\Sigma z^2$

Obtain the lines of regression for the following trivariate. ~~or~~
Fit multiple lines of regression

x	y	z	x^2	y^2	z^2	xy	yz	xz
57	8	64	3249	64	4096	456	512	3648
59	10	71	3481	100	5041	590	710	4189
49	6	53	2401	36	2809	294	318	2597
62	11	67	3844	121	4489	682	737	4154
51	8	55	2601	64	3025	408	440	2805
50	7	58	2500	49	3364	350	406	2900
55	10	77	3025	100	5929	550	770	4235
48	9	57	2304	81	3249	432	513	2736
52	10	56	2704	100	3136	520	560	2912
42	6	51	1764	36	2601	252	306	2142
61	12	76	3721	144	5776	732	912	4636
57	9	68	3249	81	4624	513	612	3876
643	106	753	34843	976	48139	5779	6796	40830

$\Sigma x = 643$ $\quad \Sigma x^2 = 34843$ $\quad \Sigma xy = 5779$

$\Sigma y = 106$ $\quad \Sigma y^2 = 976$ $\quad \Sigma yz = 6796$

$\Sigma z = 753$ $\quad \Sigma z^2 = 48139$ $\quad \Sigma xz = 40830$

Line of regression of z on x and y

$$z = a + bx + cy$$

$$\Sigma z = na + b\Sigma x + c\Sigma y$$

$$753 = 12a + b(643) + c(106) \quad \text{—} (1)$$

$$\Sigma xz = a\Sigma x + b\Sigma x^2 + c\Sigma xy$$

$$40830 = a(643) + b(34843) + c(5779) \quad \text{—} (2)$$

$$\Sigma yz = a\Sigma y + b\Sigma xy + c\Sigma y^2$$

$$6796 = a(106) + b(5779) + c(976) \quad \text{—} (3)$$

$a = 3.651$, $b = 0.854$, $c = 1.506$

Now line of regression of z on x and y is

$$\boxed{z = 3.651 + 0.854\cancel{y}x + 1.506y}$$

Line of regression of y on x and z is

$$y = a + bx + cz$$

$$\Sigma y = na + b\Sigma x + c\Sigma z$$

$$106 = 12a + b(643) + c(753)$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma xz$$

$$5779 = a(643) + b(34843) + c(40830)$$

$$\Sigma zy = a\Sigma z + b\Sigma xz + c\Sigma z^2$$

$$6796 = a(753) + b(40830) + c(48139)$$

$a = -4.56$ $b = 0.162$ $c = 0.074$

Line of regression of y on x and z is

$$\boxed{y = -4.56 + 0.162x + 0.074z}$$

Line of regression of x on y and z

$$x = a + by + cz$$

$$\Sigma x = na + b\Sigma y + c\Sigma z$$

$$643 = 12a + b(106) + c(753)$$

$$\Sigma yx = a\Sigma y + b\Sigma y^2 + c\Sigma zy$$

$$5779 = a(106) + b(976) + c(6796)$$

$$\Sigma zx = a\Sigma z + b\Sigma zy + c\Sigma z^2$$

$$40830 = a(753) + b(6796) + c(48139)$$

$a = 21.31$ $b = 1.288$ $c = 0.333$

Line of regression of x on y and z is

$$\boxed{x = 21.31 + 1.288y + 0.333z.}$$