Example

Suppose a farmer wants to know measure the production of a certain crop say, wheat when he has a document on the records of soil fertility avg rainfall of that place and the average humidity of that place / location. As we can understand that these 3 variables had a direct influence on the production of wheat. We think on a linear f^n of 3 variables explaining the variable, yield (prod.)

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

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

Remember, the relationship might be non-linear and the constants might be not be multiplicative f^n. But if both happens (linear in constant and multi) everytime we say that there is multiple regression existed in between (yield and the vars.).

Mathematical development

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

$$x_1 = a + b_2 x_2 + b_3 x_3 + \cdots + b_p x_p \quad \text{—} \quad (1)$$

Clearly from the concept of bivariate regression, x_1 is the dependent variable, while $x_2, \ldots, x_p$ $(p-1)$ are the independent variables and $a, b_2, b_3, \ldots, b_p$ are the constants to be determined.

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

If we have a data frame of size n,
where ith data point is $(x_{i1}, x_{i2}, \cdots, x_{ip})$; $i = 1(1)n$

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

p = no. of variables

point

multivariate data matrix

Where $\bar{x}_j$ = mean for j^{th} variable

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \quad , j = 1(1)p$$

(s_{jj})

s_j^2 = variance of j^{th} variable

$$= \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$$

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

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

Our next objective is to estimate the coefficients.

Using Yule's dot notation we can write ① in

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

Where x_1 = observed data point for dependent variable.

and $x_{1.23..p}$ = Predicted data point (from 1) of the dependent variable

Then, $x_{1.23..p} = a + b_2 x_2 + \cdots + b_p x_p$

Moreover, the unexplained part of the dependent variable x_1 is denoted by $e_{1.23..p}$ is error.

Thus $e_{1.23...p} = x_1 - x_{1.23...p}$

to estimate the constants, we use least square principle which minimizes the sum of squares of errors. (residual point)

$$\sum_{\alpha=1}^{n} e^2_{1.23...p\alpha} \quad \Bigg| \quad e_{1.23..p\alpha} = x_{\alpha 1} - a - b_2 x_{\alpha 2} - \cdots b_p x_{\alpha p}$$

$$\Rightarrow L = \sum_{\alpha=1}^{n} (x_{\alpha 1} - a - b_2 x_{\alpha 2} - \cdots b_p x_{\alpha p})^2$$

The normal equations are,

$$\frac{\partial L}{\partial a} = 0 \quad \Rightarrow -2 \sum_{\alpha=1}^{n} (x_{\alpha 1} - a - b_2 x_{\alpha 2} - \cdots b_p x_{\alpha p})$$

divide both side by n

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

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

value input

$$\frac{\partial L}{\partial b_2} = 0$$

$$\Rightarrow (-2) \sum_{\alpha=1}^{n} x_{\alpha 2} (x_{\alpha 1} - a - b_2 x_{\alpha 2} - \cdots - b_p x_{\alpha p}) = 0$$

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

$$- b_2 \sum_{\alpha=1}^{n} x^2_{\alpha 2} - \cdots - b_p \sum_{\alpha=1}^{n} x_{\alpha 2} x_{\alpha p} = 0$$

Divide both side by n

$$\frac{1}{n}\sum_{\alpha=1}^{n} x_{\alpha 2} x_{\alpha 1} - \bar{x}_1 \frac{1}{n}\sum_{\alpha=1}^{n} x_{\alpha 2} + b_2 \bar{x}_2 \frac{1}{n}\sum_{\alpha=1}^{n} x_{\alpha 2} + \cdots +$$

$$b_p \bar{x}_p \frac{1}{n}\sum_{\alpha=1}^{n} x_{\alpha 2} - \frac{b_2}{n}\sum_{\alpha=1}^{n} x_{\alpha 2} x_{\alpha 2} - \cdots - \frac{b_p}{n}\sum_{\alpha=1}^{n} x_{\alpha 2} x_{\alpha p}$$

$$= 0$$

$$\Rightarrow s_{21} - \bar{x}_1 \cdot \bar{x}_2 + b_2 \bar{x}_2^2 + \cdots + b_p \bar{x}_p \bar{x}_2 - b_2 s_{22} - \cdots - b_p s_{2p} = 0$$

$$\Rightarrow s_{21} = b_2 s_{22} + \cdots + b_p s_{2p} + \bar{x}_1 \bar{x}_2 - b_2 \bar{x}_2^2 - \cdots - b_p \bar{x}_p \cdot \bar{x}_2$$

$$= b_2 s_{22} + \cdots + b_p s_{2p} + \bar{x}_2 \cdot \hat{a} \quad \text{——} \quad (2)$$

Linear in variable is sometimes confused with linear in parameters but as regression theory evolves from linear model, we always look into linearity in parameters or constants.

$y = b_1 + b_2 X + e \rightarrow$ linear in constant
linear in variable

$e^y = b_1 + b_2 X + \epsilon \rightarrow$ non-linear in variable
linear in parameter

$y = b_1 + b_2 X^2 + \epsilon \rightarrow$ linear in constant
(polynomial regression)

$y = \frac{1}{b_1} + e^{b_2} X + \epsilon \rightarrow$ non-linear in constant
linear in variable.

Any non-linear in constant model can be transformed into a linear regression format with suitable transformation in old constant.

$\boxed{P_{ij} = r_{ij} s_i s_j}$

$$r = \frac{\mathrm{cov}(x_1, x_2)}{\sqrt{V(x_1) V(x_2)}}$$

8.
$$s_{21} = \text{cov}(x_2, x_1)$$

$$r_{21} = \frac{\text{cov}(x_2, x_1)}{\sqrt{V(x_1) \cdot V(x_2)}}$$

$$\Rightarrow r_{21} \cdot s_{x_1} \cdot s_2 = s_{21}$$

where, $L = \sum_{\alpha=1}^{n} e_{1.23\ldots p\alpha}^2$, $e_{1.23\ldots p_1} = x_{11} - a - b_2 x_{12} \ldots \ldots - b_p x_{1p}$

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

The normal eqn. are,

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

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

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

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

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

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

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

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

$\Rightarrow s_{12} = b_2 s_{22} + b_3 s_{23} + \cdots + b_p s_{2p}$

similarly,

$s_{13} = b_2 s_{23} + b_3 s_{33} + \cdots + b_p s_{3p}$

$\vdots$

$s_{1p} = b_2 s_{2p} + b_3 s_{3p} + \cdots + b_p s_{pp}$

System of non-homogeneous linear equ^n.

Remark : 'Linear' in regression means linear in parameters (constant).

- Any non-linear in constant model can be transformed into linear regression with suitable transformation in old constant.

- Linear regression is considered as fundamental model :

$$
\left.\begin{aligned}
a &= \bar{x}_1 - b_2 \bar{x}_2 - \cdots - b_p \bar{x}_p \\
s_{12} &= b_2 s_{22} + b_3 s_{23} + \cdots + b_p s_{2p} \\
s_{13} &= b_2 s_{23} + b_3 s_{33} + \cdots + b_p s_{3p} \\
&\vdots \\
s_{1p} &= b_2 s_{2p} + b_3 s_{3p} + \cdots + b_p s_{pp}
\end{aligned}\right\} \rightarrow (2)
$$

$[s_{ij} = Cov(x_i, x_j)$
$i = 1(1)p$
$s_{ii} = V(x_i)$
$= s_i^2]$

Define sample variance covariance matrix as,

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

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

Note that:

S is a symmetric matrix.

(2) can be rewritten in terms of system of non homogeneous eq^{n}.

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

By cramer's rule

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

In general,

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

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

$$\overline{\begin{vmatrix} s_{22} & s_{23} & \cdots & s_{2p} \\ s_{23} & s_{33} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots \\ s_{2p} & s_{3p} & \cdots & s_{pp} \end{vmatrix}}$$

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

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

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

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

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

Another representation of b_j comes from the correlation matrix R

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

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

as $r_{11} = r_{22} = \cdots r_{pp} = 1$

$$\therefore b_j = \cancel{(-1)^{j+2}\, s_1 s_1 \cdot s_2 s_3 \cdots s_{j-1} s_{j+1} \cdots s_p^2} \quad \cancel{r_{12} s_1 s_2 \;\; r_{22} s_2^2 \cdots}$$

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

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

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

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

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

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

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

Finally the multiple regression line

$$x_1 = a + b_2 x_2 + b_3 x_3 + \cdots + b_p x_p$$

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

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

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

Multiple regression line of x_1 on $x_2, x_3, \ldots, x_p$

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

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

In multiple regression line b_j & are called partial regression coefficient and mathematically denoted by

$\boxed{b_{j.23\ldots j-1, j+1 \ldots p}}$ 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 <u>measures</u> 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.

Remark

$(-\infty < b_j < \infty)$

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

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

< If x_4 dec by 1 unit then x_1 will inc. by 4 unit

If x_3 inc. by 1 " " x_1 " inc by 3.9 unit >

- Predicted value of x_1 at αth observation

$$x_{1\alpha} = \bar{x}_1 - \frac{R_{1,2}}{R_{1,1}}\frac{s_1}{s_2}(x_{2\alpha} - \bar{x}_2) - \frac{R_{1,3}}{R_{1,1}}\frac{s_1}{s_3}(x_{3\alpha} - \bar{x}_3)$$

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

- Multiple Correlation coefficient :

Def >> To measure the degree of association between the dependent variable x_1 and the set of independent variables $(x_2, x_3, \ldots, x_p)$ and extension of Pearson Correlation coefficient is introduced and that is called multiple correlation coefficient.

(How effective (meaningful) the linear regression is explaining x_1 through $x_2, x_3, \ldots, x_p$)

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

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

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

Derivation

Least square multiple regression line is

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

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

The model is

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

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

$\bar{x}_{1.2}$ -

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

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

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

Also, $\text{Cov}(x_1, x_{1.23\cdots p})$

$= \text{Cov}(x_{1.23\cdots p} + e_{1.23\cdots p};\ x_{1.23\cdots p})$

$= V(x_{1.23\cdots p}) + \text{Cov}(e_{1.23\cdots p}, x_{1.23\cdots p})$

$\text{Cov}(x+y, z)$
$= \text{Cov}(x, z)$
$+ \text{Cov}(y, z)$

Now, $\text{Cov}(e_{1.23\cdots p}, x_{1.23\cdots p})$

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

$$= \frac{1}{n}\sum_{\alpha} e_{1.23\cdots p\alpha}(a + b_2 x_{2\alpha} + \cdots + b_p x_{p\alpha} - \bar{x}_1) \quad [\because \bar{e}_{1.23\cdots p} = 0$$
(1st normal eqn$^{\underline{n}}$)

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

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

$= 0$ (from normal equations)

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

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

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

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

$$\boxed{\begin{aligned} x_1 &= x_{1.23\ldots p} + e_{1.23\ldots p} \\ V(x_1) &= V(x_{1.23\ldots p}) + V(e_{1.23\ldots p}) + 0 \\ \therefore V(x_1) &> V(x_{1.23\ldots p}) \end{aligned}}$$

In other words

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

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

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

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

$$= \frac{1}{n}\sum_\alpha (x_{1\alpha} - \bar{x}_1)\Big(\cancel{\bar{x}_1} - \frac{R_{1,2}}{R_{1,1}}\frac{s_1}{s_2}(x_{2\alpha} - \bar{x}_2) - \frac{R_{1,3}}{R_{1,1}}\frac{s_1}{s_3}(x_{3\alpha} - \bar{x}_3)$$

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

$$= -\frac{1}{n}\sum_\alpha (x_{1\alpha} - \bar{x}_1)(x_{2\alpha} - \bar{x}_2)\frac{R_{1,2}}{R_{1,1}}\frac{s_1}{s_2} - \frac{1}{n}\sum_\alpha (x_{1\alpha} - \bar{x}_1)(x_{3\alpha} - \bar{x}_3)$$

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

$$\frac{R_{1,p}}{R_{1,1}} \cdot \frac{s_1}{s_p}.$$

$$= -\left(r_{12}\, s_1 \cancel{s_2} \frac{R_{1,2}}{R_{1,1}} \frac{s_1}{\cancel{s_2}} + r_{13}\, s_1 \cancel{s_3} \frac{R_{1,3}}{R_{1,1}} \frac{s_1}{\cancel{s_3}} + \cdots \right.$$

$$\left. + r_{1p}\, s_1 \cancel{s_p} \frac{R_{1,p}}{R_{1,1}} \cdot \frac{s_1}{\cancel{s_p}} \right)$$

$\left[\because \frac{1}{n} \sum_{\alpha} (x_{1\alpha} - \bar{x}_1)(x_{2\alpha} - \bar{x}_2) = r_{12} s_1 \right.$

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

$$= -\frac{s_1^2}{R_{1,1}} \left(r_{11} R_{1,1} + r_{12} R_{1,2} + r_{13} R_{1,3} + \cdots + r_{1p} R_{1,p} - r_{11} R_{1,1} \right)$$

$$= -\frac{s_1^2}{R_{1,1}} \left(|R| - R_{1,1} \right) \qquad \left[\because r_{11} = 1 \right] = Corr(x_1, x_1)$$

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

$$\therefore r_{1.23\cdots p} = \sqrt{\frac{Cov(x_1, x_{1.23\cdots p})}{V(x_1)}}$$

$$= \sqrt{\frac{s_1^2 \left(1 - \frac{|R_1|}{R_{1,1}}\right)}{s_1^2}} = \left(1 - \frac{|R|}{R_{1,1}} \right)^{1/2}$$

Ⓢ, $r_{2.13\cdots p} = \left(1 - \frac{|R|}{R_{2,2}} \right)^{1/2}$ $\left[|R| < 1 \right.$ ↳ positive definite matrix $\left. \right]$

Prop 1 : $r_{1.23\cdots p} = 1$

$\Rightarrow V(x_1) = V(x_{1\cdot 23\cdots p})$.

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

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

$\Rightarrow$ Error is constant

$\Rightarrow$ The prediction formula is perfect.
(linear)

■ $r_{1.23\cdots p} = 0$

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

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

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

Remark

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

$r^2_{1.23\cdots p}$ is called coefficient of determination.
and generally denoted by R^2.

Coefficient of determination measures the variance of dependent variable explained by the prediction or regression equation, $x_1 = a + b_2 x_2 + \cdots + b_p x_p$

R^2 is generally denoted ~~by~~ in percentage format.

For example, $r^2_{1.23\ldots p} = 0.35$

$$\Rightarrow R^2 = 35\%$$

35 % of the total variability of dependent variable (x_1) is explained (determined) by the multiple regression line

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

65 % of the variability of x_1 remains unexpla

Remark

With the increased of independent variables in the model the value of R^2 also increases.

$$0 < R^2 < 1$$

$$x_1 = a + b_2 x_2$$

$$R_1^2$$

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

$$R_2^2 > R_1^2$$

$$\vdots$$

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

$$R_p^2 > R_2^2 > R_1^2$$

$$\boxed{\therefore R_p^2 > R_{p-1}^2 > R_{p-2}^2 > \cdots > R_2^2 > R_1^2}$$

Now to check on the value of R^2, generally we consider upto 95% of explainability.

- 60% of variable might have check whether they are present on your model or not.

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 ~~variable~~ independent variable in the model.

Remark: ~~Standard error of estimate~~.

(residual variance)

$$V(e_{1.23\cdots p})$$

$$= V(x_1) - V(x_{1.23\cdots p})$$

$$= s_1^2 - \mathrm{Cov}(x_{1.23\cdots p}, x_1)$$

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

$$= s_1^2 \frac{|R|}{R_{1,1}}$$

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

Exercise 1: Multiple Regression, Multiple and Partial correlation

Prob ① The following data relates to the ~~person~~ performance ~~of~~ of the students in the final examination (x_1), 1st term examition (x_2), 2nd term examination (x_3), 3rd term examination (x_4)

(a) Fit a least square regression equation 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 $b_2 = b_3 = b_4$ and (iii) $b_2 + b_3 + b_4 = 1$

(Restricted domain regression)

$\boxed{\text{Hint}}$ (i) $x_1 = a + b_2(x_2 + x_3 + x_4)$

$= a + b_2 z$

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

(b) Obtain the multiple corre. coeff. of x_1 on x_2, x_3, x_4

Student Roll 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	53.4	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	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	47.5	50.7
14	69.1	82.5	47.5	65.6
15	60.5	71.5	51.0	46.2

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

$$r_{12} = \frac{Cov(x_1, x_2)}{\sqrt{v(x_1)\, v(x_2)}}$$

① Mean

$\bar{x_1} = 56.806$

$\bar{x_2} = 68.513$

$\bar{x_3} = 75.48$

$\bar{x_4} = 55.23$

$cov(x_1, x_2)$

$= 3927.696 - (56.806 \times 68.513)$

$= 35.746$

② Var

$var(x_1) = \cancel{7.5772}6$

$= 57.335$

$s_2 = 6.74$ $\quad var(x_2) = 45.433$

$s_3 = 5.913$ $\quad var(x_3) = 34.965$

$s_4 = 8.372$ $\quad var(x_4) = 70.09$

$\cancel{r_{12} = \frac{35.746}{\;}}$

$r_{12} = 0.7$

$r_{13} = 0.339$

$r_{14} = -0.114$

$r_{23} = 0.579$

$r_{24} = 0.342$

$r_{34} = -0.52$

$cov(x_1, x_3)$

$= 15.1808$

$Cov(x_1, x_4)$

$= -7.0644$

$Cov(x_2, x_3) = 23.1125$

$Cov(x_2, x_4)$

$= 19.296$

$Cov(x_3, x_4)$

$= -25.744$

$$R = \begin{pmatrix} 1 & 0.7 & 0.339 & -0.114 \\ 0.7 & 1 & 0.579 & 0.342 \\ 0.339 & 0.579 & 1 & -0.52 \\ -0.114 & 0.342 & -0.52 & 1 \end{pmatrix}$$

$$\cancel{R_1 \; x_1 = \bar{x}_1 - \frac{R_{1,2}}{R_{1,1}} - \frac{R_{1,3}}{R_{1,1}} \frac{s_1}{s_3} (x_3}$$

$$R_{1,3} = (-1)^{1+3} \begin{pmatrix} 0.7 & 1 & 0.342 \\ 0.339 & 0.579 & -0.52 \\ -0.114 & 0.342 & 1 \end{pmatrix}$$

$$\cancel{= 0.7}(-0.529 - 0.279 - 0.062$$

$$= 0.188 \qquad R_{1,2} = 0.1823$$

$$R_{1,1} = 0.211$$

$$R_{1,4} = (-1)^{1+4} \begin{pmatrix} 0.7 & 1 & 0.579 \\ 0.339 & 0.579 & 1 \\ -0.114 & 0.342 & -0.52 \end{pmatrix}$$

$$= -[-0.45 + 0.062 + 0.105] = 0.283$$

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

$$- \frac{R_{1,4}}{R_{1,1}} \cdot \frac{s_1}{s_4} (x_4 - \bar{x}_4)$$

$$= 56.806 - \frac{0.182}{0.211} \times \frac{7.572}{6.74} (x_2 - 68.513)$$

$$- \frac{0.188}{0.211} \times \frac{7.572}{5.913} (x_3 - 75.48) - \frac{0.283}{0.211} \times \frac{7.572}{8.372} (x_4 - 55$$

$$= 56.806 - 0.97 x_2 + 66.392 - 1.141 x_3 + 86.121 - 1.213 x_4 + 66.9$$

$$\boxed{x_1 = 276.316 - 0.97 x_2 - 1.41 x_3 - 1.213 x_4}.$$

$$\boxed{x_1 = 17.724 + 0.88 x_2 - 0.218 x_3 - 0.0}$$

multiple corr. coeff.

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

$$\cancel{|R| = 1 \times 0.211 + (0.4 \times 0.1823) + (0.339 \times 0.188) + (0.114 \times}$$

$$\cancel{= 0.37008}$$

$$\cancel{r_{1.23\ldots p} = \left(1 - \frac{0.37008}{0.211}\right)^{1/2}}$$

$$|R| = 0.2230337 \quad , \quad R_{1,1} = \cancel{0.211}\ 0.4495$$

$$r_{1.23\ldots p} = 0.709$$

$$R = \begin{pmatrix} 1 & 0.699 & 0.339 & -0.117 \\ 0.699 & 1 & 0.586 & -0.143 \\ 0.339 & 0.586 & 1 & -0.525 \\ -0.117 & -0.143 & -0.525 & 1 \end{pmatrix}$$

$R_{1,1} = 0.4485$, $R_{1,2} = -0.3524$

$R_{1,3} = 0.0767$

$R_{1,4} = 0.0423$

$$b_3 = -\frac{0.0767}{0.4485} \times \frac{7.572}{5.913} = -0.2189$$

$$b_4 = -\frac{0.0423}{0.4485} \times \frac{7.572}{8.372} = -0.0853$$

Q.1. For three variables x_1, x_2, x_3 prove that

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

Ans :- Model $x_1 = a + b_2x_2 + b_3x_3$

$$= \bar{x}_1 - \frac{R_{1,2}}{R_{1,1}} \frac{s_1}{s_2} (x_2 - \bar{x}_2) - \frac{R_{1,3}}{R_{1,1}} \frac{s_1}{s_3} (x_3$$

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

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

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

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

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

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

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

$$r^2_{1.23} \leq 1$$

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

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

Prob 2

Suppose for a regression model, $x_1, x_2, \ldots, x_p$ correlation of x_i, x_j, $Corr(x_i, x_j) = r \ \forall (i, j)$

$i = 1(1)p \quad i \neq j$
$j = 1(1)p$

Find $r_{1.23 \cdot\cdot p}$

Ans:

$$R = \begin{pmatrix} 1 & r & r & \cdots & r \\ r & 1 & r & \cdots & r \\ \vdots & & & & \vdots \\ r & r & r & \cdots & 1 \end{pmatrix}_{p \times p}$$ Intraclass pattern matrix.

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

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

$$= (a-b)^{n-1}(a + \overline{(n-1)}b)$$

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

since $|R| > 0$

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

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

$$r_{1\cdot 23\cdots p} = \left(1 - \frac{(1-r)^{p-1}(1+\overline{p-1}\cdot r)}{(1-r)^{p-2}(1+\overline{p-2}\cdot r)}\right)^{1/2}$$

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

$$= \left(\frac{(1-r)^{p-2}\left[1+\overline{p-2}\cdot r - (1-r)(1+\overline{p-1}\cdot r)\right]}{(1-r)^{p-2}(1+\overline{p-2}\cdot r)}\right)^{1/2}$$

$$\cancel{(p-2)r} -$$

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

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

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

Prob 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 $r_{1.23\ldots p}$

$$R_{p \times p} = \left(\begin{array}{c|cccc} 1 & r & r & \ldots & r \\ \hline r & 1 & r' & \ldots & r' \\ r & r' & 1 & \ldots & r' \\ \vdots & \vdots & & & \\ r & r' & r' & \ldots & 1 \end{array}\right)_{p \times p}$$

A B C D

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

$$\cancel{|R| = |1|}$$

Partition determinant

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

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

$$= |D|\,|A - BD^{-1}C|$$

$$|R| = |1| \left| \begin{pmatrix} 1 & r' & \ldots & r' \\ r' & 1 & \ldots & r' \\ \vdots & & & \\ r' & r' & \ldots & 1 \end{pmatrix} - r \cdot 1 \cdot r' \right|$$

$$\boxed{\begin{array}{l} r' = (r \; r \ldots r) \\ r = \begin{pmatrix} r \\ r \\ \vdots \\ r \end{pmatrix} \end{array}}$$

$$= \left| \begin{pmatrix} 1 & r' & \ldots & r' \\ r' & 1 & \ldots & r' \\ \vdots & & & \\ r' & r' & \ldots & 1 \end{pmatrix} - \begin{pmatrix} r^2 & r^2 & \ldots & r^2 \\ r^2 & r^2 & \ldots & r^2 \\ \vdots & & & \\ r^2 & r^2 & \ldots & r^2 \end{pmatrix} \right|$$

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

$(P-1) \times (P$

$a = 1-r^2$

$b = r'-r^2$

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

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

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

$$r_{1.23\cdots P} = \left(1-\frac{1+(P-2)r'-(P-1)r^2}{1+(P-2)r'}\right)^{1/2}$$

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

Partial Correlation coefficient :

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

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

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

$r_{1 \cdot 23 \ldots p}$.

$$x_1 \longrightarrow \left\{ \begin{matrix} x_2 \\ x_3 \\ \vdots \\ x_p \end{matrix} \right\}$$

Def >> Partial correlation is the measure of association between two variables in a multiple regression set up when influence of the other variables is eliminated.

For a multiple regression setup $(x_1, x_2, \ldots, x_p)$ partial correlation between x_1 and x_2 is denoted by $r_{12 \cdot 34 \ldots p}$

Mathematically:

$$r_{12 \cdot 34 \ldots p} = \mathrm{Corr}\left(e_{1 \cdot 34 \ldots p}, e_{2 \cdot 34 \ldots p}\right)$$

$$= \frac{\mathrm{Cov}\left(e_{1 \cdot 34 \ldots p}, e_{2 \cdot 34 \ldots p}\right)}{\sqrt{V(e_{1 \cdot 34 \ldots p})\, V(e_{2 \cdot 34 \ldots p})}}$$

Why $r_{12.34\cdots p}$ is considered as a partial correlation

→ Think about the model

$$x_1 = a + b_3 x_3 + \cdots + b_p x_p \quad (x_1, x_3, \ldots, x_p)$$

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

If we apply least square regression line,

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

$$x_1 = \underbrace{x_{1.34\cdots p}}_{\text{explained part of } x_1 \text{ by } x_3, x_4, \ldots, x_p} + \underbrace{e_{1.34\cdots p}}_{\text{unexplained part of } x_1 \text{ by } x_3, x_4, \ldots}$$ (Model for the first eqn.)

$$x_2 = \underbrace{x_{2.34\cdots p}}_{\text{explained part of } x_2 \text{ by } x_3, x_4, \ldots, x_p} + \underbrace{e_{2.34\cdots p}}_{\text{unexplained part of } x_2 \text{ by } x_3, x_4, \ldots}$$ (Model for 2nd

Derivation of $r_{12.34\cdots p}$:

Let the data structure.

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

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

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

$$- \frac{R_{1,p}}{R_{1,1}} \frac{s_1}{s_p} (x_p - \bar{x}_p)$$

Regression model is

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

Let us think on two submodels

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

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

Borrowing the idea of multiple regression model

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

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

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

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

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

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

$$- \cdots - \frac{R^{(2)}_{1,p}}{R^{(2)}_{1,1}} \frac{s_1}{s_p} \cdot (x_p - \bar{x}_p)$$

$$x_{2.34\ldots p} = \bar{x}_2 - \frac{R^{(1)}_{2,3}}{R^{(1)}_{2,2}} \frac{s_2}{s_3} (x_3 - \bar{x}_3)$$

$$- \cdots - \frac{R^{(1)}_{2,p}}{R^{(1)}_{2,2}} \frac{s_2}{s_p} (x_p - \bar{x}_p)$$

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 $R^{(k)}$, $k = 1,$

Also we know,

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

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

Again,

$$Cov(x_1, x_{1.34\cdots p})$$
$$= \frac{1}{n}\sum_{\alpha=1}^{n}(x_{1\alpha} - \bar{x}_1)(x_{1.34\cdots p\alpha} - \bar{x}_1)$$

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

$$\boxed{\bar{x}_{1.23\cdots p} = \bar{x}_1}$$

$$\boxed{\frac{1}{n}\sum_{\alpha}(x_{1\alpha} - \bar{x}_1)(x_{3\alpha} - \bar{x}_3) = r_{13}s_1s_3}$$

$$= -\frac{R^{(2)}_{1,3}}{R^{(2)}_{1,1}} r_{13}s_1s_3\frac{s_1}{s_3} - \frac{R^{(2)}_{1,4}}{R^{(2)}_{1,1}} r_{14}s_1s_4\frac{s_1}{s_4} - \cdots - \frac{R^{(2)}_{1,p}}{R^{(2)}_{1,1}}\frac{s_1}{s_p} r_{1p}s_1s_p$$

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

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

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

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

$$= V(x_{1.34\cdots p})$$

Similarly

$$V(x_{2.34\cdots p}) = s_2^2\left(1 - \frac{|R^{(1)}|}{R^{(1)}_{2,2}}\right)$$

$$\therefore V(e_{1.34\cdots p})$$

$$= V(x_1) - V(x_{1.34\cdots p})$$

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

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

$$\boxed{\begin{aligned} &\big[x_1 = x_{1.34\cdots p} + e_{1.34\cdots p} \\ &V(x_1) = V(x_{1.34\cdots p}) + V(e_{1.34\cdots p})\big] \end{aligned}}$$

$$\therefore V(e_{2.34\cdots p}) = V(x_2) - V(x_{2.34\cdots p})$$

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

Prob 1

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

✓ Unrestricted multiple regression model

✓ multiple corr. coeff

Restricted regression model

(i) $b_2 = b_3 = b_4$

$$x_1 = a + b_2 x_2 + b_2 x_3 + b_2 x_4$$

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

$$x_1 = a + b_2 z$$

$$a = \bar{x}_1 - b_2 \bar{z}$$

$$b_2 = \frac{Cov(x_1, z)}{V(x_1)}$$

	x_1	$z = x_2 + x_3 + x_4$
		199
1	58.6	208.7
2	55.	213.2
3		203
4		185.4
5		192.3
6		189.9
7		197.9
8		212.9
9		191.5
10		224.7
11		187.1
12		200.7
13		199.1
14		183

$\bar{x}_1 = 56.806$, $V(x_1) = 57.335$

$\bar{z} = 199.223$

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

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

$= 43.8626$

$b_2 = 0.765$

$a = -95.599$

$$x_1 = -95.599 + 0.765 z$$

(iii) $b_2 + b_3 + b_4 = 1$

$$\Rightarrow b_2 = 1 - b_3 - b_4$$

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

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

$\cancel{x_1 = a + x_2 + b_3 z_1 + b_4 z_2}$ $x_1 = a + x_2(1-b_3-b_4) + b_3 x_3$

$\cancel{x_1 = \bar{x}_1}$ $a = \bar{x}_1 - \bar{x}_2 - b_3 \bar{z}_1 - b_4 \bar{z}_2$

$$= \bar{x}_1 - \bar{x}_2 - b_3(\bar{x}_3 - \bar{x}_2) - b_4(\bar{x}_4 - \bar{x}_2)$$

$$a = -11.707 - b_3(6.967) - b_4(-13.283)$$

$$a = -11.707 - 6.967 b_3 + 13.283 b_4$$

$$b_3 = -\frac{R_{1,3}}{R_{1,1}} \cdot \frac{\sqrt{V(x_1)}}{\sqrt{V(z_1)}}$$

$\text{Corr}(x_1, z_1)$

$$b_4 = -\frac{R_{1,4}}{R_{1,1}} \frac{\sqrt{V(x_1)}}{\sqrt{V(z_2)}}$$

$= \text{Corr}(x_1, x_3$

$= r_{13} - r_{12}$

$\text{Corr}(x_2, z_1)$

$= \text{Corr}(x_2, x_3 - x_2)$

$= r_{23} - 1$

$\text{Corr}(x_1, z_2)$

$= r_{14} - r_{12}$

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

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

$$b' = 1 - b_3 - b_4$$

$$b_3 = -\frac{R_{1,3}}{R_{1,1}} \frac{s_1}{s_3} = -0.2189$$

$$b_4 = -\frac{R_{1,4}}{R_{1,1}} \frac{s_1}{s_4} = -0.0853$$

x_4

$$b' = 1.3042$$

$$a = 56.806 - (1.3042 \times 68.513) - (-0.0853 \times 55.23)$$
$$- (-0.2189 \times 75.48)$$

$$= -11.3149$$

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

$$= \frac{1}{n} \sum_{\alpha=1}^{m} (e_{1.34\ldots p\alpha} \cdot e_{2.34\ldots p}) \cdot \left[\text{as } \bar{e}_{1.34\ldots p} = 0 \text{ and } \bar{e}_{2.34\ldots p} = 0\right]$$

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

$$= \frac{1}{n} \sum (x_{1\alpha} - \bar{x}_1) e_{2.34\ldots p\alpha}$$

As $\sum_{\alpha} (x_{i\alpha} - \bar{x}_i) e_{2.34\ldots p\alpha} = 0$ for $i = 3, 4, \ldots, p$

$$\mathrm{Cov}(x, y) = \frac{1}{n} \sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$$
$$= \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})$$

$x_2)$

due to normal eq^n related to

$$x_2 = a^* + b_3^* x_3 + \dots + b_p^* x_p$$

$$= \frac{1}{n}\sum (x_{1\alpha} - \bar{x}_1)\cdot e_{2.34\dots p\alpha}$$

$$= \frac{1}{n}\sum_{\alpha=1}^{n} (x_{1\alpha}-\bar{x}_1)\Big(x_{2\alpha}-\bar{x}_2 + \frac{R^{(1)}_{2,3}}{R^{(1)}_{2,2}}\frac{s_2}{s_3}(x_{3\alpha}-\bar{x}_3)+\dots+\frac{R^{(1)}_{2,p}}{R^{(1)}_{2,2}}\frac{s_2}{s_p}(x_{p\alpha}-\bar{x}_p)\Big)$$

$$= \frac{1}{n}\sum_{\alpha=1}^{n} (x_{1\alpha}-\bar{x}_1)(x_{2\alpha}-\bar{x}_2) + \frac{R^{(1)}_{2,3}}{R^{(1)}_{2,2}}\frac{s_2}{s_3}\frac{1}{n}\sum_{\alpha} (x_{1\alpha}-\bar{x}_1)(x_3-\bar{x}_3) + \dots + \frac{R^{(1)}_{2,p}}{R^{(1)}_{2,2}}\frac{s_2}{s_p}\frac{1}{n}\sum_{\alpha}(x_{1\alpha}-\bar{x}_1)(x_{p\alpha}-\bar{x}_p)$$

$$= r_{12}s_1s_2 + \frac{R^{(1)}_{2,3}}{R^{(1)}_{2,2}}\frac{s_2}{s_3}\cdot r_{13}s_1s_3+\dots+\frac{R^{(1)}_{2,p}}{R^{(1)}_{2,2}}\frac{s_2}{s_p}$$

$$= \frac{s_1s_2}{R^{(1)}_{2,2}}\Big[r_{12}R^{(1)}_{2,2} + r_{13}R^{(1)}_{2,3}+\dots+r_{1p}R^{(1)}_{2,p}$$

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

$x_2 = a + b_3 x_3 + \dots$

$= \bar{x}_2 - b_3\bar{x}_3 - \dots - b$

$+ b_3 x_3 + \dots + b$

$\cancel{\sum(x_3-\bar{x}_3)}$

$L = \sum e^2_{2.34\dots p}$

$= \sum(x_{2\alpha}-\bar{x}_2 - b_3(x_3$

$\dots b_p(x$

$\frac{\partial L}{\partial b_3} = (-2)\sum_{\alpha}(x_{3\alpha}$

$e_{2.3\dots p\alpha}$

$$= \frac{s_1 s_2}{R^{(1)}_{2,2}} \begin{vmatrix} r_{12} & r_{13} & \cdots & r_{1p} \\ r_{32} & r_{33} & \cdots & r_{3p} \\ \vdots & & \vdots & \\ r_{p2} & r_{p3} & \cdots & r_{pp} \end{vmatrix}$$

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

$= -\frac{s_1 s_2}{R^{(1)}_{2,2}} R_{2,1}$ where $R_{2,1}$ is the cofactor of r_{21} in R

Finally

$$r_{12\cdot 34\cdots p} = \frac{-\frac{s_1 s_2}{R^{(1)}_{2,2}} R_{2,1}}{\sqrt{s_1^2 \frac{|R^{(2)}|}{R^{(2)}_{1,1}} \cdot s_2^2 \frac{|R^{(1)}|}{R^{(1)}_{2,2}}}}$$

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

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

and $R^{(2)} = R_{2,2}$

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

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

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{14} & \cdots & r_{1p} \\ r_{21} & r_{22} & r_{23} & r_{34} & \cdots & r_{2p} \\ r_{31} & r_{32} & r_{33} & r_{34} & \cdots & r_{3p} \\ \vdots & & & & \vdots & \\ r_{p1} & r_{p2} & r_{p3} & r_{p4} & \cdots & r_{pp} \end{pmatrix}$$

Cofactor of $R_{1,1}$ i.e. r_{11}

Similarly first find $R^{(1)}$ then cofactor of r_{22}

As $\left(R^{(2)}_{1,1}\right) = \left(R^{(1)}_{2,2}\right)$.

- Partial correlation coefficient lies within -1 to 1 as the numerator (covariance) can be positive/negative unlike multiple correlation where the numerator (covariance $(x_1, x_{1.23\cdots p})$) turns to be $V(x_{1.23\cdots p})$.

- square of standard error of estimate (Residual variance)

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

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

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

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

* Relationship between partial regression coeff, partial corr. coeff, residual variances

$$\frac{s^2_{1,34\cdots p}}{s^2_{2.34\cdots p}} = \frac{R_{2,2} \cdot s_1^2}{R_{1,1} \cdot s_2^2}$$

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

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

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

$$= -\frac{R_{1,2}}{\sqrt{R_{1,1} R_{2,2}}} \cdot \frac{s_{1.34\cdots p}}{s_{2.34\cdots p}}$$

$$= r_{12.34\cdots p} \frac{s_{1.34\cdots p}}{s_{2.34\cdots p}}$$

[Here $R_{1,2} = R_{2,1}$ as R is symmetric matr

- Properties of residuals

$(x_1, x_2 \dots x_p)$

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

$= a + b_2 x_2 + \dots + b_p x_p$

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

From 2nd to p th normal equation

$$\sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i) e_{1.23\dots p\alpha} = 0$$

$$\Rightarrow \sum x_{i\alpha} e_{1.23\dots p\alpha} = 0$$

$L = \sum (x_{1\alpha} - a - b_2 x_{2\alpha} \dots - b_p x_{p\alpha})$

$\frac{\partial L}{\partial b_2} = 0;$

$\Rightarrow \sum (x_\alpha - c)(x_{1\alpha} \dots - b_p)$

$c \sum (x_{1\alpha} - a \dots - b$

$c \sum e_{1.23\dots}$

From the partial regression concept

① $x_1 = x_{1.34\dots p} + e_{1.34\dots p}$

$= a^* + b_3^* x_3 + \dots + b_p^* x_p + e_{1.34\dots p}$

② $x_2 = x_{2.34\dots p} + e_{2.34\dots p}$

$= a^\circ + b_3^\circ x_3 + \dots + b_p^\circ x_p + e_{2.34\dots p}$

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

$b_2 = b_{12.34\dots p}$

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

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

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

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

$$= \sum_{\alpha} (x_{1\alpha} - \bar{x}_1) e_{1.23\ldots p\alpha}$$ [Due to normal equation Others terms will be zero from the multiple regression equation]

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

$$= \sum_{\alpha} (x_{1\alpha} - a' - b'_{12.34\ldots p-1} x_{2\alpha} - b'_{13.24\ldots p-1} x_{3\alpha} - \ldots - b'_{1p-1.23\ldots p-2} x_{p-1\alpha}) e_{1.23\ldots p\alpha}$$

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

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

$$\sum_{\alpha=1}^{n} e_{1.23\ldots p-1\alpha} = 0$$

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

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

$$= \sum e_{1.23\ldots p-2\alpha} \cdot e_{1.23\ldots p-1\alpha}$$

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 after dot (secondary subscript) of the first.

$$e_{1.\underbrace{23\ldots p}}$$

(Primary subscript (suffix)) (secondary subscript)

$\boxed{\text{Result}}$ p^{th} order multiple corre... than $(p-1)^{st}$ order multiple correlation.

$$r_{1.23\cdots p} \geq r_{1.23\cdots p-1}$$

In general

$$r_{1.23\cdots p} \geq r_{1.23\cdots p-1} \geq r_{1.23\cdots p-2} \geq \cdots \geq r_{12}$$

Proof >> $V(e_{1.23\cdots p}) = s^2_{1.23\cdots p} = \frac{1}{n}\sum_{\alpha=1}^{n} e^2_{1.23\cdots p\alpha}$

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

$$x_1 = a^0 + b_2^0 x_2 + \cdots + b_{p-1}^0 x_{p-1} + e_{1.23\cdots p-1\alpha}$$

$$a = \bar{x}_{1} - b_2\bar{x}_2 - \cdots - b_{p-1}\bar{x}_{p-1}$$

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

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

$$= \sum_{\alpha=1}^{n} (x_{1\alpha} - \bar{x}_1)\, e_{1.23\cdots p\alpha} \quad \text{---} (\boxtimes)$$

Om... Uspe...

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

$$= \sum_{\alpha=1}^{n} (x_{1\alpha} - \bar{x}_1 + b_2'\bar{x}_2 + \cdots + b_{p-1}'\bar{x}_{p-1} - b_2' x_{2\alpha} - \cdots - b_{p-1}' x_{p-1}) \cdot e_{1.23\cdots p\alpha}$$

$$= \Sigma(x_{1\alpha} - \bar{x}_1) e_{1.23\cdots p\alpha} \underbrace{- b_2'\Sigma(x_{2\alpha} - \bar{x}_2) e_{1.23\cdots p\alpha} - \cdots - b_{p-1}'\Sigma(x_{p-1\alpha} - \bar{x}_{p-1}) e_{1.23\cdots p\alpha}}_{\rightarrow \text{ zero}}$$

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

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

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

$$= n \cdot s^2_{1.23\cdots p-1} \quad [\text{from } \boxed{19}]$$

$$- b_{1p.23\cdots p-1} \sum (x_{p\alpha} - \bar{x}_p) \cdot e_{1.23\cdots p-1\alpha}$$

$$= n \cdot s^2_{1.23\cdots p-1} - b_{1p.23\cdots p-1} \sum (x_{p\alpha} - \bar{x}_p - b_{p2.31\cdots p-1}(x_{2\alpha} - \bar{x}_2) - b_{p3.21\cdots p-1}(x_{3\alpha} - \bar{x}_3) - \cdots - b_{p,p-1.23\cdots p-2}(x_{p-1\alpha} - \bar{x}_{p-1})) \cdot e_{1.23\cdots p-1\alpha}$$

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

$$a^0 = \bar{x}_p - b_1^0 \bar{x}_1 - \cdots - b^0_{p-1} \bar{x}_{p-1}$$

Rough

$$\sum (x_{i\alpha} - \bar{x}_i) \cdot e_{p.12\cdots p-1\alpha}, \quad i = 1(1)p-1$$

যেহেতু $e_{1.23\cdots p-1\alpha}$ term হিসে 1 আছে তাই include করলে হবে & $e_{p.23\cdots p-1\alpha}$ লেখা হবে

$$= n s^2_{1.23\cdots p-1} - b_{1p.23\cdots p-1} \sum e_{p.23\cdots p-1\alpha} \cdot e_{1.23\cdots p-1\alpha}$$

$$= n \cdot s^2_{1.23\cdots p-1} - n b_{1p.23\cdots p-1}\, r_{1p.23\cdots p-1}\, s_{p.23\cdots p-1}\, s_{1.23\cdots p-1} \quad (*)$$

Rough

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

$$r_{12.3\cdots p} \times s_{1.34\cdots p} \times s_{2.34\cdots p} = \frac{1}{n} \sum e_{1.34\cdots p\alpha} \cdot e_{2.34\cdots p\alpha}$$

$$\Rightarrow n\, r_{12.3\cdots p} \times s_{1.34\cdots p} \times s_{2.34\cdots p} = \sum e_{1.34\cdots p\alpha} \cdot e_{2.34\cdots p\alpha}$$

$$= n \cdot s^2_{1.23\cdots p-1} - n \cdot r_{1p.23\cdots p-1} \frac{s_{1.23\cdots p-1}}{\cancel{s_{p.23\cdots p-1}}} \cdot r_{1p.23\cdots p-1} \, \cancel{s_{p.23\cdots p-1}} \, s_{1.2}$$

$$\Rightarrow \boxed{n s^2_{1.23\cdots p} = n s^2_{1.23\cdots p-1}\left(1 - r^2_{1p.23\cdots p-1}\right)}$$

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

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

$\Rightarrow \boxed{s^2_{1.23\cdots p} \leq s^2_{1.23\cdots p-1}}$ [Residual variance decreases ~~when~~ variables under model increases]

Practical

(2) Examine the following correlation matrix for internal consistency –

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

Find $r^2_{1.234}$

(3) Suppose a computer has found for a given set of values x_1, x_2, x_3 ;

$$r_{12} = 0.91 \ , \ r_{13} = 0.33 \ , \ r_{23} = 0.81$$

Examine the computer whether is free from error.

Method (i) whether $|R|$ is positive (If negative then not free from error)

(ii) $r_{1.23}$ is positive.

$\boxed{\text{Answer}}$

(2) Given correlation matrix,

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

∴ Now, we have to find R^2 i.e. $r^2_{1.234}$, that is coefficient of determination.

where $r_{1.234} = \left(1 - \dfrac{|R|}{R_{1,1}}\right)^{1/2}$

$|R| = -0.01530449$

$R_{1,1} = 0.4549578$

$\therefore r_{1.234} = \left(1 + \dfrac{0.01530449}{0.4549578}\right)^{1/2}$

$\therefore r^2_{1.234} = 1.03363$

We know, $0 \le r \le 1$ but here $r = 1.234$

So, there is no internal consistency.

(3) Given correlation matrix –

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

$r_{12} = 0.91$

$r_{13} = 0.33$

$r_{23} = 0.81$

Now, $|R| = -0.106614$

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

$R_{1,1} = 0.3439$

$$= 1.14455867$$

We see the $|R|$ is negative and also multiple correlation coefficient $r_{1.23}$ is 1.14455867 i.e

It does not belong the range 0 to 1.

So, we may conclude that the computer is not free from error.