Bivariate Regression

(X,Y)

$$Y \to \text{observed } y_i, \quad \text{predicted}$$

predicted:

$Y = a + bx \to$ Deterministic

$Y = a_0 + a_1 x + a_2 x^2$

$e_i = y_i - Y_i$

$y = Y + e \to$ Nondeter

Bivariate Regression Non deterministic model.

$y =$ dep. variable.

$X =$ indep. "

$Y =$ dep variable

$(x_2, x_3, \ldots, x_p) =$ indep. variable

$$Y = a_0 + a_2 x_2 + a_3 x_3 + \cdots + a_p x_p$$

Example:-

Suppose a farmer wants to measure the yield of produc^n of certain crop, say, wheat when he has a document on the records of soil fertility, avg rainfall of that place and the avg. humidity of the locaⁿ. As we can understand these 3 variables have a direct influence on the produc^n of wheat we think on a linear fu^n of 3 variables, explaining the variable, yield.

$$\text{yield} = a + b_1 \times \text{soil fertility} + b_2 \times \text{rainfall} + b_3 \times \text{humidity}$$

where a, b_1, b_2, b_3 are the constants to be determined.

Remember, the relationship might be non-linear and the constants might be non multiplicative.

But if the constants are multiplicative and the relationship is linear, we say that there is multiple regression rela^n existed in b/w (Yield : soil fertility, rainfall, humidity)

Mathematical Development :- Multiple Regression

Suppose we have p variables $(x_1, x_2, \ldots x_p)$ among which x_1 being influenced by $x_2, x_3 \ldots x_p$. The relationship b/w x_1 and $x_2, x_3, \ldots, x_p$ is assumed as linear as follows.

$$x_1 = a + b_2 x_2 + b_3 x_3 + \ldots + b_p x_p \quad \text{---(i)}$$

Clearly, borrowing the concept of bivariate regression, x_1 is dependent variable while $x_2, x_3, \ldots x_p$ are the bunch (set) of independent variables and $a, b_2, b_3 \ldots b_p$ are the constants to be determined.

(i) is called multiple regression line in a p dimensional space.

Suppose we have a data frame of size n where i^{th} data point is $(x_{i1}, x_{i2} \ldots x_{ip})$

$i = 1(1)n$

$$\begin{array}{ccccc} x_1 & x_2 & - & - & x_p \\ \hline \end{array}$$

$$\begin{bmatrix} (x_{11} & x_{12} & - & - & x_{1p}) \\ x_{21} & x_{22} & - & - & x_{2p} \\ \vdots & & & & \\ x_{i1} & x_{i2} & - & - & x_{ip} \\ \vdots & & & & \\ x_{n1} & x_{n2} & - & - & x_{np} \end{bmatrix}_{n \times p}$$

$\rightarrow$ 1st Data point obs^n

$\rightarrow$ Multivariate Data matrix

$$\begin{array}{cc} \overline{x_1} & \overline{x_2} \end{array}$$

where $\bar{x}_j$ = mean of j^{th} variable

$$= \frac{1}{n}\sum_{i=1}^{n} x_{ij} \quad , i = 1, 2, \ldots p.$$

s_j^2 = Variance of j^{th} variable $= \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$

$= v(x_j)$

$$s_{kj} = cov(x_k, x_j)$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_{ik} - \bar{x}_k)(x_{ij} - \bar{x}_j)$$

Clearly, $s_{jj} = Var(x_j) = s_j^2$

Our next objective is to estimate the constants/ coefficients.

Using Yule's dot notation, we can write (i) in, $x_1 = x_{1.23\ldots p} + e_{1.23\ldots p}$

where x_1 is the observed data pt. for dep. variable and $x_{1.23\ldots p}$ is the predicted data pt. (from (i)) of the dependent variable.

Then, $x_{1.23\ldots p} = a + b_2x_2 + \ldots + b_px_p$

Moreover, the unexplained part of the dependent variable x_1 is $e_{1.23\ldots p}$ (error)

[Remember the variable before the dot is the dependent variable, and the variable after the dot are the independent variable]

Thus, $e_{1.23\ldots p} = x_1 - x_{1.23\ldots p}$

To estimate the constants $a, b_2, b_3, \ldots b_p$.

We use least sq. principle which (residual pt.) minimizes the sum of square of errors.

where $L = \sum_{\alpha=1}^{n} e^2_{1.23\ldots p\alpha}$, $e_{1.23\ldots p1} = x_{11} - a - b_2 x_{12} - \cdots - b_p x_{1p}$

$= \sum_{\alpha} (x_{\alpha 1} - a - b_2 x_{\alpha 2} - \cdots - b_p x_{\alpha p})^2$ $e_{1.23\ldots p\alpha} = x_{\alpha 1} - a - b_2 x_{\alpha 2} - \cdots - b_p x_{\alpha p}$

$\therefore$ The normal eq^{n}s are,

$$\frac{\partial L}{\partial a} = 0 \Rightarrow (-2)\sum (x_{\alpha 1} - a - \cdots - b_p x_{\alpha p}) = 0$$

$$\Rightarrow \bar{x}_1 - a - b_2 \bar{x}_2 - \cdots - b_p \bar{x}_p = 0$$

$$\Rightarrow a = \bar{x}_1 - b_2 \bar{x}_2 - \cdots b_p \bar{x}_p$$

$$\frac{\partial L}{\partial b_2} = 0 \Rightarrow (-2)\sum x_{\alpha 2} \cdot (x_{\alpha 1} - a - \cdots b_p x_{\alpha p}) = 0$$

$$\sum_{\alpha} x_{\alpha 2} e_{1.23\ldots p\alpha} = 0$$

$$\sum_{\alpha} (x_{\alpha 2} - \bar{x}_2) e_{1.23\ldots p\alpha} = 0$$

$$\Rightarrow \sum x_{\alpha 2} (x_{\alpha 1} - \overline{x}_1 - b_2\overline{x}_2 - \cdots b_p \overline{x}_p - b_2 x_{\alpha 2} - \cdots - b_p x_{\alpha p}) = 0$$

$$\Rightarrow \sum_{\alpha=1}^{n} x_{\alpha 2} x_{\alpha 1} - \bar{x}_1 \sum_{\alpha=1}^{n} x_{\alpha 2} + b_2 \bar{x}_2 \sum_{\alpha=1}^{n} x_{\alpha 2} + \cdots + b_p \sum \bar{x}_p x_{\alpha 2} - b_2 \sum x_{\alpha 2}^2 - \cdots - b_p \sum x_{p\alpha} x_{\alpha 2} = 0$$

$$\Rightarrow \frac{1}{n} \sum x_{\alpha 2} x_{\alpha 1} - \frac{\bar{x}_1}{n} \sum x_{\alpha 2} + b_2 \frac{\bar{x}_2}{n} \sum x_{\alpha 2} + \cdots \frac{b_p}{n} \bar{x}_p \sum x_{\alpha 2} - \frac{b_2}{n} \sum x_{\alpha 2}^2 - \cdots - \frac{b_p}{n} \sum x_{\alpha p} x_{\alpha 2} = 0$$

$$\Rightarrow s_{12} - b_2 s_{22} - b_3 s_{23} - b_p s_{2p} = 0$$

$$\left.\begin{aligned}
&\Rightarrow s_{12} = b_2 s_{22} + b_3 s_{23} + \cdots + b_p s_{2p} \\
&\text{Similarly,} \\
&s_{13} = b_2 s_{23} + b_3 s_{33} + \cdots \; b_p s_{3p} \\
&\vdots \\
&s_{1p} = b_2 s_{2p} + b_3 s_{3p} + \cdots \; b_p s_{pp}
\end{aligned}\right\} \begin{array}{l}\text{System} \\ \text{of non-} \\ \text{homogeneous} \\ \text{linear} \\ \text{eq}^n\end{array}$$

<u>Remark:-</u>

'Linear' in regression means linear in parameters (constants). Linear in variables is sometimes confused with linear in parameters. But as regression theory evolves from linear model, we always look into linearity in parameters.

$$Y = b_1 + b_2 X + e \quad \begin{array}{l}\rightarrow \text{linear in constants} \\ \rightarrow \text{linear " variables.}\end{array}$$

$$e^Y = b_1 + b_2 X + \varepsilon \quad \begin{array}{l}\rightarrow \text{non linear in variable} \\ \rightarrow \text{linear in constant}\end{array}$$

$$Y = b_1 + b_2 X^2 + e \quad \begin{array}{l}\rightarrow \text{polynomial regression} \\ \rightarrow \text{linear in constant}\end{array}$$

$$Y = \frac{1}{b} + e^{b_2} X + \varepsilon \quad \begin{array}{l}\rightarrow \text{Non-linear in constant} \\ \rightarrow \text{Linear in variable}\end{array}$$

- Any non-linear in constant (parameters) model can be transformed into linear regression with suitable transforma^n in old constants.
- Linear regression is considered as fundamental model.

$$a = \bar{x}_1 - b_2\bar{x}_2 - \cdots - b_p\bar{x}_p$$

$$\left.\begin{aligned} s_{12} &= b_2 s_{22} + b_3 s_{23} + \cdots + b_p s_{2p} \\ s_{13} &= b_2 s_{23} + b_3 s_{33} + \cdots + b_p s_{3p} \\ &\vdots \\ s_{1p} &= b_2 s_{2p} + b_3 s_{3p} + \cdots + b_p s_{pp} \end{aligned}\right\} \rightarrow (2)$$

$\left[s_{ij} = cov(x_i, x_j) \quad i = 1(1)p \quad s_{ii} = v(x_i) = s_i^2 \right]$

Define sample variance covariance matrix as,

$$S_{p\times p} = \begin{pmatrix} v(x_1) & cov(x_1,x_2) & cov(x_1,x_3) & \cdots & cov(x_1,x_p) \\ cov(x_2,x_1) & v(x_2) & cov(x_2,x_3) & \cdots & cov(x_2,x_p) \\ \vdots & \vdots & \vdots & & \vdots \\ cov(x_p,x_1) & cov(x_p,x_2) & cov(x_p,x_3) & \cdots & v(x_p) \end{pmatrix}$$

$$= \begin{pmatrix} s_{11} & s_{12} & s_{13} & \cdots & s_{1p} \\ s_{21} & s_{22} & s_{23} & \cdots & s_{2p} \\ \vdots & & & & \vdots \\ s_{1p} & s_{2p} & s_{3p} & \cdots & s_{pp} \end{pmatrix}_{p\times p}$$

[Note that:- S is a symmetric matrix]

② can be rewritten in terms of system of non-homogeneous eq^n:

$$\begin{pmatrix} s_{12} \\ s_{13} \\ \vdots \\ \vdots \\ s_{1p} \end{pmatrix}_{(p-1)\times 1} = \begin{pmatrix} s_{22} & s_{23} & \cdots & s_{2j} & s_{2p} \\ s_{23} & s_{33} & \cdots & s_{3j} & s_{3p} \\ \vdots & & & \vdots & \vdots \\ \vdots & & & \vdots & \vdots \\ s_{2p} & s_{3p} & \cdots & s_{jp} & s_{pp} \end{pmatrix}_{(p-1)\times(p-1)} \begin{pmatrix} b_2 \\ b_3 \\ \vdots \\ \vdots \\ b_p \end{pmatrix}_{(p-1)\times 1}$$

By Cramer's rule,

$$b_2 = \frac{\begin{vmatrix} s_{12} & s_{23} & \cdots & s_{2p} \\ s_{13} & s_{33} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots \\ s_{1p} & s_{3p} & \cdots & s_{pp} \end{vmatrix}}{\begin{vmatrix} s_{22} & s_{23} & \cdots & s_{2p} \\ s_{23} & s_{33} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots \\ s_{2p} & s_{3p} & \cdots & s_{pp} \end{vmatrix}}$$

In general,

$$b_j = \frac{\begin{vmatrix} s_{22} & s_{23} & \cdots & s_{2,j-1} & s_{12} & s_{2,j+1} & \cdots & s_{2p} \\ s_{23} & s_{33} & & s_{3,j-1} & s_{13} & s_{3,j+1} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ s_{2p} & s_{3p} & & s_{p,j-1} & s_{1p} & s_{p,j+1} & \cdots & s_{pp} \end{vmatrix}}{\begin{vmatrix} s_{22} & s_{23} & \cdots & s_{2p} \\ s_{23} & s_{33} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots \\ s_{2p} & s_{3p} & \cdots & s_{pp} \end{vmatrix}}$$

$$= \frac{(-1)^{j-2}\begin{vmatrix} s_{12} & s_{22} & s_{23} & \cdots & s_{2,j-1} & s_{2,j+1} & \cdots & s_{2p} \\ s_{13} & s_{23} & s_{33} & & s_{3,j-1} & s_{3,j+1} & & s_{3p} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ s_{1p} & s_{2p} & s_{3p} & & s_{p,j-1} & s_{p,j+1} & \cdots & s_{pp} \end{vmatrix}}{\begin{vmatrix} s_{22} & s_{23} & \cdots & s_{2p} \\ s_{23} & s_{33} & \cdots & s_{3p} \\ \vdots & & & \vdots \\ s_{2p} & s_{3p} & \cdots & s_{pp} \end{vmatrix}}$$

⊛ cofactor of (i,j) element $= (-1)^{i+j}$ minor of (i,j) element.

$\therefore$ minor of (i,j) element $= (-1)^{-i+j}$ cofactor of (i,j) element.

$$= \left\langle \frac{(-1)^{j-2} \cdot (-1)^{-(1+j)} \text{ cofactor of } s_{1j}}{(-1)^{-(1+1)} \text{ cofactor of } s_{11}} \right\rangle$$

$$\Rightarrow b_j = \boxed{-\frac{S_{1,j}}{S_{1,1}}}$$

Where $S_{i,j}$ being the cofactor of (i,j) element in S.

Another representation of b_j comes from the correlation matrix R.

$$R = \begin{pmatrix} \operatorname{corr}(x_1, x_1) & \operatorname{corr}(x_1, x_2) & \cdots & \operatorname{corr}(x_1, x_p) \\ \operatorname{corr}(x_2, x_1) & \operatorname{corr}(x_2, x_2) & \cdots & \operatorname{corr}(x_2, x_p) \\ \vdots & \vdots & & \vdots \\ \operatorname{corr}(x_p, x_1) & \operatorname{corr}(x_p, x_2) & \cdots & \operatorname{corr}(x_p, x_p) \end{pmatrix}_{p \times p}$$

$$= \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{12} & 1 & r_{23} & \cdots & r_{2p} \\ \vdots & \vdots & & & \vdots \\ r_{p1} & r_{p2} & \cdots & & 1 \end{pmatrix}$$

$$\therefore\ b_j = (-1)^{j-2} \begin{vmatrix} r_{12} s_1 s_2 & r_{22} s_2^2 & r_{23} s_2 s_3 & \cdots & r_{2,j-1} s_2 s_{j-1} \end{vmatrix}$$

$$\therefore\ b_j = \frac{(-1)^{j-2}\begin{vmatrix} r_{12}s_1s_2 & s_{22}s_2^2 & \cdots & r_{2,j-1}s_2s_{j-1} & r_{2,j+1}s_2s_{j+1} & \cdots & r_{2p}s_2s_p \\ \vdots & \vdots & & & & & \vdots \\ r_{1p}s_1s_p & r_{2p}s_2s_p & \cdots & & & & r_{pp}s_ps_p \end{vmatrix}}{\begin{vmatrix} r_{22}s_2s_2 & r_{23}s_2s_3 & \cdots & r_{2p}s_2s_p \\ \vdots & \vdots & & \vdots \\ r_{2p}s_2s_p & r_{3p}s_3s_p & \cdots & r_{pp}s_ps_p \end{vmatrix}}$$

$$= \frac{(-1)^{j-2} s_1 s_2^2 s_3^2 \cdots s_{j-1}^2 s_{j+1}^2 s_j \cdots s_p^2 \begin{vmatrix} r_{12} & r_{22} & r_{23} \cdots & r_{2,j-1} & \cdots & r_{2p} \\ \vdots & & & \vdots & & \\ r_{1p} & r_{2p} & \cdots & r_{p,j-1} & \cdots & r_{pp} \end{vmatrix}}{s_2^2 s_3^2 \cdots s_j^2 \cdots s_p^2 \begin{vmatrix} r_{22} & r_{23} & \cdots & r_{2p} \\ \vdots & & & \vdots \\ r_{2p} & r_{3p} & \cdots & r_{pp} \end{vmatrix}}$$

$$\Rightarrow b_j = \frac{(-1)^{j-2}\dfrac{s_1}{s_j}(-1)^{-(1+j)} \text{ cofactor of } (1,j) \text{ element in } R}{(-1)^{-(1+1)} \text{ cofactor of } (1,1) \text{ element in } R}$$

$$= -\frac{s_1}{s_j}\,\frac{\text{cofactor of } (1,j) \text{ element in } R}{\text{cofactor of } (1,1) \text{ element in } R}$$

$$= -\frac{s_1}{s_j}\,\frac{R_{1,j}}{R_{1,1}} = \boxed{-\frac{R_{1,j}}{R_{1,1}}\cdot\frac{s_1}{s_j}}$$

Where $R_{i,j}$ being the cofactor of (i,j) element in R

<u>Finally</u> <u>the multiple regression line</u>

$$x_1 = a + b_2 x_2 + b_3 x_3 - - - - - + b_p x_p$$

$$= \bar{x}_1 - b_2\bar{x}_2 - b_3\bar{x}_3 - - - - - b_p\bar{x}_p + b_2 x_2 + b_3 x_3 + - - b_p x_p$$

$$= \bar{x}_1 + b_2(x_2 - \bar{x}_2) + b_3(x_3 - \bar{x}_3) + - - - - b_p(x_p - \bar{x}_p)$$

$$= \bar{x}_1 - \frac{R_{1,2}}{R_{1,1}}\cdot\frac{s_1}{s_2}(x_2 - \bar{x}_2) - \frac{R_{1,3}}{R_{1,1}}\frac{s_1}{s_3}(x_3 - \bar{x}_3) - - - - - \frac{R_{1,p}}{R_{1,1}}\frac{s_1}{s_p}(x_p - \bar{x}_p)$$

Multiple Regression line of x_1 on $x_2, x_3 \dots x_p$.

$$\therefore \left\langle x_1 = x_{1.23\ldots p} + e_{1.23\ldots p} \right\rangle$$

$$\left\{ x_{1.23\ldots p} = \bar{x}_1 - \frac{R_{1,2}}{R_{1,1}} \frac{s_1}{s_2}(x_2 - \bar{x}_2) - \frac{R_{1,3}}{R_{1,1}} \frac{s_1}{s_3}(x_3 - \bar{x}_3) - \cdots - \frac{R_{1,p}}{R_{1,1}} \frac{s_1}{s_p}(x_p - \bar{x}_p) \right.$$

In multiple regression line b_j s are called partial regression coefficient and mathematically denoted by $\boxed{b_{j1.23\ldots j-1,j+1\ldots p}}$ where

$$\left\langle b_{j1.23\ldots j-1, j+1\ldots p} = -\frac{R_{1,j}}{R_{1,1}} \cdot \frac{s_1}{s_j} \right\rangle$$

- We write it in a short format as b_j

<u>Why partial?</u>

- b_j measures the amount of change in dependent variable x_1 while unit amount of change occurred in independent variable x_j keeping the other ind. variables, $x_2, x_3, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p$ fixed.

<u>Remark:-</u>

$$\langle -\infty < b_j < \infty \rangle$$

$\langle x_1 = 4 - 2.2x_2 + 3.9x_3 - 4x_4 \Rightarrow$ Interpretation:-

There has a inverse relation b/w (x_1, x_2) and (x_1, x_4)

$\langle$ If x_4 dec. by 1 unit then x_1 will inc. by 4 unit.
If x_3 inc " 1 unit " x_1 " " by 3.9 unit. $\rangle$

$$x_{1.2} = \bar{x}_1 - \frac{R_{1,2}}{R_{1,1}} \frac{s_1}{s_2}(x_{22} - \bar{x}_2) - \frac{R_{1,3}}{R_{1,1}} \frac{s_1}{s_3}(x_{32} - \bar{x}_3) - \cdots - \frac{R_{1,p}}{R_{1,1}} \frac{s_1}{s_p}(x_p - \bar{x}_p)$$

Multiple Correlaⁿ coefficient :-

To measure the degree of associaⁿ b/w dep. variable x_1 and the set of indep. variables $(x_2, x_3, \ldots, x_p)$, ~~and~~ an extension of Pearson's correlaⁿ coeff. is introduced and that is called multiple correlaⁿ coeff..
(how effective (meaningful) the linear regression in explaining x_1 through, $(x_2, x_3, \ldots x_p)$).

It is denoted by, $r_{1.23\ldots p}$ and,

$$r_{1.23\ldots p} = \text{corr}(x_1, x_{1.23\ldots p})$$

$$= \frac{\text{cov}(x_1, x_{1.23\ldots p})}{\sqrt{V(x_1) \cdot V(x_{1.23\ldots p})}}$$

Derivaⁿ :-

Least square multiple regression line is,

$$x_1 = \bar{x}_1 - \frac{R_{1,2}}{R_{1,1}} \frac{s_1}{s_2}(x_2 - \bar{x}_2) - \cdots - \frac{R_{1,p}}{R_{1,1}} \frac{s_1}{s_p}(x_p - \bar{x}_p)$$

$$V(x_1) = s_{11} = s_1^2$$

The model is,

$$x_1 = x_{1.23\ldots p} + e_{1.23\ldots p}$$

$$\therefore x_{1.23\ldots p} = \bar{x}_1 - \frac{R_{1,2}}{R_{1,1}}\frac{s_1}{s_2}(x_2-\bar{x}_2) - \ldots \frac{R_{1,p}}{R_{1,1}}\frac{s_1}{s_p}(x_p-\bar{x}_p)$$

$$\Rightarrow \bar{x}_{1.23\ldots p} = \frac{1}{n}\sum_{\alpha=1}^{n} x_{1.23\ldots p\alpha}$$

$$= \frac{1}{n}\sum_{\alpha=1}^{n}\left(\bar{x}_1 - \frac{R_{1,2}}{R_{1,1}}\frac{s_1}{s_2}(x_{2\alpha}-\bar{x}_2) - \ldots \frac{R_{1,p}}{R_{1,1}}\frac{s_1}{s_p}(x_p-\bar{x}_p)\right)$$

$$= \bar{x}_1 - 0 - 0 \ldots 0$$
$$= \bar{x}_1$$

~~Also~~ Also,

$$\mathrm{cov}(x_1, x_{1.23\ldots p})$$
$$= \mathrm{cov}(x_{1.23\ldots p} + e_{1.23\ldots p}, x_{1.23\ldots p})$$
$$= V(x_{1.23\ldots p}) + \mathrm{cov}(e_{1.23\ldots p}, x_{1.23\ldots p})$$

Now, $\mathrm{cov}(e_{1.23\ldots p}, x_{1.23\ldots p})$ [From the normal eqn. $\sum e_{1.23\ldots p} = 0$]

$$= \frac{1}{n}\sum_{\alpha}(e_{1.23\ldots p\alpha} - \bar{e}_{1.23\ldots p\alpha})(x_{1.23\ldots p\alpha} - \bar{x}_1)$$

$$= \frac{1}{n}\sum_{\alpha} e_{1.23\ldots p\alpha}(a + b_2 x_{2\alpha} + \ldots b_p x_{p\alpha} - \bar{x}_1)$$

$$= \frac{1}{n}\sum_{\alpha} e_{1.23\ldots p\alpha}(\bar{x}_1 - b_2\bar{x}_2 - \ldots - b_p\bar{x}_p + b_2 x_{2\alpha} + b_3 x_{3\alpha} + \ldots b_p x_{p\alpha} - \bar{x}_1)$$

$$= \frac{1}{n}\Big[b_2\sum_{\alpha} e_{1.23\ldots p\alpha}(x_{2\alpha}-\bar{x}_2) + b_3\sum_{\alpha} e_{1.23\ldots p\alpha}(x_{3\alpha}-\bar{x}_3) + \ldots + b_p\sum_{\alpha} e_{1.23\ldots p\alpha}(x_{p\alpha}-\bar{x}_p)\Big]$$

$= 0$ [From the normal eq^n]

$$\therefore r_{1.23\ldots p} = \frac{V(x_{1.23\ldots p})}{\sqrt{V(x_1).V(x_{1.23\ldots p})}}$$

$$= \sqrt{\frac{V(x_{1.23\ldots p})}{V(x_1)}}$$

$\therefore r_{1.23\ldots p}$ is the ratio of two standard deviations, which leads to the range of $r_{1.23\ldots p}$

$\Rightarrow 0 < r_{1.23\ldots p} < 1$

In otherwards,

$$r_{1.23\ldots p} = \frac{cov(x_1, x_{1.23\ldots p})}{\sqrt{V(x_1)\,cov(x_1, x_{1.23\ldots p})}}$$

$$= \sqrt{\frac{cov(x_1, x_{1.23\ldots p})}{V(x_1)}}$$

$\therefore cov(x_1, x_{1.23\ldots p})$

$$= \frac{1}{n}\sum_{\alpha}(x_{1\alpha}-\bar{x}_1)(x_{1.23\ldots p\alpha}-\bar{x}_{1.23\ldots p})$$

$$= \frac{1}{n}\sum_{\alpha}(x_{1\alpha}-\bar{x}_1)(x_{1.23\ldots p\alpha}-\bar{x}_1)$$

$$= \frac{1}{n} \sum_{\alpha} (x_{1\alpha} - \bar{x}_1)\Big(\cancel{x_1} - \frac{R_{1,2}}{R_{1,1}} \frac{s_1}{s_2}(x_{2\alpha} - \bar{x}_2)$$

$$- \frac{R_{1,3}}{R_{1,1}} \frac{s_1}{s_3}(x_{3\alpha} - \bar{x}_3)$$

$$- \cdots \frac{R_{1,p}}{R_{1,1}} \frac{s_1}{s_p}(x_{p\alpha} - \bar{x}_p) - \cancel{\bar{x}_1}\Big)$$

$$= -\frac{1}{n} \sum_{\alpha} (x_{1\alpha} - \bar{x}_1)(x_{2\alpha} - \bar{x}_2) \frac{R_{1,2}}{R_{1,1}} \frac{s_1}{s_2}$$

$$- \frac{1}{n} \sum_{\alpha} (x_{1\alpha} - \bar{x}_1)(x_{3\alpha} - \bar{x}_3) \frac{R_{1,3}}{R_{1,1}} \frac{s_1}{s_3}$$

$$- \cdots - \frac{1}{n} \sum_{\alpha} (x_{1\alpha} - \bar{x}_1)(x_{p\alpha} - \bar{x}_p) \cdot \frac{R_{1,p}}{R_{1,1}} \frac{s_1}{s_p}$$

$$= -\Big[r_{12} s_1 s_2 \cdot \frac{R_{1,2}}{R_{1,1}} \frac{s_1}{s_2} + r_{13} s_1 s_3 \cdot \frac{R_{1,3}}{R_{1,1}} \frac{s_1}{s_3}$$

$$+ \cdots r_{1p} s_1 s_p \frac{R_{1,p}}{R_{1,1}} \frac{s_1}{s_p} \Big]$$

$$= -\frac{s_1^2}{R_{1,1}} \left(r_{12} R_{1,2} + r_{13} R_{1,3} + \cdots + r_{1p} R_{1,p} \right)$$

$$= -\frac{s_1^2}{R_{1,1}} \Big(r_{11} R_{1,1} + r_{12} R_{1,2} + \cdots r_{1p} \cdot R_{1,p}$$

$$- r_{11} R_{1,1} \Big)$$

$$= -\frac{s_1^2}{R_{1,1}} \left(|R| - R_{1,1} \right) = s_1^2 \left(1 - \frac{|R|}{R_{1,1}} \right)$$

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & \cdots & r_{2p} \\ & & & & \\ r_{p1} & r_{p2} & \cdots & \cdots & r_{pp} \end{pmatrix}$$

$$\therefore |R| = r_{11}R_{1,1} + r_{12}R_{1,2} + \cdots + r_{1p}R_{1,p}$$

$$\therefore r_{1.23\ldots p} = \sqrt{\frac{\mathrm{cov}(x_1, x_{1.23\ldots p})}{V(x_1)}}$$

$$= \sqrt{\frac{\left(1 - \frac{|R|}{R_{1,1}}\right)s_1^2}{s_1^2}}$$

$$\boxed{r_{1.23\ldots p} = \left(1 - \frac{|R|}{R_{1,1}}\right)^{1/2}}$$

Properties of multiple correla^n coeff.:-

(1) $r_{1.23\ldots p} = 1$

$\Rightarrow V(x_1) = V(x_{1.23\ldots p})$

$\Rightarrow V(e_{1.23\ldots p}) = 0$

$\Rightarrow$ There is no variability of $e_{1.23\ldots p}$

$\Rightarrow$ Error is constant $\Rightarrow$ (The predic^n formula (linear) is perfect)

(2) $r_{1.23\ldots p} = 0$

$\Rightarrow V(x_{1.23\ldots p}) = 0$

$\Rightarrow x_{1.23\ldots p} = a + b_2 x_2 + \ldots + b_p x_p = c$ is independent of $x_2, x_3, \ldots, x_p$

$\Rightarrow$ The predicⁿ formula fails to predict the dependent variable x_1.

Multiple correlaⁿ coeff:-

$$r_{1.23\ldots p} = \left(1 - \frac{|R|}{R_{1,1}}\right)^{1/2}$$

$0 < r_{1.23\ldots p} < 1 \longrightarrow = \sqrt{\frac{V(x_{1.23\ldots p})}{V(x_1)}}$

Remark:- $r^2_{1.23\ldots p} = \frac{V(x_{1.23\ldots p})}{V(x_1)}$

- $r^2_{1.23\ldots p}$ is called coefficient of determinaⁿ and generally denoted by R^2.
- Coeff. of determinaⁿ measures the variance of dependent variable explained by the prediction (regression) eqⁿ,

$$x_1 = a + b_2 x_2 + \ldots + b_p x_p$$

- R^2 is generally denoted in percentage format.

For e.g:- $r^2_{1.23\ldots p} = 0.35$

$\Rightarrow R^2 = 35\%$

35% of the total variability of dependent variable (x_1) is explained (determined) by the multiple regression line, $x_1 = a + b_2 x_2 + \dots + b_p x_p$

65% of the variability of x_1 remains unexplained.

Remark:-

- With the increase of independent variables in the model, the value of R^2 also increases.
  $\left[R_p^2 > R_{p-1}^2 > R^2_{p-2} > \dots > R_2^2 > R_1^2\right]$
- Now to check on the value of R^2, generally we consider upto 95% of explainability.
- Later on there discovered a modification of R^2 which is called adjusted R^2.
- Adjusted R^2 will not change its value significantly even if inclusion of new independent variable in the model.

Remark:- ~~Standard error of estimate~~ (Residual variance)

$$\left\{\begin{aligned} & V(e_{1.23\dots p}) \\ &= V(x_1) - V(x_{1.23\dots p}) \\ &= s_1^2 - \operatorname{cov}(x_1, x_{1.23\dots p}) \\ &= s_1^2 - s_1^2\left(1 - \frac{|R|}{R_{1,1}}\right) \\ &= \frac{|R|}{R_{1,1}} s_1^2 \end{aligned}\right\}$$

Standard error $\to$ $= \sqrt{\dfrac{|R|}{R_{1,1}} s_1^2}$

Practical: Multiple Regression, Multiple and Partial correlaⁿ

1) Q) The following data relates to the performance of a number of students in the final examinaⁿ (x_1), first term examinaⁿ (x_2), 2nd term examinaⁿ (x_3), third term examinaⁿ (x_4)

a) Fit a least square regression eqⁿ for x_1 on x_2, x_3, x_4 as

i) $x_1 = a + b_2 x_2 + b_3 x_3 + b_4 x_4$

ii) when $b_2 = b_3 = b_4$

iii) $b_2 + b_3 + b_4 = 1$

$$x_1 = a + b_2(x_2 + x_3 + x_4)$$
$$= a + b_2 Z$$

$$x_1 = a + (1 - b_3 - b_4)x_2 + b_3 x_3 + b_4 x_4$$
$$= a + x_2 + b_3(x_3 - x_2) + b_4(x_4 - x_2)$$

b) Obtain the multiple corrⁿ coeffⁿ of x_1 on x_2, x_3, x_4

Student number

Student No.	x_2	x_3	x_4	x_1
1	67.8	71.2	60.0	58.6
2	68.7	67.0	73.0	55.5
3	71.2	73.0	69.0	67.1
4	65.0	77.5	60.5	54.7
5	53.3	75.0	57.0	43.2
6	70.8	78.5	43.0	60.1
7	63.4	71.0	55.5	52.7
8	67.4	69.0	61.5	51.7
9	79.9	87.0	46.0	66.5
10	66.5	72.5	52.5	50.4
11	82.2	85.0	57.5	61.1
12	69.6	70.5	47.0	68.0
13	72.2	81.0	47.5	50.7

Student No	x_2	x_3	x_4	x_1
14	69.1	82.5	47.5	65.6
15	60.5	71.5	51.0	46.2

$$\Rightarrow R = \begin{pmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{12} & 1 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ r_{41} & r_{42} & r_{43} & 1 \end{pmatrix}$$

$\bar{x}_1 = 56.807$
$\bar{x}_2 = 68.507$
$\bar{x}_3 = 75.48$
$\bar{x}_4 = 55.233$

$\cancel{\sqrt{V(x_2)} = 1.8}$

$$r_{12} = \frac{cov(x_1, x_2)}{\sqrt{V(x_1) \cdot V(x_2)}}$$

$\sqrt{V(x_1)} = 7.572$

$= 0.699$

$\sqrt{V(x_2)} = 6.755$

$r_{13} = \cancel{0.699}\ 0.339$

$\sqrt{V(x_3)} = 5.913$

$r_{14} = -0.118$

$\sqrt{V(x_4)} = 8.37$

$r_{23} = 0.585$

$r_{24} = -0.143$

$cov(x_1, x_2) = 38.328$

$r_{34} = -0.525$

$cov(x_1, x_3) = 16.265$
$cov(x_1, x_4) = -7.984$
$cov(x_2, x_3) = 25.033$
$cov(x_2, x_4) = -8.677$
$cov(x_3, x_4) = -27.8524$

$$\therefore R = \begin{pmatrix} 1 & 0.699 & 0.339 & -0.118 \\ 0.699 & 1 & 0.585 & -0.143 \\ 0.339 & 0.585 & 1 & 0.525 \\ -0.118 & -0.143 & 0.525 & 1 \end{pmatrix}$$

$$\therefore x_1 = \bar{x}_1 - \frac{R_{1,2}}{R_{1,1}} \frac{s_1}{s_2}(x_2 - \bar{x}_2) - \frac{R_{1,3}}{R_{1,1}} \frac{s_1}{s_3}(x_3 - \bar{x}_3) - \frac{R_{1,4}}{R_{1,1}} \frac{s_1}{s_4}(x_4 - \bar{x}_4)$$

$$= 56.807 - 0.937(x_2 - 68.507) - 0.271(x_3 - 75.48) - 0.102(x_4 - 55.233)$$

$$R_{1,1} = \cancel{Q} \begin{vmatrix} 1 & 0.585 & -0.143 \\ 0.585 & 1 & 0.525 \\ -0.143 & 0.525 & 1 \end{vmatrix} = 0.27386$$

$$R_{1,2} = \begin{vmatrix} 0.699 & 0.339 & -0.118 \\ 0.585 & 1 & 0.525 \\ -0.143 & 0.525 & 1 \end{vmatrix} = 0.229$$

$$R_{1,3} = \begin{vmatrix} 0.699 & 0.339 & -0.118 \\ 1 & 0.585 & -0.143 \\ -0.143 & 0.325 & 1 \end{vmatrix} = 0.058$$

$$R_{1,4} = \begin{vmatrix} 0.699 & 0.339 & -0.118 \\ 1 & 0.585 & -0.143 \\ 0.585 & 1 & 0.525 \\ \cancel{-0.143} & \cancel{0.525} & \cancel{1} \end{vmatrix} \cancel{= 0.058} = 0.031$$

$$|R| = 0.137$$

b) $r_{1.234} = \left(1 - \frac{0.137}{0.27386}\right)^{1/2}$

$= 0.707$

Q.1) For three variables x_1, x_2, x_3 prove that, $r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1$

Answer:-

Model $x_1 = a + b_2 x_2 + b_3 x_3$

$$= \bar{x}_1 - \frac{R_{1,2}}{R_{1,1}} \frac{S_1}{S_2}(x_2 - \bar{x}_2) - \frac{R_{1,3}}{R_{1,1}} \frac{S_1}{S_3}(x_3 - \bar{x}_3)$$

where,

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{pmatrix}$$

$$\therefore r_{1.23} = \left(1 - \frac{|R|}{R_{1,1}}\right)^{1/2}$$

$$|R| = 1(1 - r_{23}^2) - r_{12}(r_{12} - r_{23}r_{13}) + r_{13}(r_{12}r_{23} - r_{13})$$

$$= 1 - r_{23}^2 - r_{12}^2 + r_{12}r_{13}r_{23} + r_{13}r_{12}r_{23} - r_{13}^2$$

$$R_{1,1} = 1 - r_{23}^2$$

$$r_{1.23} = \left(1 - \frac{1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}\right)^{1/2}$$

$$\therefore r_{1.23}^2 \leq 1$$

$$\cancel{1} - \cancel{r_{23}^2} - \cancel{1} + r_{12}^2 + \cancel{r_{23}^2} + r_{13}^2 - 2r_{12}r_{13}r_{23} \leq 1 - r_{23}^2$$

$$\Rightarrow r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1 \text{ (proved)}$$

Q.2 Suppose for a regression model $x_1, x_2, \ldots, x_p$.

$corr(x_i, x_j) = r \quad \forall (i,j) \; \underset{i \neq j}{i = 1(1)p, j = 1(1)p}$

Find $r_{1.23\ldots p}$

Answer:-

$$R = \begin{pmatrix} 1 & r & r & \cdots & r \\ r & 1 & r & \cdots & r \\ - & - & - & & - \\ - & - & - & & - \\ - & - & - & & - \\ r & r & r & \cdots & 1 \end{pmatrix}_{p \times p}$$

→ Intra-class pattern matrix

$$M = \begin{pmatrix} a & b & \cdots & b \\ b & a & b & \cdots b \\ b & b & \cdots & a \end{pmatrix}_{n \times n}$$

$$|M| = (b-a)^{n-1}(a + \overline{(n-1)} \cdot b)$$
$$= (a-b)^{n-1}(a + \overline{(n-1)}\, b)$$

Here, $a = 1$, $b = r$

$$|R| = (1-r)^{p-1}(1 + \overline{(p-1)} \cdot r)$$

$$R_{1,1} = (1-r)^{p-2}(1 + \overline{(p-2)} \cdot r)$$

$$\therefore r_{1.23\ldots p} = \left(1 - \frac{(1-r)^{p-1}(1 + \overline{p-1} \cdot r)}{(1-r)^{p-2}(1 + (p-2) \cdot r)}\right)^{1/2}$$

$$= \left(1 - \frac{(1-r)(1+pr-r)}{1+pr-2r}\right)^{1/2}$$

$$= \left(\frac{\cancel{1} + \cancel{pr} - \cancel{2r} - \cancel{1} - \cancel{pr} + \cancel{r} + \cancel{r} + pr^2 - r^2}{1+pr-2r}\right)^{1/2}$$

$$= \left(\frac{r^2(p-1)}{1+pr-2r}\right)^{1/2} = r\left(\frac{p-1}{1+pr-2r}\right)^{1/2}$$

Ans

Q.3) Suppose for a p dimensional model,

$$r_{1j} = r \quad , j = 2(1)p$$
$$r_{ij} = r' \quad i = 2(1)p \, , j = 2(1)p$$

Find, $r_{1.23....p}$

Answer:–

$$R_{p\times p} = \begin{array}{c|c} & \begin{array}{cc} A & B \end{array} \\ C & \begin{pmatrix} 1 & r & r & \cdots & r \\ r & 1 & r' & \cdots & r' \\ r & r' & 1 & \cdots & r' \\ \vdots & & & & \\ r & r' & r' & \cdots & 1 \end{pmatrix}_{p\times p} \mathcal{D} \end{array}$$

$\underset{\sim}{r}' = (r \; r \cdots r)$

$$\underset{\sim}{r} = \begin{pmatrix} r \\ r \\ \vdots \\ r \end{pmatrix}$$

$$\therefore R_{1,1} = (1-r')^{p-2}(1+(p-2)r')$$

$$|R| = |1| \left| \begin{pmatrix} 1 & r' & \cdots & r' \\ r' & 1 & \cdots & r' \\ \vdots & & & \\ r' & r' & \cdots & 1 \end{pmatrix} - \begin{pmatrix} r \\ r \\ \vdots \\ r \end{pmatrix}_{(p-1)\times 1} (r \; r \cdots r)_{(1\times p-1)} \right|$$

partitⁿ determinent

$$m = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right)$$

$$|m| = |A|\,|D - CA^{-1}B|$$
$$= |\mathcal{D}|\,|A - B\mathcal{D}^{-1}C|$$

$$= \left| \begin{pmatrix} 1 & r' & \cdots & r' \\ r' & 1 & \cdots & r' \\ \vdots & & & \\ r' & r' & \cdots & 1 \end{pmatrix}_{(p-1)\times(p-1)} - \begin{pmatrix} r^2 & r^2 & \cdots & r^2 \\ r^2 & r^2 & \cdots & r^2 \\ \vdots & & & \\ r^2 & r^2 & \cdots & r^2 \end{pmatrix}_{(p-1)\times(p-1)} \right|$$

$$= \begin{vmatrix} 1-r^2 & r'-r^2 & r'-r^2 & \cdots & r'-r^2 \\ r'-r^2 & 1-r^2 & r'-r^2 & \cdots & r'-r^2 \\ \vdots & \vdots & \vdots & & \vdots \\ r'-r^2 & r'-r^2 & r'-r^2 & \cdots & 1-r^2 \end{vmatrix}_{p-1\times p-1}$$

$$= (1-\cancel{r^2}-r'+\cancel{r^2})^{p-2}\left[1-r^2+(p-2)(r'-r^2)\right]$$

$$\therefore r_{1.23\ldots p} = \left(1 - \frac{\cancel{(1-r')^{p-2}}\left[1-r^2+(p-2)(r'-r^2)\right]}{\cancel{(1-r')^{p-2}}\left(1+\overline{(p-2)}\,r'\right)}\right)^{1/2}$$

$$= \left(\frac{\cancel{1}+\cancel{pr'}-\cancel{2r'}-\cancel{1}+r^2-\cancel{pr'}+pr^2+\cancel{2r'}-2r^2}{1+pr'-2r'}\right)^{1/2}$$

$$= r\left(\frac{p-1}{1+pr'-2r'}\right)^{1/2}$$

Partial Correlaⁿ coeff :-

$(x_1, x_2, \ldots x_p)$

$x_1 = a + b_2 x_2 + \ldots + b_p x_p$

$$x_1 = \overline{x_1} - \frac{R_{1,2}}{R_{1,1}}\frac{s_1}{s_2}(x_2-\overline{x_2}) - \ldots \frac{R_{1,p}}{R_{1,1}}\frac{s_1}{s_p}(x_p-\overline{x_p})$$

$r_{1.23\ldots p}$

$$x_1 \rightarrow \begin{Bmatrix} x_2 \\ x_3 \\ \vdots \\ x_p \end{Bmatrix}$$

Defi:- Partial correlaⁿ is the measure of associaⁿ b/w two variables in a multiple regression set up when influence of the other variables are eliminated.

For a multiple regression setup $(x_1, x_2, \ldots, x_p)$ partial correlaⁿ b/w x_1 and x_2 is denoted by $r_{12.34\ldots p}$

Mathematically,

$$r_{12.34\ldots p} = \text{Corr}\left(e_{1.34\ldots p}, e_{2.34\ldots p}\right)$$

$$= \frac{\text{Cov}\left(e_{1.34\ldots p}, e_{2.34\ldots p}\right)}{\sqrt{V\left(e_{1.34\ldots p}\right)V\left(e_{2.34\ldots p}\right)}}$$

Why $r_{12.34\ldots p}$ is considered as a partial correlaⁿ?

→ Think about the model

$$x_1 = a + b_3x_3 + \ldots + b_px_p \quad (x_1, x_3, \ldots, x_p)$$

and, $$x_2 = a^* + b_3^*x_3 + \ldots + b_p^*x_p \quad (x_2, x_3, \ldots, x_p)$$

If we apply least square regression line,

$$R^{(2)} = \begin{pmatrix} r_{11} & r_{13} & \ldots & r_{1p} \\ r_{31} & r_{33} & \ldots & r_{3p} \\ r_{p1} & r_{p3} & \ldots & r_{pp} \end{pmatrix}$$

unexplained part of x_1 by $x_3, x_4, \ldots x_p$

$$\underbrace{x_1 = x_{1.34\ldots p}}_{\text{explained part of } x_1 \text{ by } x_3, x_4, \ldots x_p} + e_{1.34\ldots p} \text{ (model for the 1st eqⁿ)}$$

unexplained part of x_2 by $x_3, x_4, \ldots x_p$

$$\underbrace{x_2 = x_{2.34\ldots p}}_{\text{explained part of } x_2 \text{ by } x_3, x_4, \ldots x_p} + e_{2.34\ldots p} \text{ (model for the 2nd eqⁿ)}$$

Derivaⁿ of $r_{12.34\ldots p}$

Let the data structure,

$$\begin{matrix} x_{11} & x_{12} & x_{13} & \ldots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \ldots & x_{2p} \\ \vdots \\ x_{n1} & x_{n2} & x_{n3} & \ldots & x_{np} \end{matrix}$$

Multiple regression line of $x_1, x_2, \ldots, x_p$

$$x_1 = \bar{x}_1 - \frac{R_{1,2}}{R_{1,1}}\frac{s_1}{s_2}(x_2-\bar{x}_2) - \frac{R_{1,3}}{R_{1,1}}\frac{s_1}{s_3}(x_3-\bar{x}_3)$$
$$- \frac{R_{1,p}}{R_{1,1}}\frac{s_1}{s_p}(x_p-\bar{x}_p)$$

<u>Regression model is</u>

$$x_1 = x_{1.23\ldots p} + e_{1.23\ldots p}$$

Let us think on two submodels,

i) x_1 on $(x_3, x_4, \ldots x_p)$

ii) x_2 on $(x_3, x_4, \ldots, x_p)$

<u>Borrowing the idea of multiple regression model</u> :-

i) $\Rightarrow x_1 = a^* + b_3^* x_3 + b_4^* x_4 + \ldots + b_p^* x_p$

$$x_1 = x_{1.34\ldots p} + e_{1.34\ldots p}$$

ii) $\Rightarrow x_2 = a^0 + b_3^0 x_3 + b_4^0 x_4 + \ldots + b_p^0 x_p$

$$x_2 = x_{2.34\ldots p} + e_{2.34\ldots p}$$

Proceeding the same way as we did in multiple regression set up,

$$x_{1.34\ldots p} = \bar{x}_1 - \frac{R^{(2)}_{1,3}}{R^{(2)}_{1,1}}\frac{s_1}{s_3}(x_3-\bar{x}_3)$$
$$- \frac{R^{(2)}_{1,p}}{R^{(2)}_{1,1}}\frac{s_1}{s_p}(x_p-\bar{x}_p)$$

$$x_{2.34\ldots p} = \bar{x}_2 - \frac{R^{(2)}_{2,3}}{R^{(2)}_{2,2}}\frac{s_2}{s_3}(x_3-\bar{x}_3)$$
$$- \frac{R^{(2)}_{2,p}}{R^{(2)}_{2,2}}\frac{s_2}{s_p}(x_p-\bar{x}_p)$$

where,

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2p} \\ \vdots \\ r_{p1} & r_{p2} & r_{p3} & \cdots & r_{pp} \end{pmatrix}$$

$R^{(1)}$ = Discarding 1st row and 1st column from R

$$R^{(1)} = \begin{pmatrix} r_{22} & r_{23} & \cdots & r_{2p} \\ r_{32} & r_{33} & \cdots & r_{3p} \\ \vdots \\ r_{p2} & r_{p3} & \cdots & r_{pp} \end{pmatrix}$$

$R^{(2)}$ = Discarding 2nd row and 2nd column from R

$$R^{(2)} = \begin{pmatrix} r_{11} & r_{13} & \cdots & r_{1p} \\ r_{31} & r_{33} & \cdots & r_{3p} \\ \vdots \\ r_{p1} & r_{p3} & \cdots & r_{pp} \end{pmatrix}$$

$R_{1,3}^{(2)}$ = cofactor of r_{13} in $R^{(2)}$

$R_{ij}^{(k)}$ = cofactor of r_{ij} in $R^{(k)}$, $k=1,2$

Also we know,

$$\text{Cov}(x_1, x_{1.34\ldots p}) = V(x_{1.34\ldots p})$$

$$\text{and } \text{Cov}(x_2, x_{2.34\ldots p}) = V(x_{2.34\ldots p})$$

Again, $\text{cov}(x_1, x_{1.34\ldots p})$

$$= \frac{1}{n}\sum_{\alpha=1}^{n}(x_{1\alpha} - \bar{x}_1)(x_{1.34\ldots p_\alpha} - \bar{x}_1)$$

$$= \frac{1}{n}\sum_{\alpha=1}^{n}(x_{1\alpha}-\bar{x}_1)\Big(\cancel{\bar{x}_1} - \frac{R^{(2)}_{1,3}}{R^{(2)}_{1,1}}\frac{s_1}{s_3}(x_{3\alpha}-\bar{x}_3)$$

$$- \cdots \frac{R^{(2)}_{1,p}}{R^{(2)}_{1,1}}\frac{s_1}{s_p}(x_{p\alpha}-\bar{x}_p)\cancel{-\bar{x}_1}\Big)$$

$$= \frac{R^{(2)}_{1,3}}{R^{(2)}_{1,1}} r_{13}s_1s_3\cdot\frac{s_1}{s_3} - \frac{R^{(2)}_{1,4}}{R^{(2)}_{1,1}} r_{14}s_1s_4\frac{s_1}{s_4}$$

$$- \cdots - \frac{R^{(2)}_{1,p}}{R^{(2)}_{1,1}}\frac{s_1}{s_p} r_{1p}s_1s_p$$

$$= \frac{-s_1^2}{R^{(2)}_{1,1}}\Big[R^{(2)}_{1,3}r_{13} + R^{(2)}_{1,4}r_{14} + \cdots + R^{(2)}_{1,p}r_{1p} + R^{(2)}_{1,1}r_{11} - r_{11}R_{1,1}^{(2)}\Big]$$

$$= \frac{-s_1^2}{R^{(2)}_{1,1}}\Big[|R^{(2)}| - R^{(2)}_{1,1}\Big]$$

$$= -s_1^2\Big[1 - \frac{|R^{(2)}|}{R^{(2)}_{1,1}}\Big]$$

$$= V(x_1\cdot 34\ldots p)$$

$$\left[\; R^{(2)} = \begin{pmatrix} r_{11} & r_{13} & r_{14} & \cdots & r_{1p} \\ r_{31} & r_{32} & r_{34} & \cdots & r_{3p} \\ \vdots & & & & \\ r_{p1} & r_{p3} & \cdots & & r_{pp} \end{pmatrix}\right.$$

similarly,

$$V(x_2\cdot 34\ldots p) = s_2^2\left(1 - \frac{|R^{(1)}|}{R^{(1)}_{2,2}}\right)$$

$$\therefore V(e_{1\cdot 34\ldots p})$$

$$= V(x_1) - V(x_{1\cdot 34\ldots p}) = s_1^2 - s_1^2\left(1 - \frac{|R^{(2)}|}{R^{(2)}_{1,1}}\right)$$

$$= \left\langle s_1^2\frac{|R^{(2)}|}{R^{(2)}_{1,1}}\right\rangle$$

$$\left[\begin{array}{l} x_1 = x_{1\cdot 34\ldots p} + e_{1\cdot 34\ldots p} \\ V(x_1) = V(x_{1\cdot 34\ldots p}) + V(e_{1\cdot 34\ldots p}) \end{array}\right.$$

$$v(e_{2.34\ldots p}) = v(x_2) - v(x_{2.34\ldots p})$$

$$= \left\langle \frac{s_2^2\,|R^{(1)}|}{R^{(1)}_{2,2}} \right\rangle$$

Prob 1

$$x_1 = a + b_2x_2 + b_3x_3 + b_4x_4$$

- unrestricted multiple regression model
- multiple corr. coeff

Restricted regression model

i) $b_2 = b_3 = b_4$

$$x_1 = a + b_2x_2 + b_2x_3 + b_2x_4$$

$$\Rightarrow x_1 = a + b_2(x_2 + x_3 + x_4)$$

$$\Rightarrow x_1 = a + b_2 Z$$

$$\left\langle \begin{array}{l} a = \bar{x}_1 - b_2\bar{Z} \\ b_2 = \dfrac{cov(x_1, Z)}{V(x_1)} \end{array} \right\rangle$$

	x_1	$Z = x_2 + x_3 + x_4$
1	58.6	199
2	55.5	208.7
3		213.2
4		203
5		185.4
6		192.3
7		189.9
8		197.9
9		212.9
10		191.5
11		224.7
12		187.1
13		200.7
14		199.1
15		183

$\bar{x}_1 = 56.807$, $V(x_1) = 57.335$

$cov(x_1, x_2 + x_3 + x_4)$

$= cov(x_1, x_2) + cov(x_1, x_3) + cov(x_1, x_4)$

$= 43.8626$

$b_2 = 0.765$

$a = -95.599$

$x_1 = -95.599 + 0.765 Z$

ii) $b_2 + b_3 + b_4 = 1$

$\Rightarrow b_2 = 1 - b_3 - b_4$

$$x_1 = a + (1 - b_3 - b_4)x_2 + b_3 x_3 + b_4 x_4$$
$$= a + x_2 + b_3(x_3 - x_2) + b_4(x_4 - x_2)$$
$$\doteq a + x_2 (1 - b_3 - b_4) + b_3 x_3 + b_4 x_4$$

$$x_1 = a + b' x_2 + b_3 x_3 + b_4 x_4$$

$$a = \bar{x}_1 - b'\bar{x}_2 - b_3\bar{x}_3 - b_4\bar{x}_4$$

$$1 - b_3 - b_4 = b' = -\frac{R_{1,2}}{R_{1,1}}\frac{s_1}{s_2}$$

$$b_4 = \frac{R_{1,4}}{R_{1,1}}\frac{s_1}{s_4}$$

$$b_3 = \frac{-R_{1,3}}{R_{1,1}}\frac{s_1}{s_3},$$

$b_3 = -0.218$

$b_4 = -0.086$, $b' = 1.304$

$a = 56.807 - (1.$

$\mathrm{cov}(e_{1.34\ldots p}, e_{2.34\ldots p})$

$$= \frac{1}{n}\sum_{\alpha=1}^{n}\left(e_{1.34\ldots p_\alpha} - \bar{e}\right)\left(e_{2.34\ldots p_\alpha} - \bar{e}\right)$$

$$= \frac{1}{n}\sum_{\alpha=1}^{n} e_{1.34\ldots p_\alpha}\, e_{2.34\ldots p_\alpha}$$ as $\left\{\begin{array}{l}\bar{e}_{1.34\ldots p}=0 \\ \text{and } \bar{e}_{2.34\ldots p}=0\end{array}\right\}$

$$= \frac{1}{n}\sum_{\alpha=1}^{n}\left(x_{1\alpha}-\bar{x}_1+\frac{R_{1,3}^{(2)}}{R_{1,1}^{(2)}}\frac{s_1}{s_3}(x_{3\alpha}-\bar{x}_3)+\cdots+\frac{R_{1,p}^{(2)}}{R_{1,1}^{(2)}}\frac{s_1}{s_p}(x_{p\alpha}-\bar{x}_p)\right)\cdot e_{2.34\ldots p_\alpha}$$

$$= \frac{1}{n}\sum_{\alpha=1}^{n}(x_{1\alpha}-\bar{x}_1)\, e_{2.34\ldots p_\alpha}$$

As $\sum_\alpha (x_{i\alpha}-\bar{x}_i)\, e_{2.34\ldots p_\alpha}=0$ for, $i=3,4,\ldots p$ due to normal eq^{n}s. related to $x_2 = a^* + b_3^* x_3 + \ldots + b_p^* x_p$

$$= \frac{1}{n}\sum_{\alpha=1}^{n}(x_{1\alpha}-\bar{x}_1)\left(x_{2\alpha}-\bar{x}_2+\frac{R_{2,3}^{(1)}}{R_{2,2}^{(1)}}\frac{s_2}{s_3}(x_{3\alpha}-\bar{x}_3)+\cdots+\frac{R_{2,p}^{(1)}}{R_{2,2}^{(1)}}\frac{s_2}{s_p}(x_{p\alpha}-\bar{x}_p)\right)$$

$$= \frac{1}{n}\sum_{\alpha=1}^{n}(x_{1\alpha}-\bar{x}_1)(x_{2\alpha}-\bar{x}_2) + \frac{R_{2,3}^{(1)}}{R_{2,2}^{(1)}}\frac{s_2}{s_3}\frac{1}{n}\sum_\alpha (x_{1\alpha}-\bar{x}_1)(x_{3\alpha}-\bar{x}_3)$$

$$+\cdots \frac{R_{2,p}^{(1)}}{R_{2,2}^{(1)}}\frac{s_2}{s_p}\frac{1}{n}\sum_\alpha (x_{1\alpha}-\bar{x}_1)(x_{p\alpha}-\bar{x}_p)$$

$$= r_{12}s_1s_2 + \frac{R_{2,3}^{(1)}}{R_{2,2}^{(1)}}\frac{s_2}{\cancel{s_3}}r_{13}s_1\cancel{s_3}+\cdots+\frac{R_{2,p}^{(1)}}{R_{2,2}^{(1)}}\frac{s_2}{\cancel{s_p}}\cdot r_{1p}s_1\cancel{s_p}$$

$$= \frac{s_1 s_2}{R_{2,2}^{(1)}}\left[r_{12}R_{2,2}^{(1)} + r_{13}R_{2,3}^{(1)} + \cdots + r_{1p}R_{2,p}^{(1)}\right]$$

The expression inside the bracket is the determinant obtainable from $R^{(1)}$ by replacing its first row $(r_{22}, r_{23}, \ldots, r_{2p})$ by $(r_{12}, r_{13}, \ldots, r_{1p})$

$$= \frac{S_1 S_2}{R_{2,2}^{(1)}} \begin{vmatrix} r_{12} & r_{13} & \cdots & r_{1p} \\ r_{32} & r_{33} & \cdots & r_{3p} \\ \vdots & & & \\ r_{p2} & r_{p3} & \cdots & r_{pp} \end{vmatrix}$$

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1p} \\ \cancel{r_{21}} & \cancel{r_{22}} & \cancel{r_{23}} & \cdots & \cancel{r_{2p}} \\ r_{31} & r_{32} & r_{33} & \cdots & r_{3p} \\ \vdots & & & & \\ r_{p1} & r_{p2} & r_{p3} & \cdots & r_{pp} \end{pmatrix}$$

$$= -\frac{S_1 S_2}{R_{2,2}^{(1)}} R_{2,1}$$ where $R_{2,1}$ is the cofactor of r_{21} in R.

Finally,

$$r_{12.34\ldots p} = \frac{-\dfrac{\cancel{S_1}\cancel{S_2}}{R_{2,2}^{(1)}} R_{2,1}}{\sqrt{\cancel{S_1^2} \dfrac{|R^{(2)}|}{R_{1,1}^{(2)}} \cdot \cancel{S_2^2} \dfrac{|R^{(1)}|}{R_{2,2}^{(1)}}}}$$

$$= -\frac{R_{2,1} \left[R_{1,1}^{(2)}\right]^{1/2}}{\left[R_{2,2}^{(1)}\right]^{1/2} \left[|R^{(2)}|\right]^{1/2} \left[|R^{(1)}|\right]^{2/2}}$$

But, $R^{(1)} = R_{1,1}$ & $R^{(2)} = R_{2,2}$

$$\therefore = -\frac{R_{2,1}}{\sqrt{R_{1,1}R_{2,2}}} \cdot \frac{\left(R_{1,1}^{(2)}\right)^{1/2}}{\left(R_{2,2}^{(1)}\right)^{1/2}}$$

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2p} \\ r_{31} & r_{32} & r_{33} & \cdots & r_{3p} \\ \vdots & & & & \\ r_{p1} & r_{p2} & r_{p3} & \cdots & r_{pp} \end{pmatrix} \to R_{2,2}^{(1)} = R_{1,1}^{(2)}$$

$$= -\frac{R_{2,1}}{\sqrt{R_{1,1} \cdot R_{2,2}}} \left[\because R_{1,1}^{(2)} = R_{2,2}^{(1)}\right]$$

- Partial correla^n coeff. lies within -1 to 1 as the numerator ~~denominator~~ (covariance) can be positive/negative, unlike multiple correla^n where the numerator (covariance, $cov(x_1, x_{1.23\ldots p})$) turns to be $V(x_{1.23\ldots p})$.
- Square of Standard error of estimate (residual variance)

$$s^2_{1.34\ldots p} = V(e_{1.34\ldots p}) = \frac{|R^{(2)}|}{R_{1,1}^{(2)}} s_1^2 = \frac{R_{2,2}}{R_{1,1}^{(2)}} s_1^2$$

$$s^2_{2.34\ldots p} = V(e_{2.34\ldots p}) = \frac{|R^{(1)}|}{R_{2,2}^{(1)}} s_2^2 = \frac{R_{1,1}}{R_{2,2}^{(1)}} s_2^2$$

and $$\boxed{r_{12.34\ldots p} = \frac{-R_{2,1}}{\sqrt{R_{1,1}R_{2,2}}}}$$

$$\boxed{b_{12.34\ldots p} = -\frac{R_{1,2}}{R_{1,1}}\frac{s_1}{s_2}}$$

• Relaⁿ ship b/w partial regression coeff, partial correlaⁿ coeff. and residuals:-

$$\frac{s^2_{1.34\ldots p}}{s^2_{2.34\ldots p}} = \frac{R_{2,2}}{R_{1,1}}\frac{s_1^2}{s_2^2} \quad \left[\because R^{(2)}_{1,1} = R^{(1)}_{2,2}\right]$$

$$\Rightarrow \frac{s_{1.34\ldots p}}{s_{2.34\ldots p}} = \sqrt{\frac{R_{2,2}}{R_{1,1}}}\cdot\frac{s_1}{s_2}$$

$$\Rightarrow \frac{s_{1.34\ldots p}}{s_{2.34\ldots p}}\frac{\sqrt{R_{1,1}}}{\sqrt{R_{2,2}}} = \frac{s_1}{s_2}$$

$$\therefore b_{12.34\ldots p} = -\frac{R_{1,2}}{R_{1,1}}\frac{s_{1.34\ldots p}}{s_{2.34\ldots p}}\cdot\frac{\sqrt{R_{1,1}}}{\sqrt{R_{2,2}}}$$

$$= -\frac{R_{1,2}}{\sqrt{R_{1,1}R_{2,2}}}\cdot\frac{s_{1.34\ldots p}}{s_{2.34\ldots p}} \quad \text{Ans} \quad [R_{1,2} = R_{2,1}]$$

$$= \boxed{r_{12.34\ldots p}\cdot\frac{s_{1.34\ldots p}}{s_{2.34\ldots p}}}$$

Properties of residuals:-

$$(x_1, x_2, \dots, x_p)$$

$$x_1 = x_{1.23\dots p} + e_{1.23\dots p}$$

{From 2nd to pth normal eqn s} $\Rightarrow$ $$\sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i) e_{1.23\dots p_\alpha} = 0, \quad i = 2, 3, \dots, p \quad \text{---} (1)$$

$$\Rightarrow \sum_{\alpha=1}^{n} x_{i\alpha} \cdot e_{1.23\dots p_\alpha} = 0$$

From the partial regression concept,

$$x_1 = x_{1.34\dots p} + e_{1.34\dots p}$$
$$= a^* + b_3^* x_3 + \dots + b_p^* x_p + e_{1.34\dots p}$$
$$x_2 = x_{2.34\dots p} + e_{2.34\dots p}$$
$$= a^0 + b_3^0 x_3 + \dots + b_p^0 x_p + e_{2.34\dots p}$$

From (1),

$$\sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i) e_{1.34\dots p_\alpha} = 0, \quad i = 3, 4, \dots, p$$

From (2),

$$\sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i) e_{2.34\dots p_\alpha} = 0, \quad i = 3, 4, \dots, p$$

$$\sum_{\alpha=1}^{n} e_{1.23\dots p\alpha}^2$$

$$= \sum_{\alpha=1}^{n} e_{1.23\dots p_\alpha} \cdot e_{1.23\dots p_\alpha}$$

$$= \sum_{\alpha=1}^{n} \left(x_{1\alpha} - \bar{x}_1 - b_{12.34\dots p}(x_{2\alpha} - \bar{x}_2) - b_{13.24\dots p}(x_{3\alpha} - \bar{x}_3) - \dots - b_{1p.23\dots p-1}(x_{p\alpha} - \bar{x}_p)\right) e_{1.23\dots p_\alpha}$$

$$= \sum_{\alpha} (x_{1\alpha} - \bar{x}_1) e_{1.23\dots p_\alpha} - b_{12.34\dots p} \sum (x_{2\alpha} - \bar{x}_2) e_{1.23\dots p_\alpha} \dots$$
$$\dots - b_{1p.23\dots p-1} \sum (x_{p\alpha} - \bar{x}_p) e_{1.23\dots p_\alpha}$$

$$= \sum_{\alpha} (x_{1\alpha} - \bar{x}_1) e_{1.23\dots p_\alpha} = \sum_{\alpha=1}^{n} x_{1\alpha} e_{1.23\dots p_\alpha}$$

$$\sum_{\alpha}\left(x_{1\alpha} - a' - b'_{12.34\ldots p-1}\, x_{2\alpha} - b_{13.24\ldots p-1}\, x_{3\alpha} - \cdots - b_{1p-1.23\ldots p-2}\, x_{(p-1)\alpha}\right) e_{1.23\ldots p\alpha}$$

Think on a least square multiple regression line of x_1 on $x_2, x_3, \ldots, x_{p-1}$.

$$x_1 = a' + b'_2 x_2 + b'_3 x_3 + \cdots + b'_{p-1} x_{p-1}$$

$$\begin{cases} \sum_{\alpha=1}^{n} e_{1.23\ldots p-1\alpha} = 0 \\ \sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)\, e_{1.23\ldots p-1\alpha} = 0, \quad i = 2,3,\ldots p-1 \end{cases}$$

$$= \boxed{\sum_{\alpha=1}^{n} e_{1.23\ldots p-1\alpha}\, e_{1.23\ldots p\alpha}}$$

Statement:-

The sum of product of any two residuals in which all the subscripts after dot of the first term occur among the subscripts after dot of the seconds is unaltered, if we omit any subscript ~~of~~ after dot of the first.

$e_{1.23\ldots p}$

primary subscript (suffix)

secondary suffix

Result:

~~pth ord~~

p^{th} order multiple correla^n is always higher than $(p-1)^{st}$ order multiple correla^n.

$$\boxed{r_{1.23\ldots p} \geq r_{1.23\ldots p-1}}$$

In general $\langle r_{1.23\ldots p} \geq r_{1.23\ldots p-1} \geq r_{1.23\ldots p-2} \geq \cdots \geq r_{12} \rangle$

Proof:

$$V(e_{1.23\ldots p}) = s^2_{1.23\ldots p}$$

$$= \frac{1}{n}\sum e^2_{1.23\ldots p\alpha}$$

$$\Rightarrow n s^2_{1.23\ldots p} = \sum_{\alpha=1}^{n} e^2_{1.23\ldots p\alpha}$$

$$= \sum_{\alpha=1}^{n} e_{1.23\ldots p-1\alpha} \cdot e_{1.23\ldots p\alpha}$$

$$= \sum_{\alpha=1}^{n} (x_{1\alpha} - \bar{x}_1)\, e_{1.23\ldots p\alpha} \ldots$$ (We ~~have to~~ will use it)

$$\sum_{\alpha=1}^{n} (x_1 - a' - b_2' x_2 - \ldots - b'_{p-1} x_{p-1}) \cdot e_{1.23\ldots p\alpha}$$

$$\Rightarrow \sum_{\alpha} \left(x_{1\alpha} - \bar{x}_1 + b_2'\bar{x}_2 + b_3'\bar{x}_3 + \ldots b'_{p-1}\bar{x}_{p-1} - b_2' x_{2\alpha} - \ldots - b_p' x_{p-1\alpha}\right) e_{1.23\ldots p\alpha}$$

$$\Rightarrow \sum (x_{1\alpha} - \bar{x}_1)\, e_{1.23\ldots p\alpha} - b_2 \underbrace{\sum (x_{2\alpha} - \bar{x}_2)\, e_{1.23\ldots p\alpha}}_{\to 0} - \ldots - b'_{p-1} \underbrace{\sum (x_{(p-1)\alpha} - \bar{x}_{p-1}) \cdot e_{1.23\ldots p\alpha}}_{\to 0}$$

Again,
$$n s^2_{1.23\ldots p} = \sum_{\alpha} e_{1.23\ldots p-1\alpha}\, e_{1.23\ldots p\alpha}$$

$$= \sum_{\alpha} e_{1.23\ldots p-1\alpha} \big(x_{1\alpha} - \bar{x}_1 - b_{12.34\ldots p}(x_{2\alpha} - \bar{x}_2) - b_{13.24\ldots p}(x_{3\alpha} - \bar{x}_3) - \ldots - b_{1p.23\ldots p-1}(x_{p\alpha} - \bar{x}_p)\big)$$

$$= \sum_{\alpha} (x_{1\alpha} - \bar{x}_1)\, e_{1.23\ldots \overline{p-1}\alpha} - 0 - \ldots - 0 - b_{1p.23\ldots p-1} \sum (x_{p\alpha} - \bar{x}_p) \cdot e_{1.23\ldots \overline{p-1}\alpha}$$

$$= n.s^2_{1.23\ldots p-1} - b_{1p.23\ldots p-1} \sum (x_{p\alpha} - \bar{x}_p).\, e_{1.23\ldots \overline{p-1}\alpha}$$

[From ■]

$$= ns^2_{1.23\ldots p-1} - b_{1p.23\ldots p-1} \sum \Big(x_{p\alpha} - \bar{x}_p - b_{p2.34\ldots p-1}(x_{2\alpha} - \bar{x}_2) - b_{p3.24\ldots p-1}(x_{3\alpha} - \bar{x}_3) - \ldots - b_{p,p-1.23\ldots p-2}(x_{\overline{p-1}\alpha} - \bar{x}_{p-1})\Big) \cdot (e_{1.23\ldots \overline{p-1}\alpha})$$

Rough work

$$x_p = a^0 + b_1^0 x_1 + b_2^0 x_2 + \ldots + b_{p-1}^0 x_{p-1} + e_{p.12\ldots p-1}$$

$$a^0 = \bar{x}_p - b_1^0 x_1 - \ldots - b_{p-1}^0 \bar{x}_{p-1}$$

$$\sum (x_{i\alpha} - \bar{x}_i)\, e_{p.12\ldots(p-1)\alpha} = 0, \; i = 1, \ldots, p-1$$

$$\sum (x_{i\alpha} - \bar{x}_i)\, e_{1.23\ldots p-1\,\alpha} = 0$$

→ It comes from,

$$x_1 = a^* + b_2^* x_2 + \ldots + b_{p-1}^* x_{p-1}$$

$$= ns^2_{1.23\ldots p-1} - b_{1p.23\ldots p-1} \sum e_{p.23\ldots \overline{p-1}\alpha}\, e_{1.23\ldots \overline{p-1}\alpha}$$

$$= ns^2_{1.23\ldots p-1} - nb_{1p.23\ldots p-1}\, r_{1p.23\ldots p-1}\, s_{p.23\ldots p-1}.\, s_{1.23\ldots p-1}$$

Rough work

$$r_{12.34\ldots p} = \frac{cov(e_{1.34\ldots p}, e_{2.34\ldots p})}{s_{1.34\ldots p}\, s_{2.34\ldots p}}$$

Rough work

$cov(e_{1.34\ldots p}, e_{2.34\ldots p}) = r_{12.34\ldots p} \times s_{1.34\ldots p} \times s_{2.34\ldots p}$

$= \frac{1}{n} \sum e_{1.34\ldots p\alpha} e_{2.34\ldots p\alpha}$

$\cancel{n s^2_{1.23}} = n s^2_{1.23\ldots p-1} - n \cdot r_{1p.23\ldots p-1} \frac{s_{1.23\ldots p-1}}{s_{p.23\ldots p-1}} \times \frac{r_{1p.23\ldots p-1} \times s_{p.23\ldots p-1}}{s_{1.23\ldots p-1}}$

$$\boxed{n s^2_{1.23\ldots p} = n s^2_{1.23\ldots p-1}\left[1 - r^2_{1p.23\ldots p-1}\right]}$$

Since, $0 < r^2_{1p.23\ldots p-1} < 1$

$0 < 1 - r^2_{1p.23\ldots p-1} < 1$

$$\Rightarrow \boxed{s^2_{1.23\ldots p} \leq s^2_{1.23\ldots p-1}}$$

⊛ residual variance dec. as the variable under model is inc.

practical

② Examine the following correlaⁿ matrix for internal consistency,

$$\begin{pmatrix} 1 & 0.175 & 0.812 & 0.463 \\ 0.175 & 1 & 0.712 & 0.139 \\ 0.812 & 0.712 & 1 & 0.268 \\ 0.463 & 0.139 & 0.268 & 1 \end{pmatrix}_{4\times4}$$

We need to find, $r^2_{1.234}$.

③ Suppose a computer has found for a given set of values of x_1, x_2, x_3 ~~are~~, $r_{12} = 0.91$, $r_{13} = 0.33$, $r_{23} = 0.81$. Examine ~~the~~ whether the computer is free from error.

i) $|R| > 0$

ii) $r_{1.23} > 0$

Sol^n.

2) The correlation matrix is given by,

$$R = \begin{pmatrix} 1 & 0.175 & 0.812 & 0.463 \\ 0.175 & 1 & 0.712 & 0.139 \\ 0.812 & 0.712 & 1 & 0.268 \\ 0.463 & 0.139 & 0.268 & 1 \end{pmatrix}$$

$$\therefore \det(R) = |R| = -0.0153$$

$$R_{1,1} = (-1)^2\, 0.455 = 0.455$$

We know that,

$$\therefore r^2_{1.234} = \left(1 - \frac{|R|}{R_{1,1}}\right)$$

$$= 1.0336 > 1$$

$$\therefore r^2_{1.234} > 1$$

But we know that,

$$0 < r^2_{1.234} \leq 1$$

$\therefore$ There exist no internal consistency.

3) We have,

$r_{21} = r_{12} = 0.91, r_{31} = r_{13} = 0.33, r_{32} = r_{23} = 0.81$

$\therefore$ The correla^n matrix will be

$$R = \begin{pmatrix} 1 & 0.91 & 0.33 \\ 0.91 & 1 & 0.81 \\ 0.33 & 0.81 & 1 \end{pmatrix}$$

$\therefore |R| = -0.10661 < 0$

and, $R_{1,1} = 1 - 0.81^2 = 0.3439$

$\therefore r^2_{1.23} = 1 - \frac{|R|}{R_{1,1}}$

$= 1.31 > 1$

We know that, $0 < r^2_{1.23} \leq 1$

Here, $|R| < 0$ and $r^2_{1.23} > 1$

$\therefore$ The computer is not free from error.

$n s^2_{1.23\ldots p} = n s^2_{1.23\ldots p-1}\left(1 - r^2_{1p.23\ldots p-1}\right)$ $\left[s^2_{1.23\ldots p} < s^2_{1.23\ldots p-1}\right]$ — A

We know, $s^2_{1.23\ldots p} = V(e_{1.23\ldots p})$

$= (1 - r^2_{1.23\ldots p}) s_1^2$

$r^2_{1.23\ldots p} = \frac{V(x_{1.23\ldots p})}{V(x_1)}$

$= \frac{V(x_1) - V(e_{1.23\ldots p})}{V(x_1)}$

$V(x_1) = s_1^2$

$V(e_{1.23\ldots p}) = s^2_{1.23\ldots p}$

Replacing the above in A,

$$(1-r^2_{1.23\ldots p})s_1^2 < (1-r^2_{1.23\ldots p-1})s_1^2$$

$$\Rightarrow r^2_{1.23\ldots p} > r^2_{1.23\ldots p-1}$$

$$\Rightarrow \boxed{r_{1.23\ldots p} > r_{1.23\ldots p-1}}$$

In general,

$$r_{1.23\ldots p} > r_{1.23\ldots p-1} > r_{1.23\ldots p-2} > \ldots > r_{1.23} > r_{12}$$

Recall

$$ns^2_{1.23\ldots p} = ns^2_{1.23\ldots p-1}(1-r^2_{1p.23\ldots p-1})$$

$$\Rightarrow n(1-r^2_{1.23\ldots p})s_1^2 = n(1-r^2_{1p.23\ldots p-1})s^2_{1.23\ldots p-2}(1-r^2_{1p-1.23\ldots p-2})$$

$$\Rightarrow n(1-r^2_{1.23\ldots p})s_1^2 = n(1-r^2_{1p.23\ldots p-1})(1-r^2_{1p-1.23\ldots p-2})(1-r^2_{1p-2.23\ldots p-3})s^2_{1.23\ldots p-3}$$

$$\Rightarrow (1-r^2_{1.23\ldots p})s_1^2 = (1-r^2_{1p.23\ldots p-1})(1-r^2_{1p-1.23\ldots p-2})\ldots(1-r^2_{13.2})(1-r^2_{12})s_1^2$$

$$\Rightarrow (1-r^2_{1.23\ldots p}) = (1-r^2_{1p.23\ldots p-1})(1-r^2_{1p-1.23\ldots p-2})\ldots(1-r^2_{13.2})(1-r^2_{12})$$

Corollary: Show that if multiple correlaⁿs is zero, then all other correlaⁿ (partial and total) will vanish.

proof:

$[\because r^2_{1.23\ldots p}=0]$ $1 = (1-r^2_{1p.23\ldots p-1})(1-r^2_{1p-1.23\ldots p-2})\ldots(1-r^2_{12})$

To hold the above equality each factor must be one (1) as, $0 \le r^2_{1p.23\ldots p-1} \le 1$

$0 \le r^2_{1p-1.23\ldots p-2} \le 1$

$\therefore 1 - r^2_{1p.23\ldots p-1} = 1$

$\Rightarrow \boxed{r^2_{1p.23\ldots p-1} = 0}$

$0 \le r^2_{12} \le 1$

Similarly, $\boxed{r^2_{12}=0} \to \boxed{r_{12}=0}$ From that we can also say, $\boxed{r_{13}=0}$, $\boxed{r_{14}=0}$, ... $\boxed{r_{1p}=0}$

Other types of partial correlaⁿs will ~~be~~ also be 0.

⊛ $b_{12.345}$ $r_{12.345} = \text{corr}^n(x_1, x_2 / x_3, x_4, x_5$ ~~remain fixed~~) (eliminate the effect of x_3, x_4, x_5 from the reg. model)

$b_{21.345} \ne b_{12.345}$ $r_{21.345} = r_{12.345}$

→ partial regression coefficient

→ [It means the change of x_2 while the changing of x_1 by one unit and the effect of x_3, x_4, x_5 remain fixed]

$$\boxed{b_{12.34\ldots p} = -\frac{R_{1,2}}{R_{1,1}}\frac{s_1}{s_2}}$$

$$\boxed{b_{21.34\ldots p} = -\frac{R_{2,1}}{R_{2,2}}\frac{s_2}{s_1}}$$

$$b_{12.34\ldots p} \times b_{21.34\ldots p} = \frac{R_{1,2} \times R_{2,1}}{R_{1,1} \cdot R_{2,2}} = R^2_{1,2}/R_{1,1} \cdot R_{2,2}$$

$\boxed{r^2_{12.34\ldots p}}$ ←

(*)
$$R_{4\times 4} = \begin{pmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{12} & 1 & r_{23} & r_{24} \\ r_{13} & r_{23} & 1 & r_{34} \\ r_{14} & r_{24} & r_{34} & 1 \end{pmatrix}$$

$$R_{1,2} = \begin{vmatrix} r_{12} & r_{23} & r_{24} \\ r_{13} & 1 & r_{34} \\ r_{14} & r_{34} & 1 \end{vmatrix}, \quad R_{2,1} = \begin{vmatrix} r_{12} & r_{13} & r_{14} \\ r_{23} & 1 & r_{34} \\ r_{24} & r_{34} & 1 \end{vmatrix}$$

$$\left\{ \begin{array}{l} R'_{1,2} = R_{2,1} \\ |R'_{1,2}| = |R_{2,1}| \end{array} \right\}$$

<u>Problem:</u>

Prove that if $r_{1j} = 0$, $j = 2(1)p$ then,

$r_{1.23\ldots p} = 0$.

<u>proof:</u>

$r_{12} = r_{13} = \ldots = r_{1p} = 0$

$$R = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & r_{23} & \cdots & r_{2p} \\ 0 & r_{32} & 1 & \cdots & r_{3p} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & r_{p2} & r_{p3} & \cdots & 1 \end{pmatrix}$$

$$r_{1.23\ldots p} = \left(1 - \frac{|R|}{R_{1,1}}\right)^{1/2}$$

Here, $|R| = R_{1,1}$

$$\therefore 1 - \frac{|R|}{R_{1,1}} = 0$$

$$\therefore \boxed{r_{1.23\ldots p} = 0} \text{ (proved)}$$

Problem: Suppose for a 3 variable model (x_1, x_2, x_3), $r_{12} = 0$, $r_{12.3} = 0$, does it imply $r_{1.23} = 0$?

proof:

$$R = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{pmatrix}$$

$$r_{12.3} = \frac{-R_{2,1}}{\sqrt{R_{1,1} R_{2,2}}} = 0$$

$$R_{2,1} = -(r_{12} - r_{13} \cdot r_{23}) = \begin{vmatrix} r_{12} & r_{13} \\ r_{23} & 1 \end{vmatrix}$$

$$= 0$$

$$\Rightarrow r_{12} = r_{23} \cdot r_{13}$$

$$\Rightarrow \because \boxed{r_{12} = 0} \quad \therefore \boxed{r_{23} \cdot r_{13} = 0}$$

$$\therefore r_{23} = 0 \,/\, r_{13} = 0$$

or, $r_{23} = r_{13} = 0$

$$\therefore r_{1.23} = \left(1 - \frac{|R|}{R_{1,1}}\right)^{1/2},$$

$$R = \begin{pmatrix} 1 & 0 & r_{13} \\ 0 & 1 & 0 \\ r_{13} & 0 & 1 \end{pmatrix}$$

We consider,
$r_{23} = 0$,
$r_{13} \neq 0$

$|R| = 1 - r_{13}^2$

$R_{1,1} = 1$

$\therefore |R| \neq R_{1,1}$

$\therefore r_{1.23} \neq 0$

$\therefore$ In general we can not say that if $r_{12} = 0$ and $r_{12.3} = 0$ then $r_{1.23}$ will also be equal to 0.

Result: Higher order regression coeff. in terms of lower order regression coeff.

$$\sum_{\alpha=1}^{n} e_{1.34\ldots p\alpha}\, e_{2.34\ldots p\alpha}$$

$$= \sum_{\alpha=1}^{n} (x_{1\alpha} - x_{1.34\ldots p\alpha})\, e_{2.34\ldots p\alpha}$$

$$= \sum_{\alpha=1}^{n} \Big(x_{1\alpha} - \bar{x}_1 - b_{13.4\ldots p}(x_{3\alpha} - \bar{x}_3) - \ldots - b_{1p.34\ldots(p-1)} (x_{p\alpha} - \bar{x}_p)\Big) e_{2.34\ldots p\alpha}$$

$$= \sum_{\alpha=1}^{n} x_{1\alpha} \cdot e_{2.34\ldots p\alpha}$$

$$\boxed{\begin{array}{l} \sum e_{2.34\ldots p\alpha} = 0 \\ \sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)\, e_{2.34\ldots p\alpha} = 0 \\ i = 3(1)p \end{array}}$$

$$= \sum_{\alpha=1}^{n} \Big(x_{1\alpha} - \bar{x}_1 - b_{13.4\ldots p-1}(x_{3\alpha} - \bar{x}_3) - \ldots - b_{1p-1.45\ldots p-2}(x_{p-1\alpha} - \bar{x}_{p-1})\Big) e_{2.34\ldots p\alpha}$$

$$= \sum_{\alpha=1}^{n} e_{1.34\ldots(p-1)\alpha}\, e_{2.34\ldots p\alpha}$$

$$= \sum_{\alpha=1}^{n} e_{1.34\ldots(p-1)\alpha} \left[x_{2\alpha} - x_{2.34\ldots p\alpha}\right]$$

$$= \sum_{\alpha=1}^{n} e_{1.34\ldots(p-1)\alpha} \Big[x_{2\alpha} - \bar{x}_2 - b_{23.4\ldots p}(x_{3\alpha} - \bar{x}_3) - \ldots - b_{2p.34\ldots(p-1)}(x_{p\alpha} - \bar{x}_p)\Big]$$

$$= \sum_{\alpha=1}^{n} e_{1.34\ldots(p-1)\alpha} \cdot x_{2\alpha} - 0 - 0 - \ldots - b_{2p.34\ldots p-1} \sum e_{1.34\ldots(p-1)\alpha} (x_{p\alpha} - \bar{x}_p)$$

$$= \sum e_{1.34\ldots(p-1)\alpha} \Big[x_{2\alpha} - \bar{x}_2 - b_{23.4\ldots(p-1)}(x_{3\alpha} - \bar{x}_3) - \ldots - b_{2(p-1)}(x_{(p-1)\alpha} - \bar{x}_{p-1})\Big]$$

$$- b_{2p.34\ldots p-1} \sum e_{1.34\ldots(p-1)\alpha}(x_{p\alpha} - \bar{x}_p)$$

$$= \sum e_{1.34\ldots(p-1)\alpha}\, e_{2.34\ldots(p-1)\alpha} - b_{2p.34\ldots p-1} \sum e_{1.34\ldots(p-1)\alpha}\, e_{p.34\ldots(p-1)\alpha}$$

Dividing both sides by n,

$$\Rightarrow \begin{cases} cov(e_{1.34\ldots p}, e_{2.34\ldots p}) = cov(e_{1.34\ldots p-1}, e_{2.34\ldots p-1}) \\ \qquad - b_{2p.34\ldots p-1} cov(e_{1.34\ldots p-1}, e_{p.34\ldots p-1}) \end{cases}$$

Now,

L.H.S $Cov(e_{1.34\ldots p}, e_{2.34\ldots p})$

$$= b_{21.34\ldots p}\, s^2_{1.34\ldots p}$$

Now,

R.H.S

$$Cov(e_{1.34\ldots p-1}, e_{2.34\ldots p-1}) - b_{2p.34\ldots p-1} Cov(e_{1.34\ldots p-1}, e_{p.34\ldots p-1})$$

$$= b_{21.34\ldots p-1}\, s^2_{1.34\ldots p-1} - b_{2p.34\ldots p-1} b_{p1.34\ldots p-1} \cdot s^2_{1.34\ldots p-1}$$

$\Rightarrow$ L.H.S = R.H.S

$$b_{21.34\ldots p}\, s^2_{1.34\ldots p} = (b_{21.34\ldots p-1} - b_{2p.34\ldots p-1} \cdot b_{p1.34\ldots p-1})\, s^2_{1.34\ldots p-1}$$

$$\Rightarrow b_{21.34\ldots p} = \frac{(b_{21.34\ldots p-1} - b_{2p.34\ldots p-1} \cdot b_{p1.34\ldots p-1})\, s^2_{1.34\ldots p-1}}{s^2_{1.34\ldots p}}$$

$$\therefore \boxed{b_{21.34\ldots p} = \frac{b_{21.34\ldots p-1} - b_{2p.34\ldots p-1} \cdot b_{p1.34\ldots p-1}}{1 - r^2_{1p.3\ldots p-1}}}$$

Using, $s^2_{1.23\ldots p} = (1 - r^2_{1p.23\ldots p-1})\, s^2_{1.23\ldots p-1}$

$\Rightarrow s^2_{1.3\ldots p} = (1 - r^2_{1p.3\ldots p-1})\, s^2_{1.34\ldots p-1}$

Corollary:-

H.W. Suppose x_1, x_2, x_3 are three variables with associatn. Express $b_{12.3}$ in terms of subsequent bivariate regression coeffn.

H.W Find what the value of $r_{1.23\ldots p}$ will be if the independent variables are pairwise uncorrelated.

$$R = \begin{pmatrix} 1 & 0 & r_{13} & r_{14} & \cdots & r_{1p} \\ 0 & 1 & 0 & r_{24} & \cdots & r_{2p} \\ r_{31} & 0 & 1 & 0 & \cdots & r_{3p} \\ \vdots \\ r_{p1} & r_{p2} & - & \cdots & 0 & 1 \end{pmatrix}$$

H.W
Fundamental
11.9 (301), 11.13

Practical

The following constants are obtain from measurement on length (mm) (x_1), volume (c.c) (x_2), weight (gm) (x_3) of 300 eggs.

$\bar{x}_1 = 55.95$, $\bar{x}_2 = 51.48$, $\bar{x}_3 = 56.03$
$s_1 = 2.26$, $s_2 = 4.39$, $s_3 = 4.41$
$r_{12} = 0.578$, $r_{13} = 0.581$, & $r_{23} = 0.974$

i) Obtain the linear regression eq^n of egg wt on length and volume. Hence estimate the wt of an egg whose length is 58 mm and volume is 52.5 cc.

ii) Give a measure of the usefullness of the above regression eq^n as a predicting formula.

iii) Compute the partial correla^n coeff^n of wt and volume eliminating the effect of length.

Sol^n:-

i) The regression eq^n is,

$$x_3 = a + b_1 x_1 + b_2 x_2$$

$$\therefore a = \bar{x}_3 - b_1\bar{x}_1 - b_2\bar{x}_2$$

$$b_j = -\frac{R_{3,j}}{R_{3,3}}\frac{s_3}{s_j}, \quad j = 1, 2$$

$$R = \begin{pmatrix} 1 & 0.578 & 0.581 \\ 0.578 & 1 & 0.974 \\ 0.581 & 0.974 & 1 \end{pmatrix}, \quad b_1 = -\frac{R_{3,1}}{R_{3,3}}\frac{s_3}{s_1}, \quad R_{3,1} = -0.018, \quad R_{3,3} = 0.666$$

$$b_1 = 0.05$$

$$b_2 = -\frac{R_{3,2}}{R_{3,3}} \cdot \frac{s_3}{s_2}, \quad R_{3,2} = -0.638, \quad R_{3,3} = 0.666$$

$$= \cancel{1.869} = 0.962$$

$$\therefore a = \cancel{-42.984}$$

$$\therefore x_3 = \cancel{-42.984} + 0.05x_1 + 1.869x_2$$

putting, $x_1 = 58$, $x_2 = 52.5$ we will get,

$$x_3 = \cancel{58.0385}$$

ii) $\cancel{|R| = (1 - 0.974^2) - 0.578(0.578 - 0.974 \times 0.581) + 0.581(0.578 \times 0.974 - 0.581)}$

$$\therefore a = 3.709$$

$$\therefore x_3 = 3.709 + 0.05x_1 + 0.962x_2$$

$$x_1 = 58, \quad x_2 = 52.5$$

$$\therefore \boxed{x_3 = 57.114}$$

ii) $r^2_{3.12} = R^2 = \left(1 - \frac{|R|}{R_{3,3}}\right) \quad \therefore |R| = 0.034$

$$= 0.949 \text{ (Ans)}$$

(95% of the total variance of wt is explained through the multiple regression line by ht. and volume.)

iii) $r_{23.1} = -\frac{R_{3,2}}{\sqrt{R_{2,2} \times R_{3,3}}} = \frac{0.638}{\sqrt{0.662 \times 0.666}}$

$R_{2,2} = 0.662$

$$= 0.961$$

~~H.W 1) We know that,~~

$$\cancel{\mathrm{cov}(e_{1.34\ldots p}, e_{2.34\ldots p}) = \mathrm{cov}(e_{1.34\ldots p-1}, e_{2.34\ldots p-1}) - b_{2p.34\ldots p-1}\cdot \mathrm{cov}(e_{1.34\ldots p-1}, e_{p.34\ldots p-1})}$$

H.W

1) We know that,

$$\mathrm{cov}(e_{1.34\ldots p}, e_{2.34\ldots p}) = \mathrm{cov}(e_{1.34\ldots p-1}, e_{2.34\ldots p-1}) - b_{2p.34\ldots p-1}\,\mathrm{cov}(e_{1.34\ldots p-1}, e_{p.34\ldots p-1})$$

Now in that question we have 3 variables x_1, x_2, x_3.

$$\therefore \mathrm{cov}(e_{1.3}, e_{2.3}) = \mathrm{cov}(e_1, e_2) - b_{23}\,\mathrm{cov}(e_1, e_3)$$

$$\Rightarrow b_{12.3}\, s_{23}^2 = b_{12} s_2^2 - b_{23} b_{13} s_3^2$$

$$\Rightarrow b_{12.3} = \left\langle \frac{b_{12} s_2^2 - b_{23} b_{13} s_3^2}{s_{23}^2} \right\rangle$$

2)

$$R = \begin{pmatrix} 1 & 0 & r_{13} & r_{14} & \cdots & r_{1p} \\ 0 & 1 & 0 & r_{24} & \cdots & r_{2p} \\ r_{31} & 0 & 1 & 0 & \cdots & r_{3p} \\ \vdots & & & & & \\ r_{p1} & r_{p2} & \cdots & & 0 & 1 \end{pmatrix}_{p\times p}$$

3) a) $a_1x_1 + a_2x_2 + a_3x_3 = K$

$\Rightarrow a_1x_1 + a_2x_2 = K - a_3x_3$

Let, $V(x_i) = s_i^2 \ \forall i = 1(1)3$

$cov(x_i, x_j) = r_{ij} s_i s_j, \quad i \neq j$

$\Rightarrow V(a_1x_1 + a_2x_2) = V(K - a_3x_3)$

$\Rightarrow a_1^2 V(x_1) + a_2^2 V(x_2) + 2a_1a_2 \, cov(x_1, x_2) = a_3^2 V(x_3)$

$\Rightarrow a_1^2 s_1^2 + a_2^2 s_2^2 + 2a_1a_2 r_{12} s_1 s_2 = a_3^2 s_3^2$

$$\Rightarrow r_{12} = \left\langle \frac{a_3 s_3^2 - a_1^2 s_1^2 - a_2^2 s_2^2}{2a_1a_2 s_1 s_2} \right\rangle \text{ (Ans)}$$

Similarly we can find r_{23} and r_{13}.

b) $$r_{12.3} = \frac{cov(x_1, x_2)}{\sqrt{V(x_1) \cdot V(x_2)}}$$

$a_1x_1 + a_2x_2 = K - a_3x_3$

$\Rightarrow a_1x_1 + a_2x_2 = K_1$ [∵ x_3 is ~~will be~~ fixed]

$$= \frac{cov\left(x_1, \frac{K_1 - a_1x_1}{a_2}\right)}{\sqrt{V(x_1) \cdot V\left(\frac{K_1 - a_1x_1}{a_2}\right)}}$$

$x_1 = \frac{K}{a_1} - \frac{a_2}{a_1}x_2 - \frac{a_3}{a_1}x_3$

$b_{12.3} = -\frac{a_2}{a_1}$

$x_2 = \frac{K}{a_2} - \frac{a_1}{a_2}x_1 - \frac{a_3x_3}{a_2}$

$$= \frac{-\frac{a_1}{a_2} V(x_1)}{\sqrt{\frac{a_1^2}{a_2^2} V(x_1) \cdot V(x_1)}}$$

$b_{21.3} = -\frac{a_1}{a_2}$

$b_{12.3} \times b_{21.3} = r_{12.3}^2 \Rightarrow r_{12.3} = \pm 1$

$$= \frac{-\frac{a_1}{a_2}}{\left|\frac{a_1}{a_2}\right|} \cdot \frac{V(x_1)}{V(x_1)}$$

$$= \begin{cases} +1 & \text{if } a_1 \,\&\, a_2 \text{ are of opposite sign} \\ -1 & \text{if } a_1 \,\&\, a_2 \text{ are of same sign} \end{cases}$$

c) $x = u_1 + u_2 + \dots + u_n + v_1 + v_2 + \dots + v_s$

$y = u_1 + u_2 + \dots + u_n + \omega_1 + \omega_2 + \dots + \omega_t$

$V(u_i) = V(v_j) = V(\omega_k) = 1$

$Cov(u_i, u_j) = Cov(v_i, v_j) = Cov(\omega_i, \omega_j) = 0$

$Cov(u, v) = Cov(u, \omega) = Cov(v, \omega) = 0$

$$\therefore r_{xy} = \frac{Cov(x,y)}{\sqrt{V(x) \cdot V(y)}} \quad \text{--- (i)}$$

$Cov(x,y) = Cov(u_1 + u_2 + \dots + u_n + v_1 + v_2 + \dots + v_s, u_1 + u_2 + \dots + u_n + \omega_1 + \omega_2 + \dots + \omega_t)$

$= V(u_1) + V(u_2) + \dots + V(u_n) + 0 + \dots$

$= \underbrace{1 + 1 + \dots + 1}_{n \text{ times}}$

$\Rightarrow Cov(x,y) = n$ --- (ii)

$V(x) = V(u_1 + u_2 + \dots + u_n + v_1 + v_2 + \dots + v_s)$

$= V(u_1) + V(u_2) + \dots + V(u_n) + V(v_1) + V(v_2) + \dots + V(v_s)$

$\Rightarrow V(x) = n + s$ --- (iii) $\left[\because V(u_i) = V(v_j) = 1 \text{ and } V(u_i, u_j) = V(v_i, v_j) = V(u, v) = 0\right]$

similarly,

$\Rightarrow V(y) = n + t$ --- (iv)

putting the value of (ii) (iii) and (iv) in (i) we will get $r_{xy} = \frac{r}{\sqrt{(r+s)(r+t)}}$ (proved)

2)

$$R = \left(\begin{array}{c|cccc} 1 & r_{12} & r_{13} & \cdots & r_{1p} \\ \hline r_{12} & 1 & 0 & \cdots & 0 \\ r_{13} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & & \\ r_{1p} & 0 & 0 & \cdots & 1 \end{array}\right)$$

$\because$ independent variables are pairwise uncorrelated.

$\therefore r_{ij} = 0 \quad \forall i \neq j, \; i,j = 2(1)p$

$$= \begin{pmatrix} 1 & A \\ B & I_{p-1 \times p-1} \end{pmatrix}$$

$$\therefore R_{1,1} = (-1)^{1+1} \begin{vmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \end{vmatrix}_{p-1 \times p-1} = 1$$

$$\therefore |R| = |I| \, |1 - AI^{-1}B|$$

$$= |I| \, |1 - AB|$$

$$= 1 \, |1 - AB|$$

$$= 1 - \sum_{i=2}^{p} r_{1i}^2$$

$$\therefore \frac{|R|}{R_{1,1}} = 1 - \sum_{i=2}^{p} r_{1i}^2$$

$$\therefore \left(1 - \frac{|R|}{R_{1,1}}\right) = \sum_{i=2}^{p} r_{1i}^2$$

$$\therefore r_{1.23\ldots p}^2 = \sum_{i=2}^{p} r_{1i}^2$$

$$\Rightarrow r_{1.23\ldots p} = \left(\sum_{i=2}^{p} r_{1i}^2\right)^{1/2}$$

(Ans)

Q. ~~Suppose~~

$$R = \begin{pmatrix} 1 & r & r & \cdots & r \\ r & 1 & r & \cdots & r \\ \vdots & & & & \vdots \\ r & r & r & \cdots & 1 \end{pmatrix}$$

Show that, if r be negative then

$$r \geq -\frac{1}{p-1}$$

$\Rightarrow$

$$r_{12.34\ldots p} = \frac{-R_{2,1}}{\sqrt{R_{1,1} \times R_{2,2}}}$$

$$R = \begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix}$$

$$R_{2,1} = (-1)^3 [r - r^2]$$

$$R_{1,1} = 1 - r^2$$

$$R_{2,2} = 1 - r^2$$

$$r_{12.3} = \frac{r - r^2}{1 - r^2} = \frac{r(1-r)}{(1-r)(1+r)} = \frac{r}{1+r}$$

$$\left\langle r_{12.34\ldots p} = \frac{r}{1+(p-2)r} \right\rangle$$

$$R = \begin{pmatrix} 1 & r & r & r \\ r & 1 & r & r \\ r & r & 1 & r \\ r & r & r & 1 \end{pmatrix}$$

$$r_{12.34} = \frac{r[1-r]^2}{1+2r^3-3r^2} = \frac{r\cancel{[1-r]^2}}{[1+2r]\cancel{[1-r]^2}} = \frac{r}{1+2r}$$

$$((1-2r+r^2)(1+2r)$$
$$1 - \cancel{2r} + r^2$$
$$+\cancel{2r} - 4r^2 + 2r^3$$

$$R_{2,1} = (-1)^3 \begin{vmatrix} r & r & r \\ r & 1 & r \\ r & r & 1 \end{vmatrix}$$

$$= -1\left[r(1-r^2) - r(r-r^2) + r(r^2-r)\right]$$

$$= -1\left[r - \cancel{r^3} - r^2 + \cancel{r^3} + r^3 - r^2\right]$$

$$= -1\left[r(1-r) - r^2[1-r]\right]$$

$$= -[1-r][r-r^2]$$

$$= -r[1-r]^2$$

$$R_{2,2} = \begin{vmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{vmatrix}$$

$$= 1(1-r^2) - r(r-r^2) + r(r^2-r)$$

$$= 1 - r^2 - r^2 + r^3 + r^3 - r^2$$

$$= 1 + 2r^3 - 3r^2$$

$$R_{1,1} = 1(1-r^2) - r(r-r^2) + r(r^2-r)$$

$$= 1 - r^2 - r^2 + r^3 + r^3 - r^2$$

$$= 1 - 3r^2 + 2r^3$$

In general we can write

$$r_{12.34\ldots p} = \frac{r}{1+(p-2)r}$$

$$R = \begin{pmatrix} a & b & b & \cdots & b \\ b & a & b & \cdots & b \\ \vdots & & & & \\ b & b & b & \cdots & a \end{pmatrix}$$

$$|A| = (a-b)^{n-1}(a+(n-1)b)$$

$$R_{2,1} = \left|\begin{array}{c|cccc} r & r & r & \cdots & r \\ \hline r & 1 & r & \cdots & r \\ r & r & 1 & \cdots & r \\ \vdots & & & & \\ r & r & r & \cdots & 1 \end{array}\right|$$

A, B, C, D

$$|R| = |A|\,|D - CA^{-1}B|$$

$$\therefore R_{2,1} = -r \left| \begin{pmatrix} 1 & r & \cdots & r \\ r & 1 & \cdots & r \\ \vdots & & & \\ r & r & \cdots & 1 \end{pmatrix} - \begin{pmatrix} r \\ r \\ \vdots \\ r \end{pmatrix} \left(\frac{1}{r}\right) \begin{pmatrix} r & r & \cdots & r \end{pmatrix} \right|$$

$$= -r \left| \begin{pmatrix} 1 & r & \cdots & r \\ r & 1 & \cdots & r \\ \vdots & & & \\ r & r & \cdots & 1 \end{pmatrix} - \begin{pmatrix} r & r & \cdots & r \\ r & r & \cdots & r \\ \vdots & & & \\ r & r & \cdots & r \end{pmatrix} \right|$$

$$= -r \begin{vmatrix} 1-r & 0 & \cdots & 0 \\ 0 & 1-r & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1-r \end{vmatrix}$$

$$= -r(1-r)^{p-2}$$

$$R_{1,1} = (1-r)^{p-2}\left[1+(p-2)r\right]$$

$$\therefore r_{12.34\ldots p} = \frac{+r(1-r)^{p-2}}{(1-r)^{p-2}\left[1+(p-2)r\right]} = \frac{r}{1+(p-2)r}$$

$$\therefore \; -1 \le r_{12.3\ldots p} \le 0$$

$$\Rightarrow -1 \le \frac{r}{1+(p-2)r}$$

$$\Rightarrow -1 - pr + 2r \le r$$

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

$$\Rightarrow r \ge -\frac{1}{p-1}$$

---

the use of the following statistic,

$$\left\langle D_n = \sup_x \left| F_n(x) - F_X(x) \right| \right\rangle$$

Under H_0, $\boxed{D_n = \sup_x \left| F_n(x) - F_0(x) \right|}$

$$= \max_x \left(\left| F_n(x) - F_0(x) \right|, \left| F_n(x-\epsilon) - F_0(x) \right| \right)$$

where, ϵ is a very small positive quantity.

D_n is called one sample Kolmogorov-Smirnov test statistic.

The directional derivatives,

$D_n^+ = \sup_x \left[F_n(x) - F_X(x) \right]$ and

$D_n^- = \sup_x \left[F_X(x) - F_n(x) \right]$ are called

one-sided Kolmogorov-Smirnov test statistics.