Multivariate Analysis

MULTIVARIATE ANALYSIS

Multivariate Data:– Investigators seeking to understand social or physical phenomena generally to collect simultaneous measurements on many variables (characters) for each distinct individual, because social or physical phenomena are complex in nature. The measurements are all recorded for each distinct individual. For example, the data may relate to the scores obtained by each of a number of students in three subjects: Math, Physics and Statistics. Another example, the data consists of rating over a course of treatment for patients undergoing radiotheraphy; variables rated include sore throat, sleep, food consumption, appetite, skin-reaction. Data of this type are called multivariate data because they are simultaneous measurements on many variables. Like bivariate case, multivariate data may also be arranged into a frequency distribution for p variables $x_1, x_2, \dots, x_p$ with $k_1, k_2, \dots, k_p$ classes, respectively, the joint frequency distribution will have $k_1 \times k_2 \times \dots \times k_p$ cell frequency; the frequency in a cell being the number of individuals belonging simultaneously to the corresponding x_1-class, x_2-class, ..., x_p-class. From this joint distn. of p-variables, we can obtain the marginal distn. of any p_1 variables $(1 \le p_1 \le p-1)$ and the conditional distribution of any p_1 variables for given values p_2 of the other variables $\{(p_1, p_2 \geq 1)$ and $(p_1+p_2) \le p\}$. These marginal and conditional distn.s can be obtained in a way similar to that in the bivariate case.

Suppose $x_{i\alpha}$ denotes the value of the variable x_i on the individual α, $i=1,2,\dots,p$, $\alpha = 1,2,\dots,n$. These n multivariate observations can be displayed as a data matrix X of p rows and n columns, i.e. $X = ((x_{i\alpha}))_{p\times n}$.

The useful descriptive statistics measuring location, dispersion and correlation are:

$$\bar{x}_i = \frac{1}{n}\sum_{\alpha=1}^{n} x_{i\alpha},$$

$$s_{ij} = \begin{cases} \frac{1}{n}\sum_{\alpha=1}^{n}(x_{i\alpha}-\bar{x}_i)(x_{j\alpha}-\bar{x}_j) & \text{for } i \neq j \\ \frac{1}{n}\sum_{\alpha=1}^{n}(x_{i\alpha}-\bar{x}_i)^2 & \text{for } i=j \end{cases}$$

$$= \begin{cases} \mathrm{cov}(x_i, x_j) & \text{for } i \neq j \\ \mathrm{Var}(x_i) & \text{for } i=j. \end{cases}$$

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} = \frac{\mathrm{cov}(x_i, x_j)}{s.d.(x_i)\, s.d.(x_j)}.$$

We can represent the statistics as:

$$\bar{x} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{pmatrix}, \quad S_{p\times p} = \begin{pmatrix} s_{11} & s_{12} & s_{13} & \cdots & s_{1p} \\ s_{21} & s_{22} & s_{23} & \cdots & s_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & s_{p3} & \cdots & s_{pp} \end{pmatrix} \text{ and}$$

$$R_{p\times p} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ r_{31} & r_{32} & \cdots & r_{3p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{pmatrix}.$$

$\bar{x}_{p\times 1}$, $S_{p\times p}$ and $R_{p\times p}$ are called mean vector, variance-covariance (dispersion) matrix and correlation matrix, respectively, of the variables $x_1, x_2, \ldots, x_p$.

Result:- A variance-covariance matrix is non-negative definite.

Proof:- Let $S_{p\times p} = ((s_{ij}))$ be the variance-covariance matrix of the variables $x_1, x_2, \ldots, x_p$ with the mean vector $\bar{\underset{\sim}{x}} = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_p \end{pmatrix}$. To prove that $S_{p\times p}$ is non-negative definite, we are to show that $\underset{1\times p}{l'}\ \underset{p\times p}{S}\ \underset{p\times 1}{l} \geq 0$ for every vector $\underset{\sim}{l} = \begin{pmatrix} l_1 \\ \vdots \\ l_p \end{pmatrix}$.

Now,

$$\underset{\sim}{l}' S \underset{\sim}{l} = \sum_{i=1}^{p} \sum_{j=1}^{p} s_{ij} l_i l_j$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{p} l_i l_j \cdot \frac{1}{n} \sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)$$

$$= \frac{1}{n} \sum_{\alpha=1}^{n} \sum_{i=1}^{p} \sum_{j=1}^{p} l_i l_j (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)$$

$$= \frac{1}{n} \sum_{\alpha=1}^{n} \left(\sum_{i=1}^{p} l_i (x_{i\alpha} - \bar{x}_i) \right) \left(\sum_{j=1}^{p} (x_{j\alpha} - \bar{x}_j) \right)$$

$$= \frac{1}{n} \sum_{\alpha=1}^{n} \left(\sum_{i=1}^{p} l_i (x_{i\alpha} - \bar{x}_i) \right)^2 = \frac{1}{n} \sum_{\alpha=1}^{n} \left(\underset{\sim}{l}' (\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}}) \right), \quad \alpha = 1(1)n.$$

$$= \frac{1}{n} \sum_{\alpha=1}^{n} z_\alpha^2,$$

$$\geq 0 \text{ for every } \underset{\sim}{l} \cdots\cdots (1)$$

Now from ①,

$\underset{\sim}{l}' S \underset{\sim}{l} = 0$ for some $\underset{\sim}{l} \neq \underset{\sim}{0}$

iff $l'(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}}) = \underset{\sim}{0}$ for some $\underset{\sim}{l} \neq \underset{\sim}{0}$ and $\forall \alpha = 1(1)n$.

i.e. iff $(\underset{\sim}{x}_\alpha - \bar{\underset{\sim}{x}})'\underset{\sim}{l} = \underset{\sim}{0}$ for some $\underset{\sim}{l} \neq \underset{\sim}{0}$ and $\forall \alpha = 1(1)n$.

i.e. iff
$$\begin{pmatrix} x_{11} - \bar{x}_1 & x_{21} - \bar{x}_2 & x_{31} - \bar{x}_3 & \cdots & x_{p1} - \bar{x}_p \\ x_{12} - \bar{x}_1 & x_{22} - \bar{x}_2 & x_{32} - \bar{x}_3 & \cdots & x_{p2} - \bar{x}_p \\ \vdots \\ x_{1n} - \bar{x}_1 & x_{2n} - \bar{x}_2 & x_{3n} - \bar{x}_3 & \cdots & x_{pn} - \bar{x}_p \end{pmatrix} \begin{pmatrix} l_1 \\ l_2 \\ \vdots \\ l_p \end{pmatrix} = \underset{\sim}{0} \text{ for some } \underset{\sim}{l} \neq \underset{\sim}{0}.$$

i.e. iff The columns of the matrix $((x_{i\alpha} - \bar{x}_i))$ are linearly dependent.

i.e. iff $\exists$ some deviation $(x_{i\alpha} - \bar{x}_i)$ which is a linear combination of the other $(p-1)$ deviations (over the n observations).

i.e. iff $\exists$ some variable which is a exact linear function of the other $(p-1)$ variables x_i over the n observations $\{\underset{\sim}{x}_\alpha ; \alpha = 1(1)n\}$.

<u>Fact</u>:- A covariance matrix is positive semidefinite.

<u>Proof</u>:- Let Σ be the covariance matrix of a random vector $\underset{\sim}{X}$ with mean vector $\underset{\sim}{\mu}$. Then $\Sigma = E[(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})']$

Now, let $\underset{\sim}{v}$ be any vector. We have to show that $\underset{\sim}{v}'\Sigma\underset{\sim}{v} \geq 0$.

But
$$\underset{\sim}{v}'\Sigma\underset{\sim}{v} = \underset{\sim}{v}' E[(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})']\underset{\sim}{v}$$
$$= E[\underset{\sim}{v}'(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})'\underset{\sim}{v}] = E[Y^2] \geq 0$$

where, $Y = \underset{\sim}{v}'(\underset{\sim}{X} - \underset{\sim}{\mu})$.

'=' holds iff $Y^2 = 0$ with probability 1.

i.e. iff $\underset{\sim}{v}'(\underset{\sim}{X} - \underset{\sim}{\mu}) = 0$, <u>or</u>, $\underset{\sim}{v}'\underset{\sim}{X} = \underset{\sim}{v}'\underset{\sim}{\mu}$.

i.e. iff $\exists$ a linear combination of the elements of $\underset{\sim}{X}$ which is equal to its mean with probability 1,

i.e. iff there is a variate which is degenerate in this sense of being a constant random variable.

Except when this happens, the covariance matrix is positive definite, not just positive semidefinite.

MULTIVARIATE FREQUENCY DISTRIBUTION

Multivariate Data:- In some investigations, data may be collected for the given set of individuals, on a no. of variables at the same time. Suppose we have 'p' variables $x_1, x_2, \ldots, x_p$. The values of the variables for the α^{th} individual may be denoted by $x_{1\alpha}, x_{2\alpha}, \ldots, x_{p\alpha}$, $\alpha = 1(1)n$.

Notations:- Let the vector variable be

$$\underset{\sim}{x} = (x_1, x_2, \ldots, x_p)'$$

Then, $\underset{\sim}{\bar{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)'$ is called the mean vector of $\underset{\sim}{x}$ and

$$\bar{x}_i = \frac{1}{n}\sum_{\alpha=1}^{n} x_{i\alpha}.$$

Define,

$$S_{ij} = \sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)$$

and

$$s_{ij} = \frac{1}{n}\sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)$$

$$= \begin{cases} Cov(x_i, x_j), & i \neq j \\ var(x_i), & i = j \end{cases}$$

Then $S = ((s_{ij}))_{p\times p}$ is called the 'variance-covariance matrix of $\underset{\sim}{x}$' or 'dispersion matrix of $\underset{\sim}{x}$'. The matrix $nS = ((S_{ij}))_{p\times p}$ is called the SSSP matrix (sum of squares and sum of product matrix).

Theorem 1:- Every dispersion matrix is non-negative definite.

Proof:- Note that, $\frac{1}{n}\sum_{\alpha=1}^{n}\left\{a_1(x_{1\alpha} - \bar{x}_1) + a_2(x_{2\alpha} - \bar{x}_2) + \cdots + a_p(x_{p\alpha} - \bar{x}_p)\right\}^2 \geq 0$, for all $a_i\,(i = 1(1)p)$

$$\Rightarrow \sum_{i=1}^{p}\left\{\frac{1}{n} a_i^2 \sum_{\alpha=1}^{n}(x_{i\alpha} - \bar{x}_i)^2\right\} + \sum_{i\neq j=1}^{p} a_i a_j \left\{\frac{1}{n}\sum_{\alpha=1}^{n}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)\right\} \geq 0 \;\forall a_i$$

$$\Rightarrow \sum_{i=1}^{p} a_i^2 . s_{ii} + \sum_{i\neq j=1}^{p} a_i a_j . s_{ij} \geq 0 \;\forall a_i$$

$\Rightarrow \underset{\sim}{a}' S \underset{\sim}{a} \geq 0 \;\forall \underset{\sim}{a}$, where $\underset{\sim}{a} = (a_1, a_2, \ldots, a_p)'$ and $S = ((s_{ij}))_{p\times p}$ is the dispersion matrix.

$\therefore$ S is n.n.d.

Corollary:—

(1) When S is p.s.d., then $\exists$ a non-null $\underset{\sim}{a} \ni \underset{\sim}{a}' S \underset{\sim}{a} = 0$

$\Leftrightarrow$ for some $\underset{\sim}{a} \neq \underset{\sim}{0}$,

$a_1(x_{1\alpha} - \bar{x}_1) + a_2(x_{2\alpha} - \bar{x}_2) + \cdots\cdots + a_p(x_{p\alpha} - \bar{x}_p) = 0 \ \forall \alpha = 1(1)n.$

$\Leftrightarrow$ for some $\underset{\sim}{a}' = (a_1, a_2, \ldots, a_p) \neq \underset{\sim}{0}'$, $\sum_{i=1}^{p} a_i(x_{i\alpha} - \bar{x}_i) = 0, \alpha = 1(1)n.$

$\Leftrightarrow$ variables are linearly related. [Here, $r(S) < p$]

(2) When S is p.d., then $\nexists$ a non null $\underset{\sim}{a} \ni \underset{\sim}{a}' S \underset{\sim}{a} = 0$

$\Leftrightarrow$ ~~for some~~ $\nexists \ \underset{\sim}{a} \neq \underset{\sim}{0} \ni \sum_{i=1}^{p} a_i(x_{i\alpha} - \bar{x}_i) = 0, \ \alpha = 1(1)n.$

$\Leftrightarrow \nexists \ \underset{\sim}{a} \neq \underset{\sim}{0}, \ b \neq 0$ (a constant) $\ni \sum_{i=1}^{p} a_i x_{i\alpha} = b, \ \alpha = 1(1)n.$

[Here $r(S) = p$.]

(3) Consider the matrix $((x_{i\alpha}))_{\substack{i=1,2,\ldots,p \\ \alpha=1,2,\ldots,n}}$. Define ~~the pxp matrix~~ $\bar{x}_i = \frac{1}{n}\sum_{\alpha=1}^{n} x_{i\alpha}$ and $s_{ij} = \frac{1}{n}\sum_{\alpha=1}^{n}(x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)$. Then the $p \times p$ matrix $((s_{ij}))$ is:

(i) p.d. when the rank $((x_{i\alpha} - \bar{x}_i)) = p$.

(ii) p.s.d. when the rank $((x_{i\alpha} - \bar{x}_i)) < p$, in which case $\exists$ constants $a_1, a_2, \ldots, a_p \ni \sum_{i=1}^{p} a_i(x_{i\alpha} - \bar{x}_i) = 0, \ \alpha = 1(1)n.$

Multiple Regression:— The theory of regression is concerned with the prediction of one or more variables $(y_1, y_2, \ldots, y_q)$ on the basis of information provided by either measurements on concomitant variables $(x_1, x_2, \ldots, x_p) = \underset{\sim}{x}'$. It is customary to call the latter independent or predictor variables and the former dependent or criterion variables. Prediction is needed in several ~~prediction~~ practical situations. A meterologist wants to forecast weather several hours ahead on the basis of suitable atmosphere measurements taken at a point in time.

In all these situations, the criteria are some variables in the future which we sought to be predicted by the available measurements for taking disizions. How should the predictors be chosen?

In probabilistic approach, the conditional mean of y given $(x_1, x_2, \ldots, x_p)$ is called the regression equation.

Notations:– Let, $\underset{\sim}{x} = (x_1, x_2, \ldots, x_p)'$

$$= (x_{(1)}, \underset{\sim}{x}^{(2)})'$$

and mean vector $\underset{\sim}{\bar{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)'$

$$= (\bar{x}_1, \underset{\sim}{\bar{x}}^{(2)})'$$

The dispersion matrix is

$$S = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ s_{31} & s_{32} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix} \quad [\text{where, } s_{ij} = s_{ji} \ \forall \ i, j]$$

$$= \begin{pmatrix} s_{11} & \underset{\sim}{s}_{(1)}' \\ \underset{\sim}{s}_{(1)} & S_2 \end{pmatrix}$$

Clearly, S_2 is the dispersion mtx of $(x_2, x_3, \ldots, x_p)$ and it is assumed to be non-singular (p.d.).

The correlation matrix is
$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{12} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{pmatrix}$$

Note that both the matrices S and R are symmetric.

Note that $R = DSD$, where $D = \text{diag}\left(\frac{1}{\sqrt{s_{11}}}, \frac{1}{\sqrt{s_{22}}}, \ldots, \frac{1}{\sqrt{s_{pp}}}\right)$

$$= \begin{pmatrix} \frac{1}{\sqrt{s_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{s_{pp}}} \end{pmatrix} \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{s_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{s_{pp}}} \end{pmatrix}$$

$$\therefore |R| = |D||S||D| = |D|^2|S|$$

$$\Rightarrow |S| = s_{11}s_{22}\cdots s_{pp} \cdot |R|$$

When the regression function is linear in predictor variables, it has been studied extensively.

[In probabilistic approach, $E(X_1 | \underset{\sim}{X}^{(2)}) = M(\underset{\sim}{X}^{(2)})$, is called the conditional mean of X_1 given $\underset{\sim}{X}^{(2)}$, is also called the regression function. The regression is said to be linear or non-linear according as the function $M(\cdot)$ is linear or non-linear.]

Let us assume that the regression of x_1 on $x_2, x_3, \ldots, x_p$ is linear whether the true regression is linear or not, i.e., assuming the regression equation as:

$$f(\underset{\sim}{x}^{(2)}) = a + b_2 x_2 + b_3 x_3 + \cdots + b_p x_p$$

Since the MSE is minimum, i.e.,

$\sum_{\alpha=1}^{n} \{x_{1\alpha} - f(x_{2\alpha}, x_{3\alpha}, \ldots, x_{p\alpha})\}^2$ is minimum when $f(\underset{\sim}{x}^{(2)})$ is the regression function, hence the constants $a, b_2, \ldots, b_p$ are determined by minimising

$$S^2 = \sum_{\alpha=1}^{n} (x_{1\alpha} - a - b_2 x_{2\alpha} - \cdots - b_p x_{p\alpha})^2 \quad \text{w.r.t. } a, b_2, b_3, \ldots, b_p$$

The Normal equations are:

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

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

$$\Rightarrow \begin{cases} \sum_{\alpha=1}^{n} (x_{1\alpha} - a - b_2 x_{2\alpha} - \cdots - b_p x_{p\alpha}) = 0 \\ \sum_{\alpha=1}^{n} (x_{1\alpha} - a - b_2 x_{2\alpha} - \cdots - b_p x_{p\alpha}) x_{i\alpha} = 0 \quad \forall\ p = 2(1)p. \end{cases}$$

$$\Rightarrow \begin{cases} \bar{x}_1 = a + b_2 \bar{x}_2 + \cdots + b_p \bar{x}_p \\ \Leftrightarrow a = \bar{x}_1 - (b_2 \bar{x}_2 + b_3 \bar{x}_3 + \cdots + b_p \bar{x}_p) \\ \quad = \bar{x}_1 - \underset{\sim}{b}' \bar{\underset{\sim}{x}}^{(2)}, \text{ where } \underset{\sim}{b}' = (b_2, b_3, \ldots, b_p) \\ \sum_{\alpha} x_{1\alpha} x_{i\alpha} = a \sum_{\alpha} x_{i\alpha} + b_2 \sum_{\alpha} x_{2\alpha} x_{i\alpha} + \cdots + b_i \sum_{\alpha} x_{i\alpha}^2 + \cdots \\ \quad \cdots + b_p \sum_{\alpha} x_{p\alpha} x_{i\alpha}, \ \forall\ i = 2(1)p. \end{cases}$$

$$\Rightarrow \begin{cases} \bar{x}_1 = a + b_2\bar{x}_2 + \cdots + b_p\bar{x}_p \\ \sum_\alpha x_{1\alpha} x_{i\alpha} = \{\bar{x}_1 - (b_2\bar{x}_2 + \cdots + b_p\bar{x}_p)\} \sum_\alpha x_{i\alpha} + b_2 \sum_\alpha x_{2\alpha} x_{i\alpha} \\ \qquad + \cdots + b_p \sum_\alpha x_{p\alpha} x_{i\alpha}, \ \forall\ i = 2(1)p. \end{cases}$$

$$\Rightarrow \sum_{\alpha=1}^{n} (x_{1\alpha} - \bar{x}_1)(x_{i\alpha} - \bar{x}_i) = b_2 \sum_\alpha (x_{2\alpha} - \bar{x}_2)(x_{i\alpha} - \bar{x}_i) + \cdots + b_p \sum_\alpha (x_{p\alpha} - \bar{x}_p)(x_{i\alpha} - \bar{x}_i) \ \forall\ i = 2(1)p$$

$$\Rightarrow s_{1i} = b_2 s_{2i} + b_3 s_{3i} + \cdots + b_p s_{pi}, \ \forall i = 2(1)p.$$

$$\Leftrightarrow \begin{matrix} s_{12} = b_2 s_{22} + b_3 s_{32} + \cdots + b_p s_{p2} \\ s_{13} = b_2 s_{23} + b_3 s_{33} + \cdots + b_p s_{p3} \\ \vdots \\ s_{1p} = b_2 s_{2p} + b_3 s_{3p} + \cdots + b_p s_{pp} \end{matrix}$$

$$\Leftrightarrow \underset{\sim}{s}_{(1)} = \begin{pmatrix} s_{22} & s_{32} & \cdots & s_{p2} \\ s_{23} & s_{33} & \cdots & s_{p3} \\ \vdots & \vdots & & \vdots \\ s_{2p} & s_{3p} & \cdots & s_{pp} \end{pmatrix} \begin{pmatrix} b_2 \\ b_3 \\ \vdots \\ b_p \end{pmatrix}$$

$$\Leftrightarrow \underset{\sim}{s}_{(1)} = S_2 \underset{\sim}{b}$$

or, $\underset{\sim}{b} = S_2^{-1} \underset{\sim}{s}_{(1)}$, since S_2 is assumed to be non-singular.

Hence, $f(\underset{\sim}{x}^{(2)}) = a + \underset{\sim}{b}' \underset{\sim}{x}^{(2)}$, where $a = \bar{x}_1 - \underset{\sim}{b}' \bar{\underset{\sim}{x}}^{(2)}$ and $\underset{\sim}{b} = S_2^{-1} \cdot \underset{\sim}{s}_{(1)}$ is the multiple linear regression of x_1 on $\underset{\sim}{x}^{(2)} = (x_2, \ldots, x_p)'$

Define, $X_{1.23\ldots p} = a + \underset{\sim}{b}' \underset{\sim}{x}^{(2)}$
$= a + b_2 x_2 + \cdots + b_p x_p$, as the part of x_1 explained by the multiple linear regression of x_1 on $(x_2, x_3, \ldots, x_p)$.

Then we write $x_1 = X_{1.23\ldots p} + e_{1.23\ldots p}$, where $e_{1.23\ldots p}$ is the residual part of x_1 corresponding to its multiple linear regression.

For the α^{th} individual,

$$x_{1\alpha} = X_{1.23\ldots p, \alpha} + e_{1.23\ldots p, \alpha}$$

Theorem 1:- $\sum_{\alpha=1}^{n} e_{1.23\ldots p,\alpha} = 0$ and $e_{1.23\ldots p}$ is uncorrelated with every predictor variable and hence with multiple linear regression equation.

Proof:- Note that, $e_{1.23\ldots p} = x_{1\alpha} - X_{1.23\ldots p,\alpha}$

$$= x_{1\alpha} - (a + b_2 x_{2\alpha} + \cdots + b_p x_{p\alpha})$$

From the 1st Normal equation, we have,

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

$$\Rightarrow \sum_{\alpha} e_{1.23\ldots p} = 0.$$

From the other normal equations,

$$\sum_{\alpha=1}^{n} e_{1.23\ldots p,\alpha} \cdot x_{i\alpha} = 0 \quad , \forall\ i = 2(1)p$$

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

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

$$\Rightarrow \text{Cov}(e_{1.23\ldots p}, x_i) = 0 \ \forall\ i = 2(1)p$$

Again, $\text{Cov}(e_{1.23\ldots p}, X_{1.23\ldots p})$

$$= \text{Cov}\left(e_{1.23\ldots p}, a + \sum_{i=2}^{p} b_i x_i\right)$$

$$= 0 + \text{Cov}\left(e_{1.23\ldots p}, \sum_{i=2}^{p} b_i x_i\right)$$

$$= \sum_{i=2}^{p} b_i \cdot \text{Cov}(e_{1.23\ldots p}, x_i)$$

$$= 0.$$

Hence the residual is uncorrelated with the multiple linear regression.

Theorem 2. Variance $(e_{1\cdot 23\ldots p}) = s_{11} - \underset{\sim}{s}'_{(1)} S_2^{-1} \underset{\sim}{s}_{(1)}$

$$= \frac{|S|}{|S_2|} = \frac{1}{s^{11}} = \frac{|R|}{R_{11}} s_{11} = \frac{s_{11}}{r^{11}}$$

where $S^{-1} = ((s^{ij}))$; R_{11} = cofactor of r_{11} in R ;
$R^{-1} = ((r^{ij}))$

[The symbols have their usual meaning]

Proof:- $Var(e_{1\cdot 23\ldots p})$

$= cov(e_{1\cdot 23\ldots p}, e_{1\cdot 23\ldots p})$

$= cov(e_{1\cdot 23\ldots p}, x_1 - X_{1\cdot 23\ldots p})$

$= cov(e_{1\cdot 23\ldots p}, x_1)$, since $cov(e_{1\cdot 23\ldots p}, X_{1\cdot 23\ldots p}) = 0$.

$= cov(x_1 - X_{1\cdot 23\ldots p}, x_1)$

$= cov(x_1, x_1) - cov(x_1, X_{1\cdot 23\ldots p})$

$= var(x_1) - cov\left(x_1, a + \sum_{i=2}^{p} b_i x_i\right)$

$= s_{11} - \sum_{i=2}^{p} b_i \cdot cov(x_1, x_i)$

$= s_{11} - \sum_{i=2}^{p} b_i \cdot s_{1i}$

$= s_{11} - \underset{\sim}{b}' \underset{\sim}{s}_{(1)}$

$= s_{11} - \underset{\sim}{s}'_{(1)} S_2^{-1} \underset{\sim}{s}_{(1)}$, since we have from normal equations $S_2 \underset{\sim}{b} = \underset{\sim}{s}_{(1)}$ and S_2 being symmetric, $(S_2^{-1})' = S_2^{-1}$.

Note that, $S = \begin{pmatrix} s_{11} & \underset{\sim}{s}'_{(1)} \\ \underset{\sim}{s}_{(1)} & S_2 \end{pmatrix}$

and $|S| = |S_2| \left(s_{11} - \underset{\sim}{s}'_{(1)} S_2^{-1} \underset{\sim}{s}_{(1)} \right)$, since we have assumed that S_2 is non-singular.

$\therefore |S| = |S_2| \cdot Var(e_{1\cdot 23\ldots p})$

$\Rightarrow var(e_{1\cdot 23\ldots p}) = \frac{|S|}{|S_2|}$

$$= \frac{|S|}{\text{cofactor of } s_{11} \text{ in } S} = \frac{1}{s^{11}},$$

where $S^{-1} = \left(\left(\frac{\text{cofactor of } s_{ji}}{|S|}\right)\right)$

$= ((s^{ij}))$, say.

We have $|S| = s_{11} s_{22} \cdots s_{pp} |R|$; Again $|S_2| = s_{22} s_{33} \cdots s_{pp} \cdot R_{11}$, where R_{11} is the cofactor of r_{11} in R.

$\therefore Var(e_{1.23\ldots p}) = \frac{|S|}{|S_2|}$

$= \frac{s_{11}s_{22}\cdots s_{pp}}{s_{22}s_{33}\cdots s_{pp}} \cdot \frac{|R|}{R_{11}}$

$= \frac{s_{11}|R|}{R_{11}} = \frac{s_{11}}{R_{11}/|R|}$

$= \frac{s_{11}}{r^{11}}$, where $R^{-1} = ((r^{ij}))$, say.

Problems:-

(1) If $x_1, x_2, \ldots, x_p$ are p variables $\ni$ the correlation coefficient between each pair of components is r. S.T. $-\frac{1}{p-1} \le r \le 1$.

ANS:- The correlation matrix is

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

Clearly, R is n.n.d.

[$var(\sum_{i=1}^{p} a_i x_i) \geq 0 \quad \forall \underset{\sim}{a} = (a_1, a_2, \ldots, a_p)$

$\Leftrightarrow \sum_i \sum_j a_i a_j Cov(x_i, x_j) \geq 0$

$\Leftrightarrow \sum_i \sum_j a_i a_j r_{ij} \sqrt{s_{ii}} \sqrt{s_{jj}} \geq 0$

$\Leftrightarrow \sum_i \sum_j r_{ij} u_i u_j \geq 0$, where $u_i = a_i\sqrt{s_{ii}}$, $u_j = a_j\sqrt{s_{jj}}$

$\Leftrightarrow \underset{\sim}{u}' R \underset{\sim}{u} \geq 0 \quad \forall \underset{\sim}{u}$

$\Leftrightarrow$ R is n.n.d.]

$\therefore$ All principle minors are ≥ 0

$\therefore \begin{vmatrix} 1 & r \\ r & 1 \end{vmatrix} \geq 0, \quad \begin{vmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{vmatrix} \geq 0, \ldots\ldots, \begin{vmatrix} 1 & r & r & \cdots & r \\ r & 1 & r & \cdots & r \\ r & r & 1 & \cdots & r \\ \vdots & \vdots & \vdots & & \vdots \\ r & r & r & \cdots & 1 \end{vmatrix}_{p\times p} \geq 0$

Now, $\begin{vmatrix} 1 & r \\ r & 1 \end{vmatrix} \geq 0 \Rightarrow 1 - r^2 \geq 0$

$\Rightarrow r^2 \le 1$

$\Rightarrow -1 \le r \le 1$

$\Rightarrow (1-r)^{p-1}[1+(p-1)r] \geq 0$

$\Rightarrow 1+(p-1)r \geq 0 \quad [\because (1-r) \geq 0]$

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

$\therefore -\frac{1}{p-1} \le r \le 1$. (Proved)

Problem (2):– If $R = ((r_{ij}))$ is the correlation matrix of $(x_1, x_2, \ldots, x_p)$, S.T. $|R| \leq 1$.

Ans:– The correlation mtx is

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{pmatrix} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix} = \begin{pmatrix} 1 & \underset{\sim}{r}_{(1)}' \\ \underset{\sim}{r}_{(1)} & R_2 \end{pmatrix}$$

$$\therefore |R| = |R_2| \left(1 - \underset{\sim}{r}_{(1)}' R_2^{-1} \underset{\sim}{r}_{(1)}\right)$$

$\leq |R_2|$, since R is p.d.

$\Rightarrow R_2$ is p.d.

$\Rightarrow R_2^{-1}$ is p.d.

$\Rightarrow \underset{\sim}{r}_{(1)}' \cdot R_2^{-1} \cdot \underset{\sim}{r}_{(1)} \geq 0 \;\; \forall \; \underset{\sim}{r}_{(1)}$

$$\Rightarrow |R| \leq |R_2| = \begin{vmatrix} 1 & \underset{\sim}{r}_{(2)}' \\ \underset{\sim}{r}_{(2)} & R_3 \end{vmatrix}$$

$$\leq |R_3| \cdots\cdots \leq \begin{vmatrix} 1 & r \\ r & 1 \end{vmatrix} = 1 - r^2 \leq 1.$$

or, $|R| \leq 1$ (Proved)

Remark:– It can be shown that $|S| \leq s_{11} s_{22} \cdots s_{pp}$

$$\left[\because |R| = \frac{|S|}{s_{11} s_{22} \cdots s_{pp}} \text{ and } |R| \leq 1. \right]$$

Problem (3):– Each of the variables x, y, z has mean 0, variance 1 while $ax + by + cz = 0$. Show that $a^4 + b^4 + c^4 \leq 2(b^2c^2 + c^2a^2 + a^2b^2)$

Also obtain the dispersion matrix.

Solution:– We have $ax + by + cz = 0$

$\Rightarrow ax + by = -cz$

$\Rightarrow Var(ax + by) = Var(-cz)$

$\Rightarrow a^2 Var(x) + b^2 Var(y) + 2ab \; cov(x, y) = c^2 Var(z)$

$\Rightarrow a^2 + b^2 + 2ab \, r_{xy} = c^2$

$$\Rightarrow \frac{c^2 - (a^2 + b^2)}{2ab} = r_{xy}$$

$$\Rightarrow \left[\frac{c^2 - (a^2 + b^2)}{2ab}\right]^2 = r_{xy}^2 \leq 1.$$

$\Rightarrow c^4 + (a^2 + b^2)^2 - 2c^2(a^2 + b^2) \leq 4a^2b^2$

$\Rightarrow c^4 + a^4 + b^4 + 2a^2b^2 - 2c^2a^2 - 2c^2b^2 \leq 4a^2b^2$

$\Rightarrow a^4 + b^4 + c^4 \leq 2(a^2b^2 + b^2c^2 + c^2a^2)$

(Proved)

Multiple Regression:— In many practical cases, predicted values of a response (dependent) variable obtained from a single predictor (independent) variable, via a regression model, are too imprecise to be useful. The main reason is that the single predictor (independent) variable is one of the many potential predictor variables affecting the response variable in important ways. In such a situation, a model containing important predictor variables will be more useful because it will predict the values of the response variable more precisely. For example, in predicting the rainfall at a place in a year, it is appropriate to include three or four things as the predictor variables.

Suppose one of the p variables, $x_1, x_2, \dots, x_p$; say, x_1, is the response variable of interest and the others are the predictor variables. We are to predict a value of x_1 for given values of $x_2, x_3, \dots, x_p$ via a regression model. We assume that the relationship between x_1 and the set $\{x_2, \dots, x_p\}$ is, at least in an approximate sense, represented by a linear equation of the form

$$x_1 = a + b_2 x_2 + b_3 x_3 + \dots + b_p x_p, \quad \text{------------} (1)$$

where, a and b_i's are unknown coefficients.

We determine the unknowns $a, b_2, b_3, \dots, b_p$ on the basis of the n multivariate observations $(x_{1\alpha}, x_{2\alpha}, \dots, x_{p\alpha})$; $\alpha = 1(1)n$, by the method of least squares. In this method, $a, b_2, \dots, b_p$ are determined so that the error sum of squares

$$S^2(a, b_2, \dots, b_p) = \sum_{\alpha=1}^{n} \left(x_{1\alpha} - a - b_2 x_{2\alpha} - \dots - b_p x_{p\alpha}\right)^2 \text{ is minimum.}$$

The normal equations, obtained by equating the partial derivatives of $S^2(a, b_2, \dots, b_p)$ with respect to $a, b_2, \dots, b_p$ to zero, are

$$\left.\begin{aligned}
\sum_\alpha x_{1\alpha} &= na + b_2 \sum_\alpha x_{2\alpha} + b_3 \sum_\alpha x_{3\alpha} + \dots + b_p \sum_\alpha x_{p\alpha}, \\
\sum_\alpha x_{2\alpha} x_{1\alpha} &= a \sum_\alpha x_{2\alpha} + b_2 \sum_\alpha x_{2\alpha}^2 + b_3 \sum_\alpha x_{2\alpha} x_{3\alpha} + \dots + b_p \sum_\alpha x_{2\alpha} x_{p\alpha}, \\
\sum_\alpha x_{3\alpha} x_{1\alpha} &= a \sum_\alpha x_{3\alpha} + b_2 \sum_\alpha x_{3\alpha} x_{2\alpha} + b_3 \sum_\alpha x_{3\alpha}^2 + \dots + b_p \sum_\alpha x_{3\alpha} x_{p\alpha}, \\
&\vdots \\
\sum_\alpha x_{p\alpha} x_{1\alpha} &= a \sum_\alpha x_{p\alpha} + b_2 \sum_\alpha x_{p\alpha} x_{2\alpha} + b_3 \sum_\alpha x_{p\alpha} x_{3\alpha} + \dots + b_p \sum_\alpha x_{p\alpha}^2 .
\end{aligned}\right\} (2)$$

(14)

The first equation gives, on being divided by n,

$$\bar{x}_1 = a + b_2\bar{x}_2 + b_3\bar{x}_3 + \cdots + b_p\bar{x}_p \;, \quad \cdots\cdots (3)$$

which shows incidentally that the mean point $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p)$ necessarily satisfies the prediction equation.

Multiplying (3) by $n\bar{x}_2, n\bar{x}_3, \dots, n\bar{x}_p$ and subtracting the result from the second, third, ..., pth equation, respectively, of the system (2), we have $(p-1)$ equations determining the b's, viz.

$$\left.\begin{aligned} S_{21} &= b_2 S_{22} + b_3 S_{23} + \cdots + b_p S_{2p}, \\ S_{31} &= b_2 S_{32} + b_3 S_{33} + \cdots + b_p S_{3p}, \\ &\vdots \\ S_{p1} &= b_2 S_{p2} + b_3 S_{p3} + \cdots + b_p S_{pp}, \end{aligned}\right\} \longrightarrow (4)$$

where, $S_{ij} = \sum_\alpha x_{i\alpha}x_{j\alpha} - n\bar{x}_i\bar{x}_j = \sum_\alpha (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)$.

Taking $s_{ij} = \frac{1}{n} \times S_{ij} = \begin{cases} Cov(x_i, x_j) & \forall\, i \neq j \\ Var(x_i) & \forall\, i = j \end{cases}$

Then (4) reduces to,

$$\begin{pmatrix} s_{21} \\ s_{31} \\ \vdots \\ s_{p1} \end{pmatrix} = \begin{pmatrix} s_{22} & s_{23} & \cdots & s_{2p} \\ s_{32} & s_{33} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots \\ s_{p2} & s_{p3} & \cdots & s_{pp} \end{pmatrix} \begin{pmatrix} b_2 \\ b_3 \\ \vdots \\ b_p \end{pmatrix} \longrightarrow (5)$$

The $p \times p$ matrix $S = \begin{pmatrix} s_{22} & s_{23} & \cdots & s_{2p} \\ s_{32} & s_{33} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots \\ s_{p2} & s_{p3} & \cdots & s_{pp} \end{pmatrix}$ is non-singular (i.e. is of rank p, i.e. full rank)

and $S = (s_{ij})_{p\times p}$ is called variance-covariance (or, dispersion) matrix of $x_1, \dots, x_p$. This non-singularity of the dispersion matrix implies that the system of equation (5) has the unique solution

$$\begin{pmatrix} b_2 \\ b_3 \\ \vdots \\ b_p \end{pmatrix} = \begin{pmatrix} s_{22} & s_{23} & \cdots & s_{2p} \\ s_{32} & s_{33} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots \\ s_{p2} & s_{p3} & \cdots & s_{pp} \end{pmatrix}^{-1} \begin{pmatrix} s_{21} \\ s_{31} \\ \vdots \\ s_{p1} \end{pmatrix} \longrightarrow (6)$$

so that $b_j = s_{21}s^{j2} + s_{31}s^{j3} + \cdots + s_{p1}s^{jp}$

Alternatively, we can write b_j as $\longrightarrow$ (for $j = 2, 3, \ldots, p$)

$$b_j = \frac{\begin{vmatrix} s_{22} & s_{23} & \cdots & s_{2(j-1)} & s_{21} & s_{2(j+1)} & \cdots & s_{2p} \\ s_{32} & s_{33} & \cdots & s_{3(j-1)} & s_{31} & s_{3(j+1)} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ s_{p2} & s_{p3} & \cdots & s_{p(j-1)} & s_{p1} & s_{p(j+1)} & \cdots & s_{pp} \end{vmatrix}}{\begin{vmatrix} s_{22} & s_{23} & \cdots & s_{2p} \\ s_{32} & s_{33} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots \\ s_{p2} & s_{p3} & \cdots & s_{pp} \end{vmatrix}} \quad \text{(by Cramer's Rule)}$$

$$= (-1)^{j-2} \frac{\begin{vmatrix} s_{21} & s_{22} & s_{23} & \cdots & s_{2(j-1)} & s_{2(j+1)} & \cdots & s_{2p} \\ s_{31} & s_{32} & s_{33} & \cdots & s_{3(j-1)} & s_{3(j+1)} & \cdots & s_{3p} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ s_{p1} & s_{p2} & s_{p3} & \cdots & s_{p(j-1)} & s_{p(j+1)} & \cdots & s_{pp} \end{vmatrix}}{\begin{vmatrix} s_{22} & s_{23} & \cdots & s_{2p} \\ s_{32} & s_{33} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots \\ s_{p2} & s_{p3} & \cdots & s_{pp} \end{vmatrix}}$$

[putting the column $\begin{pmatrix} s_{21} \\ s_{31} \\ \vdots \\ s_{p1} \end{pmatrix}$ in the 1st place introduces <u>or</u> removes $(j-2)$ inversions]

$$= (-1)^{j-2} \times \frac{s_1}{s_j} \cdot \frac{\begin{vmatrix} r_{21} & r_{22} & r_{23} & \cdots & r_{2(j-1)} & r_{2(j+1)} & \cdots & r_{2p} \\ r_{31} & r_{32} & r_{33} & \cdots & r_{3(j-1)} & r_{3(j+1)} & \cdots & r_{3p} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & r_{p3} & \cdots & r_{p(j-1)} & r_{p(j+1)} & \cdots & r_{pp} \end{vmatrix}}{\begin{vmatrix} r_{22} & r_{23} & \cdots & r_{2p} \\ r_{32} & r_{33} & \cdots & r_{3p} \\ \vdots & \vdots & & \vdots \\ r_{p2} & r_{p3} & \cdots & r_{pp} \end{vmatrix}} \quad \text{------}(*)$$

as $s_{ij} = r_{ij} \cdot s_i s_j$; s_i, s_j being the s.d.s of x_i and x_j respectively and r_{ij} being correlation coefficient of x_i and x_j.

$$= (-1)^{j-2} \cdot \frac{s_1}{s_j} \times \frac{(-1)^{j+1} R_{1j}}{R_{11}} = (-1)^{2j-1} \times \frac{s_1}{s_j} \times \frac{R_{1j}}{R_{11}}$$

$$= -\frac{R_{1j}}{R_{11}} \times \frac{s_1}{s_j} \text{ for } j = 2, 3, \ldots, p. ;$$

where, R_{ij} is the co-factor of r_{ij} in the correlation matrix $R = (r_{ij})_{p \times p}$ of the p-values $x_1, \ldots, x_p$.

We see that the determent in the numerator of (*) is the minor of r_{ij} in $R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{pmatrix}$ and hence is $(-1)^{1+j} \times$ co-factor of r_{ij}, while the determinant in the denominator is the minor of r_{11}.

Putting these b_j values in (3), we get,

$$a = \bar{x}_1 + \sum_{j=2}^{p} \left(\frac{s_1}{s_j} \times \frac{R_{1j}}{R_{11}} \right) \bar{x}_j$$

Thus the prediction equation (called the multiple regression equation of x_1 on $x_2, x_3, \ldots, x_p$) becomes

$$x_{1.23\ldots p} = \bar{x}_1 + \sum_{j=2}^{p} \left(\frac{s_1}{s_j} \times \frac{R_{1j}}{R_{11}} \right) \bar{x}_j - \sum_{j=2}^{p} \left(\frac{s_1}{s_j} \times \frac{R_{1j}}{R_{11}} \right) x_j$$

$$= \bar{x}_1 - \sum_{j=2}^{p} \left(\frac{s_1}{s_j} \times \frac{R_{1j}}{R_{11}} \right) (x_j - \bar{x}_j) \quad \text{--------- (7)}$$

The co-efficient $b_j = -\frac{s_1}{s_j} \times \frac{R_{1j}}{R_{11}}$ in (7) gives the amount by which the predicted value $X_{1.23\ldots p}$ increases when x_j is increased by a unit amount, the other independent variable being kept fixed. The coefficient b_j is called the <u>Partial regression coefficient</u> of x_1 on x_j, and usually written as $b_{1j.23\ldots \overline{j-1}\,\overline{j+1}\ldots p}$.

Some useful Results :-

(1) $\bar{X}_{1.23\ldots p} = \bar{x}_1$, where $\bar{X}_{1.23\ldots p} = \frac{1}{n} \sum_{\alpha=1}^{n} X_{1.23\ldots p\alpha}$ being the sample mean of the predicted values of x_1.

Proof:- We have

$$X_{1.23\ldots p} = \bar{x}_1 - \sum_{j=2}^{p} \left(\frac{s_1}{s_j} \times \frac{R_{1j}}{R_{11}} \right) (x_j - \bar{x}_j) ;$$

giving

$$\bar{X}_{1.23\ldots p} = \frac{1}{n} \sum_{\alpha=1}^{n} X_{1.23\ldots p\alpha} = \frac{1}{n} \sum_{\alpha=1}^{n} \left[\bar{x}_1 - \sum_{j=2}^{p} \frac{s_1}{s_j} \times \frac{R_{1j}}{R_{11}} (x_{j\alpha} - \bar{x}_j) \right]$$

$$= \bar{x}_1 - \frac{1}{n} \sum_{j=2}^{p} \frac{s_1}{s_j} \times \frac{R_{1j}}{R_{11}} \sum_{\alpha=1}^{n} (x_{j\alpha} - \bar{x}_j)$$

$$= \bar{x}_1 , \quad \text{because} \quad \sum_{\alpha=1}^{n} (x_{j\alpha} - \bar{x}_j) = 0 \;\; \forall\, j = 2(1)p.$$

Cor.: → This result implies that the means of the residual term $x_{1.23\ldots p} = x_1 - X_{1.23\ldots p}$ is $\bar{x}_{1.23\ldots p} = \bar{x}_1 - \bar{X}_{1.23\ldots p} = 0$.

(2) $Var(X_{1.23\ldots p}) = s_1^2\left(1 - \frac{|R|}{R_{11}}\right)$

Proof:– $Cov(x_1, X_{1.23\ldots p})$

$= Cov(X_{1.23\ldots p} + x_{1.23\ldots p}, X_{1.23\ldots p})$

$= Var(X_{1.23\ldots p}) + cov(x_{1.23\ldots p}, X_{1.23\ldots p})$

$= Var(X_{1.23\ldots p})$; because ——

$n\,cov(x_{1.23\ldots p}, X_{1.23\ldots p})$

$$= \sum_{\alpha=1}^{n} \{ x_{1.23\ldots p\alpha} (X_{1.23\ldots p\alpha} - \bar{X}_{1.23\ldots p})$$

$$= \sum_{\alpha=1}^{n} X_{1.23\ldots p\alpha} x_{1.23\ldots p\alpha} - \sum_{\alpha=1}^{n} \bar{X}_{1.23\ldots p} x_{1.23\ldots p\alpha}$$

$$= \sum_{\alpha=1}^{n} X_{1.23\ldots p\alpha} x_{1.23\ldots p\alpha}, \text{ because } \bar{X}_{1.23\ldots p} = 0.$$

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

$$= a\sum_{\alpha}(x_{1\alpha} - a - b_2 x_{2\alpha} - \cdots - b_p x_{p\alpha}) + b_2\left(\sum_{\alpha} x_{2\alpha}(x_{1\alpha} - a - b_2 x_{2\alpha} - \cdots - b_p x_{p\alpha})\right) + \cdots + b_p \sum_{\alpha} x_{p\alpha}(x_{1\alpha} - a - b x_{2\alpha} - \cdots - b_p x_{p\alpha})$$

$= 0$, as the i^{th} term in the sum is zero for the i^{th} normal equation $\forall\, i = 2, 3, \ldots, p$.

Now, $Cov(x_1, X_{1.23\ldots p}) = \frac{1}{n}\sum_{\alpha=1}^{n}(x_{1\alpha} - \bar{x}_1)(x_{1.23\ldots p\alpha} - \bar{x}_1)$ as $\bar{x}_1 = \bar{X}_{1.23\ldots p}$;

$$= \frac{1}{n}\sum_{\alpha=1}^{n}(x_{1\alpha} - \bar{x}_1)\left\{(-1)\sum_{j=2}^{p}\frac{s_1}{s_j}\times\frac{R_{1j}}{R_{11}}(x_{j\alpha} - \bar{x}_j)\right\}$$

$$= (-1)\sum_{j=2}^{p}\frac{s_1}{s_j}\times\frac{R_{1j}}{R_{11}}\times\frac{1}{n}\sum_{\alpha=1}^{n}(x_{1\alpha} - \bar{x}_1)(x_{j\alpha} - \bar{x}_j)$$

$$= (-1)\sum_{j=2}^{p}\frac{s_1}{s_j}\times\frac{R_{1j}}{R_{11}}\cdot s_{1j} \quad [\because s_{1j} = r_{1j}\cdot s_1 s_j]$$

$$= (-1)\sum_{j=2}^{p}\frac{s_1}{s_j}\times\frac{R_{1j}}{R_{11}}\cdot r_{1j}\cdot s_1 s_j = (-1)\frac{s_1^2}{R_{11}}\sum_{j=2}^{p} R_{1j} r_{1j}$$

$$= (-1)\frac{s_1^2}{R_{11}}(|R| - r_{11}R_{11}) \quad \left[\because |R| = \sum_{j=1}^{p} R_{1j} r_{1j}\right]$$

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

Best Linear Predictor:

Theorem: The multiple linear regression of x_1 on $\underset{\sim}{x}^{(2)}$ is the linear function having maximum correlation with x_1 among all the linear functions of $\underset{\sim}{x}^{(2)}$.

Proof: Let us consider a class of linear functions of $\underset{\sim}{x}^{(2)}$ as:

$$\left\{ L(\underset{\sim}{x}^{(2)}) : L(\underset{\sim}{x}^{(2)}) = l_0 + \underset{\sim}{l}'\underset{\sim}{x}^{(2)} = l_0 + \sum_{i=2}^{p} l_i x_i \right\}$$

Now,

$$r^2(x_1, L(\underset{\sim}{x}^{(2)})) = \frac{\left[\operatorname{cov}(x_1, L(\underset{\sim}{x}^{(2)}))\right]^2}{\operatorname{Var}(x_1) \cdot \operatorname{Var}(L(\underset{\sim}{x}^{(2)}))}$$

$$= \frac{\left[\operatorname{cov}\left(x_1, l_0 + \sum_{i=2}^{p} l_i x_i\right)\right]^2}{s_{11} \cdot \operatorname{Var}\left(l_0 + \sum_{i=2}^{p} l_i x_i\right)}$$

$$= \frac{\left[\sum_{i=2}^{p} l_i \cdot \operatorname{cov}(x_1, x_i)\right]^2}{s_{11} \cdot \sum_{i=2}^{p}\sum_{j=2}^{p} l_i \cdot l_j \operatorname{cov}(x_i, x_j)}$$

$$= \frac{\left(\sum_{i=2}^{p} l_i s_{1i}\right)^2}{s_{11} \cdot \sum_{i=2}^{p}\sum_{j=2}^{p} l_i l_j s_{ij}}$$

$$= \frac{(\underset{\sim}{l}' \underset{\sim}{s}_{(1)})^2}{s_{11} \cdot \underset{\sim}{l}' S_2 \underset{\sim}{l}}$$

$$= \frac{(\underset{\sim}{l}' S_2 \underset{\sim}{b})^2}{s_{11} \cdot \underset{\sim}{l}' S_2 \underset{\sim}{l}} \quad \text{——— (1)}$$

[Since we have $S_2 \underset{\sim}{b} = \underset{\sim}{s}_{(1)}$, from normal equations]

Note that, S_2 is p.d. matrix.

Thus, for some non-singular matrix P, we can write $S_2 = PP^T$.

$\therefore (\underset{\sim}{l}' S_2 \underset{\sim}{b})^2 = (\underset{\sim}{l}' PP' \underset{\sim}{b})^2 = (\underset{\sim}{u}' \underset{\sim}{v})^2$, where $\underset{\sim}{u} = P'\underset{\sim}{l}$ and $\underset{\sim}{v} = P'\underset{\sim}{b}$.

$\leq (\underset{\sim}{u}'\underset{\sim}{u})(\underset{\sim}{v}'\underset{\sim}{v})$ [from Cauchy-Schwarz inequality]

$$\Rightarrow (\underset{\sim}{l}' S_2 \underset{\sim}{b})^2 \leq (\underset{\sim}{l}' PP' \underset{\sim}{l})(\underset{\sim}{b}' PP' \underset{\sim}{b})$$
$$= (\underset{\sim}{l}' S_2 \underset{\sim}{l})(\underset{\sim}{b}' S_2 \underset{\sim}{b})$$

From (1), we can write $r^2\left(x_1, L\left(\underset{\sim}{x}^{(2)}\right)\right) \leq \dfrac{\underset{\sim}{b}' S_2 \underset{\sim}{b}}{s_{11}}$ —— (2)

Note that,

$r^2\left(x_1, M\left(\underset{\sim}{x}^{(2)}\right)\right) = r^2\left(x_1, a + \underset{\sim}{b}'\underset{\sim}{x}^{(2)}\right)$, where $M\left(\underset{\sim}{x}^{(2)}\right) = a + \underset{\sim}{b}'\underset{\sim}{x}^{(2)} = a + \sum_{i=2}^{p} b_i x_i$ is the multiple linear regression equation.

$$= \frac{(\underset{\sim}{b}' S_2 \underset{\sim}{b})^2}{s_{11} \cdot (\underset{\sim}{b}' S_2 \underset{\sim}{b})}$$

$$= \frac{\underset{\sim}{b}' S_2 \underset{\sim}{b}}{s_{11}} \quad \text{[from (1)]}$$

From (2), $r^2\left(x_1, L\left(\underset{\sim}{x}^{(2)}\right)\right) \leq r^2\left(x_1, M\left(\underset{\sim}{x}^{(2)}\right)\right)$.

Hence, the multiple linear regression equation $M\left(\underset{\sim}{x}^{(2)}\right)$ has the maximum correlation with x_1.

<u>Multiple correlation coefficient:</u> the maximum correlation coefficient between any linear function of x_1 and $\underset{\sim}{x}^{(2)}$ is known as the multiple correlation coefficient between x_1 and $\underset{\sim}{x}^{(2)}$ and is denoted by

$$r_{1,23\cdots p} = +\sqrt{r^2\left(x_1, M\left(\underset{\sim}{x}^{(2)}\right)\right)}$$

Now, $r_{1.23\cdots p} = +\sqrt{r^2\left(x_1, M\left(\underset{\sim}{x}^{(2)}\right)\right)}$

$$= +\sqrt{\frac{\underset{\sim}{b}' S_2 \underset{\sim}{b}}{s_{11}}}$$

$$= +\sqrt{\frac{Var\left(M\left(\underset{\sim}{x}^{(2)}\right)\right)}{Var(x_1)}}$$

$$= \frac{s.d.\left(M\left(\underset{\sim}{x}^{(2)}\right)\right)}{s.d.(x_1)}$$

$$= \frac{s.d.(X_{1.23\ldots p})}{s.d.(x_1)} \geq 0$$

Clearly, $0 \leq r_{1.23\ldots p} \leq 1$, since $r^2_{1.23\ldots p}$ is nothing but the product moment correlation coefficient between x_1 and $M\left(\underset{\sim}{x}^{(2)}\right)$, hence $r^2_{1.23\ldots p} \leq 1$.

Theorem: Define $Var(e_{1.23\ldots p}) = s^2_{1.23\ldots p}$. Then,

$$r^2_{1.23\ldots p} = 1 - \frac{s^2_{1.23\ldots p}}{s_{11}}.$$

Proof:

$$r^2_{1.23\ldots p} = r^2\left(x_1, M(\underset{\sim}{x}^{(2)})\right)$$

$$= \frac{var(X_{1.23\ldots p})}{var(x_1)}$$

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

$$= \frac{Var(x_1) + var(e_{1.23\ldots p}) - 2cov(x_1, e_{1.23\ldots p})}{var(x_1)}$$

$$= \frac{var(x_1) + var(e_{1.23\ldots p}) - 2cov(x_1 - X_{1.23\ldots p}, e_{1.23\ldots p})}{var(x_1)}$$

[Since $e_{1.23\ldots p}$ is uncorrelated with the multiple linear regression equation $X_{1.23\ldots p}$]

$$= \frac{var(x_1) + var(e_{1.23\ldots p}) - 2.cov(e_{1.23\ldots p}, e_{1.23\ldots p})}{var(x_1)}$$

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

$$= 1 - \frac{s^2_{1.23\ldots p}}{s}$$

Remark: 1. $s^2_{1\cdot 23\ldots p} = s_{11} - \underset{\sim}{s}'_{(1)} S_2^{-1} \underset{\sim}{s}_{(1)}$, then

$$r^2_{1\cdot 23\ldots p} = 1 - \frac{s_{11} - \underset{\sim}{s}'_{(1)} S_2^{-1} \underset{\sim}{s}_{(1)}}{s_{11}}$$

$$= \frac{\underset{\sim}{s}'_{(1)} S_2^{-1} \underset{\sim}{s}_{(1)}}{s_{11}}$$

$$= 1 - \frac{1}{r^{11}}$$

2. $s^2_{1\cdot 23\ldots p} = s_{11}\left(1 - r^2_{1\cdot 23\ldots p}\right)$

3. $r^2_{1\cdot 23\ldots p} = 1$ iff $s^2_{1\cdot 23\ldots p} = 0$

i.e., iff $var(e_{1\cdot 23\ldots p}) = 0$

i.e., x_1 is totally explained by the multiple linear regression equation

and, $r^2_{1\cdot 23\ldots p} = 0$ iff $\frac{\underset{\sim}{b}' S_2 \underset{\sim}{b}}{s_{11}} = 0$

i.e., iff $\underset{\sim}{k} = \underset{\sim}{0}$

i.e., the multiple linear regression equation fails to predict x_1. $\left[M\left(\underset{\sim}{x}^{(2)}\right) = a + \underset{\sim}{0}'\underset{\sim}{x}^{(2)} = a\right]$

Note that, $r^2_{1\cdot 23\ldots p} = \frac{var(X_{1\cdot 23\ldots p})}{Var(x_1)}$ is the proportion of variability of x_1 explained by the multiple linear regression of x_1 on $\underset{\sim}{x}^{(2)}$.

Multiple Correlation:— Having a fitted multiple regression equation of the response variation x_1 on the predictor variables $x_2, x_3, \ldots, p$ we may be interested in studying the precision of the predicted values of x_1. Common sense indicates that the smaller the prediction errors, the higher the correlation between the observed and predicted values of x_1 and vice versa. It is to be noted that higher joint influence of $x_2, \ldots, x_p$ on x_1, means smaller prediction errors. So, we may consider the product moment correlation coefficient between x_1 and $X_{1.23\ldots p}$ as a measure of the joint influence of $x_2, x_3, \ldots, x_p$ on x_1 via linear regression. This measure is called the multiple correlation coefficient of x_1 on $x_2, x_3, \ldots, x_p$ and is denoted by $r_{1.23\ldots p}$, giving

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

$$= \frac{Var(X_{1.23\ldots p})}{\sqrt{Var(x_1)}\sqrt{Var(X_{1.23\ldots p})}}$$

$$= \sqrt{Var(X_{1.23\ldots p})/Var(x_1)}$$

$$= s_1\left(1-\frac{|R|}{R_{11}}\right)^{1/2}\cdot\frac{1}{s_1}, \text{as } Var(X_{1.23\ldots p}) = s_1^2\left(1-\frac{|R|}{R_{11}}\right)$$

$$= \left(1-\frac{|R|}{R_{11}}\right)^{1/2}$$

Some Useful Results:—

(1) $0 \le r_{1.23\ldots p} \le 1$.

Proof:— The coefficient $r_{1.23\ldots p}$, being the product moment correlation coefficient between x_1 and $X_{1.23\ldots p}$, must lies between -1 and $+1$. But the covariance between x_1 and $X_{1.23\ldots p}$ is equal to the variance of $X_{1.23\ldots p}$, giving that the covariance$(x_1, X_{1.23\ldots p})$ is a non-negative quantity. Hence, we always have

$$0 \le r_{1.23\ldots p} \le 1$$

(2)

$\text{Cov}(x_i, x_{1\cdot 23\ldots p}) = 0$ for $i = 2, 3, \ldots, p$, giving that
$\text{Cov}(X_{1\cdot 23\ldots p}, x_{1\cdot 23\ldots p}) = 0$

Proof:- Consider

$$n\,\text{Cov}(x_2, x_{1\cdot 23\ldots p}) = \sum_{\alpha=1}^{n} x_{2\alpha} \cdot x_{1\cdot 23\ldots p\alpha} \quad \text{as } \bar{x}_{1\cdot 23\ldots p} = 0$$

$$= \sum_{\alpha=1}^{n} x_{2\alpha}\left(x_{1\alpha} - X_{1\cdot 23\ldots p\alpha}\right)$$

$$= \sum_{\alpha} x_{2\alpha}\left\{x_{1\alpha} - \left(a + b_2 x_{2\alpha} + b_3 x_{3\alpha} + \cdots + b_p x_{p\alpha}\right)\right\},$$

a and b_i are the solutions of the normal equation in (2).

$$= \sum_{\alpha} x_{2\alpha} x_{1\alpha} - \left[a\sum_{\alpha} x_{2\alpha} + b_2 \sum_{\alpha} x_{2\alpha}^2 + b_3 \sum_{\alpha} x_{2\alpha} x_{3\alpha} + \cdots + b_p \sum_{\alpha} x_{2\alpha} x_{p\alpha}\right]$$

$= 0$, because of the 2nd normal equation in (2).

Similarly, for other $i = 3, \ldots, p$, giving that the residual part $x_{1\cdot 23\ldots p}$ is uncorrelated with each of the predictor variables x_i, $i = 2, 3, \ldots, p$. So, $X_{1\cdot 23\ldots p}$ being a linear function of $x_2, x_3, \ldots, x_p$, itself is uncorrelated with $x_{1\cdot 23\ldots p}$.

(3) $s^2_{1\cdot 23\ldots p} = \left(1 - r^2_{1\cdot 23\ldots p}\right) s_1^2$, where $s^2_{1\cdot 23\ldots p} = \text{Var}(x_{1\cdot 23\ldots p})$.

Proof:- We have $x_1 = X_{1\cdot 23\ldots p} + x_{1\cdot 23\ldots p}$, giving

$\text{Var}(x_1) = \text{Var}(X_{1\cdot 23\ldots p}) + \text{Var}(x_{1\cdot 23\ldots p})$; because $X_{1\cdot 23\ldots p}$ and $x_{1\cdot 23\ldots p}$ are uncorrelated

or, $s_1^2 = s_1^2\left(1 - \frac{|R|}{R_{11}}\right) + s^2_{1\cdot 23\ldots p}$.

or, $s^2_{1\cdot 23\ldots p} = \frac{|R|}{R_{11}} s_1^2$ $\quad \left[\because r_{1\cdot 23\ldots p} = \left(1 - \frac{|R|}{R_{11}}\right)^{1/2}\right]$

$= \left(1 - r^2_{1\cdot 23\ldots p}\right) s_1^2$

The equation indicates that the residual variance $s^2_{1\cdot 23\ldots p}$ is a strictly decreasing function of the multiple correlation coefficient $r_{1\cdot 23\ldots p}$ that ranges from 0 to 1.

When $r_{1.23\cdots p} = 1$, $s^2_{1.23\cdots p} = 0$, implying that $x_{1\alpha} = X_{1.23\cdots p\alpha}$ for each α and in this case the multiple regression equation may be viewed as a perfect predicting formula.

When $r_{1.23\cdots p} = 0$, then $s^2_{1.23\cdots p} = s_1^2$, giving that $Var(X_{1.23\cdots p}) = 0$, which indicates that $X_{1.23\cdots p\alpha} = \bar{x}_1$ for each α, an equation independent of $x_2, x_3, \ldots, x_p$ and hence the equation fails completely as a predicting formula.

Conclusion:– So, $r_{1.23\cdots p}$ may be used as a measure of the efficiency of the multiple regression equation in predicting x_1. The quantity $r^2_{1.23\cdots p}$, which is called the co-efficient of determination for the regression equation, may also be taken as such a measure.

(4) $r^2_{1.23\cdots p} = 1 - \dfrac{Var(x_{1.23\cdots p})}{Var(x_1)}$.

Proof:– We have already proved that $Cov(x_1, X_{1.23\cdots p}) = Var(X_{1.23\cdots p})$ which gives that

$$r^2_{1.23\cdots p} = \frac{Var(X_{1.23\cdots p})}{Var(x_1)}$$

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

$$= 1 - \frac{Var(x_{1.23\cdots p})}{Var(x_1)}$$

Partial Correlation :—

Sometimes two variables x_1 and x_2 are correlated due to the effect of a 3rd variable x_3 on either or both x_1 and x_2. In these cases to study the relationship between x_1 and x_2, it may be desirable to calculate the correlation between x_1 and x_2 after eliminating (or partialling out) the effect of the third variable x_3. This corelation is called 'Partial Correlation' (or, net correlation) between x_1 and x_2 eleminating the effect of x_3. As an example, to understand whether the relationship between sales and advertising-expenditure is strong or not, one may calculate the partial correlation between sales and advertising-expenditure eliminating the effect of price. We generalise this notion for p variables, and consider that x_1 and x_2 are correlated due to the influence of a group of $(p-2)$ variables $x_3, x_4, \dots, x_p$, on both x_1 and x_2. And study the partial correlation between x_1 and x_2, eliminating the effects $x_3, x_4, \dots, x_p$.

Considering the least-square regression equation of x_1 on $x_3, x_4, \dots, x_p$ and that of x_2 on $x_3, x_4, \dots, x_p$; we may write:

$$x_1 = X_{1\cdot 34\dots p} + x_{1\cdot 34\dots p}$$

and, $$x_2 = X_{2\cdot 34\dots p} + x_{2\cdot 34\dots p}$$

Here, $X_{1\cdot 34\dots p}$ and $X_{2\cdot 34\dots p}$ are predicted values and $x_{1\cdot 34\dots p}$ and $x_{2\cdot 34\dots p}$ are errors in prediction. As both $x_{1\cdot 34\dots p}$ and $x_{2\cdot 34\dots p}$ are uncorrelated with the predictor variables $x_3, x_4, \dots, x_p$; these errors may be looked upon as the parts of x_1 and x_2, respectively, which are free from the influence of the group of variables $x_3, \dots, x_p$. Hence the simple correlation coefficient between $x_{1\cdot 34\dots p}$ and $x_{2\cdot 34\dots p}$ may be considered as a measure of partial correlation between x_1 and x_2, eliminating the effect of $x_3, x_4, \dots, x_p$. It is known as Partial Correlation coefficient and is denoted by $r_{12\cdot 34\dots p}$.

Thus, assuming $Var(x_{1\cdot 34\cdots p}) > 0$ and $Var(x_{2\cdot 34\cdots p}) > 0$, so that R_{11} and R_{22} are both positive definite, we have

$$r_{12\cdot 34\cdots p} = \frac{Cov(x_{1\cdot 34\cdots p}, x_{2\cdot 34\cdots p})}{\sqrt{Var(x_{1\cdot 34\cdots p})\, Var(x_{2\cdot 34\cdots p})}} \qquad \text{—— (1)}$$

According to our notation,

$$x_{1\cdot 34\cdots p} = x_1 - X_{1\cdot 34\cdots p} = (x_1 - \bar{x}_1) + \sum_{j=3}^{p} \frac{s_1}{s_j} \times \frac{R_{1j}^{(2)}}{R_{11}^{(2)}} (x_j - \bar{x}_j);$$

where, $R_{ij}^{(2)}$ is the co-factor of r_{ij} in $R^{(2)}$, the determinant obtained from R by deleting the 2nd row and the 2nd column.

$\therefore Cov(x_{1\cdot 34\cdots p}, x_{2\cdot 34\cdots p})$

$= Cov(x_1, x_{2\cdot 34\cdots p}) + 0$ as $Cov(x_j, x_{2\cdot 34\cdots p}) = 0$ for $j = 3, 4, \ldots, p$.

$$= Cov(x_1, x_2) + \sum_{j=3}^{p} \frac{s_2}{s_j} \times \frac{R_{2j}^{(1)}}{R_{22}^{(1)}} \cdot Cov(x_1, x_j)$$

$$= r_{12} s_1 s_2 + \sum_{j=3}^{p} \frac{s_2}{s_j} \times \frac{R_{2j}^{(1)}}{R_{22}^{(1)}} \times r_{1j} \cdot s_1 s_j$$

$$= s_1 s_2 \left(r_{12} + \sum_{j=3}^{p} r_{1j} \cdot \frac{R_{2j}^{(1)}}{R_{22}^{(1)}} \right)$$

$$= \frac{s_1 s_2}{R_{22}^{(1)}} \left(\sum_{j=2}^{p} r_{1j} R_{2j}^{(1)} \right)$$

$$= -s_1 s_2 \cdot \frac{R_{12}}{R_{22}^{(1)}}; \qquad \text{—— (2)}$$

because, $\sum_{j=2}^{p} r_{1j} R_{2j}^{(1)}$ = determinant of the matrix obtained from $R^{(1)}$ by replacing its first row by $(r_{12}, r_{13}, \ldots, r_{1p})$

$$= \begin{vmatrix} r_{12} & r_{13} & \cdots & r_{1p} \\ r_{32} & r_{33} & \cdots & r_{3p} \\ \vdots & \vdots & & \vdots \\ r_{p2} & r_{p3} & \cdots & r_{pp} \end{vmatrix}$$

= Minor of [redacted] r_{21} in R

= Minor of r_{12} in R

$= -R_{12}$

Further, similar to the result

$$Var(x_{1\cdot 23\cdots p}) = \frac{|R|}{R_{11}} \cdot s_1^2 ,$$

we have,

$$\left. \begin{aligned} Var(x_{1\cdot 34\cdots p}) &= \frac{|R^{(2)}|}{R_{11}^{(2)}} \cdot s_1^2 \text{ and} \\ Var(x_{2\cdot 34\cdots p}) &= \frac{|R^{(1)}|}{R_{22}^{(1)}} \cdot s_2^2 \end{aligned} \right\} \text{———} (3)$$

From (1), (2), (3); we get $\longrightarrow$

$$r_{12\cdot 34\cdots p} = -s_1 s_2 \frac{R_{12}}{R_{22}^{(1)}} \times \left(\frac{R_{11}^{(2)} R_{22}^{(1)}}{|R^{(2)}|\,|R^{(1)}|} \right)^{1/2} \cdot \frac{1}{s_1 s_2}$$

$$= -\frac{R_{12}}{R_{22}^{(1)}} \times \left(\frac{R_{22}^{(1)} R_{22}^{(1)}}{R_{11} R_{22}} \right)^{1/2}$$

$$= -\frac{R_{12}}{\sqrt{R_{11} R_{22}}}$$

as $|R^{(1)}| = R_{11}$, $|R^{(2)}| = R_{22}$, $R_{11}^{(2)} = R_{22}^{(1)}$.

Unlike the multiple correlation coefficient $r_{1\cdot 23\cdots p}$, the partial correlation coefficient $r_{12\cdot 34\cdots p}$ lies in $[-1, 1]$.

Particular Case:- $(p = 3)$

For 3 variables,

$$R_{3\times 3} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, \text{ which gives}$$

$$-R_{12} = \begin{vmatrix} r_{21} & r_{23} \\ r_{31} & r_{33} \end{vmatrix} = r_{12} - r_{13} r_{23},$$

$$R_{11} = r_{22} - r_{23}^2 = 1 - r_{23}^2$$

and, $R_{22} = 1 - r_{13}^2$

So,

$$r_{12\cdot 3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{1 - r_{13}^2}\,\sqrt{1 - r_{23}^2}}.$$

Some Useful Results:-

(a) $b_{12\cdot 34\cdots p} = r_{12\cdot 34\cdots p} \cdot \dfrac{s_{1\cdot 23\cdots p}}{s_{2\cdot 134\cdots p}}$,

Proof:- We know that

$$b_{12\cdot 34\cdots p} = -\frac{R_{12}}{R_{11}} \times \frac{s_1}{s_2}$$

$$r_{12\cdot 34\cdots p} = \frac{-R_{12}}{\sqrt{R_{11}R_{22}}}$$

$$s^2_{1\cdot 234\cdots p} = \frac{|R|}{R_{11}} \cdot s_1^2$$

$$s^2_{2\cdot 134\cdots p} = \frac{|R|}{R_{22}} \cdot s_2^2$$

Hence,
$$r_{12\cdot 34\cdots p} \cdot \frac{s_{1\cdot 23\cdots p}}{s_{2\cdot 13\cdots p}} = -\frac{R_{12}}{\sqrt{R_{11}R_{22}}} \times \frac{|R|^{1/2} s_1}{R_{11}^{1/2}} \times \frac{R_{22}^{1/2}}{|R|^{1/2} s_2}$$

$$= -\frac{R_{12}}{R_{11}} \times \frac{s_1}{s_2} = b_{12\cdot 34\cdots p}$$

(b) $b_{12\cdot 34\cdots p} = r_{12\cdot 34\cdots p} \times \dfrac{s_{1\cdot 34\cdots p}}{s_{2\cdot 34\cdots p}}$

[It is a relation of the same form as $b_{12} = r_{12} \cdot \dfrac{s_1}{s_2}$.]

Proof:- As
$$s^2_{1\cdot 34\cdots p} = \frac{|R^{(2)}|}{R_{11}^{(2)}} s_1^2 \quad \text{and}$$

$$s^2_{2\cdot 34\cdots p} = \frac{|R^{(1)}|}{R_{22}^{(1)}} s_2^2 ,$$

$$r_{12\cdot 34\cdots p} \times \frac{s_{1\cdot 34\cdots p}}{s_{2\cdot 3\cdots p}} = -\frac{R_{12}}{\sqrt{R_{11}R_{22}}} \times \frac{|R^{(2)}|^{1/2}}{\left(R_{11}^{(2)}\right)^{1/2}} \cdot s_1 \times \frac{\left(R_{22}^{(1)}\right)^{1/2}}{|R^{(1)}|^{1/2} s_2}$$

$$= -\frac{s_1}{s_2} \times \frac{R_{12}}{\sqrt{R_{11}R_{22}}} \times \frac{R_{22}^{1/2}}{R_{11}^{1/2}}$$

$$= -\frac{s_1}{s_2} \times \frac{R_{12}}{R_{11}}$$

$$= b_{12\cdot 34\cdots p} .$$

because, $|R^{(1)}| = R_{11}$, $|R^{(2)}| = R_{22}$, $R_{11}^{(2)} = R_{22}^{(1)}$

Note:- This result may be expressed as

$$b_{12\cdot34\cdots p} = r_{12\cdot34\cdots p}\cdot\frac{s_{1\cdot34\cdots p}}{s_{2\cdot34\cdots p}}$$

$$= \frac{\mathrm{Cov}(x_{1\cdot34\cdots p}, x_{2\cdot34\cdots p})}{s_{1\cdot34\cdots p}\cdot s_{2\cdot34\cdots p}} \times \frac{s_{1\cdot34\cdots p}}{s_{2\cdot34\cdots p}}$$

$$= \frac{\mathrm{Cov}(x_{1\cdot34\cdots p}, x_{2\cdot34\cdots p})}{s^2_{2\cdot34\cdots p}}$$

or, $\mathrm{Cov}(x_{1\cdot34\cdots p}, x_{2\cdot34\cdots p}) = s^2_{2\cdot34\cdots p}\cdot b_{12\cdot34\cdots p}$

(c) $b_{12\cdot34\cdots p}\cdot b_{21\cdot34\cdots p} = r^2_{12\cdot34\cdots p}$, similar to the result $b_{12}\,b_{21} = r_{12}^2$.

Proof:- We have already proved that

$$b_{12\cdot34\cdots p} = r_{12\cdot34\cdots p}\cdot\frac{s_{1\cdot34\cdots p}}{s_{2\cdot34\cdots p}}$$

This implies, $b_{21\cdot34\cdots p} = r_{12\cdot34\cdots p}\cdot\dfrac{s_{2\cdot34\cdots p}}{s_{1\cdot34\cdots p}}$

Thus, $b_{12\cdot34\cdots p}\,b_{21\cdot34\cdots p} = r^2_{12\cdot34\cdots p}$.

here $b_{12\cdot34\cdots p}$, $b_{21\cdot34\cdots p}$ and $r^2_{12\cdot34\cdots p}$ have the same sign.

(d) (i) $s^2_{1.23\ldots p} = (1 - r^2_{1p.23\ldots \overline{p-1}}) \cdot s^2_{1.23\ldots \overline{p-1}}$

(ii) $r_{1.23\ldots p} \geq r_{1.23\ldots \overline{p-1}}$

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

Proof :- $\text{Var}(x_{1.23\ldots p}) = \text{Cov}(x_{1.23\ldots p}, x_{1.23\ldots p})$

$$= \text{Cov}\left[x_{1.23\ldots \overline{p-1}} - \sum_{j=2}^{p-1} c_j (x_j - \bar{x}_j) - b_{1p.23\ldots \overline{p-1}}(x_p - \bar{x}_p),\ x_{1.23\ldots p}\right],$$

where, $c_j = b_{1j.23\ldots \overline{j-1}\,\overline{j+1}\ldots p} - b_{1j.23\ldots \overline{j-1}\,\overline{j+1}\ldots \overline{p-1}}$.

$= \text{Cov}(x_{1.23\ldots \overline{p-1}}, x_{1.23\ldots p})$ as $\text{Cov}(x_{1.23\ldots p}, x_j) = 0 \ \forall j = 2(1)p$.

$= \text{Cov}(x_{1.23\ldots \overline{p-1}}, x_{1.23\ldots \overline{p-1}}) - b_{1p.23\ldots \overline{p-1}} \text{Cov}(x_{1.23\ldots \overline{p-1}}, x_p)$

as $\text{Cov}(x_{1.23\ldots \overline{p-1}}, x_j) = 0 \ \forall j = 2(1)\overline{p-1}$.

$= \text{Var}(x_{1.23\ldots \overline{p-1}}) - b_{1p.23\ldots \overline{p-1}} \text{Cov}(x_{1.23\ldots \overline{p-1}}, x_{p.23\ldots \overline{p-1}})$,

as $x_{p.23\ldots \overline{p-1}} = (x_p - \bar{x}_p) + \sum_{j=2}^{p-1} b_{pj.23\ldots \overline{j-1}\,\overline{j+1}\ldots \overline{p-1}} (x_j - \bar{x}_j)$

giving $\text{Cov}(x_{1.23\ldots \overline{p-1}}, x_{p.23\ldots \overline{p-1}}) = \text{Cov}(x_{1.23\ldots \overline{p-1}}, x_p)$

$= \text{Var}(x_{1.23\ldots \overline{p-1}}) - b_{1p.23\ldots \overline{p-1}}\, b_{p1.23\ldots \overline{p-1}} \cdot s^2_{1.23\ldots \overline{p-1}}$

$\therefore$ $s^2_{1.23\ldots p} = s^2_{1.23\ldots \overline{p-1}} - r^2_{1p.23\ldots p-1} \cdot s^2_{1.23\ldots \overline{p-1}}$

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

Hence (i) is proved.

Scanned by CamScanner

Equation ① gives:—

$s^2_{1\cdot 23\cdots p} \le s^2_{1\cdot 23\cdots \overline{p-1}}$ —② , because $0 \le r^2_{1p\cdot 23\cdots \overline{p-1}} \le 1$.

or, $(1-r^2_{1\cdot 23\cdots p}) s_1^2 \le (1-r^2_{1\cdot 23\cdots \overline{p-1}}) s_1^2$, because $s^2_{1\cdot 23\cdots p} = (1-r^2_{1\cdot 23\cdots p}) s_1^2$

or, $r^2_{1\cdot 23\cdots p} \ge r^2_{1\cdot 23\cdots \overline{p-1}}$

or, $r_{1\cdot 23\cdots p} \ge r_{1\cdot 23\cdots \overline{p-1}}$ because $0 \le r_{1\cdot 23\cdots p} \le 1$ —③

Hence, (ii) is proved.

Note:— Inequalities ② & ③ indicate that by introducing an addition predictor variable in the multiple regression equation, one may expect to improve its usefulness as a predicting formula.

Applying ① successively to $s^2_{1\cdot 23\cdots \overline{p-1}}, s^2_{1\cdot 23\cdots \overline{p-2}}, \cdots, s^2_{1\cdot 2}$; we get

$$s^2_{1\cdot 23\cdots p} = (1-r^2_{1p\cdot 23\cdots \overline{p-1}})\, s^2_{1\cdot 23\cdots \overline{p-1}}$$

$$= (1-r^2_{1p\cdot 23\cdots \overline{p-1}})(1-r^2_{1\overline{p-1}\cdot 23\cdots \overline{p-2}})\, s^2_{1\cdot 23\cdots \overline{p-2}}$$

$$\vdots$$

$$= (1-r^2_{1p\cdot 23\cdots \overline{p-1}})(1-r^2_{1\overline{p-1}\cdot 23\cdots \overline{p-2}})\cdots\cdots(1-r^2_{13\cdot 2})\, s^2_{1\cdot 2}$$

$$= (1-r^2_{1p\cdot 23\cdots \overline{p-1}})(1-r^2_{1\overline{p-1}\cdot 23\cdots \overline{p-2}})\cdots(1-r^2_{13\cdot 2})(1-r^2_{12})\, s_1^2$$

or, $(1-r^2_{1\cdot 23\cdots p}) = (1-r^2_{12})(1-r^2_{13\cdot 2})\cdots\cdots(1-r^2_{1\overline{p-1}\cdot 23\cdots \overline{p-2}})(1-r^2_{1p\cdot 23\cdots \overline{p-1}})$

Hence (iii) is proved.

(e) If $a_1x_1 + a_2x_2 + \cdots\cdots + a_px_p = k$, then what will be the partial correlation coefficient of $(p-2)$ orders? What will be the multiple correlation coefficient $r_{1\cdot 23\cdots p}$?

[The variables, being linearly related, S is singular (p.s.d.)
i.e. $|S| = 0$.
$\Rightarrow r_{1\cdot 23\cdots p} = 1 - \frac{|S|}{|S_2|} = 1$.]

Partial correlation: In multivariate data analysis, the study of the degree to which two variables, say x_1 and x_2, may be related when the influence of the other variables, $x_3, x_4, \ldots, x_p$ is eliminated from both of them is of interest. Though the study is concerned about two primary variables x_1 and x_2 but the other $p-2$ variables are also taken into consideration because of their possible relationship with x_1 and x_2. In practice, we usually eliminate the linear effect of $(x_3, x_4, \ldots, x_p)$ from x_1 and x_2.

Let us define, $X_{1\cdot 34\ldots p}$ be the part of x_1 explained by the multiple linear regression of x_1 on $(x_3, x_4, \ldots, x_p)$ and $X_{2\cdot 34\ldots p}$ be the part of x_2 explained by the multiple linear regression of x_2 on $(x_3, x_4, \ldots, x_p)$. We donte

$$\underset{\sim}{x}^{p\times 1} = (x_1, x_2, \ldots, x_p)' = (x_1, x_2, \underset{\sim}{x}^{(3)'})', \text{ and}$$

$$\bar{\underset{\sim}{x}} = (\bar{x}_1, \bar{x}_2, \bar{\underset{\sim}{x}}^{(3)'})',$$

$$\text{Dispersion matrix } (S) = \begin{pmatrix} s_{11} & s_{12} & \underset{\sim}{s}'_{(13)} \\ s_{12} & s_{22} & \underset{\sim}{s}'_{(23)} \\ \underset{\sim}{s}_{(13)} & \underset{\sim}{s}_{23} & S_3 \end{pmatrix}$$

Again, let $X_{1\cdot 34\ldots p} = a + \underset{\sim}{b}'\underset{\sim}{x}^{(3)}$; where $a = \bar{x}_1 - \underset{\sim}{b}'\bar{\underset{\sim}{x}}^{(3)}$,

$b = S_3^{-1}\underset{\sim}{s}_{(13)}$

also let $X_{2\cdot 34\ldots p} = a^* + \underset{\sim}{b}^{*'}\underset{\sim}{x}^{(3)}$; where $a^* = \bar{x}_2 - \underset{\sim}{b}^{*'}\bar{\underset{\sim}{x}}^{(3)}$,

$b^* = S_3^{-1}\underset{\sim}{s}_{(23)}$.

It may be noted that the residuals $e_{1\cdot 34\ldots p} = x_1 - X_{1\cdot 34\ldots p}$ and $e_{2\cdot 34\ldots p} = x_2 - X_{2\cdot 34\ldots p}$ may be regarded as the parts of x_1 and x_2 uninfluenced by $\underset{\sim}{x}^{(3)}$, the reason is;

$$\text{cov}(e_{1\cdot 34\ldots p}, x_j) = 0 \quad \forall\ j = 3(1)p$$

$$\text{Cov}(e_{2\cdot 34\ldots p}, x_j) = 0 \quad \forall\ j = 3(1)p.$$

<u>Definition:</u> The product moment correlation coefficient between $e_{1\cdot 34\cdots p}$ and $e_{2\cdot 34\cdots p}$ is called the partial correlation coefficient of x_1 and x_2, eliminating the linear effect of $\underset{\sim}{x}^{(3)}$ from both of them and denoted by $r_{12\cdot 34\cdots p}$.

Thus, be definition, $r_{12\cdot 34\cdots p} = \dfrac{\operatorname{cov}(e_{1\cdot 34\cdots p},\, e_{2\cdot 34\cdots p})}{\sqrt{\operatorname{Var}(e_{1\cdot 34\cdots p})}\sqrt{\operatorname{Var}(e_{2\cdot 34\cdots p})}}$

We note that $\operatorname{Var}(e_{1\cdot 34\cdots p}) = \begin{vmatrix} s_{11} & \underset{\sim}{s}'_{(13)} \\ \underset{\sim}{s}_{(13)} & S_3 \end{vmatrix} \Big/ |S_3|$

$$= \frac{\text{co-factor of } s_{22} \text{ in } S}{|S_3|}$$

and, $\operatorname{Var}(e_{2\cdot 34\cdots p}) = \begin{vmatrix} s_{22} & \underset{\sim}{s}'_{(23)} \\ \underset{\sim}{s}_{23} & S_3 \end{vmatrix} \Big/ |S_3|$

$$= \frac{\text{co-factor of } s_{11} \text{ in } S}{|S_3|}$$

Now, $\operatorname{Cov}(e_{1\cdot 34\cdots p},\, e_{2\cdot 34\cdots p}) = \operatorname{cov}(e_{1\cdot 34\cdots p},\, x_2 - X_{2\cdot 34\cdots p})$

$$= \operatorname{cov}(e_{1\cdot 34\cdots p},\, x_2) = \operatorname{cov}(x_1 - X_{1\cdot 34\cdots p},\, x_2)$$

$$= \operatorname{Cov}(x_1, x_2) - \operatorname{Cov}(x_2, X_{1\cdot 34\cdots p})$$

$$= \operatorname{cov}(x_1, x_2) - \operatorname{cov}(x_2,\, a + \underset{\sim}{b}'\underset{\sim}{x}^{(3)})$$

$$= s_{12} - \sum_{i=3}^{p} b_i \cdot \operatorname{Cov}(x_i, x_2)$$

$$= s_{12} - \sum_{i=3}^{p} b_i s_{i2}$$

$$= s_{12} - \underset{\sim}{b}'\underset{\sim}{s}_{(23)} = s_{12} - \underset{\sim}{s}'_{(13)} S_3^{-1} \underset{\sim}{s}_{(23)}$$

$$= \frac{\begin{vmatrix} s_{12} & \underset{\sim}{s}'_{(23)} \\ \underset{\sim}{s}_{(23)} & S_3 \end{vmatrix}}{|S_3|}$$

$$= (-1)^{2+1} \cdot \frac{\text{cofactor of } s_{21} \text{ in } S}{|S_3|}$$

$$\therefore\ r_{12\cdot 34\cdots p} = \frac{-\text{cofactor of } s_{21} \text{ in } S}{\sqrt{(\text{cofactor of } s_{22} \text{ in } S)\cdot(\text{cofactor of } s_{11} \text{ in } S)}}$$

$$= \frac{-S_{21}}{\sqrt{S_{11}\cdot S_{22}}}, \text{ where } S_{ij} \text{ is the cofactor of } s_{ij} \text{ in } S.$$

$$= -\frac{S_{21}/|S|}{\sqrt{\frac{S_{11}}{|S|}\cdot\frac{S_{22}}{|S|}}}$$

$$= -\frac{s^{21}}{\sqrt{s^{11}\cdot s^{22}}}, \text{ where } S^{-1} = ((s^{ij})) = \left(\left(\frac{S_{ij}}{|S|}\right)\right)$$

Problem: Let $\underset{\sim}{x}$ be a vector variable with mean $\underset{\sim}{\bar{x}}$ and dispersion matrix S,

$$S = s^2\begin{pmatrix} 1 & r & r & \cdots & r \\ r & 1 & r & \cdots & r \\ \vdots & \vdots & \vdots & & \vdots \\ r & r & r & \cdots & 1 \end{pmatrix}_{p\times p}$$

Show that for any $(\beta_1, \beta_2, \ldots, \beta_p)$,

$$\frac{1}{n}\sum_{\alpha=1}^{n}\left(x_{1\alpha} - \beta_1 - \beta_2 x_{2\alpha} - \cdots - \beta_p x_{p\alpha}\right)^2 \geq s^2\left[1 - \frac{r^2(p-1)}{1+(p-2)r}\right]$$

Solution: $\text{var}(e_{1.23\ldots p}) = s^2_{1.23\ldots p} = \frac{|S|}{|S_2|}$

and $|S| = (s^2)^p (1-r)^{p-1}\{1+(p-1)r\}$

$$S_2 = s^2\begin{pmatrix} 1 & r & r & \cdots & r \\ r & 1 & r & \cdots & r \\ \vdots & \vdots & \vdots & & \vdots \\ r & r & r & & 1 \end{pmatrix}_{\overline{p-1}\times\overline{p-1}} \qquad \therefore |S_2| = (s^2)^{p-1}(1-r)^{p-2} \times \{1+(p-2)r\}$$

By definition of multiple linear regression,

$$\frac{1}{n}\sum_{i=1}^{n}\left(x_{1\alpha} - \beta_1 - \beta_2 x_{2\alpha} - \cdots - \beta_p x_{p\alpha}\right)^2 \geq s^2_{1.23\ldots p}$$

$$\geq \frac{s^2(1-r)\{1+(p-1)r\}}{\{1+(p-2)r\}}$$

Problem: Let $\underset{\sim}{x}^{p\times 1}$ be a vector variable with mean vector $\bar{\underset{\sim}{x}}$ and correlation matrix R. Then show that for any $(b_1, b_2, \ldots, b_p)$,

$$\sum_{\alpha=1}^{n} (x_{1\alpha} - b_1 - b_2 x_{2\alpha} - \cdots - b_p x_{p\alpha})^2 \geq \frac{\sum_{i=1}^{n} (x_{1\alpha} - \bar{x}_1)^2}{r^{11}},$$

where r^{11} is the $(1,1)^{th}$ element of R^{-1} (see next page).

Hints: Consider any linear function $L(\underset{\sim}{x}^{(3)}) = b_1 + b_2 x_2 + \cdots + b_p x_p$, $b_i \in \mathbb{R}\ \forall i$.

Note that, the multiple linear regression of x_1 on $\underset{\sim}{x}^{(3)}$ is obtained by minimizing

$$\sum_{\alpha=1}^{n} \left\{ x_{1\alpha} - L(x_{2\alpha}, x_{3\alpha}, \ldots, x_{p\alpha}) \right\}^2$$

$$= \sum_{\alpha=1}^{n} (x_{1\alpha} - b_1 - b_2 x_{2\alpha} - \cdots - b_p x_{p\alpha})^2$$

By definition of multiple linear regression,

$$\sum_{\alpha=1}^{n} (x_{1\alpha} - b_1 - b_2 x_{2\alpha} - \cdots - b_p x_{p\alpha})^2$$

$$\geq \sum_{\alpha=1}^{n} (x_{1\alpha} - X_{1.23\ldots p,\alpha})^2$$

$$= n.\mathrm{Var}(e_{1.23\ldots p})$$

$$= n.\, s^2_{1.23\ldots p}$$

$$= n.\frac{s_{11}}{r^{11}}.$$

Problem: Find the value of $r_{1.23\ldots p}$ if the independent variables are pairwise uncorrelated.

Solution: Here, the independent variables are $x_2, x_3, \ldots, x_p$.
Since $x_2, x_3, \ldots, x_p$ are pairwise uncorrelated then

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

The correlation matrix is given by

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1p} \\ r_{12} & r_{22} & r_{23} & \cdots & r_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{1p} & r_{2p} & r_{3p} & \cdots & r_{pp} \end{pmatrix}$$

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

$$= \begin{pmatrix} 1 & \underset{\sim}{r}'_{(1)} \\ \underset{\sim}{r}_{(1)} & I_{(p-1)} \end{pmatrix}, \text{ where } \underset{\sim}{r}'_{(1)} = (r_{12}, r_{13}, \ldots, r_{1p})$$

$$\therefore |R| = |I| \, |1 - \underset{\sim}{r}'_{(1)} \cdot I^{-1} \underset{\sim}{r}_{(1)}|$$

$$= 1 - \underset{\sim}{r}'_{(1)} r_{(1)}$$

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

$$\therefore r^2_{1.23\ldots p} = 1 - \frac{1}{\frac{R_{11}}{|R|}}$$

$$= 1 - \frac{1}{1 \Big/ \left[1 - \sum_{i=2}^{p} r_{1i}^2\right]}$$

$$\text{or, } r_{1.23\ldots p} = \sqrt{\sum_{i=2}^{p} r_{1i}^2}$$

Partial Regression Coefficient:- If the multiple regression equation of x_1 on $x_2, x_3, \dots, x_p$ is $X_{1\cdot 23\dots p} = a + b_2 x_2 + \dots + b_p x_p$, a linear function of $\underset{\sim}{x}^{(2)} = (x_2, x_3, \dots, x_p)$, then $a = X_{1\cdot 23\dots p}$ when $\underset{\sim}{x}^{(2)} = \underset{\sim}{0}$ and b_j is the amount by which $X_{1\cdot 23\dots p}$ increases for a unit increment in the value of x_j, the other variables $x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_p$ being kept fixed; b_j is called the partial regression coefficient of x_1 on x_j for fixed $x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_p$.

$$\left[\begin{aligned} X_{1\cdot 23\dots p} &= a, \text{ when } \underset{\sim}{x}^{(2)} = \underset{\sim}{0} \\ &= a + b_j, \text{ when } \underset{\sim}{x}^{(2)} = (x_2, \dots, x_j, \dots, x_p) = (0, 0, \dots, 1, 0, \dots, 0) \end{aligned}\right]$$

Notation: 'b_j' is often written more explicitly as $b_{1j\cdot 23\dots(j-1)(j+1)\dots p}$

Formula for b_j: Normal equation: $S_2 \underset{\sim}{b} = \underset{\sim}{s}_{(1)}$

$$\Rightarrow \begin{pmatrix} s_{22} & s_{23} & \cdots & s_{2p} \\ s_{32} & s_{33} & \cdots & s_{3p} \\ \vdots & \vdots & & \vdots \\ s_{p2} & s_{p3} & \cdots & s_{pp} \end{pmatrix} \begin{pmatrix} b_2 \\ b_3 \\ \vdots \\ b_p \end{pmatrix} = \begin{pmatrix} s_{12} \\ s_{13} \\ \vdots \\ s_{1p} \end{pmatrix}$$

We have $b_2 s_{j2} + b_3 s_{j3} + \dots + b_p s_{jp} = s_{1j} \;\forall\; j = 2(1)p$ — (1)

Note that,

$$S_{11} s_{j1} + S_{12} s_{j2} + S_{13} s_{j3} + \dots + S_{1p} s_{jp} = 0, \; j = 2(1)p,$$

where S_{ij} is the cofactor of s_{ij} in S.

i.e., $$\left(-\frac{S_{12}}{S_{11}}\right) s_{j2} + \left(-\frac{S_{13}}{S_{11}}\right) s_{j3} + \dots + \left(-\frac{S_{1p}}{S_{11}}\right) s_{jp} = s_{1j} \quad \text{— (2)}$$

Combining (1) and (2), we get,

$$b_j = -\frac{S_{1j}}{S_{11}}$$

$$= \frac{-\sqrt{s_{11}\, s_{22} \cdots s_{(j-1)(j-1)}}\sqrt{s_{jj}\, s_{(j+1)(j+1)} \cdots s_{pp}}}{s_{22} \cdots s_{jj} \cdots s_{pp} \cdot R_{11}} R_{ij}$$

$$= -\frac{R_{1j}}{R_{11}} \cdot \sqrt{\frac{s_{11}}{s_{jj}}}$$

Result 1. Show that $b_{12\cdot34\cdots p} = r_{12\cdot34\cdots p} \cdot \frac{s_{1\cdot234\cdots p}}{s_{2\cdot134\cdots p}}$.

Proof; $b_{12\cdot34\cdots p} = -\frac{S_{12}}{\sqrt{S_{11}\, S_{22}}} \cdot \sqrt{\frac{S_{22}}{S_{11}}}$

$$= r_{12\cdot34\cdots p} \sqrt{\frac{S_{22}/|S|}{S_{11}/|S|}}$$

$$= r_{12\cdot34\cdots p} \sqrt{\frac{|S|/S_{11}}{|S|/S_{22}}}$$

$$= r_{12\cdot34\cdots p} \frac{s_{1\cdot234\cdots p}}{s_{2\cdot134\cdots p}}$$

Result 2. Show that $b_{12\cdot34\cdots p} = r_{12\cdot34\cdots p} \cdot \frac{s_{1\cdot34\cdots p}}{s_{2\cdot34\cdots p}}$.

Proof: Try yourself.

Problem:- Show that r_{12}, r_{23}, r_{13} must satisfy the inequality

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

Solution:- We know that $|R| \geq 0$

$$\text{i.e.} \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix} \geq 0$$

$$\Rightarrow \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix} \geq 0$$

$$\Rightarrow 1(1 - r_{23}^2) - r_{12}(r_{12} - r_{13}r_{23}) + r_{13}(r_{12}r_{23} - r_{13}) \geq 0.$$

$$\Rightarrow 1 - r_{23}^2 - r_{12}^2 + r_{12}r_{13}r_{23} + r_{12}r_{13}r_{23} - r_{13}^2 \geq 0$$

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

Problem:- Suppose x_1, x_2, x_3 satisfy the relation

$$a_1x_1 + a_2x_2 + a_3x_3 = K.$$

State what partial correlation coefficient will be?

Solution:- Let $Var(x_i) = s_i^2$, $i = 1, 2, 3$.

Now $a_1x_1 + a_2x_2 + a_3x_3 = K$

or, $a_1x_1 + a_2x_2 = K - a_3x_3$

or, $Var(a_1x_1 + a_2x_2) = Var(K - a_3x_3)$

$$\Rightarrow a_1s_1^2 + a_2s_2^2 + 2a_1a_2 \cdot r_{12}s_1s_2 = a_3^2s_3^2$$

$$\Rightarrow r_{12} = \frac{a_3^2s_3^2 - a_1^2s_1^2 - a_2^2s_2^2}{2a_1a_2s_1s_2}$$

Similarly,

$$r_{13} = \frac{a_2^2s_2^2 - a_1^2s_1^2 - a_3^2s_3^2}{2a_1a_3s_1s_3}$$

$$r_{23} = \frac{a_1^2s_1^2 - a_2^2s_2^2 - a_3^2s_3^2}{2a_2a_3s_2s_3}$$

Now $r_{12\cdot3}$ = the product moment correlation coefficient between x_1 and x_2 after eliminating the linear effect of x_3 from both of x_1 and x_2.

= the product moment correlation coefficient between the x_1 and x_2 where x_3 is fixed and $a_1x_1 + a_2x_2 + a_3x_3 = k$.

$$= \frac{Cov(x_1, x_2)}{\sqrt{Var(x_1)}\sqrt{Var(x_2)}}, \text{ where } a_1x_1 + a_2x_2 + a_3x_3 = k$$

$\Rightarrow a_1x_1 + a_2x_2 = k - a_3x_3$

$\Rightarrow a_1x_1 + a_2x_2 = k^*$ (say)

$$= \frac{Cov\left(x_1, \frac{k^*}{a_2} - \frac{a_1}{a_2}x_1\right)}{\sqrt{Var(x_1)}\sqrt{Var\left(\frac{k^*}{a_2} - \frac{a_1}{a_2}x_1\right)}}$$

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

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

∴ Partial correlation coefficient are -1 if a_1, a_2, a_3 are of the same sign.

* ————— *

Multi-variate Analysis.

Multivariate analysis is a branch of Statistics where we study several variables simultaneously.

Notions of Multivariate distn.s.

Here, a p-dimensional random variable $\underset{\sim}{X}$ is a vector : $\underset{\sim}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$

whose elements are unidimensional r.v.'s and $\underset{\sim}{x}$, a realization of $\underset{\sim}{X}$, $\underset{\sim}{x} \in R^p$.

Here $\underset{\sim}{X}$ is discrete, continuous or mixed.

(i) If $\underset{\sim}{X}$ is discrete, the jt pmf of $X_1, \ldots, X_p$ is given by

$$P(X_1 = x_1, \ldots, X_p = x_p) = p(x_1, \ldots, x_p) \text{ or } p(\underset{\sim}{x}), \text{ say.}$$

A function $p(\underset{\sim}{x})$ satisfying the conditions

a) $p(\underset{\sim}{x}) \geq 0 \ \forall \ \underset{\sim}{x} \in R^p$.

b) $\sum_{\underset{\sim}{x}} p(\underset{\sim}{x}) = 1$.

is called the pmf of $\underset{\sim}{X}$.

The marginal pmf of $(X_1, \ldots, X_{p_1})$, $p_1 < p$ is

$$q(x_1, \ldots, x_{p_1}) = \sum p(x_1, \ldots, x_{p_1}, x_{p_1+1}, \ldots, x_p)$$

where the sum is taken over all possible values of $x_{p_1+1}, \ldots, x_p$.

The conditional pmf of $(X_1, \ldots, X_{p_1})$ given $X_{p_1+1} = x_{p_1+1}, \ldots, X_p = x_p$, is.

$$P(X_1 = x_1, \ldots, X_{p_1} = x_{p_1} \mid X_{p_1+1} = x_{p_1+1}, \ldots, X_p = x_p)$$

$$= \frac{p(x_1, \ldots, x_{p_1}, x_{p_1+1}, \ldots, x_p)}{q(x_{p_1+1}, \ldots, x_p)}$$

$$= l(x_1, \ldots, x_{p_1} \mid x_{p_1+1}, \ldots, x_p).$$

(ii) If $\underset{\sim}{X}$ is continuous, then the joint pdf of $X_1, \ldots, X_p$ is $f(x_1, \ldots, x_p)$ say, which satisfies

a) $f(\underset{\sim}{x}) \geq 0 \;\; \forall \; \underset{\sim}{x} \in R^p$

b) $\int_{R^p} f(\underset{\sim}{x}) \, d\underset{\sim}{x} = 1$.

The marginal pdf of $(X_1, \ldots, X_{p_1})$, $p_1 < p$, is $g(x_1, \ldots, x_{p_1}) = \int_{x_{p_1+1}=-\infty}^{\infty} \cdots \int_{x_p=-\infty}^{\infty} f(\underset{\sim}{x}) \, dx_{p_1+1} \cdots dx_p$.

The conditional pdf of $(X_1, \ldots, X_{p_1})$ given $X_{p_1+1} = x_{p_1+1}, \ldots, X_p = x_p$ is

$$h(x_1, \ldots, x_{p_1} \mid x_{p_1+1}, \ldots, x_p) = \frac{f(\underset{\sim}{x})}{g(x_{p_1+1}, \ldots, x_p)}$$

(iii) The cdf of a random vector $\underset{\sim}{X}$ is

$$F(x_1, \ldots, x_p) = P(X_1 \leq x_1, \ldots, X_p \leq x_p).$$

$$= \begin{cases} \sum_{t_1=-\infty}^{x_1} \cdots \sum_{t_p=-\infty}^{x_p} p(t_1, \ldots, t_p) & \text{if } \underset{\sim}{X} \text{ is discrete.} \\ \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f(t_1, \ldots, t_p)\, dt_p \cdots dt_1 & \text{if } \underset{\sim}{X} \text{ is continuous} \end{cases}$$

The marginal cdf of $(X_1, \ldots, X_{p_1})$, $p_1 < p$ is

$G(x_1, \ldots, x_{p_1})$

$$= \lim_{x_{p_1+1} \to +\infty} \cdots \lim_{x_p \to +\infty} F(x_1, \ldots, x_{p_1}, x_{p_1+1}, \ldots, x_p).$$

$$= F(x_1, \ldots, x_{p_1}, \infty, \ldots \infty).$$

If $\underset{\sim}{X}$ is absolutely continuous then

$$\frac{\partial^p F(x_1, \ldots, x_p)}{\partial x_1 \partial x_2 \cdots \partial x_p} = f(x_1, \ldots, x_p), \text{ if } F(\underset{\sim}{x}) \text{ is continuous (almost everywhere)}$$

(iv) The set of variables $(X_1, \ldots, X_{p_1})$ is indept. to the set $(X_{p_1+1}, \ldots, X_p)$ iff

$$F(x_1, \ldots, x_{p_1}, x_{p_1+1}, \ldots, x_p)$$

$$= F(x_1, \ldots, x_{p_1}, \infty, \infty, \ldots, \infty)\, F(\infty, \ldots, \infty, x_{p_1+1}, \ldots x_p)$$

Moments of multidimensional variate Expectation:

Let $\underset{\sim}{X}^{p\times 1}$ denote a column vector of random components X_i, $i = 1(1)p$. Then the expectation of $\underset{\sim}{X}$ is defined as.

$$E(\underset{\sim}{X}) = E\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} = \underset{\sim}{\mu}, \text{ say}.$$

Variance - covariance :-

Define $cov(X_i, X_j) = E[(X_i - E(X_i))(X_j - E(X_j))]$

$= \sigma_{ij}$, say.

If $i=j$ $\sigma_{ii} =$ ~~var(X_i)~~ V

We extend the variance notion to the p-dimensional random vector $\underset{\sim}{X}$ by the following matrix:

$$E[(\underset{\sim}{X} - E(\underset{\sim}{X}))(\underset{\sim}{X} - E(\underset{\sim}{X}))^T]$$

$$= E[(\underset{\sim}{X} - \underset{\sim}{\mu})(\underset{\sim}{X} - \underset{\sim}{\mu})^T]$$

$$= E\left[\begin{pmatrix} X_1 - \mu_1 \\ \vdots \\ X_p - \mu_p \end{pmatrix}(X_1 - \mu_1, \dots, X_p - \mu_p)\right]$$

$$= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & & & \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} = \Sigma, \text{ say}.$$

Here, we assume $E((v_{ij})) = ((E v_{ij}))$

The symmetric matrix Σ is called the variance-covariance matrix or dispersion.

matrix of $\underset{\sim}{X}$.

Moments:

We define

$$E\left(h(X_1,\dots,X_p)\right) = \begin{cases} \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} h(x_1,\dots,x_p)\, f(x_1,\dots,x_p)\, dx_1\dots dx_p & \text{if } \underset{\sim}{X} \text{ is continuous} \\ \sum_{x_1}\cdots\sum_{x_p} h(x_1,\dots,x_p)\, p(x_1,\dots,x_p) & \text{if } \underset{\sim}{X} \text{ is discrete} \end{cases}$$

(*) Now, we find the mean & variance of a linear combination of $X_1,\dots,X_p$ —

$\underset{\sim}{a}'\underset{\sim}{X} = \sum_1^p a_i X_i$ is a linear combination of $X_1,\dots,X_p$.

Now, $E(\underset{\sim}{a}'\underset{\sim}{X})$

$$= E\left(\sum_1^p a_i X_i\right)$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left(\sum_1^p a_i x_i\right) f(x_1,\dots,x_p)\, dx_1\dots dx_p$$

[assuming $\underset{\sim}{X}$ is continuous]

$$= \sum_1^p \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} a_i x_i\, f(x_1,\dots,x_p)\, dx_1\dots dx_p$$

$$= \sum_1^p a_i E(X_i)$$

$$= \sum_1^p a_i \mu_i$$

$$= \underset{\sim}{a}'\underset{\sim}{\mu}$$

[proof of the discrete case is similar]

Now, $V(\underset{\sim}{a}'\underset{\sim}{X})$

$$= E(\underset{\sim}{a}'\underset{\sim}{X} - \underset{\sim}{a}'\underset{\sim}{\mu})^2$$

$$= E\left(\sum_{1}^{p} a_i (x_i - \mu_i)\right)^2$$

$$= E\left\{ \sum_{1}^{p} \sum_{1}^{p} a_i a_j (x_i - \mu_i)(x_j - \mu_j) \right\}$$

$$= \sum_{1}^{p} \sum_{1}^{p} a_i a_j \, E\{(x_i - \mu_i)(x_j - \mu_j)\} \qquad \left[\because E\left(\sum_{1}^{p} a_i x_i\right) = \sum_{1}^{p} a_i E(x_i)\right]$$

$$= \sum_{1}^{p} \sum_{1}^{p} a_i a_j \sigma_{ij}$$

$$= \underset{\sim}{a}' \Sigma \underset{\sim}{a}$$

Remarks:

① The covariance betn $\underset{\sim}{a}'\underset{\sim}{x}$ & $\underset{\sim}{b}'\underset{\sim}{x}$ is

$$\text{cov}(\underset{\sim}{a}'\underset{\sim}{x}, \underset{\sim}{b}'\underset{\sim}{x}) = \text{Cov}\left(\sum a_i x_i, \sum b_i x_i\right)$$

$$= \sum_i \sum_j a_i b_j \sigma_{ij} = \underset{\sim}{a}' \Sigma \underset{\sim}{b}$$

② Consider the transformations

$$\underset{\sim}{y} = A^{r \times p} \underset{\sim}{x}^{p \times 1}, \quad \underset{\sim}{z} = B^{s \times p} \underset{\sim}{x}^{p \times 1}$$

$$\text{Disp}(\underset{\sim}{y}) = \text{cov}(\underset{\sim}{y}, \underset{\sim}{y}) = E\left\{(\underset{\sim}{y} - E(\underset{\sim}{y}))(\underset{\sim}{y} - E(\underset{\sim}{y}))'\right\}$$

$$= E\left\{(A\underset{\sim}{x} - A\underset{\sim}{\mu})(A\underset{\sim}{x} - A\underset{\sim}{\mu})'\right\} \qquad \left[\text{where } E(\underset{\sim}{y}) = E(A\underset{\sim}{x}) = A\underset{\sim}{\mu} \text{ (prove it)}\right]$$

$$= E\left[A(\underset{\sim}{x} - \underset{\sim}{\mu})(\underset{\sim}{x} - \underset{\sim}{\mu})' A'\right]$$

$$= A\, E\left[(\underset{\sim}{x} - \underset{\sim}{\mu})(\underset{\sim}{x} - \underset{\sim}{\mu})'\right] A'$$

$$= A \Sigma A'$$

Similarly, $\text{Disp}(\underset{\sim}{z}) = \text{cov}(\underset{\sim}{z}, \underset{\sim}{z}) = B \Sigma B'$

$\text{cov}(\underset{\sim}{Y}, \underset{\sim}{Z})$

$= A\Sigma B'$ (prove it)

Correlation matrix:

Define $\rho_{ij} = \dfrac{\text{cov}(X_i, X_j)}{\sqrt{V(X_i)\,V(X_j)}}$

Then the matrix of correlation coeffs is

$$R = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{pmatrix}$$

$$= D\left(\frac{1}{\sqrt{\sigma_{ii}}}\right) \Sigma \, D\left(\frac{1}{\sqrt{\sigma_{ii}}}\right)$$

where $D\left(\frac{1}{\sqrt{\sigma_{ii}}}\right) = \begin{pmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{pmatrix}$

Theorem:

Any variance-covariance matrix is n.n.d, and every n.n.d matrix is the variance-covariance matrix of some random vector.

[CU]

<u>proof:</u> <u>First part:</u>

Let $\Sigma^{p\times p} = (\sigma_{ij})$ be the variance-covariance matrix of a random vector $\underset{\sim}{X}^{p\times 1}$.

Note that $V(\underset{\sim}{a}'\underset{\sim}{x}) \geq 0 \quad \forall \underset{\sim}{a}$

$\Leftrightarrow \quad \sum_i \sum_j a_i a_j \sigma_{ij} \geq 0 \quad \forall \underset{\sim}{a}$

$\Leftrightarrow \quad \underset{\sim}{a}' \Sigma \underset{\sim}{a} \geq 0 \quad \forall \underset{\sim}{a}$

$\Rightarrow \quad \Sigma$ is n.n.d.

Remark: ① Σ is psd.

$\Leftrightarrow \quad \underset{\sim}{a}' \Sigma \underset{\sim}{a} = 0$ for some $\underset{\sim}{a} \neq \underset{\sim}{0}$

$\Leftrightarrow \quad Var(\underset{\sim}{a}'\underset{\sim}{x}) = 0$ for some $\underset{\sim}{a} \neq \underset{\sim}{0}$

$\Leftrightarrow \quad \underset{\sim}{a}'\underset{\sim}{x} = c$ for some $\underset{\sim}{a} \neq \underset{\sim}{0}$, with probability unity

$\Leftrightarrow \quad \underset{\sim}{a}'(\underset{\sim}{x} - \underset{\sim}{\mu}) = 0$ for some $\underset{\sim}{a} \neq \underset{\sim}{0}$, "

$[\because c = E(\underset{\sim}{a}'\underset{\sim}{x}) = \underset{\sim}{a}'\underset{\sim}{\mu}]$

If Σ is p.d. then there do not exist $\underset{\sim}{a} \neq \underset{\sim}{0}$ & c such that $\underset{\sim}{a}'\underset{\sim}{x} = c$.

② If $\underset{\sim}{a}'\underset{\sim}{x} = c$ for some $\underset{\sim}{a} \neq \underset{\sim}{0}$ then Σ is p.s.d $\Rightarrow |\Sigma| = 0$ and the distn is called a singular distn ~~is known as~~ (in that case $\rho(\Sigma) < p$).

If $|\Sigma| \neq 0$, the distn is known as nonsingular distn.

2nd part: Let $\Sigma^{p \times p}$ be an n.n.d matrix. Then their exists a matrix $B^{k \times p}$ $(k \leq p)$ such that $\Sigma = B^T B$

Consider a R.V. $\underset{\sim}{Y}^{k \times 1}$ with $E(\underset{\sim}{Y}) = 0$ & $disp(\underset{\sim}{Y}) = I_k$.

Then, $\underset{\sim}{x}^{p\times 1} = \underset{\sim}{\mu}^{p\times 1} + B^T \underset{\sim}{y}$ has the mean vector as

$$E(\underset{\sim}{x}) = \underset{\sim}{\mu} + B^T E(\underset{\sim}{y}) = \underset{\sim}{\mu}$$

& the dispersion matrix

$$\text{Disp}(\underset{\sim}{x}) = \text{Disp}(\underset{\sim}{\mu} + B^T \underset{\sim}{y})$$

$$= \text{Disp}(B^T \underset{\sim}{y}) \quad [\because \text{dispersion is indept of origin}]$$

[$\underset{\sim}{\mu}$ is unnecessary for the proof]

$$= B^T \, \text{Disp}(\underset{\sim}{y}) \, B$$

$$= B^T I_K B$$

$$= B^T B$$

$$= \Sigma$$

Result ①:

Let $\underset{\sim}{x}$ be a R.V. with mean $\underset{\sim}{\mu}$ & dispersion matrix Σ; A is a real matrix, then show that

$$E(\underset{\sim}{x}' A \underset{\sim}{x}) = tr(A\Sigma) + \underset{\sim}{\mu}' A \underset{\sim}{\mu}$$

proof: $E(\underset{\sim}{x}' A \underset{\sim}{x}) = E\left(tr(\underset{\sim}{x}' A \underset{\sim}{x})\right)$

$$= E\left(tr(A \underset{\sim}{x}\underset{\sim}{x}')\right) \quad [\because tr(AB) = tr(BA)]$$

$$= tr\left[A E(\underset{\sim}{x}\underset{\sim}{x}')\right]$$

$$= tr\left[A(\Sigma + \underset{\sim}{\mu}\underset{\sim}{\mu}')\right]$$

$$\left[\because \Sigma = E(\underset{\sim}{x}-\underset{\sim}{\mu})(\underset{\sim}{x}-\underset{\sim}{\mu})' = E(\underset{\sim}{x}\underset{\sim}{x}') - \underset{\sim}{\mu}\underset{\sim}{\mu}'\right]$$

$$= tr(A\Sigma) + tr(A\underset{\sim}{\mu}\underset{\sim}{\mu}')$$

$$= tr(A\Sigma) + tr(\underset{\sim}{\mu}' A \underset{\sim}{\mu})$$

$$= tr(A\Sigma) + \underset{\sim}{\mu}' A \underset{\sim}{\mu}$$

Result (2): In some R.V. $\underset{\sim}{X}^{p\times 1}$, [CU]

$$E\left((\underset{\sim}{X}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{X}-\underset{\sim}{\mu})\right)$$

$$= E\left(tr\left\{(\underset{\sim}{X}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{X}-\underset{\sim}{\mu})\right\}\right)$$

$$= tr\left[\Sigma^{-1} E\left\{(\underset{\sim}{X}-\underset{\sim}{\mu})(\underset{\sim}{X}-\underset{\sim}{\mu})'\right\}\right]$$

$$= tr\left[\Sigma^{-1}\Sigma\right] = tr\, I_p = p$$

Problem:

① S.T. all the characteristic roots of a dispersion matrix of a R.V. are nonnegative.

② S.T. any dispersion matrix Σ can be written as BB where B is nnd.

Hint: $\exists$ an orthogonal matrix P $\ni$

$$P'\Sigma P = \begin{pmatrix} d_1 & & 0 \\ & d_2 & \\ 0 & & \ddots \; d_n \end{pmatrix}$$

Here $d_i \geq 0 \; \forall i$ $\because \Sigma$ is nnd.

$$\Rightarrow \Sigma = P\begin{pmatrix} \sqrt{d_1} & & 0 \\ & \sqrt{d_2} & \\ 0 & & \ddots \; \sqrt{d_n} \end{pmatrix} P'P \begin{pmatrix} \sqrt{d_1} & & 0 \\ & \sqrt{d_2} & \\ 0 & & \ddots \; \sqrt{d_n} \end{pmatrix} P'$$

$$= BB$$

where

$$B = P\begin{pmatrix} \sqrt{d_1} & & 0 \\ & \sqrt{d_2} & \\ 0 & & \ddots \; \sqrt{d_n} \end{pmatrix} P'$$

B is n.n.d since P is n.s. & $\sqrt{d_i} \geq 0 \; \forall i = 1(1)n$.

⊛ Regression theory:

General concept of regression:

The concept of regression is concerned with the prediction of one or more variables $(y_1, \cdots, y_q)$ on the basis of the information provided by other measurements or concomitant variables $(x_1, \cdots, x_p)$. It is customary to call the latter as indept or predictor variables & the former as dept or criterion variables.

We are naturally interested in the question "how should the predictors be chosen?"

Consider a single criterion variable X_1 & $p-1$ indept variables $(X_2, \cdots, X_p)' = \underset{\sim}{X}^{(2)}$
Let $f(\underset{\sim}{X}^{(2)}) = f(X_2, \cdots, X_p)$ be a predictor of X_1.

(i) Minimum MSE predictor: (Theorem 2)

Let $M(\underset{\sim}{X}^{(2)}) = E(X_1 | \underset{\sim}{X}^{(2)})$.
Then $E(X_1 - f(\underset{\sim}{X}^{(2)}))^2$ is minimized when

$$f(\underset{\sim}{X}^{(2)}) = M(\underset{\sim}{X}^{(2)})$$

<u>proof:</u> $E(X_1 - f(\underset{\sim}{X}^{(2)}))^2$

$$= E(X_1 - M(\underset{\sim}{X}^{(2)}))^2 + E(M(\underset{\sim}{X}^{(2)}) - f(\underset{\sim}{X}^{(2)}))^2 + 2E\left\{(X_1 - M(\underset{\sim}{X}^{(2)}))(M(\underset{\sim}{X}^{(2)}) - f(\underset{\sim}{X}^{(2)}))\right\}$$

Now,

$$E\left\{(X_1 - M(\underset{\sim}{X}^{(2)}))(M(\underset{\sim}{X}^{(2)}) - f(\underset{\sim}{X}^{(2)}))\right\}$$

$$= E_{\underset{\sim}{X}^{(2)}} E_{X_1|\underset{\sim}{X}^{(2)}} \left\{(X_1 - M(\underset{\sim}{X}^{(2)}))(M(\underset{\sim}{X}^{(2)}) - f(\underset{\sim}{X}^{(2)}))\right\}$$

$$= E_{\underset{\sim}{X}^{(2)}}\left[\left(M(\underset{\sim}{X}^{(2)}) - f(\underset{\sim}{X}^{(2)})\right) E_{X_1|\underset{\sim}{X}^{(2)}}\left(X_1 - M(\underset{\sim}{X}^{(2)})\right)\right]$$

$$= E_{\underset{\sim}{X}^{(2)}}\left[\left(M(\underset{\sim}{X}^{(2)}) - f(\underset{\sim}{X}^{(2)})\right)\left\{\underbrace{E(X_1|\underset{\sim}{X}^{(2)}) - M(\underset{\sim}{X}^{(2)})}_{0}\right\}\right]$$

$$= 0$$

$$\therefore \quad E(X_1 - f(\underset{\sim}{X}^{(2)}))^2$$

$$= E\left(X_1 - M(\underset{\sim}{X}^{(2)})\right)^2 + E\left(M(\underset{\sim}{X}^{(2)}) - f(\underset{\sim}{X}^{(2)})\right)^2$$

$$\geq E\left(X_1 - M(\underset{\sim}{X}^{(2)})\right)^2$$

∴ The lower bound of $E(X_1 - f(\underset{\sim}{X}^{(2)}))^2$ is attained when $f(\underset{\sim}{X}^{(2)}) = M(\underset{\sim}{X}^{(2)})$, so the best choice of the predictor which minimizes the MSE is $M(\underset{\sim}{X}^{(2)})$, the conditional mean of X_1 given $\underset{\sim}{X}^{(2)}$ is called the regression of X_1 on $\underset{\sim}{X}^{(2)}$.

(ii) Predictor having maximum correlation with the criterion (Theorem 3):

Let $M(\underset{\sim}{X}^{(2)}) = E(X_1 | \underset{\sim}{X}^{(2)})$.

Then $\rho(X_1, M(\underset{\sim}{X}^{(2)}))$ is non-negative &

$\rho(X_1, M(\underset{\sim}{X}^{(2)})) \geq |\rho(X_1, f(\underset{\sim}{X}^{(2)}))|$ for any f^n $f(\underset{\sim}{X}^{(2)})$.

proof: For any f^n $f(\underset{\sim}{X}^{(2)})$,

$\text{cov}(X_1, f(\underset{\sim}{X}^{(2)}))$

$$= E\left[(X_1 - E(X_1))(f(\underset{\sim}{X}^{(2)}) - E(f(\underset{\sim}{X}^{(2)})))\right]$$

$$= E\left[\{f(\underset{\sim}{X}^{(2)}) - E(f(\underset{\sim}{X}^{(2)}))\} X_1\right]$$

$$= E_{\underset{\sim}{X}^{(2)}}\left[E_{X_1|\underset{\sim}{X}^{(2)}}\left[\{f(\underset{\sim}{X}^{(2)}) - E(f(\underset{\sim}{X}^{(2)}))\} X_1\right]\right]$$

$$= E_{\underset{\sim}{X}^{(2)}}\left[\{f(\underset{\sim}{X}^{(2)}) - E(f(\underset{\sim}{X}^{(2)}))\} E_{X_1|\underset{\sim}{X}^{(2)}}(X_1)\right]$$

$$= E_{\underset{\sim}{X}^{(2)}}\left[\{f(\underset{\sim}{X}^{(2)}) - E(f(\underset{\sim}{X}^{(2)}))\} M(\underset{\sim}{X}^{(2)})\right]$$

$$= \operatorname{cov}\left(f(\underset{\sim}{X}^{(2)}), M(\underset{\sim}{X}^{(2)})\right).$$

When $f(\underset{\sim}{X}^{(2)}) = M(\underset{\sim}{X}^{(2)})$,

$$\operatorname{cov}\left(X_1, M(\underset{\sim}{X}^{(2)})\right) = \operatorname{cov}\left(M(\underset{\sim}{X}^{(2)}), M(\underset{\sim}{X}^{(2)})\right)$$

$$= \operatorname{Var}(M(\underset{\sim}{X}^{(2)}))$$

$$= \sigma_M^2 \quad \text{(say)}.$$

Now, $\rho(X_1, M(\underset{\sim}{X}^{(2)}))$

$$= \frac{\operatorname{cov}(X_1, M(\underset{\sim}{X}^{(2)}))}{\sigma_{X_1}\sigma_M} = \frac{\sigma_M^2}{\sigma_{X_1}\sigma_M} = \frac{\sigma_M}{\sigma_{X_1}} \geq 0.$$

Again, $\rho^2(X_1, f(\underset{\sim}{X}^{(2)}))$

$$= \frac{\operatorname{cov}^2(X_1, f(\underset{\sim}{X}^{(2)}))}{\sigma_{X_1}^2 \sigma_f^2}$$

$$= \frac{\text{cov}^2(f(\underset{\sim}{X}^{(2)}), M(\underset{\sim}{X}^{(2)}))}{\sigma_f^2 \sigma_M^2} \cdot \frac{\sigma_M^2}{\sigma_{X_1}^2}$$

$$= \rho^2(f(\underset{\sim}{X}^{(2)}), M(\underset{\sim}{X}^{(2)})) \cdot \rho^2(X_1, M(\underset{\sim}{X}^{(2)}))$$

$$\leq \rho^2(X_1, M(\underset{\sim}{X}^{(2)})) \qquad [\because \rho^2(f, M) \leq 1]$$

$$\Leftrightarrow |\rho(X_1, f(\underset{\sim}{X}^{(2)}))| \leq \rho(X_1, M(\underset{\sim}{X}^{(2)}))$$

Equality holds iff $\rho^2(f, M) = 1$,
ie iff $f(\underset{\sim}{X}^{(2)})$ is a linear f^n of $M(\underset{\sim}{X}^{(2)})$.
Again, the regression of X_1 on $\underset{\sim}{X}^{(2)}$ is the answer.

The maximum correlation of $\rho(X_1, f(\underset{\sim}{X}^{(2)}))$, ie $\rho(X_1, M(\underset{\sim}{X}^{(2)}))$ is called the correlation ratio & it is denoted by $\eta_{X_1 \cdot \underset{\sim}{X}^{(2)}}$.
Clearly $\eta_{X_1 \cdot \underset{\sim}{X}^{(2)}} \leq 1$.

We define $\eta^2_{X_1 \cdot \underset{\sim}{X}^{(2)}} = \frac{\sigma_M^2}{\sigma_{X_1}^2}$ as the correlation ratio.

We observe that

$$\sigma_{X_1}^2 = E(X_1 - E(X_1))^2$$

$$= E(X_1 - M(\underset{\sim}{X}^{(2)}))^2 + E(M(\underset{\sim}{X}^{(2)}) - E(X_1))^2 \quad \text{[Prove it]}$$

$$= \sigma_{1.23\ldots p}^2 + \sigma_M^2$$

$$\Rightarrow \eta^2_{X_1 \cdot \underset{\sim}{X}^{(2)}} = 1 - \frac{\sigma_{1.23\ldots p}^2}{\sigma_{X_1}^2} \leq 1$$

Clearly $\eta^2_{X_1 \cdot \underset{\sim}{X}^{(2)}} \to 1$ as $\sigma^2_{1.23\ldots p} \to 0$
(ie error variance $\to 0$).

& $\eta^2_{x_1 \cdot \underset{\sim}{x}^{(2)}} = 0$, when $\sigma^2_{1.23\ldots p} = \sigma^2_{x_1}$, ie when there is no redution in error due to the use of $M(\underset{\sim}{x}^{(2)})$ as a predicting formula.

The conditional mean of x_1 given $\underset{\sim}{x}^{(2)}$ is called the regression of x_1 on $\underset{\sim}{x}^{(2)}$. The regression is called linear or non linear according as the f^n $M(\underset{\sim}{x}^{(2)}) = E(x_1 | \underset{\sim}{x}^{(2)})$ is linear or not.

Linear Regression:

Let $\underset{\sim}{x}^{p\times 1}$ be a R.V. which is partitioned as $\underset{\sim}{x} = \begin{pmatrix} x_1 \\ \underset{\sim}{x}^{(2)} \end{pmatrix}$ with mean $E(\underset{\sim}{x}) = \begin{pmatrix} \mu_1 \\ \underset{\sim}{\mu}^{(2)} \end{pmatrix}$

& $Disp(\underset{\sim}{x}) = \left(\begin{array}{c|ccc} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \hline \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{array}\right) = \begin{pmatrix} \sigma_{11} & \underset{\sim}{\sigma}_{(1)}' \\ \underset{\sim}{\sigma}_{(1)} & \Sigma_2 \end{pmatrix}$

Assuming that dis^{tn} of $\underset{\sim}{x}^{(2)}$ is non singular ie $Disp(\underset{\sim}{x}^{(2)}) = \Sigma_2$ is p.d.

We approximate the regression of x_1 on $\underset{\sim}{x}^{(2)}$ as a linear f^n assuming regression is linear whether the regression $E(x_1 | \underset{\sim}{x}^{(2)})$ is actually linear or not.

Scanned by CamScanner

—

Since $E(X_1 - f(\underset{\sim}{X}^{(2)}))^2$ is minimum when $f(\underset{\sim}{X}^{(2)}) = M(\underset{\sim}{X}^{(2)})$, the regression of X_1 on $\underset{\sim}{X}^{(2)}$, we can consider an arbitrary f^n

$$\alpha + \beta_2 X_2 + \cdots + \beta_p X_p = \alpha + \underset{\sim}{\beta}' \underset{\sim}{X}^{(2)} \quad \text{where}$$

$\underset{\sim}{\beta}' = (\beta_2, \ldots, \beta_p)$; and determine the constants α & $\underset{\sim}{\beta}$ by minimizing $E(X_1 - \alpha - \beta_2 X_2 - \cdots - \beta_p X_p)^2 = S^2$, say.

$$\therefore \frac{\partial S^2}{\partial \alpha} = 0 \Leftrightarrow E(X_1 - \alpha - \beta_2 X_2 - \cdots - \beta_p X_p) = 0$$

$$\Leftrightarrow \mu_1 - \alpha - \beta_2 \mu_2 - \cdots - \beta_p \mu_p = 0$$

$$\Leftrightarrow \alpha = \mu_1 - \underset{\sim}{\beta}' \underset{\sim}{\mu}^{(2)}.$$

$$\& \quad \frac{\partial S^2}{\partial \beta_j} \Leftrightarrow E\{(X_1 - \alpha - \beta_2 X_2 - \cdots - \beta_p X_p) X_j\} = 0$$

$$\Leftrightarrow E(X_1 X_j) = \alpha E(X_j) + \sum_{i=2}^{p} \beta_i E(X_i X_j)$$

$$\Leftrightarrow \sigma_{1j} + \mu_1 \mu_j = (\mu_1 - \underset{\sim}{\beta}' \underset{\sim}{\mu}^{(2)}) \mu_j + \sum_{i=2}^{p} \beta_i (\sigma_{ij} + \mu_i \mu_j)$$

$$\Leftrightarrow \sigma_{1j} = \sum_{i=2}^{p} \beta_i \sigma_{ij}, \quad j = 2(1)p$$

$$\therefore \begin{pmatrix} \sigma_{12} \\ \sigma_{13} \\ \vdots \\ \sigma_{1p} \end{pmatrix} = \begin{pmatrix} \sigma_{22} & \sigma_{32} & \cdots & \sigma_{p2} \\ \sigma_{23} & \sigma_{33} & \cdots & \sigma_{p3} \\ \vdots & \vdots & & \vdots \\ \sigma_{2p} & \sigma_{3p} & \cdots & \sigma_{pp} \end{pmatrix} \begin{pmatrix} \beta_2 \\ \beta_3 \\ \vdots \\ \beta_p \end{pmatrix}$$

$$\Leftrightarrow \underset{\sim}{\sigma}_{(1)} = \Sigma_2 \underset{\sim}{\beta}$$

$\Leftrightarrow \quad \underset{\sim}{\beta} = \Sigma_2^{-1} \underset{\sim}{\sigma}_{(1)} \qquad [\because \Sigma_2 \text{ is pd}]$

Hence $\hat{\alpha} + \hat{\underset{\sim}{\beta}}' \underset{\sim}{x}^{(2)} = X_{1\cdot 23\cdots p}$, say ; where $\alpha = \mu_1 - \underset{\sim}{\beta}' \underset{\sim}{\mu}^{(2)}$,

$\underset{\sim}{\beta} = \Sigma_2^{-1} \underset{\sim}{\sigma}_{(1)}$, is the multiple linear regression.

Here $X_{1\cdot 23\cdots p}$ is the part of X_1 explained by the multiple linear regression of X_1 on $\underset{\sim}{X}^{(2)}$, we can define

$$X_1 = X_{1\cdot 23\cdots p} + e_{1\cdot 23\cdots p}$$

where $e_{1\cdot 23\cdots p}$ is the part of X_1 remaining unexplained by the multiple linear regression.

Theorem 3':

$E(e_{1\cdot 23\cdots p}) = 0$ & error is uncorrelated with the predictor variable & hence with the multiple linear regression.

proof: From 1st normal eq^{n},

$$E(X_1 - \alpha - \beta_2 X_2 - \cdots - \beta_p X_p) = 0$$

$$\Leftrightarrow E(X_1 - X_{1\cdot 23\cdots p}) = 0$$

$$\Leftrightarrow E(e_{1\cdot 23\cdots p}) = 0$$

Now, $cov(e_{1\cdot 23\cdots p}, X_j) = E(e_{1\cdot 23\cdots p} X_j) \qquad j = 2(1)p$

$= E[(X_1 - \alpha - \beta_2 X_2 - \cdots - \beta_p X_p) X_j]$

$= 0$, by normal eq^{n}s

Again, $cov(e_{1\cdot 23\cdots p}, X_{1\cdot 23\cdots p})$

$= cov(e_{1\cdot 23\cdots p}, \alpha + \sum_{j=2}^{p} \beta_j X_j)$

$= \beta_j \, cov(e_{1.23\cdots p}, x_j) = 0.$

Theorem: 4:

$$Var(e_{1.23\cdots p}) = \sigma_{11} - \underset{\sim}{\sigma}_{(1)}' \Sigma_2^{-1} \underset{\sim}{\sigma}_{(1)}$$

$$= \frac{|\Sigma|}{|\Sigma_2|} = \frac{1}{\sigma^{11}} = \frac{\sigma_{11}|R|}{R_{11}} = \frac{\sigma_{11}}{\rho^{11}}$$

(The symbols have usual meanings)

proof:

$$Var(e_{1.23\cdots p}) = cov(e_{1.23\cdots p}, e_{1.23\cdots p})$$

$$= cov(X_1 - X_{1.23\cdots p}, e_{1.23\cdots p})$$

$$= cov(X_1, e_{1.23\cdots p}) - cov(X_{1.23\cdots p}, e_{1.23\cdots p})$$

$$= cov(X_1, e_{1.23\cdots p})$$ [$\because$ error is uncorrelated with multiple regression]

$$= cov(X_1, X_1 - X_{1.23\cdots p})$$

$$= Var(X_1) - cov(X_1, X_{1.23\cdots p})$$

$$= \sigma_{11} - cov\left(X_1, \alpha + \sum_{j=2}^{p} \beta_j X_j\right)$$

$$= \sigma_{11} - \sum_{j=2}^{p} \beta_j \, cov(X_1, X_j)$$

$$= \sigma_{11} - \sum_{j=2}^{p} \beta_j \, \sigma_{1j}$$

$$= \sigma_{11} - \underset{\sim}{\sigma}_{(1)}' \underset{\sim}{\beta}$$

$$= \sigma_{11} - \underset{\sim}{\sigma}_{(1)}' \Sigma_2^{-1} \underset{\sim}{\sigma}_{(1)}$$

Note that $\Sigma = \begin{pmatrix} \sigma_{11} & \underset{\sim}{\sigma}_{(1)}' \\ \underset{\sim}{\sigma}_{(1)} & \Sigma_2 \end{pmatrix}$

$$\Rightarrow |\Sigma| = |\Sigma_2| \left(\sigma_{11} - \underset{\sim}{\sigma}_{(1)}' \Sigma_2^{-1} \underset{\sim}{\sigma}_{(1)}\right)$$

$$\Rightarrow \quad \sigma_{11} - \underset{\sim}{\sigma}_{(1)}' \Sigma_2^{-1} \underset{\sim}{\sigma}_{(1)} = \frac{|\Sigma|}{|\Sigma_2|}$$

$$\Rightarrow \quad Var(e_{1\cdot 23\cdots p}) = \frac{|\Sigma|}{|\Sigma_2|}$$

Define $\Sigma^{-1} = \left(\left(\frac{\Sigma_{ij}}{|\Sigma|}\right)\right)$ [where Σ_{ij} is the cofactor of σ_{ij} in Σ

$= ((\sigma^{ij}))$ & $\Sigma_{ij} = \Sigma_{ji}$]

Here $\sigma^{11} = \frac{\Sigma_{11}}{|\Sigma|} = \frac{|\Sigma_2|}{|\Sigma|}$

$$\therefore \quad V(e_{1\cdot 23\cdots p}) = \frac{|\Sigma|}{|\Sigma_2|} = \frac{1}{\sigma^{11}}$$

Note that

$R = D\Sigma D$ where $D = diag\left(\frac{1}{\sqrt{\sigma_{11}}}, \dots, \frac{1}{\sqrt{\sigma_{pp}}}\right)$

$$\Rightarrow \quad |R| = |D|^2 |\Sigma|$$

$$\Leftrightarrow \quad |\Sigma| = \sigma_{11} \cdots \sigma_{pp} |R|$$

Now, $R_2 = D^* \Sigma_2 D^*$ where $D^* = diag\left(\frac{1}{\sqrt{\sigma_{22}}}, \dots, \frac{1}{\sqrt{\sigma_{pp}}}\right)$

$\Leftrightarrow \quad |\Sigma_2| = \sigma_{22} \cdots \sigma_{pp} |R_2|$ where $R = \begin{pmatrix} 1 & \underset{\sim}{r}_{(1)}' \\ \underset{\sim}{r}_{(1)} & R_2 \end{pmatrix}$.

$$\therefore \quad V(e_{1\cdot 23\cdots p}) = \frac{|\Sigma|}{|\Sigma_2|} = \frac{\sigma_{11}|R|}{|R_2|}$$

$$= \frac{\sigma_{11}|R|}{R_{11}}$$

[R_{ij} = cofactor of (i,j)th element in R]

Define $R^{-1} = \left(\left(\frac{R_{ij}}{|R|}\right)\right) = ((\rho^{ij}))$, say.

Then, $\rho^{11} = \frac{R_{11}}{|R|}$

$$\therefore \quad V(e_{1\cdot 23\cdots p}) = \frac{\sigma_{11}}{\rho^{11}}$$

Theorem 5:

The correlation coeff betⁿ X_1 & the multiple linear regression of X_1 on $\underset{\sim}{X}^{(2)}$, ie $X_{1.23\cdots p}$, is the maximum among all linear fⁿs of $\underset{\sim}{X}^{(2)}$ like $L(\underset{\sim}{X}^{(2)}) = l_0 + l_2 x_2 + \cdots + l_p x_p$.

proof: $\rho^2(X_1, L(\underset{\sim}{X}^{(2)}))$

$$= \frac{\text{cov}^2[X_1, l_0 + \sum_{j=2}^{p} l_j X_j]}{\text{Var}(X_1)\,\text{Var}(l_0 + \sum_j l_j X_j)}$$

$$= \frac{\left\{\sum_{j=2}^{p} l_j \,\text{cov}(X_1, X_j)\right\}^2}{\sigma_{11} \cdot \underset{\sim}{l}' \Sigma_2 \underset{\sim}{l}}$$

$$= \frac{\left\{\sum_{j=2}^{p} l_j \sigma_{1j}\right\}^2}{\sigma_{11}(\underset{\sim}{l}' \Sigma_2 \underset{\sim}{l})} \quad [\underset{\sim}{l}' = (l_2, \ldots, l_p)]$$

$$= \frac{(\underset{\sim}{l}' \underset{\sim}{\sigma}_{(1)})^2}{\sigma_{11}(\underset{\sim}{l}' \Sigma_2 \underset{\sim}{l})}$$

$$= \frac{(\underset{\sim}{l}' \Sigma_2 \underset{\sim}{\beta})^2}{\sigma_{11}(\underset{\sim}{l}' \Sigma_2 \underset{\sim}{l})} \quad \cdots (*) \quad [\underset{\sim}{\sigma}_{(1)} = \Sigma_2 \underset{\sim}{\beta} \text{ where } X_{1.23\cdots p} = \alpha + \underset{\sim}{\beta}' \underset{\sim}{X}^{(2)}]$$

$$= \frac{(\underset{\sim}{l}' P'P \underset{\sim}{\beta})^2}{\sigma_{11}(\underset{\sim}{l}' \Sigma_2 \underset{\sim}{l})} \quad [\because \Sigma_2 \text{ is pd}]$$

$$= \frac{\{(P\underset{\sim}{l})'(P\underset{\sim}{\beta})\}^2}{\sigma_{11}(\underset{\sim}{l}' \Sigma_2 \underset{\sim}{l})}$$

$$\leq \frac{[(P\underset{\sim}{l})'(P\underset{\sim}{l})]\,[(P\underset{\sim}{\beta})'(P\underset{\sim}{\beta})]}{\sigma_{11}(\underset{\sim}{l}' \Sigma_2 \underset{\sim}{l})} \quad [\text{By C-S inequality } |\underset{\sim}{a}\cdot\underset{\sim}{b}| \leq |\underset{\sim}{a}||\underset{\sim}{b}|]$$

$$= \frac{\{\underset{\sim}{l}'(P'P)\underset{\sim}{l}\}\{\underset{\sim}{\beta}'(P'P)\underset{\sim}{\beta}\}}{\sigma_{11}\,(\underset{\sim}{l}'\Sigma_2\underset{\sim}{l})}$$

$$= \frac{(\underset{\sim}{l}'\Sigma_2\underset{\sim}{l})(\underset{\sim}{\beta}'\Sigma_2\underset{\sim}{\beta})}{\sigma_{11}\,(\underset{\sim}{l}'\Sigma_2\underset{\sim}{l})}$$

$$= \frac{\underset{\sim}{\beta}'\Sigma_2\underset{\sim}{\beta}}{\sigma_{11}}$$

From ⊗,

$$\rho^2(X_1, X_{1\cdot 23\ldots p})$$

$$= \rho^2(X_1, \alpha + \underset{\sim}{\beta}'\underset{\sim}{X}^{(2)})$$

$$= \frac{\underset{\sim}{\beta}'\Sigma_2\underset{\sim}{\beta}}{\sigma_{11}} \geq \rho^2(X_1, L(\underset{\sim}{X}^{(2)}))$$

Multiple correlation coeff:

The maximum correlation coeff betn X_1 & any linear fn of $\underset{\sim}{X}^{(2)}$ is defined as multiple correlation coeff betn X_1 & $\underset{\sim}{X}^{(2)}$ and denoted by $\rho_{1\cdot 23\ldots p}$ & it is given by

$$\rho_{1\cdot 23\ldots p} = \sqrt{\frac{\underset{\sim}{\beta}'\Sigma_2\underset{\sim}{\beta}}{\sigma_{11}}} = \frac{SD(X_{1\cdot 23\ldots p})}{SD(X_1)}$$

Where $X_{1\cdot 23\ldots p} = \alpha + \underset{\sim}{\beta}'\underset{\sim}{X}^{(2)}$ is the multiple regression (linear) of X_1 on $\underset{\sim}{X}^{(2)}$.

Clearly $0 \leq \rho_{1\cdot 23\ldots p} \leq 1$.

Theorem 6:

$$\rho^2_{1\cdot 23\ldots p} = 1 - \frac{\sigma^2_{1\cdot 23\ldots p}}{\sigma_{11}}, \text{ where } \sigma^2_{1\cdot 23\ldots p} = V(e_{1\cdot 23\ldots p})$$

proof: $X_1 = X_{1\cdot 23\ldots p} + e_{1\cdot 23\ldots p}$

$\therefore V(X_1) = V(X_{1\cdot 23\ldots p}) + V(e_{1\cdot 23\ldots p})$

$[\because \text{cov}(e_{1\cdot 23\ldots p}, X_{1\cdot 23\ldots p}) = 0]$

$$\therefore \rho^2_{1\cdot 23\ldots p} = \frac{V(X_{1\cdot 23\ldots p})}{V(X_1)}$$

$$= 1 - \frac{V(e_{1\cdot 23\ldots p})}{V(X_1)}$$

$$= 1 - \frac{\sigma^2_{1\cdot 23\ldots p}}{\sigma_{11}}$$

Remark: ① We can write

$$\rho^2_{1\cdot 23\ldots p} = 1 - \frac{|\Sigma|}{|\Sigma_2|\sigma_{11}}$$

$$= \frac{\underset{\sim}{\sigma}'_{(1)} \Sigma_2^{-1} \underset{\sim}{\sigma}_{(1)}}{\sigma_{11}}$$

$$= 1 - \frac{1}{\sigma^{11}\sigma_{11}}$$

$$= 1 - \frac{|R|}{R_{11}} = 1 - \frac{1}{\rho^{11}}$$

A measure of usefulness of the least square linear regression of X_1 on $\underset{\sim}{X}^{(2)}$ is $\rho^2_{1\cdot 23\ldots p} = \frac{V(X_{1\cdot 23\ldots p})}{V(X_1)}$

② $\rho^2_{1\cdot 23\ldots p} = 1 \Leftrightarrow \sigma^2_{1\cdot 23\ldots p} = 0$

$\Leftrightarrow \text{var}(X_{1\cdot 23\ldots p}) = \text{var}(X_1)$

$\Leftrightarrow$ The variability of X_1 is completely explained by the multiple linear regression of X_1 on $\underset{\sim}{X}^{(2)}$.

$$\Rightarrow \beta' \Sigma_2 \beta = 0$$

$\rho^2_{1.23\ldots p} = 0 \iff \underset{\sim}{\beta} = \underset{\sim}{0}$, since Σ_2 is p.d.

$\iff$ The multiple regression fails to predict X_1.

Partial correlation coefficient:

Let

$$\underset{\sim}{X}^{p\times 1} = \begin{pmatrix} X_1 \\ X_2 \\ \underset{\sim}{X}^{(3)} \end{pmatrix}, \quad \underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \underset{\sim}{\mu}^{(3)} \end{pmatrix}, \quad \Sigma = \left(\begin{array}{cc:c} \sigma_{11} & \sigma_{12} & \underset{\sim}{\sigma}'_{(13)} \\ \sigma_{21} & \sigma_{22} & \underset{\sim}{\sigma}'_{(23)} \\ \hdashline \underset{\sim}{\sigma}_{(13)} & \underset{\sim}{\sigma}_{(23)} & \Sigma_3 \end{array}\right)$$

Suppose, we wish to know the correlation betn X_1 & X_2, after eliminating the effect of $\underset{\sim}{X}^{(3)}$. In practice, we eliminate the linear effect of $\underset{\sim}{X}^{(3)}$ from both of them.

Define $X_{1.34\ldots p}$ = the part of X_1 explained by the multiple linear regression of X_1 on $\underset{\sim}{X}^{(3)}$.

& $X_{2.34\ldots p}$ = the part of X_2 explained by the multiple linear regression of X_2 on $\underset{\sim}{X}^{(3)}$.

Then, $e_{1.34\ldots p} = X_1 - X_{1.34\ldots p}$

= part of X_1 uninfluenced by $\underset{\sim}{X}^{(3)}$

$e_{2.34\ldots p} = X_2 - X_{2.34\ldots p}$

= part of X_2 uninfluenced by $\underset{\sim}{X}^{(3)}$.

Now, we define the partial correlation coeff betn X_1 & X_2 after eliminating the linear effect of $\underset{\sim}{X}^{(3)}$ from both of them as the correlation coeff betn $e_{1\cdot 34\cdots p}$ & $e_{2\cdot 34\cdots p}$, ie

$$\rho_{12\cdot 34\cdots p} = \frac{\mathrm{cov}(e_{1\cdot 34\cdots p}, e_{2\cdot 34\cdots p})}{\sqrt{\mathrm{Var}(e_{1\cdot 34\cdots p})\,\mathrm{Var}(e_{2\cdot 34\cdots p})}}$$

is the partial correlm coeff betn X_1 & X_2 after eliminating the (linear) effect of $\underset{\sim}{X}^{(3)}$.

Let $X_{1\cdot 34\cdots p} = \alpha + \underset{\sim}{\beta}' \underset{\sim}{x}^{(3)}$ where $\alpha = \mu_1 - \underset{\sim}{\beta}' \underset{\sim}{\mu}^{(3)}$

& $\underset{\sim}{\sigma}_{(13)} = \Sigma_3 \underset{\sim}{\beta}$

& $X_{2\cdot 34\cdots p} = \alpha^* + \beta^{*\prime} \underset{\sim}{x}^{(3)}$ where $\alpha^* = \mu_2 - \beta^{*\prime} \underset{\sim}{\mu}^{(3)}$

& $\underset{\sim}{\sigma}_{(23)} = \Sigma_3 \underset{\sim}{\beta}^*$

Now,

$$\mathrm{Var}(e_{1\cdot 34\cdots p}) = \sigma_{11} - \underset{\sim}{\sigma}_{(13)}' \Sigma_3^{-1} \underset{\sim}{\sigma}_{(13)} = \frac{\begin{vmatrix} \sigma_{11} & \underset{\sim}{\sigma}_{(13)}' \\ \underset{\sim}{\sigma}_{(13)} & \Sigma_3 \end{vmatrix}}{|\Sigma_3|} = \frac{\Sigma_{22}}{|\Sigma_3|}$$

$$\& \ \mathrm{Var}(e_{2\cdot 34\cdots p}) = \sigma_{22} - \underset{\sim}{\sigma}_{(23)}' \Sigma_3^{-1} \underset{\sim}{\sigma}_{(23)} = \frac{\Sigma_{11}}{|\Sigma_3|}$$

$$\mathrm{Cov}(e_{1\cdot 34\cdots p}, e_{2\cdot 34\cdots p})$$

$$= \mathrm{cov}(X_1 - X_{1\cdot 34\cdots p}, e_{2\cdot 34\cdots p})$$

$$= \mathrm{cov}(X_1, e_{2\cdot 34\cdots p})$$ [∵ error is uncorrelated with the multiple linear regression]

$$= \operatorname{cov}(X_1, X_2 - X_{2\cdot 34\cdots p})$$

$$= \sigma_{12} - \operatorname{cov}(X_1, \alpha^* + \underset{\sim}{\beta}^* \underset{\sim}{X}^{(3)})$$

$$= \sigma_{12} - \underset{\sim}{\beta}^{*'} \underset{\sim}{\sigma}_{(13)}$$

$$= \sigma_{12} - \underset{\sim}{\sigma}'_{(23)} \Sigma_3^{-1} \underset{\sim}{\sigma}_{(13)}.$$

$$= -\frac{\Sigma_{12}}{|\Sigma_3|}$$

Hence,

$$\rho_{12\cdot 34\cdots p} = -\frac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$$

$$= -\frac{\sigma^{12}}{\sqrt{\sigma^{11}\sigma^{22}}}$$

$$= -\frac{R_{12}}{\sqrt{R_{11}R_{22}}}$$

$$= -\frac{\rho^{12}}{\sqrt{\rho^{11}\rho^{22}}}$$

In general,

$$\rho_{ij\cdot 12\cdots \overline{i-1}\,\overline{i+1}\cdots\overline{j-1}\,\overline{j+1}\cdots p} = -\frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$$

Partial regression coeff:

If the multiple regression eqn of X_1 on $\underset{\sim}{X}^{(2)}$ is

$E(X_1 | \underset{\sim}{x}^{(2)}) = \alpha + \beta_2 x_2 + \cdots + \beta_p x_p$, a linear fn of $\underset{\sim}{x}^{(2)}$, then

$$\alpha = E(X_1 | \underset{\sim}{x}^{(2)} = \underset{\sim}{0});$$

$$\alpha + \beta_j = E(X_1 | \underset{\sim}{x}^{(2)} = \underset{\sim}{e}_j),$$

ie β_j is the amount by which the conditional mean increases for a unit increment in x_j, keeping the other variables fixed.

β_j, written more explicitly

$$\beta_{1j \cdot 23 \cdots j-1\, j+1 \cdots p}$$

is called the partial regression coeff of X_1 on X_j, keeping the other variables fixed.

Then, the multiple linear regression of X_1 on $\underset{\sim}{X}^{(2)}$ is

$$X_{1 \cdot 23 \cdots p} = \alpha + \underset{\sim}{\beta}' \underset{\sim}{X}^{(2)}$$

where $\underset{\sim}{\sigma}_{(1)} = \Sigma_2 \underset{\sim}{\beta}$

and $\alpha = \mu_1 - \underset{\sim}{\beta}' \underset{\sim}{\mu}^{(2)}$

Note that

$$\underset{\sim}{\sigma}_{(1)} = \Sigma_2 \underset{\sim}{\beta}$$

$\Leftrightarrow \quad \sigma_{1i} = \beta_2 \sigma_{2i} + \beta_3 \sigma_{3i} + \cdots + \beta_p \sigma_{pi}$; $i = 2(1)p$

Scanned by CamScanner

Again,

$$\sigma_{1i}\Sigma_{11} + \sigma_{2i}\Sigma_{21} + \sigma_{3i}\Sigma_{31} + \cdots + \sigma_{pi}\Sigma_{p1} = 0 \quad \forall\, i = 2(1)p$$

$$\Leftrightarrow \sigma_{1i} = \left(-\frac{\Sigma_{21}}{\Sigma_{11}}\right)\sigma_{2i} + \left(-\frac{\Sigma_{31}}{\Sigma_{11}}\right)\sigma_{3i} + \cdots + \left(-\frac{\Sigma_{p1}}{\Sigma_{11}}\right)\sigma_{pi}$$

Comparing, we get

$$\beta_j = -\frac{\Sigma_{j1}}{\Sigma_{11}}, \quad j = 2(1)p$$

$$= -\frac{\Sigma_{1j}}{\Sigma_{11}}$$

$$= -\frac{R_{1j}/\sqrt{\sigma_{11}\sigma_{jj}}}{R_{11}/\sigma_{11}}$$

$$= -\frac{R_{1j}}{R_{11}}\sqrt{\frac{\sigma_{11}}{\sigma_{jj}}}$$

Hence, the multiple linear regression of X_1 on $\underset{\sim}{X}^{(2)}$ is

$$X_{1\cdot 23\ldots p} = \mu_1 - \sum_{j=2}^{p} \frac{R_{1j}\sqrt{\sigma_{11}}}{R_{11}\sqrt{\sigma_{jj}}}(X_j - \mu_j).$$

Result ③:

$$\beta_{12\cdot 34\ldots p} = \rho_{12\cdot 34\ldots p} \cdot \frac{\sigma_{1\cdot 34\ldots p}}{\sigma_{2\cdot 34\ldots p}}$$

$$= \rho_{12\cdot 34\ldots p} \cdot \frac{\sigma_{1\cdot 234\ldots p}}{\sigma_{2\cdot 134\ldots p}}$$

Result ④:

$$\beta_{12\cdot 34\ldots p} \cdot \beta_{21\cdot 34\ldots p} = \rho^2_{12\cdot 34\ldots p}$$

Relationship betn multiple correlation coeff & partial correlation coeffs of different orders:

$$(1-\rho^2_{1\cdot 23\ldots p}) = (1-\rho^2_{12})(1-\rho^2_{13\cdot 2})\cdots(1-\rho^2_{1p\cdot 24\ldots \overline{p-1}})$$

$$= (1-\rho_{1p})(1-\rho_{1\overline{p-1}\cdot p})\cdots(1-\rho^2_{12\cdot 34\ldots p})$$

proof: $V(e_{1\cdot 23\ldots p}) = cov(e_{1\cdot 23\ldots p}, e_{1\cdot 23\ldots p})$

$$= cov(X_1, e_{1\cdot 23\ldots p})$$

$$= cov(X_1 - X_{1\cdot 23\ldots \overline{p-1}}, e_{1\cdot 23\ldots p})$$

$$= cov(e_{1\cdot 23\ldots \overline{p-1}}, (X_1-\mu_1) - \beta_{12\cdot 34\ldots p}(X_2-\mu_2) - \cdots - \beta_{1p\cdot 34\ldots \overline{p-1}}(X_p-\mu_p))$$

$$= cov(e_{1\cdot 23\ldots \overline{p-1}}, X_1) - \beta_{1p\cdot 34\ldots \overline{p-1}}\, cov(e_{1\cdot 23\ldots \overline{p-1}}, X_p)$$

$$= \sigma^2_{1\cdot 23\ldots \overline{p-1}} - \rho_{1p\cdot 23\ldots \overline{p-1}} \cdot \frac{\sigma_{1\cdot 23\ldots \overline{p-1}}}{\sigma_{p\cdot 23\ldots \overline{p-1}}} \rho_{1p\cdot 2\ldots \overline{p-1}}\, \sigma_{1\cdot 23\ldots \overline{p-1}}\, \sigma_{p\cdot 12\ldots \overline{p-1}}$$

$$= \sigma^2_{1\cdot 23\ldots p-1} - \rho^2_{1p\cdot 23\ldots \overline{p-1}}\, \sigma^2_{1\cdot 23\ldots \overline{p-1}}$$

$$= (1-\rho^2_{1p\cdot 23\ldots \overline{p-1}})\, \sigma^2_{1\cdot 23\ldots \overline{p-1}}$$

$$\Rightarrow (1-\rho^2_{1.23\cdots p})\sigma_{11}$$

$$= (1-\rho^2_{1.23\cdots \overline{p-1}})\sigma_{11}(1-\rho^2_{1p.23\cdots \overline{p-1}})$$

$$\Rightarrow (1-\rho^2_{1.23\cdots p}) = (1-\rho^2_{1p.23\cdots \overline{p-1}})(1-\rho^2_{1.23\cdots \overline{p-1}})$$

Using this repeatedly, we get the result:

$$(1-\rho^2_{1.23\cdots p}) = (1-\rho_{12}^2)(1-\rho^2_{13.2})\cdots(1-\rho^2_{1p.234\cdots \overline{p-1}})$$

Problems:

① Let $\underset{\sim}{x}$ be a random vector with mean $\underset{\sim}{\mu}$ & dispersion matrix Σ. S.T.

[CU] $P[(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}) > \lambda] < \frac{p}{\lambda}$ for $\lambda > 0$

Soln: By Markov's inequality,

$$P[(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}) > \lambda]$$

$$< \frac{E[(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})]}{\lambda}$$

$$= \frac{E[tr\{(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\}]}{\lambda}$$

$$= \frac{E[tr\{\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})(\underset{\sim}{x}-\underset{\sim}{\mu})'\}]}{\lambda}$$

$$= \frac{tr[\Sigma^{-1}E\{(\underset{\sim}{x}-\underset{\sim}{\mu})(\underset{\sim}{x}-\underset{\sim}{\mu})'\}]}{\lambda}$$

$$= \frac{tr[\Sigma^{-1}\Sigma]}{\lambda} = \frac{p}{\lambda}$$

Alternative: $\Sigma^{p\times p}$ is pd $\Rightarrow \exists$ n.s $B \ni \Sigma = BB^T$.

Define $\underset{\sim}{Y} = B^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu})$

$$E(\underset{\sim}{Y}) = 0, \quad \text{Disp}(\underset{\sim}{Y}) = B^{-1}\Sigma(B^{-1})^T$$

$$= B^{-1}BB^T(B^{-1})^T$$

$$= I_p$$

Now,

$$P[(\underset{\sim}{X} - \underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu}) > \lambda]$$

$$= P[\underset{\sim}{Y}'\underset{\sim}{Y} > \lambda]$$

$$< \frac{E(\underset{\sim}{Y}'\underset{\sim}{Y})}{\lambda}$$

$$= \frac{E\left(\sum_{i=1}^{p} Y_i^2\right)}{\lambda} = \frac{p}{\lambda}$$

② Let $(Y, X_1, \ldots, X_p)'$ be any $(p+1)$ component random vector. Show, making suitable assumption that

i) $\rho^2_{y\cdot 123\ldots p} \geq 1 - \frac{1}{\sigma_y^2} E(Y - l_0 - \sum_1^p l_i x_i)^2$

ii) $\rho^2_{y\cdot 12\ldots p} \geq \rho^2_{y\cdot 23\ldots p}$

In each of the above cases comment on the case of equality

Hint: (i) Consider an arbitrary linear f^n $l_0 + l_1 X_1 + \dots + l_p X_p$ as a predictor of Y.
Now, consider the problem of minimization of the MSE

$$E\left(Y - l_0 - \sum_1^p l_i X_i\right)^2 .$$

We know that the linear f^n

$$\beta_0 + \underset{\sim}{\beta}' \underset{\sim}{x} = X_{y \cdot 12 \cdots p},$$

obtained by minimizing the MSE, is the multiple linear regression of Y on $\underset{\sim}{X}$.

$$E\left(Y - l_0 - \sum_1^p l_i X_i\right)^2 \geq E\left(Y - \beta_0 - \underset{\sim}{\beta}' \underset{\sim}{x}\right)^2$$

$$= \sigma^2_{y \cdot 123 \cdots p}$$

Now, $\rho^2_{y \cdot 123 \cdots p} = 1 - \dfrac{\sigma^2_{y \cdot 12 \cdots p}}{\sigma_y^2}$ (prove it)

$$\geq 1 - \frac{1}{\sigma_y^2} E\left(Y - l_0 - \sum_1^p l_i X_i\right)^2$$

(ii) $(1 - \rho^2_{y \cdot 12 \cdots p}) = (1 - \rho^2_{1y \cdot 23 \cdots p})(1 - \rho^2_{y \cdot 23 \cdots p})$

$\leq (1 - \rho^2_{y \cdot 23 \cdots p})$ since $0 \leq 1 - \rho^2_{1y \cdot 23 \cdots p} \leq 1$

Hence the proof.

"=" case: (ii) "=" iff $\rho^2_{1y \cdot 23 \cdots p} = 0$

$\Rightarrow$ There is no use of X_1 in the prediction of Y.

③ If $a_1X_1+a_2X_2+\cdots+a_pX_p = K$ (Constant), then find $\rho_{12\cdot34\cdots p}$ & $\rho_{1\cdot23\cdots p}$.

Soln: $E(X_1 | X_2=x_2, \cdots, X_p=x_p)$

$$= \frac{1}{a_1}[K - a_2x_2 - \cdots - a_px_p]$$

So, the regression is linear.
Thus, the multiple linear regression of X_1 on $X_2, \cdots, X_p$ will also be

$$X_{1\cdot23\cdots p} = \frac{1}{a_1}[K - a_2X_2 - \cdots - a_pX_p] \quad \cdots (1)$$

So, $e_{1\cdot23\cdots p} = X_1 - X_{1\cdot23\cdots p} = 0$

$\Rightarrow V(e_{1\cdot23\cdots p}) = 0 \Rightarrow \rho_{1\cdot23\cdots p} = 1$

Since it is a case of linear regression, so partial correlⁿ betⁿ X_1 & X_2 eliminating the effect $X_3, X_4, \cdots, X_p$ is equivalent to correlⁿ betⁿ X_1 & X_2 keeping $X_3, X_4, \cdots, X_p$ fixed, ie correlⁿ betⁿ X_1 & X_2 with

$$a_1X_1 + a_2X_2 = \text{constant}$$

So, this partial correlⁿ will be +1 or −1 according as a_1 & a_2 are of opposite sign or same sign.

Alternatively:

$$\beta_{12\cdot34\cdots p} = -\frac{a_2}{a_1} \quad [\text{From } (1)]$$

Similarly $\beta_{21\cdot34\cdots p} = -\frac{a_1}{a_2}$

$\Rightarrow \rho_{12\cdot34\cdots p} = 1$ or -1 according as a_1 & a_2 are of opposite sign or same sign

[$\because \rho^2_{12\cdot34\cdots p} = \beta_{12\cdot34\cdots p} \cdot \beta_{21\cdot34\cdots p}$ & sign of $\rho_{12\cdot34\cdots p}$ is same as that of $\beta_{12\cdot34\cdots p}$]

Multivariate Normal Dis^{tn}:

The pdf of a univariate normal is in the form $f(x) = Ke^{-\frac{1}{2}a(x-l)^2}$

$= Ke^{-\frac{1}{2}(x-l)a(x-l)}$ if $x \in R$.

Where $l \in R$, $a>0$ & $K>0$.

Generalizing concept, the pdf of a multivariate normal is taken as

$$f(\underset{\sim}{x}) = ke^{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{l})'A(\underset{\sim}{x}-\underset{\sim}{l})} \quad , \text{ if } \underset{\sim}{x} \in R^p$$

Where $\underset{\sim}{l} \in R^p$, A is a p.d. matrix & $k>0$.

Our object is to find the constant $\underset{\sim}{l}$ & A in terms of the moments of the multivariate normal distn. Let $\underset{\sim}{X}$ be a multivariate normal variate with the pdf

$$f(\underset{\sim}{x}) = ke^{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{l})'A(\underset{\sim}{x}-\underset{\sim}{l})} \quad , \underset{\sim}{x} \in R^p$$

Now, $\int_{R^p} f(\underset{\sim}{x})\,d\underset{\sim}{x} = 1$

$$\Rightarrow K\int_{R^p} e^{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{l})'A(\underset{\sim}{x}-\underset{\sim}{l})}\,d\underset{\sim}{x} = 1$$

$$\Rightarrow K\int_{R^p} e^{-\frac{1}{2}\underset{\sim}{y}'\underset{\sim}{y}} \frac{1}{|J|}\,d\underset{\sim}{y} = 1$$

[$\because$ A is pd $\exists$ a n.s P $\ni$ $PP^T = \Sigma$
Let $\underset{\sim}{y} = P^T(\underset{\sim}{x}-\underset{\sim}{l})$
Then $|J(\frac{\underset{\sim}{y}}{\underset{\sim}{x}})|$
$= ||P'|| = \sqrt{|PP'|}$
$= \sqrt{|A|}$

$$\Rightarrow \frac{K}{\sqrt{|A|}}\int_{R^p} e^{-\frac{1}{2}\underset{\sim}{y}'\underset{\sim}{y}}\,d\underset{\sim}{y} = 1$$

$$\Rightarrow \frac{K}{\sqrt{|A|}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}\sum_1^p y_i^2} dy_1 \cdots dy_p = 1$$

$$\Rightarrow \frac{K}{\sqrt{|A|}} \prod_{i=1}^{p} \left\{ \int_{-\infty}^{\infty} e^{-\frac{1}{2}y_i^2} dy_i \right\} = 1$$

$$\Rightarrow \frac{K}{\sqrt{|A|}} (\sqrt{2\pi})^p = 1$$

$$\Rightarrow K = \frac{\sqrt{|A|}}{(\sqrt{2\pi})^p}$$

Now, the pdf of $\underset{\sim}{y}$ is

$$g_{\underset{\sim}{y}}(\underset{\sim}{y}) = \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2}\underset{\sim}{y}'\underset{\sim}{y}}, \quad \underset{\sim}{y} \in \mathbb{R}^p$$

Here,

$$E(Y_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} y_i \left(\frac{1}{\sqrt{2\pi}}\right)^p e^{-\frac{1}{2}\sum_1^p y_i^2} dy_1 \cdots dy_p$$

$$= \left(\int_{-\infty}^{\infty} y_i \left(\frac{1}{\sqrt{2\pi}}\right) e^{-\frac{1}{2}y_i^2} dy_i \right) \prod_{j \neq i} \left\{ \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2}y_j^2} dy_j}{(\sqrt{2\pi})} \right\}$$

$$= 0 \quad \forall i$$

$$\text{cov}(Y_i, Y_j) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} y_i y_j \left(\frac{1}{\sqrt{2\pi}}\right)^p e^{-\frac{1}{2}\sum_1^p y_i^2} dy_1 \cdots dy_p \quad [\because E(Y_i) = 0]$$

$$= \begin{cases} \left(\int_{-\infty}^{\infty} y_i^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_i^2} dy_i \right) \left(\prod_{j \neq i} \left\{ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_j^2} dy_j \right\} \right) & \text{if } i = j \\ \left(\int_{-\infty}^{\infty} y_i \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_i^2} dy_i \right) \left(\int_{-\infty}^{\infty} y_j \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_j^2} dy_j \right) \left(\prod_{k \neq i \neq j} \cdots \right) & \text{if } i \neq j \end{cases}$$

$$= \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases}$$

$\therefore \quad E(\underset{\sim}{Y}) = \underset{\sim}{0} \quad \& \quad Disp(\underset{\sim}{Y}) = I_p$.

Now, $E(\underset{\sim}{Y}) = \underset{\sim}{0}$

$\Rightarrow \quad P'(E(\underset{\sim}{x}) - \underset{\sim}{l}) = \underset{\sim}{0}$

$\Rightarrow \quad \underset{\sim}{l} = E(\underset{\sim}{x}) = \underset{\sim}{\mu}$, say $\quad [\because P' \text{ is n.s}]$

$D(\underset{\sim}{Y}) = I_p$

$\Rightarrow \quad P' D(\underset{\sim}{x}) P = I_p$

$\Rightarrow \quad D(\underset{\sim}{x}) = (P')^{-1} P^{-1} = (PP')^{-1} = A^{-1}$

$\Rightarrow \quad A = \Sigma^{-1}$ where $\Sigma = D(\underset{\sim}{x})$.

Hence, $K = \dfrac{1}{(\sqrt{2\pi})^p \sqrt{|\Sigma|}}$

Defn:

A random vector $\underset{\sim}{X}^{p\times 1}$ is said to have a multivariate normal distn with mean vector $E(\underset{\sim}{X}) = \underset{\sim}{\mu}$ & dispersion matrix Σ, if the pdf of $\underset{\sim}{X}$ is

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \frac{1}{(2\pi)^{p/2} \sqrt{|\Sigma|}} e^{-\frac{1}{2}(\underset{\sim}{x} - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu})}, \quad \underset{\sim}{x} \in \mathbb{R}^p$$

We write $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$.

Remarks:

(1) The pdf $f_{\underset{\sim}{X}}(\underset{\sim}{x})$ is maximum if the exponent is minimum, ie $(\underset{\sim}{x} - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{x} - \underset{\sim}{\mu})$ is minimum.

Note that

$$(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}) \begin{cases} >0 & \text{if } \underset{\sim}{x} \neq \underset{\sim}{\mu} \\ =0 & \text{if } \underset{\sim}{x} = \underset{\sim}{\mu} \end{cases}$$

since Σ^{-1} is p.d.

Hence $\underset{\sim}{x}=\underset{\sim}{\mu}$ is the mode of the distn.

Let $Q(\underset{\sim}{x}) = (\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})$

Then, $\frac{\partial Q(\underset{\sim}{x})}{\partial(\underset{\sim}{x})} = \underset{\sim}{0}$ at $\underset{\sim}{x}=\underset{\sim}{\mu}$

(2) Note that the exponent

$$\begin{aligned} Q(\underset{\sim}{x}) &= (\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}) \\ &= \underset{\sim}{x}'\Sigma^{-1}\underset{\sim}{x} - \underset{\sim}{x}'\Sigma^{-1}\underset{\sim}{\mu} - \underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{x} \\ &\quad + \underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{\mu} \\ &= \underset{\sim}{x}'\Sigma^{-1}\underset{\sim}{x} - 2\underset{\sim}{x}'\Sigma^{-1}\underset{\sim}{\mu} + \underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{\mu} \end{aligned}$$

Let $Q^*(\underset{\sim}{x})$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p} a_{ij}x_i x_j + \sum_{i=1}^{p} l_i x_i + c$$

be the exponent of a multivariate normal distn's pdf.

Note that

$$Q^*(\underset{\sim}{x}) = \underset{\sim}{x}'A\underset{\sim}{x} + \underset{\sim}{x}'\underset{\sim}{l} + c$$

Comparing, we get,

$$A = \Sigma^{-1} \Rightarrow \Sigma = A^{-1}$$

Theorem 7:

If $\underset{\sim}{X}^{p\times 1} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then for a n.s matrix $P^{p\times p}$,

$$P'\underset{\sim}{X} \sim N_p(P'\underset{\sim}{\mu}, P'\Sigma P)$$

proof: The pdf of $\underset{\sim}{x}$ is

$$f(\underset{\sim}{x}) = \frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})}, \quad \underset{\sim}{x} \in \mathbb{R}^p.$$

Let $\underset{\sim}{Y} = P'\underset{\sim}{X}$.

Now, $\left|J\left(\frac{\underset{\sim}{y}}{\underset{\sim}{x}}\right)\right| = ||P'|| = \sqrt{|P'P|}$

The pdf of $\underset{\sim}{Y}$ is

$$g(\underset{\sim}{y}) = \frac{1}{(\sqrt{2\pi})^p\sqrt{|\Sigma|}} e^{-\frac{1}{2}\{(P')^{-1}\underset{\sim}{y}-\underset{\sim}{\mu}\}'\Sigma^{-1}\{(P')^{-1}\underset{\sim}{y}-\underset{\sim}{\mu}\}} \frac{1}{\sqrt{|P'P|}}$$

$$= \frac{1}{(\sqrt{2\pi})^p\sqrt{|P'\Sigma P|}} e^{-\frac{1}{2}(\underset{\sim}{y}-P'\underset{\sim}{\mu})'P^{-1}\Sigma^{-1}(P')^{-1}(\underset{\sim}{y}-P'\underset{\sim}{\mu})}.$$

$$= \frac{1}{(\sqrt{2\pi})^p\sqrt{|P'\Sigma P|}} e^{-\frac{1}{2}(\underset{\sim}{y}-P'\underset{\sim}{\mu})'(P'\Sigma P)^{-1}(\underset{\sim}{y}-P'\underset{\sim}{\mu})}.$$

$\therefore \underset{\sim}{Y} \sim N_p(P'\underset{\sim}{\mu}, P'\Sigma P)$.

Moment generating function:

The mgf of a p dimensional R.V. $\underset{\sim}{x}$ is defined as

$$M_{\underset{\sim}{x}}(\underset{\sim}{t}) = E(e^{\underset{\sim}{t}'\underset{\sim}{x}}),$$

provided it exists, where $\underset{\sim}{t}$ belongs to a region containing the origin as an interior point.

Theorem 8:

If $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then the mgf of $\underset{\sim}{x}$ is

$M_{\underset{\sim}{x}}(\underset{\sim}{t}) = e^{\underset{\sim}{t}'\underset{\sim}{\mu} + \frac{1}{2}\underset{\sim}{t}'\Sigma\underset{\sim}{t}}$, where $\underset{\sim}{t}$ belongs to a region containing the origin as an interior point.

proof: $M_{\underset{\sim}{x}}(\underset{\sim}{t}) = E(e^{\underset{\sim}{t}'\underset{\sim}{x}})$

$$= \int_{R^p} e^{\underset{\sim}{t}'\underset{\sim}{x}} \frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})} d\underset{\sim}{x}$$

$$= \frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma|}} \int_{R^p} e^{-\frac{1}{2}\{\underset{\sim}{x}'\Sigma^{-1}\underset{\sim}{x} - 2\underset{\sim}{x}'\Sigma^{-1}\underset{\sim}{\mu} + \underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{\mu} - 2\underset{\sim}{t}'\underset{\sim}{x}\}} d\underset{\sim}{x}$$

$$= \frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma|}} \int_{R^p} e^{-\frac{1}{2}\{\underset{\sim}{x}'\Sigma^{-1}\underset{\sim}{x} - 2\underset{\sim}{x}'\Sigma^{-1}(\underset{\sim}{\mu}+\Sigma\underset{\sim}{t}) + \underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{\mu}\}} d\underset{\sim}{x}$$

$$= \frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma|}} \int_{R^p} e^{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu}-\Sigma\underset{\sim}{t})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}-\Sigma\underset{\sim}{t}) + \underset{\sim}{t}'\underset{\sim}{\mu} + \frac{1}{2}\underset{\sim}{t}'\Sigma\underset{\sim}{t}} d\underset{\sim}{x}$$

$$\Big[\because\ \underset{\sim}{x}'\Sigma^{-1}\underset{\sim}{x} - 2\underset{\sim}{x}\Sigma^{-1}(\underset{\sim}{\mu}+\Sigma\underset{\sim}{t}) + \underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{\mu}$$

$$= \underset{\sim}{x}'\Sigma^{-1}\underset{\sim}{x} - 2\underset{\sim}{x}\Sigma^{-1}(\underset{\sim}{\mu}+\Sigma\underset{\sim}{t}) + (\underset{\sim}{\mu}+\Sigma\underset{\sim}{t})'\Sigma^{-1}(\underset{\sim}{\mu}+\Sigma\underset{\sim}{t})$$

$$+ \underset{\sim}{\mu}'\Sigma^{-1}\underset{\sim}{\mu} - (\underset{\sim}{\mu}+\Sigma\underset{\sim}{t})'\Sigma^{-1}(\underset{\sim}{\mu}+\Sigma\underset{\sim}{t})$$

$$= (\underset{\sim}{x}-\underset{\sim}{\mu}-\Sigma\underset{\sim}{t})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}-\Sigma\underset{\sim}{t}) - \underset{\sim}{\mu}\Sigma^{-1}\Sigma\underset{\sim}{t}$$

$$-\underset{\sim}{t}'\Sigma\Sigma^{-1}\underset{\sim}{\mu} - \underset{\sim}{t}'\Sigma\Sigma^{-1}\Sigma\underset{\sim}{t}$$

$$= (\underset{\sim}{x}-\underset{\sim}{\mu}-\Sigma\underset{\sim}{t})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}-\Sigma\underset{\sim}{t}) - 2\underset{\sim}{t}'\underset{\sim}{\mu} - \underset{\sim}{t}'\Sigma\underset{\sim}{t}\ \Big]$$

$$= e^{\underset{\sim}{t}'\underset{\sim}{\mu}+\frac{1}{2}\underset{\sim}{t}'\Sigma\underset{\sim}{t}} \int_{\mathbb{R}^p} n_p(\underset{\sim}{x}\,|\,\underset{\sim}{\mu}+\Sigma\underset{\sim}{t}, \Sigma)\, d\underset{\sim}{x}$$

$$= e^{\underset{\sim}{t}'\underset{\sim}{\mu}+\frac{1}{2}\underset{\sim}{t}'\Sigma\underset{\sim}{t}}$$

[where $n_p(\underset{\sim}{x}\,|\,\underset{\sim}{\mu}, \Sigma)$ is the pdf of $N_p(\underset{\sim}{\mu}, \Sigma)$]

Example 1:

Prove theorem 7 using mgf technique.

Hint: $M_{\underset{\sim}{y}}(\underset{\sim}{t}) = E(e^{\underset{\sim}{t}'y})$

$$= E\left(e^{(P\underset{\sim}{t})'\underset{\sim}{x}}\right)$$

$$= E\left(e^{\underset{\sim}{u}'\underset{\sim}{x}}\right), \quad \underset{\sim}{u} = P\underset{\sim}{t}$$

$$= e^{\underset{\sim}{u}'\underset{\sim}{\mu}+\frac{1}{2}\underset{\sim}{u}'\Sigma\underset{\sim}{u}}$$

$$= e^{\underset{\sim}{t}'(P'\underset{\sim}{\mu})+\frac{1}{2}\underset{\sim}{t}'(P'\Sigma P)\underset{\sim}{t}}$$

which is the mgf of $N_p(P'\underset{\sim}{\mu}, P'\Sigma P)$.

Theorem 9:

Let $\underset{\sim}{x}^{p\times 1} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ & $B^{q\times p}$ $(q \le p)$ be a matrix of rank q.

Then, $\underset{\sim}{y}^{q\times 1} = B\underset{\sim}{x} \sim N_q(B\underset{\sim}{\mu}, B\Sigma B^T)$

proof: $M_{\underset{\sim}{y}}(\underset{\sim}{t})$

$[\rho(B\Sigma B^T) = \rho(B^T B\Sigma)$
$= \rho(B^T B)$
$= \rho(B) = q$
$\Rightarrow (B\Sigma B^T)^{q\times q}$ is n.s.$]$

$$= E(e^{\underset{\sim}{t}'\underset{\sim}{y}})$$

$$= E(e^{\underset{\sim}{t}'B\underset{\sim}{x}})$$

$$= E(e^{(B'\underset{\sim}{t})'\underset{\sim}{x}})$$

$$= E(e^{\underset{\sim}{u}'\underset{\sim}{x}}) \qquad \underset{\sim}{u}^{p\times 1} = \underset{\sim}{B}'\underset{\sim}{t}$$

$$= e^{\underset{\sim}{u}'\underset{\sim}{\mu} + \frac{1}{2}\underset{\sim}{u}'\Sigma\underset{\sim}{u}}$$

$$= e^{(B'\underset{\sim}{t})'\underset{\sim}{\mu} + \frac{1}{2}(B'\underset{\sim}{t})'\Sigma(B'\underset{\sim}{t})}$$

$$= e^{\underset{\sim}{t}'(B\underset{\sim}{\mu}) + \frac{1}{2}\underset{\sim}{t}'(B\Sigma B')\underset{\sim}{t}}$$

, which is the mgf of $N_q(B\underset{\sim}{\mu}, B\Sigma B^T)$

Hence by uniqueness of mgf,

$$\underset{\sim}{y} \sim N_q(B\underset{\sim}{\mu}, B\Sigma B^T).$$

Theorem 10:

If $\underset{\sim}{x}^{p\times 1} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then we can write $\underset{\sim}{x} = \underset{\sim}{\mu} + P\underset{\sim}{y}$, where $PP^T = \Sigma$ & $\underset{\sim}{y} \sim N_p(\underset{\sim}{0}, I_p)$.

proof: The pdf of $\underset{\sim}{x}$ is

$$f_{\underset{\sim}{x}}(\underset{\sim}{x}) = \frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})}, \ \underset{\sim}{x} \in \mathbb{R}^p.$$

Since Σ is p.d., there exists a n.s. P such that $PP^T = \Sigma$.

Let $\underset{\sim}{Y} = P^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu})$.

Note that $\left| J\left(\frac{\underset{\sim}{y}}{\underset{\sim}{x}}\right) \right| = \left| |P^{-1}| \right| = \sqrt{|(P^{-1})^T||P^{-1}|}$

$$= \sqrt{|(PP^T)^{-1}|}$$

$$= \sqrt{|\Sigma^{-1}|}$$

$$= \frac{1}{\sqrt{|\Sigma|}}.$$

The pdf of $\underset{\sim}{Y}$ is

$$g_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma|}} e^{-\frac{1}{2}\underset{\sim}{y}'P'(PP')^{-1}P\underset{\sim}{y}} \cdot \sqrt{|\Sigma|}$$

$$= \frac{1}{(2\pi)^{p/2}} e^{-\frac{1}{2}\underset{\sim}{y}'\underset{\sim}{y}}$$

$\therefore \underset{\sim}{Y} \sim N_p(\underset{\sim}{0}, I_p)$, where $\underset{\sim}{X} = \underset{\sim}{\mu} + P\underset{\sim}{Y}$
& $\Sigma = PP^T$.

<u>Corollary</u>:

If $\underset{\sim}{X}^{p\times 1} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then
$(\underset{\sim}{X} - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{X} - \underset{\sim}{\mu}) \sim \chi^2_p$.

<u>proof</u>: If $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ then we have
$\underset{\sim}{X} - \underset{\sim}{\mu} = \underline{P}\underset{\sim}{Y}$ where $PP' = \Sigma$
& $\underset{\sim}{Y} \sim N_p(0, I_p)$.

$\therefore$ pdf of $\underset{\sim}{Y}$ is

$$g_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \frac{1}{(2\pi)^{p/2}} e^{-\frac{1}{2}\underset{\sim}{y}'\underset{\sim}{y}} = \frac{1}{(2\pi)^{p/2}} e^{-\frac{1}{2}\sum_1^p y_i^2}$$

$\Leftrightarrow Y_i \overset{iid}{\sim} N(0,1), i = 1(1)p$.

Now,

$$(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})$$

$$= \underset{\sim}{y}'\underset{\sim}{y}$$

$$= \sum_{1}^{p} y_i^2$$

$$\sim \chi^2_p$$

Let $\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, \Sigma)$.

Consider the following partition

$$\underset{\sim}{X}^{p\times 1} = \begin{pmatrix} \underset{\sim}{X}^{(1)\, p_1\times 1} \\ \underset{\sim}{X}^{(2)} \end{pmatrix}, \quad \underset{\sim}{\mu}^{p\times 1} = \begin{pmatrix} \underset{\sim}{\mu}^{(1)} \\ \underset{\sim}{\mu}^{(2)} \end{pmatrix}$$

$$\& \quad Disp(\underset{\sim}{x}) = \Sigma^{p\times p} = \begin{pmatrix} \Sigma_{11}^{p_1\times p_1} & \Sigma_{12} \\ \Sigma_{21}\,(=\Sigma_{12}') & \Sigma_{22} \end{pmatrix}$$

$$= \begin{pmatrix} cov(\underset{\sim}{x}^{(1)}, \underset{\sim}{x}^{(1)}) & cov(\underset{\sim}{x}^{(1)}, \underset{\sim}{x}^{(2)}) \\ cov(\underset{\sim}{x}^{(2)}, \underset{\sim}{x}^{(1)}) & cov(\underset{\sim}{x}^{(2)}, \underset{\sim}{x}^{(2)}) \end{pmatrix}$$

Now, consider the following theorems –

~~Therome~~

Theorem 11:

If $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then the necessary & sufficient condition for $\underset{\sim}{X}^{(1)}$ & $\underset{\sim}{X}^{(2)}$ to be indept is that $\Sigma_{12} = 0$.

proof: Let $\Sigma_{12} = \text{cov}(\underset{\sim}{X}^{(1)}, \underset{\sim}{X}^{(2)}) = 0$

Then the pdf of $\underset{\sim}{X}^{p\times 1}$ is

$$n_p(\underset{\sim}{x} | \underset{\sim}{\mu}, \Sigma)$$

$$= \frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})}, \quad \underset{\sim}{x} \in \mathbb{R}^p$$

Now, $\Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0^T & \Sigma_{22} \end{pmatrix}$

Hence $|\Sigma| = |\Sigma_{11}||\Sigma_{22}|$

& $\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix}$

Now, the exponent becomes

$$(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})$$

$$= \begin{pmatrix} \underset{\sim}{x}^{(1)}-\underset{\sim}{\mu}^{(1)} \\ \underset{\sim}{x}^{(2)}-\underset{\sim}{\mu}^{(2)} \end{pmatrix}' \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} \underset{\sim}{x}^{(1)}-\underset{\sim}{\mu}^{(1)} \\ \underset{\sim}{x}^{(2)}-\underset{\sim}{\mu}^{(2)} \end{pmatrix}$$

$$= (\underset{\sim}{x}^{(1)}-\underset{\sim}{\mu}^{(1)})'\Sigma_{11}^{-1}(\underset{\sim}{x}^{(1)}-\underset{\sim}{\mu}^{(1)}) + (\underset{\sim}{x}^{(2)}-\underset{\sim}{\mu}^{(2)})'\Sigma_{22}^{-1}(\underset{\sim}{x}^{(2)}-\underset{\sim}{\mu}^{(2)})$$

Hence,

$$n(\underset{\sim}{x} | \underset{\sim}{\mu}, \Sigma) = \frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma_{11}|}} e^{-\frac{1}{2}(\underset{\sim}{x}^{(1)}-\underset{\sim}{\mu}^{(1)})'\Sigma_{11}^{-1}(\underset{\sim}{x}^{(1)}-\underset{\sim}{\mu}^{(1)})}$$

$$\times \frac{1}{(2\pi)^{\frac{p-p_1}{2}}\sqrt{|\Sigma_{22}|}} e^{-\frac{1}{2}(\underset{\sim}{x}^{(2)}-\underset{\sim}{\mu}^{(2)})'\Sigma_{22}^{-1}(\underset{\sim}{x}^{(2)}-\underset{\sim}{\mu}^{(2)})}$$

Hence $\underset{\sim}{X}^{(1)}, \underset{\sim}{X}^{(2)}$ are indept.

Note that $\underset{\sim}{X}^{(1)} \sim N_{p_1}(\underset{\sim}{\mu}^{(1)}, \Sigma_{11})$

& $\underset{\sim}{X}^{(2)} \sim N_{p-p_1}(\underset{\sim}{\mu}^{(2)}, \Sigma_{22})$

Let $\underset{\sim}{X}^{(1)}$ & $\underset{\sim}{X}^{(2)}$ are indept

[$\Leftrightarrow$ X_i & X_j are indept, $i = 1(1)p_1$, $j = \overline{p_1+1}(1)p$] ??

$$Cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] \quad , \; i = 1(1)p, \; j = \overline{p_1+1}(1)p.$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j) f(x_1, \ldots, x_p)\, dx_1 \cdots dx_p$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j) f_1(x_1, \ldots, x_{p_1}) f_2(x_{p_1+1}, \ldots, x_p)\, dx_1 \cdots dx_p$$

$$= \left\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_i - \mu_i) f_1(x_1, \ldots, x_{p_1})\, dx_1 \cdots dx_{p_1} \right\}$$

$$\times \left\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (x_j - \mu_j) f_2(x_{p_1+1}, \ldots, x_p)\, dx_{p_1+1} \cdots dx_p \right\}$$

$$= \{E(X_i) - \mu_i\}\{E(X_j) - \mu_j\}$$

$$= 0$$

$\therefore \sigma_{ij} = 0 \quad \forall \; i = 1(1)p_1, \; j = \overline{p_1+1}(1)p.$

$\Rightarrow \Sigma_{12} = \Sigma_{21}^{T} = 0.$

<u>Theorem 12:</u>

If $\underset{\sim}{X}^{p \times 1} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then any subvector is also a multivariate normal with mean vector & dispersion matrix obtained by taking the corresponding components

Scanned by CamScanner

of $\underset{\sim}{\mu}$ & Σ. In particular,

$$\underset{\sim}{X}^{(2)} \sim N_{p-p_1}(\underset{\sim}{\mu}^{(2)}, \Sigma_{22}).$$

proof: [From the previous theorem, we have following facts:

If $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then

$\underset{\sim}{X}^{(1)} \sim N_{p_1}(\underset{\sim}{\mu}^{(1)}, \Sigma_{11})$

& $\underset{\sim}{X}^{(2)} \sim N_{p-p_1}(\underset{\sim}{\mu}^{(2)}, \Sigma_{22})$ if $\Sigma_{12} = 0$]

Consider the transformation

$$\underset{\sim}{Y}^{(1)} = \underset{\sim}{X}^{(1)} + M\underset{\sim}{X}^{(2)}$$

$$\underset{\sim}{Y}^{(2)} = \underset{\sim}{X}^{(2)}$$

Where M is such that

$$\text{cov}(\underset{\sim}{Y}^{(1)}, \underset{\sim}{Y}^{(2)}) = 0.$$

$$\Rightarrow \quad E[\underset{\sim}{Y}^{(1)} - E(\underset{\sim}{Y}^{(1)})][\underset{\sim}{Y}^{(2)} - E(\underset{\sim}{Y}^{(2)})]' = 0$$

$$\Rightarrow \quad E[(\underset{\sim}{X}^{(1)} - \underset{\sim}{\mu}^{(1)}) + M(\underset{\sim}{X}^{(2)} - \underset{\sim}{\mu}^{(2)})][\underset{\sim}{X}^{(2)} - \underset{\sim}{\mu}^{(2)}]' = 0$$

$$\Rightarrow \quad E[\underset{\sim}{X}^{(1)} - \underset{\sim}{\mu}^{(1)}][\underset{\sim}{X}^{(2)} - \underset{\sim}{\mu}^{(2)}]' + M\,E[\underset{\sim}{X}^{(2)} - \underset{\sim}{\mu}^{(2)}][\underset{\sim}{X}^{(2)} - \underset{\sim}{\mu}^{(2)}]' = 0$$

$$\Rightarrow \quad \text{cov}(\underset{\sim}{X}^{(1)}, \underset{\sim}{X}^{(2)}) + M\,\text{cov}(\underset{\sim}{X}^{(2)}, \underset{\sim}{X}^{(2)}) = 0$$

$$\Rightarrow \quad \Sigma_{12} + M\Sigma_{22} = 0$$

$$\Rightarrow \quad M = -\Sigma_{12}\Sigma_{22}^{-1}$$

Hence the transformation becomes

$$\underset{\sim}{Y}^{(1)} = \underset{\sim}{X}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{X}^{(2)}$$

$$\underset{\sim}{Y}^{(2)} = \underset{\sim}{X}^{(2)}$$

$$\Leftrightarrow \quad \underset{\sim}{Y} = \begin{pmatrix} \underset{\sim}{Y}^{(1)} \\ \underset{\sim}{Y}^{(2)} \end{pmatrix} = \begin{pmatrix} I_{p_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p-p_1} \end{pmatrix} \begin{pmatrix} \underset{\sim}{X}^{(1)} \\ \underset{\sim}{X}^{(2)} \end{pmatrix}$$

$$\Leftrightarrow \quad \underset{\sim}{Y} = P\underset{\sim}{X}, \text{ say, where } P = \begin{pmatrix} I_p & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p-p_1} \end{pmatrix} \text{ is n.s.}$$

Hence by theorem 7,

$$\underset{\sim}{Y}^{p\times 1} = P\underset{\sim}{X} \sim N_p(P\underset{\sim}{\mu}, P\Sigma P')$$

Note that

$$P\underset{\sim}{\mu} = \begin{pmatrix} I_{p_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p-p_1} \end{pmatrix} \begin{pmatrix} \underset{\sim}{\mu}^{(1)} \\ \underset{\sim}{\mu}^{(2)} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{\mu}^{(2)} \\ \underset{\sim}{\mu}^{(2)} \end{pmatrix}$$

$$\& \quad P\Sigma P' = \begin{pmatrix} I_p & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p-p_1} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I_p & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p-p_1} \end{pmatrix}'$$

$$= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}$$

$$\text{Hence, } \begin{pmatrix} \underset{\sim}{Y}^{(1)} \\ \underset{\sim}{Y}^{(2)} \end{pmatrix} \sim N_p\left(\begin{pmatrix} \underset{\sim}{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{\mu}^{(2)} \\ \underset{\sim}{\mu}^{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \right)$$

Here, $cov(\underset{\sim}{Y}^{(1)}, \underset{\sim}{Y}^{(2)}) = 0$

Hence, we have

$$\underset{\sim}{Y}^{(2)} = \underset{\sim}{X}^{(2)} \sim N_{p-p_1}(\underset{\sim}{\mu}^{(2)}, \Sigma_{22}).$$

Conditional distn:

Theorem 13:

If $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$, then the conditional distn of $\underset{\sim}{X}^{(1)}$ given $\underset{\sim}{X}^{(2)} = \underset{\sim}{x}^{(2)}$ is

$$N_{p_1}\left(\underset{\sim}{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}^{(2)} - \underset{\sim}{\mu}^{(2)}), \Sigma_{11\cdot 2}\right)$$

Where $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

proof: Consider the following transformation-

$$\underset{\sim}{Y}^{(1)} = \underset{\sim}{X}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{X}^{(2)}$$

$$\underset{\sim}{Y}^{(2)} = \underset{\sim}{X}^{(2)}$$

Then $\begin{pmatrix} \underset{\sim}{Y}^{(1)} \\ \underset{\sim}{Y}^{(2)} \end{pmatrix} \sim N_p\left[\begin{pmatrix} \underset{\sim}{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{\mu}^{(2)} \\ \underset{\sim}{\mu}^{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11\cdot 2} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right]$

Hence the pdf of $\begin{pmatrix} \underset{\sim}{Y}^{(1)} \\ \underset{\sim}{Y}^{(2)} \end{pmatrix}$ is

$n(\underset{\sim}{y}^{(1)}, \underset{\sim}{y}^{(2)})$

$= n(\underset{\sim}{y}^{(1)} \mid \underset{\sim}{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{\mu}^{(2)}, \Sigma_{11\cdot 2})\, n(\underset{\sim}{y}^{(2)} \mid \underset{\sim}{\mu}^{(2)}, \Sigma_{22})$ ... (*)

[$\because$ $\underset{\sim}{Y}^{(1)} \sim N_{p_1}(\underset{\sim}{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{\mu}^{(2)}, \Sigma_{11\cdot 2})$

& $\underset{\sim}{Y}^{(2)} \sim N_{p-p_1}(\underset{\sim}{\mu}^{(2)}, \Sigma_{22})$ indepdtly,

as $\text{cov}(\underset{\sim}{Y}^{(1)}, \underset{\sim}{Y}^{(2)}) = 0$]

The pdf of $\begin{pmatrix} \underset{\sim}{X}^{(1)} \\ \underset{\sim}{X}^{(2)} \end{pmatrix}$ will be obtained from (*) by replacing $\underset{\sim}{y}^{(1)}$ by $\underset{\sim}{x}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{x}^{(2)}$ & $\underset{\sim}{y}^{(2)}$ by $\underset{\sim}{x}^{(2)}$ since the Jacobian of the transformation is unity $\left[|J(\frac{\underset{\sim}{y}}{\underset{\sim}{x}})| = \left\| \begin{matrix} I_{p_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p-p_1} \end{matrix} \right\| = 1 \right]$.

Hence the pdf of $\begin{pmatrix} \underset{\sim}{X}^{(1)} \\ \underset{\sim}{X}^{(2)} \end{pmatrix}$ is

$$n(\underset{\sim}{x}^{(1)}, \underset{\sim}{x}^{(2)})$$

$$= n\left(\underset{\sim}{x}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{x}^{(2)} \,\middle|\, \underset{\sim}{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{\mu}^{(2)}, \Sigma_{11\cdot 2}\right)$$

$$\times\, n\left(\underset{\sim}{x}^{(2)} \,\middle|\, \underset{\sim}{\mu}^{(2)}, \Sigma_{22}\right).$$

Hence the conditional pdf of $\underset{\sim}{X}^{(1)}$ given $\underset{\sim}{X}^{(2)} = \underset{\sim}{x}^{(2)}$ is

$$f_{\underset{\sim}{X}^{(1)} | \underset{\sim}{X}^{(2)} = \underset{\sim}{x}^{(2)}}(\underset{\sim}{x}^{(1)})$$

$$= \frac{n(\underset{\sim}{x}^{(1)}, \underset{\sim}{x}^{(2)})}{n(\underset{\sim}{x}^{(2)} \,|\, \underset{\sim}{\mu}^{(2)}, \Sigma_{22})}$$

$$= n\left(\underset{\sim}{x}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{x}^{(2)} \,\middle|\, \underset{\sim}{\mu}^{(2)} - \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{\mu}^{(2)}, \Sigma_{11\cdot 2}\right)$$

$$= \frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma_{11\cdot 2}|}}\, e^{-\frac{1}{2}\left[\left(\underset{\sim}{x}^{(1)} - \underset{\sim}{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}^{(2)} - \underset{\sim}{\mu}^{(2)})\right)' \Sigma_{11\cdot 2}^{-1} \left(\underset{\sim}{x}^{(1)} - \underset{\sim}{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}^{(2)} - \underset{\sim}{\mu}^{(2)})\right)\right]}$$

Hence,

$$\underset{\sim}{X}^{(1)} | \underset{\sim}{X}^{(2)} = \underset{\sim}{x}^{(2)} \sim N_{p_1}\left(\underset{\sim}{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}^{(2)} - \underset{\sim}{\mu}^{(2)}), \Sigma_{11\cdot 2}\right)$$ ✓

<u>Remarks:</u>

(i) Note that

$$E(\underset{\sim}{X}^{(1)} | \underset{\sim}{X}^{(2)} = \underset{\sim}{x}^{(2)}) = \underset{\sim}{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}^{(2)} - \underset{\sim}{\mu}^{(2)}),$$ a linear f^n of $\underset{\sim}{x}^{(2)}$.

Hence the regression of $\underset{\sim}{X}^{(1)}$ on $\underset{\sim}{X}^{(2)}$ is linear.

Also, the dispersion matrix of $\underset{\sim}{X}^{(1)}$ given $\underset{\sim}{X}^{(2)} = \underset{\sim}{x}^{(2)}$ is $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ which is

Indep^t of $\underset{\sim}{x}^{(2)}$.
Hence the conditional dist^n of $\underset{\sim}{x}^{(1)}$ given $\underset{\sim}{x}^{(2)}$ is homoscedastic.

<u>Problems:</u>

(4) Suppose $\underset{\sim}{X} = (X_1, X_2, X_3) \sim N_3(\underset{\sim}{0}, \Sigma)$.

[1993 3(a)] Where $\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}$

S.T for any $c > 0$,

$$P\left[(X_2^2 + c)\rho^2 - 2(X_1X_2 + X_2X_3)\rho + X_1^2 + X_2^2 + X_3^2 - c \le 0\right]$$

$$= \int_0^c \frac{1}{\sqrt{2\pi}} e^{-y/2} y^{1/2}\, dy$$

<u>Sol^n:</u> Since $\underset{\sim}{X} \sim N_3(0, \Sigma)$,

$$(\underset{\sim}{X} - \underset{\sim}{0})' \Sigma^{-1} (\underset{\sim}{X} - \underset{\sim}{0}) \sim \chi^2_3.$$

Hence $P[\underset{\sim}{X}' \Sigma^{-1} \underset{\sim}{X} \le c] = \int_0^c f_{\chi^2_3}(y)\, dy$

$$= \int_0^c \frac{1}{\sqrt{2\pi}} e^{-y/2} y^{1/2}\, dy$$

Now, $\underset{\sim}{X}' \Sigma^{-1} \underset{\sim}{X} \le c$

$$\Leftrightarrow (X_2^2 + c)\rho^2 - 2(X_1X_2 + X_2X_3)\rho + X_1^2 + X_2^2 + X_3^2 - c \le 0.$$

(5) Suppose $(X_1, \cdots, X_p)^T \sim N_p(\underset{\sim}{\mu}, \Sigma)$, where

[2005 3(a)] $\underset{\sim}{\mu} = (1, 1, \ldots, 1)^T$, $\Sigma = \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 2 & 2 & 2 & \cdots & 2 & 2 \\ 1 & 2 & 3 & 3 & \cdots & 3 & 3 \\ 1 & 2 & 3 & 4 & \cdots & 4 & 4 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 2 & 3 & 4 & \cdots & p-1 & p \end{pmatrix}$

Then, S.T. $Q = (X_2 - X_1)^2 + (X_3 - X_2)^2 + \cdots + (X_p - X_{p-1})^2$ has a chi-square dist^n.

Sol^n: Let $Y_i = X_i - X_{i-1}$, $i = 2(1)p$.

$$\therefore \underset{\sim}{Y}^{\overline{p-1}\times 1} = \begin{pmatrix} Y_2 \\ Y_3 \\ \vdots \\ Y_p \end{pmatrix}$$

$$= \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}^{\overline{p-1}\times p} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_p \end{pmatrix}$$

$$= B^{\overline{p-1}\times p} \underset{\sim}{X}^{p\times 1} \text{, say.}$$

Note that $\rho(B^{\overline{p-1}\times p}) = p-1$. [$\rho(A)$: rank of A]

Hence,

$\underset{\sim}{Y} = B\underset{\sim}{X} \sim N_{p-1}(B\underset{\sim}{\mu}, B\Sigma B^T)$ [Theorem 9]

It can be shown that

$$B\underset{\sim}{\mu} = \underset{\sim}{0}$$
$$B\Sigma B^T = I_{p-1}$$

Hence, $\underset{\sim}{Y} \sim N_{p-1}(\underset{\sim}{0}, I_{p-1})$

$\Rightarrow$ $Y_2, Y_3, \ldots, Y_p$ are iid $N_1(0,1)$

$\Rightarrow$ $Q = \sum_{i=2}^{p} Y_i^2 \sim \chi^2_{p-1}$

Theorem 14:

$\underset{\sim}{X}^{p\times 1} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ iff $\underset{\sim}{l}'\underset{\sim}{x} \sim N_1(\underset{\sim}{l}'\underset{\sim}{\mu}, \underset{\sim}{l}'\Sigma\underset{\sim}{l}) \; \forall \; \underset{\sim}{l}^{p\times 1} \in \mathbb{R}^p$

proof: If: Let $\underset{\sim}{l}'\underset{\sim}{x} \sim N_1(\underset{\sim}{l}'\underset{\sim}{\mu}, \underset{\sim}{l}'\Sigma\underset{\sim}{l})$

Then the mgf of $\underset{\sim}{l}'\underset{\sim}{x}$ is

$$E\left(e^{t(\underset{\sim}{l}'\underset{\sim}{x})}\right)$$

$$= e^{t(\underset{\sim}{l}'\underset{\sim}{\mu}) + \frac{1}{2}t^2(\underset{\sim}{l}'\Sigma\underset{\sim}{l})} \quad \forall t$$

Putting $t=1$,

$$E\left(e^{\underset{\sim}{l}'\underset{\sim}{x}}\right) = e^{\underset{\sim}{l}'\underset{\sim}{\mu} + \frac{1}{2}(\underset{\sim}{l}'\Sigma\underset{\sim}{l})} \quad \forall \underset{\sim}{l}$$

$\Leftrightarrow$ The mgf of $\underset{\sim}{x}$

$$= e^{\underset{\sim}{l}'\underset{\sim}{\mu} + \frac{1}{2}\underset{\sim}{l}'\Sigma\underset{\sim}{l}}$$

$$= \text{mgf of } N_p(\underset{\sim}{\mu}, \Sigma).$$

By uniqueness property of mgf,

$$\underset{\sim}{x} \sim N_p(\underset{\sim}{\mu}, \Sigma)$$

Only if: Let $\underset{\sim}{X}^{p\times 1} \sim N_p(\underset{\sim}{\mu}, \Sigma)$.

Consider a vector $k\underset{\sim}{t}$ where k is a constant.

The mgf of $\underset{\sim}{x}$ is

$$E\left(e^{(k\underset{\sim}{t})'\underset{\sim}{x}}\right)$$

$$= e^{(k\underset{\sim}{t})'\underset{\sim}{\mu} + \frac{1}{2}(k\underset{\sim}{t})'\Sigma(k\underset{\sim}{t})} \quad \forall \underset{\sim}{t} \in \mathbb{R}^p.$$

$$\Leftrightarrow E\left(e^{k(\underset{\sim}{t}'\underset{\sim}{x})}\right) = e^{k(\underset{\sim}{t}'\underset{\sim}{\mu}) + \frac{1}{2}k^2(\underset{\sim}{t}'\Sigma\underset{\sim}{t})}$$

$$\Leftrightarrow \text{mgf of } \underset{\sim}{t}'\underset{\sim}{x} = e^{k(\underset{\sim}{t}'\underset{\sim}{\mu}) + \frac{1}{2}k^2(\underset{\sim}{t}'\Sigma\underset{\sim}{t})} \quad \forall \underset{\sim}{t} \in \mathbb{R}^p$$

$$= \text{mgf of } N_1(\underset{\sim}{t}'\underset{\sim}{\mu}, \underset{\sim}{t}'\Sigma\underset{\sim}{t})$$

Hence $\underset{\sim}{t}'\underset{\sim}{x} \sim N_1(\underset{\sim}{t}'\underset{\sim}{\mu}, \underset{\sim}{t}'\Sigma\underset{\sim}{t}) \quad \forall \; t \in \mathbb{R}^p$.

Multiple & Partial Correl$^{\text{n}}$ Coeffs:

The regression of $\underset{\sim}{X}^{(1)}$ on $\underset{\sim}{X}^{(2)}$ is

$$E(\underset{\sim}{X}^{(1)} | \underset{\sim}{X}^{(2)})$$

$$= \underset{\sim}{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{X}^{(2)} - \underset{\sim}{\mu}^{(2)})$$

Define the residual variables by

$$\underset{\sim}{X}_{1\cdot2} = \underset{\sim}{X}^{(1)} - \underset{\sim}{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{X}^{(2)} - \underset{\sim}{\mu}^{(2)})$$

Note that $E(\underset{\sim}{X}_{1\cdot2}) = \underset{\sim}{0}$ &

$$E[(\underset{\sim}{X}^{(2)} - \underset{\sim}{\mu}^{(2)})\underset{\sim}{X}'_{1\cdot2}]$$

$$= E\left[(\underset{\sim}{X}^{(2)} - \underset{\sim}{\mu}^{(2)})\left\{(\underset{\sim}{X}^{(1)} - \underset{\sim}{\mu}^{(1)}) - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{X}^{(2)} - \underset{\sim}{\mu}^{(2)})\right\}^T\right]$$

$$= \Sigma_{21} - \Sigma_{22}\Sigma_{22}^{-1}\Sigma_{21} = 0$$

ie the residual variables are uncorrelated with the fixed set variables.

Again, the covariance matrix of $\underset{\sim}{X}_{1\cdot2}$ is

$$E(\underset{\sim}{X}_{1\cdot2}\underset{\sim}{X}'_{1\cdot2})$$

$$= E\left[\left\{X^{(1)} - \underset{\sim}{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{X}^{(2)} - \underset{\sim}{\mu}^{(2)})\right\}\underset{\sim}{X}'_{1\cdot2}\right]$$

$$= E\left[(\underset{\sim}{X}^{(1)} - \underset{\sim}{\mu}^{(1)})\underset{\sim}{X}^T_{1\cdot2}\right]$$ since the residuals are uncorrelated with $\underset{\sim}{X}^{(2)}$.

$$= E\left[(\underset{\sim}{X}^{(1)} - \underset{\sim}{\mu}^{(1)})\left\{\underset{\sim}{X}^{(1)} - \underset{\sim}{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{X}^{(2)} - \underset{\sim}{\mu}^{(2)})\right\}^T\right]$$

$$= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Sigma_{11\cdot2}$$

= covariance matrix of the conditional dis$^{\text{n}}$ of $\underset{\sim}{X}^{(1)}$ given $\underset{\sim}{X}^{(2)}$.

Hence the elements of the covariance matrix of the of the conditional dis$^{\text{n}}$

of $\underset{\sim}{X}^{(1)}$ given $\underset{\sim}{x}^{(2)}$ are the partial variances & covariances.

Let the $(i,j)^{th}$ element of

$$\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

be denoted by $\sigma_{ij\cdot p_1+1\cdots p}$.

Then the partial correlⁿ coeff betⁿ X_i & X_j with all the members of the 2nd set held constant is

$$\rho_{ij\cdot p_1+1\cdots p} = \frac{\sigma_{ij\cdot p_1+1\cdots p}}{\sqrt{\sigma_{ii\cdot p_1+1\cdots p}}\sqrt{\sigma_{jj\cdot p_1+1\cdots p}}}, \quad i,j = 1(1)p_1, \; i \neq j.$$

Note that $E(\underset{\sim}{X}^{(1)} \mid \underset{\sim}{X}^{(2)} = \underset{\sim}{x}^{(2)})$

$$= \underset{\sim}{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}^{(2)} - \underset{\sim}{\mu}^{(2)})$$

Consider the component X_i, $i \in \{1, 2, \cdots, p_1\}$,

$$E(X_i \mid \underset{\sim}{X}^{(2)} = \underset{\sim}{x}^{(2)}) = \alpha + \underset{\sim}{\beta}'\underset{\sim}{x}^{(2)}$$

where $\underset{\sim}{\beta}' = \underset{\sim}{\sigma}_{(i)}'\Sigma_{22}^{-1}$ where $\underset{\sim}{\sigma}_{(i)}'$ is the ith row of Σ_{12}

& $\alpha = \mu_i - \underset{\sim}{\beta}'\underset{\sim}{\mu}^{(2)}$

The multiple correlⁿ coeff

$$\rho_{i\cdot p_1+1\cdots p} = \sqrt{\frac{\underset{\sim}{\beta}'\Sigma_{22}\underset{\sim}{\beta}}{\sigma_{ii}}} = \sqrt{\frac{\underset{\sim}{\sigma}_{(i)}'\Sigma_{22}^{-1}\underset{\sim}{\sigma}_{(i)}}{\sigma_{ii}}}$$

$$\Rightarrow 1 - \rho^2_{i\cdot p_1+1\cdots p} = \frac{\sigma_{ii} - \underset{\sim}{\sigma}_{(i)}'\Sigma_{22}^{-1}\underset{\sim}{\sigma}_{(i)}}{\sigma_{ii}}$$

$$= \frac{|\Sigma^*|}{\sigma_{ii}|\Sigma_{22}|}$$

where $\Sigma^* = \begin{pmatrix} \sigma_{ii} & \underset{\sim}{\sigma}_{(i)}' \\ \underset{\sim}{\sigma}_{(i)} & \Sigma_{22} \end{pmatrix}$

Problems

⑥ Let $\underset{\sim}{x} \sim N_p$ ~~$(\underset{\sim}{\mu}, \Sigma)$~~ $(\underset{\sim}{0}, \sigma^2 I_p)$ and $P_1^{m\times p}$ is a matrix such that $P_1 P_1^T = I_m$, then show that $\underset{\sim}{Y} = P_1 \underset{\sim}{x} \sim N_m(0, \sigma^2 I_m)$ is indeptly dist'd with $\frac{1}{\sigma^2}(\underset{\sim}{x}'\underset{\sim}{x} - \underset{\sim}{Y}'\underset{\sim}{Y}) \sim \chi^2_{p-m}$.

Soln: Note that

$P_1 P_1^T = I_m$, here $P_1^{m\times p}$ is a semi-orthogonal matrix.

Hence we can find $P_2^{p-m\times p}$ such that $PP^T = I_p$ where $P^{p\times p} = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}$

Let $\underset{\sim}{Z} = P\underset{\sim}{x} = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}\underset{\sim}{x} = \begin{pmatrix} P_1\underset{\sim}{x} \\ P_2\underset{\sim}{x} \end{pmatrix} = \begin{pmatrix} \underset{\sim}{Y} \\ \underset{\sim}{W} \end{pmatrix}$ where $\underset{\sim}{W} = P_2\underset{\sim}{x}$.

Since P is n.s,

$$\underset{\sim}{Z} = P\underset{\sim}{x} \sim N_p(\underset{\sim}{0}, \sigma^2 I_p)$$

ie $\begin{pmatrix} \underset{\sim}{Y} \\ \underset{\sim}{W} \end{pmatrix} \sim N_p(\underset{\sim}{0}, \sigma^2 I_p)$

$\Rightarrow \underset{\sim}{Y}^{m\times 1} \sim N_m(\underset{\sim}{0}, \sigma^2 I_m)$

& $\underset{\sim}{W} \sim N_{p-m}(0, \sigma^2 I_{p-m})$ indeptly.

Now, $\underset{\sim}{Z}'\underset{\sim}{Z} = \underset{\sim}{x}'P'P\underset{\sim}{x} = \underset{\sim}{x}'\underset{\sim}{x}$

$$\Rightarrow \underset{\sim}{x}'\underset{\sim}{x} = \underset{\sim}{Z}'\underset{\sim}{Z} = \begin{pmatrix} \underset{\sim}{Y} \\ \underset{\sim}{W} \end{pmatrix}'\begin{pmatrix} \underset{\sim}{Y} \\ \underset{\sim}{W} \end{pmatrix}$$

$$= (\underset{\sim}{Y}' \ \underset{\sim}{W}')\begin{pmatrix} \underset{\sim}{Y} \\ \underset{\sim}{W} \end{pmatrix}$$

$$= \underset{\sim}{Y}'\underset{\sim}{Y} + \underset{\sim}{W}'\underset{\sim}{W}$$

$\Rightarrow \underset{\sim}{W}'\underset{\sim}{W} = \underset{\sim}{x}'x -$ ~~$\underset{\sim}{Z}'\underset{\sim}{Z}$~~ $\underset{\sim}{Y}'\underset{\sim}{Y}$

Since $\underset{\sim}{W} \sim N_{p-m}(\underset{\sim}{0}, \sigma^2 I_{p-m})$

$$\Rightarrow \frac{\underset{\sim}{W}'\underset{\sim}{W}}{\sigma^2} \sim \chi^2_{p-m}$$

$$\Rightarrow \frac{1}{\sigma^2}(\underset{\sim}{X}'\underset{\sim}{X} - \underset{\sim}{Y}'\underset{\sim}{Y}) = \frac{\underset{\sim}{W}'\underset{\sim}{W}}{\sigma^2} \sim \chi^2_{p-m}$$

Which is indeptly dis^{td} with

$$\underset{\sim}{Y} \sim N_m(0, \sigma^2 I_m)$$

(7) Let $\underset{\sim}{Y}_\alpha \sim N_p(a_\alpha \underset{\sim}{\mu}, \Sigma)$, $\alpha = 1(1)n$ indeptly.
S.T. $B = \sum_{\alpha=1}^{n} \underset{\sim}{Y}_\alpha \underset{\sim}{Y}_\alpha' - \underset{\sim}{Z}\underset{\sim}{Z}'$ & $\underset{\sim}{Z}\underset{\sim}{Z}'$ are indeptly dis^{td} where $\underset{\sim}{Z} = \sum_{\alpha=1}^{n} \left(\frac{a_\alpha}{\sqrt{\sum_1^n a_\alpha^2}}\right) \underset{\sim}{Y}_\alpha$

Soln: Let $\underset{\sim}{Z}_\alpha = \sum_{\beta=1}^{n} b_{\alpha\beta} \underset{\sim}{Y}_\beta \quad \alpha = 1(1)n.$

Where $b_{n\beta} = \frac{a_\beta}{\sqrt{\sum_{\alpha=1}^{n} a_\alpha^2}}$ & $B = (b_{\alpha\beta})$ is orthogonal

Now, $\sum_{\alpha=1}^{n} \underset{\sim}{Z}_\alpha \underset{\sim}{Z}_\alpha'$

$$= \sum_{\alpha=1}^{n} \left(\sum_{\beta=1}^{n} b_{\alpha\beta} \underset{\sim}{Y}_\beta\right)\left(\sum_{\gamma=1}^{n} b_{\alpha\gamma} \underset{\sim}{Y}_\gamma'\right)$$

$$= \sum_\beta \sum_\gamma \sum_{\alpha=1}^{n} (b_{\alpha\beta} b_{\alpha\gamma}) \underset{\sim}{Y}_\beta \underset{\sim}{Y}_\gamma'$$

$$= \sum_\beta \sum_\gamma \delta_{\beta\gamma} \underset{\sim}{Y}_\beta \underset{\sim}{Y}_\gamma' \quad \text{since } B^T B = I_n$$

[The $(\beta,\gamma)^{th}$ element of $B^T B$ is

$$C_{\beta\gamma} = \sum_{\alpha=1}^{n} b'_{\beta\alpha} b_{\alpha\gamma}$$
$$= \sum_{\alpha=1}^{n} b_{\alpha\beta} b_{\alpha\gamma}$$

$\because B^T B = I_n$
$\Rightarrow \sum_{\alpha=1}^{n} b_{\alpha\beta} b_{\alpha\gamma} = \delta_{\beta\gamma}$]

$$= \sum_{\beta} \underset{\sim}{Y}_\beta \underset{\sim}{Y}'_\beta \quad \text{since } \delta_{\beta\gamma} = \begin{cases} 1 & \text{if } \beta = \gamma \\ 0 & \text{if } \beta \neq \gamma \end{cases}$$

Note that

$$\underset{\sim}{Z}_n = \sum_{\beta=1}^{n} \left(\frac{a_\beta}{\sqrt{\sum_1^n a_\alpha^2}} \right) \underset{\sim}{Y}_\beta$$

$$= \underset{\sim}{Z}$$

Since $\underset{\sim}{Z}_\alpha$'s are the linear combination of $\underset{\sim}{Y}_\alpha$'s, the set $\{\underset{\sim}{Z}_\alpha\}$ have a jt normal distn.

Now,

$$\text{cov}(\underset{\sim}{Z}_\alpha, \underset{\sim}{Z}_\gamma)$$

$$= \text{cov}\left(\sum_{\beta=1}^{n} b_{\alpha\beta} \underset{\sim}{Y}_\beta, \sum_{\epsilon=1}^{n} b_{\gamma\epsilon} \underset{\sim}{Y}_\epsilon \right)$$

$$= \sum_{\beta} \sum_{\epsilon} b_{\alpha\beta} b_{\gamma\epsilon} \text{cov}(\underset{\sim}{Y}_\beta, \underset{\sim}{Y}_\epsilon)$$

$$= \sum_{\beta} \sum_{\epsilon} b_{\alpha\beta} b_{\gamma\epsilon} (\delta_{\beta\epsilon} \cdot \Sigma)$$

$$= \left(\sum_{\beta=1}^{n} b_{\alpha\beta} b_{\gamma\beta} \right) \Sigma \qquad \text{since } \delta_{\beta\epsilon} = \begin{cases} 1 & \text{if } \beta = \epsilon \\ 0 & \text{o.w} \end{cases}$$

$$= \delta_{\alpha\gamma} \Sigma$$

For $\alpha \neq \gamma$, $\text{cov}(\underset{\sim}{Z}_\alpha, \underset{\sim}{Z}_\gamma) = 0$, ie $\underset{\sim}{Z}_\alpha$'s are indept.

Now,

$$Q = \sum_{\alpha=1}^{n} \underset{\sim}{Y}_\alpha \underset{\sim}{Y}'_\alpha - \underset{\sim}{Z} \underset{\sim}{Z}'$$

$$= \sum_{\alpha=1}^{n} \underset{\sim}{Z}_\alpha \underset{\sim}{Z}'_\alpha - \underset{\sim}{Z}_n \underset{\sim}{Z}'_n$$

$$= \sum_{\alpha=1}^{n-1} \underset{\sim}{Z}_\alpha \underset{\sim}{Z}'_\alpha \quad \&$$

$\underset{\sim}{Z}\underset{\sim}{Z}' = \underset{\sim}{Z}_n \underset{\sim}{Z}'_n$ are indeptly distd.

Scanned by CamScanner

Multinomial dis$^{\text{n}}$:

Let $X_1, X_2, \dots, X_k$ be jtly dis^{td} with pmf

$$f(\underset{\sim}{x}) = \begin{cases} \dfrac{m!}{x_1!\, x_2! \cdots x_k!} \, p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} & \text{for nonnegative integral values } x_1, \dots, x_k \text{ such that } \sum_1^k x_i = m \\ 0 & \text{o.w.} \end{cases}$$

where the parameters p_i are such that

$$p_i > 0 \quad \forall \; i = 1(1)k$$

$$\& \quad \sum_1^k p_i = 1$$

Then $(X_1, \dots, X_k)$ is said to follow a multinomial dis$^{\text{n}}$.

The mgf of $\underset{\sim}{X}$ is

$$M_{\underset{\sim}{X}}(\underset{\sim}{t}) = E\left(e^{\sum_1^k t_i x_i}\right)$$

$$= \sum_{x_1, \dots, x_k} e^{\sum_1^k t_i x_i} \frac{m!}{\prod_1^k x_i!} \prod_1^k p_i^{x_i}$$

$$= \sum_{x_1, \dots, x_k} \frac{m!}{\prod_1^k x_i!} \prod_1^k (p_i e^{t_i})^{x_i}$$

$$= \left(\sum_1^k p_i e^{t_i}\right)^m$$

Note that,

$$\frac{\partial M_{\underset{\sim}{x}}(\underset{\sim}{t})}{\partial t_i} = m\left(p_1 e^{t_1} + \dots + p_i e^{t_i} + \dots + p_k e^{t_k}\right)^{m-1} p_i e^{t_i}$$

$$\frac{\partial^2 M_{\underset{\sim}{x}}(\underset{\sim}{t})}{\partial t_i^2} = m(m-1)\left(p_1 e^{t_1} + \dots + p_k e^{t_k}\right)^{m-2} \{p_i e^{t_i}\}^2$$

Scanned by CamScanner

$$+ \; m p_i (p_1 e^{t_1} + \dots + p_k e^{t_k})^{m-1} e^{t_i}$$

$$\frac{\partial^2 M_{\underset{\sim}{X}}(\underset{\sim}{t})}{\partial t_i \partial t_j} = m p_i e^{t_i} (m-1)(p_1 e^{t_1} + \dots + p_k e^{t_k})^{m-2} p_j e^{t_j} \quad [i \neq j]$$

Evaluating the derivatives at $\underset{\sim}{t} = \underset{\sim}{0}$, we get

$$E(X_i) = m p_i$$

$$E(X_i^2) = m(m-1) p_i^2 + m p_i$$

$$\Rightarrow \quad V(X_i) = m p_i (1 - p_i)$$

$$E(X_i X_j) = m(m-1) p_i p_j$$

$$\Rightarrow \quad \text{cov}(X_i, X_j) = -m p_i p_j, \; i \neq j$$

$$\therefore \rho_{X_i, X_j} = -\sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}}, \quad i \neq j$$

The dispersion matrix of $\underset{\sim}{X} = (X_1, \dots, X_k)$ is

$$\Sigma = \begin{pmatrix} m p_1(1-p_1) & -m p_1 p_2 & \dots & -m p_1 p_k \\ -m p_2 p_1 & m p_2(1-p_2) & \dots & -m p_2 p_k \\ \vdots & \vdots & & \vdots \\ -m p_k p_1 & -m p_k p_2 & \dots & m p_k(1-p_k) \end{pmatrix}$$

Now

$$|\Sigma| = m^k \begin{vmatrix} p_1(1-p_1) & -p_1 p_2 & \dots & -p_1 p_k \\ -p_2 p_1 & p_2(1-p_2) & \dots & -p_2 p_k \\ \vdots & \vdots & & \vdots \\ -p_k p_1 & -p_k p_2 & \dots & p_k(1-p_k) \end{vmatrix}$$

$$= m^K \begin{vmatrix} 0 & -p_1p_2 & \cdots & -p_1p_K \\ 0 & p_2(1-p_2) & \cdots & -p_2p_K \\ \vdots & \vdots & & \vdots \\ 0 & -p_Kp_2 & \cdots & p_K(1-p_K) \end{vmatrix} \quad \left[\text{By } c_1' \to \sum_1^K c_i\right]$$

$$= 0 \qquad \left[\because \sum_1^K p_i = 1\right]$$

Hence, the above form of the dis^n is singular since $|\Sigma^{K\times K}| = 0$. In fact $\text{rank}(\Sigma^{K\times K}) = K-1$. To avoid the difficulties associated with singularity, we consider the jt dis^n of $K-1$ r.v.s, say, $X_1, \ldots, X_{K-1}$ with pmf

$$f(x_1, \ldots, x_{K-1}) = \begin{cases} \dfrac{m!}{x_1! \cdots x_{K-1}! \left(m - \sum_1^{K-1} x_i\right)!} p_1^{x_1} \cdots p_{K-1}^{x_{K-1}} \left(1 - \sum_1^{K-1} p_i\right)^{m - \sum_1^{K-1} x_i} & \text{for nonnegative integral values } x_1, \ldots, x_{K-1} \text{ such that } \sum_1^{K-1} x_i \le m \\ 0 & \text{o.w} \end{cases}$$

Then, the dis^n of $(X_1, X_2, \ldots, X_{K-1})$ is said to follow a multinomial dis^n with parameters $m, p_1, \ldots, p_{K-1}$, such that $p_i > 0 \ \forall i$ & $\sum_1^{K-1} p_i < 1$

The mgf of the above multinomial dis^n is

$M_{X_1, \ldots, X_{K-1}}(t_1, \ldots, t_{K-1})$

$$= E\left(e^{\sum_1^{K-1} t_i X_i}\right)$$

$$= \sum_{x_1, \ldots, x_{K-1}} e^{\sum_1^{K-1} t_i x_i} \frac{m!}{\prod_1^{K-1} x_i! \left(m - \sum_1^{K-1} x_i\right)!} \prod_1^{K-1} p_i^{x_i} \left(1 - \sum_1^{K-1} p_i\right)^{m - \sum_1^{K-1} x_i}$$

$$= \sum_{\substack{x_1, \ldots, x_{k-1} \\ \ni \sum_1^{k-1} x_i \le m}} \frac{m!}{\prod_1^{k-1} x_i! \, (m - \sum_1^{k-1} x_i)!} \prod_1^{k-1} (p_i e^{t_i})^{x_i} \left(1 - \sum_1^{k-1} p_i\right)^{m - \sum_1^{k-1} x_i}$$

$$= \left(p_1 e^{t_1} + \cdots + p_{k-1} e^{t_{k-1}} + 1 - p_1 - \cdots - p_{k-1}\right)^m$$

The marginal mgf of X_i is

$$M_{X_i}(t_i) = M_{X_1, \ldots, X_{k-1}}(0, 0, \ldots, 0, t_i, 0, \ldots, 0)$$

$$= \left(p_i e^{t_i} + 1 - p_i\right)^m$$

$$= \text{mgf of } Bin(m, p_i)$$

By uniqueness property of mgf,

$$X_i \sim Bin(m, p_i), \quad i = 1(1)\overline{k-1}$$

The mgf of the marginal distn of any subset of q r.v.s, say, $X_1, \ldots, X_q$, $q < k-1$, is

$$M_{X_1, \ldots, X_q, \ldots, X_{k-1}}(t_1, \ldots, t_q, 0, \ldots, 0)$$

$$= \left(p_1 e^{t_1} + \cdots + p_q e^{t_q} + 1 - p_1 - \cdots - p_q\right)^m$$

which the mgf of multinomial distn with parameters $m, p_1, \ldots, p_q$.

Hence any subset of a multinomial r.v's is also a multinomial random vector.

Conditional distⁿ:

The conditional distⁿ of $X_1, \dots, X_q$ given a set of values of $X_{q+1}, \dots, X_{k-1}$, say, $X_{q+1} = x_{q+1}, \dots, X_{k-1} = x_{k-1}$, is given by the conditional pmf

$$f_{X_1,\dots,X_q | X_{q+1},\dots,X_{k-1}}(x_1, \dots, x_q | x_{q+1}, \dots, x_{k-1})$$

$$= \frac{f_{X_1,\dots,X_{k-1}}(x_1,\dots,x_{k-1})}{f_{X_{q+1},\dots,X_{k-1}}(x_{q+1},\dots,x_{k-1})}$$

$$= \frac{\frac{m!}{x_1!\cdots x_q!\,x_{q+1}!\cdots x_{k-1}!\,(m-\sum_1^{k-1} x_i)!}\, p_1^{x_1}\cdots p_q^{x_q}\cdots p_{k-1}^{x_{k-1}}\left(1-\sum_1^{k-1} p_i\right)^{m-\sum_1^{k-1} x_i}}{\frac{m!}{x_{q+1}!\cdots x_{k-1}!\,(m-\sum_{q+1}^{k-1} x_i)!}\, p_{q+1}^{x_{q+1}}\cdots p_{k-1}^{x_{k-1}}\left(1-\sum_{q+1}^{k-1} p_i\right)^{m-\sum_{q+1}^{k-1} x_i}}$$

$$= \frac{(m-\sum_{q+1}^{k-1} x_i)!}{x_1!\cdots x_q!\,(m-\sum_1^{k-1} x_i)!}\left(\frac{p_1}{Q}\right)^{x_1}\cdots\left(\frac{p_q}{Q}\right)^{x_q}\left(1-\sum_1^{q}\left(\frac{p_i}{Q}\right)\right)^{(m-\sum_1^{k-1} x_i)}$$

where $Q = 1-\sum_{q+1}^{k-1} p_i$

$$\left[1-\sum_1^{k-1} p_i = \left(1-\sum_{q+1}^{k-1} p_i - \sum_1^{q} p_i\right) = \left(Q-\sum_1^{q} p_i\right)\right]$$

which is the pmf of a multinomial distⁿ with parameters $m-\sum_{q+1}^{k-1} x_i, \frac{p_1}{Q}, \dots, \frac{p_q}{Q}$ where $Q = 1-\sum_{q+1}^{k-1} p_i$.

Hence the conditional distⁿ of X_1 given $X_2 = x_2, \dots, X_{k-1} = x_{k-1}$ is binomial distⁿ with parameters $m-\sum_2^{k-1} x_i$ & $\frac{p_1}{1-\sum_2^{k-1} p_i}$

Hence,

$E(X_1 \mid X_2 = x_2, \ldots, X_{K-1} = x_{K-1})$

$$= \left(m - \sum_2^{K-1} x_i\right)\left(\frac{p_1}{1 - \sum_2^{K-1} p_i}\right)$$

which is linear in x_i's.

[The regression eqn of any r.v. on the others is linear in multinomial set up]

and $V(X_1 \mid X_2 = x_2, \ldots, X_{K-1} = x_{K-1})$

$$= \left(m - \sum_2^{K-1} x_i\right)\left(\frac{p_1}{1 - \sum_2^{K-1} p_i}\right)\left(1 - \frac{p_1}{1 - \sum_2^{K-1} p_i}\right)$$

Problems:

(8) Suppose $(X_1, \ldots, X_K)$ follows a multinomial distn with parameters m & $p_1, \ldots, p_K$, such that $\sum_1^K X_i = m$ & $\sum_1^K p_i = 1$. S.T. the square of the multiple correln coeff of X_1 on $X_2, \ldots, X_{K-1}$ is

$$\rho^2_{1.23\ldots \overline{K-1}} = \frac{p_1(p_2 + p_3 + \cdots + p_{K-1})}{(1-p_1)(1 - p_2 - p_3 - \cdots - p_{K-1})}$$

. Also find $\rho_{1.23\ldots K}$

Hint: Find Σ.

$$\rho^2_{1.23\ldots K-1} = 1 - \frac{|\Sigma|}{\sigma_{11}|\Sigma_2|}$$

(9) Suppose $(X_1, \ldots, X_K)$ follows a multinomial distn with parameters m & $p_1, \ldots, p_K$ such that $\sum_1^K X_i = m$ & $\sum_1^K p_i = 1$. S.T the partial correln coeff betn X_1 & X_2 when the variables $X_3, \ldots, X_q$ are held fixed

is $\rho_{12.34\ldots q} = -\sqrt{\dfrac{p_1 p_2}{(1-p_2-\cdots-p_q)(1-p_1-p_3-\cdots-p_q)}}$

$, q = 3(1)k-1.$

Soln: Here the distn of $(X_1,\ldots,X_k)$ is singular. Then the distn of $(X_1,\ldots,X_{k-1})$, where $\sum_1^{k-1} X_i \le m$ & $\sum_1^{k-1} p_i < 1$, is nonsingular. The marginal distn of $(X_1,\ldots,X_q)$ is a multinomial with parameters $m, p_1,\ldots,p_q$, $q \le k-1$.

Note that

$$E(X_1 \mid X_2 = x_2,\ldots,X_q = x_q)$$

$$= (m - x_2 - x_3 - \cdots - x_q) \cdot \frac{p_1}{1-\sum_{i}^{k-q-1} p_i}$$

Again, since the regression is linear,

$$E(X_1 \mid X_2 = x_2,\ldots,X_q = x_q)$$

$$= \alpha + \underset{\sim}{\beta}' \underset{\sim}{x}^{(2)}$$

$$= \alpha + \sum_{j=2}^{q} \beta_{1j.2\ldots j-1\, j+1\ldots q}\, x_j$$

$$\therefore \beta_{12.34\ldots q} = -\frac{p_1}{1-p_2-p_3-\cdots-p_q}$$

Similarly,

$$\beta_{21.34\ldots q} = -\frac{p_2}{1-p_1-p_3-\cdots-p_q}$$

$$\therefore \rho_{12.34\ldots q} = -\sqrt{\beta_{12.34\ldots q} \cdot \beta_{21.34\ldots q}}$$

$$= -\sqrt{\frac{p_1 p_2}{(1-p_2-\cdots-p_q)(1-p_1-p_3-\cdots-p_q)}}$$

Since $\beta_{12.34\ldots q}$, $\beta_{21.34\ldots q}$ & $\rho_{12.34\ldots q}$ have the same sign.

Alt: Use $\rho_{12.34\ldots q} = -\dfrac{\Sigma_{12}}{\sqrt{\Sigma_{11}\Sigma_{22}}}$

Where Σ_{ij} is the cofactor of $(ij)^{th}$ element of dispersion matrix of $(x_1, \ldots, x_q)$

Ellipsoid of concentration:

Let $\underset{\sim}{x}^{p\times 1}$ be a random vector with mean $\underset{\sim}{\mu}$ & dispersion matrix Σ. Our problem is to compare the variability of $\underset{\sim}{x}^{p\times 1}$ with another random vector $\underset{\sim}{y}^{p\times 1}$ with the mean $\underset{\sim}{\mu}$ & dispersion matrix Σ!

Case I: Let X be a r.v. with mean μ & variance σ^2. Let U be a r.v. uniformly distd in the interval $(\mu - K\sigma, \mu + K\sigma)$ such that U has the same mean & variance as that of X.

Note that $E(U) = \mu = E(X)$

& $V(U) = \dfrac{K^2\sigma^2}{3}$.

But $V(U) = V(X)$

$\Rightarrow \dfrac{K^2\sigma^2}{3} = \sigma^2$

$\Rightarrow K = \sqrt{3}$.

Clearly, the interval $(\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma)$ can be interpreted as the geometrical representation of concentration of X.

Let $L = \{u : \mu - \sqrt{3}\sigma < u < \mu + \sqrt{3}\sigma\}$, the interval L is called the line of concentration of X with mean μ & variance σ^2.

Another r.v. Y with mean μ & variance σ'^2 has the line of concentration.

$$L' = \{u : \mu - \sqrt{3}\sigma' < u < \mu + \sqrt{3}\sigma'\}$$

If $\sigma \gtrless \sigma'$, then accordingly, $L \supset L'$ or $L \subset L'$ and we say that Y has greater or smaller concentration than that of X.

Case II: Ellipsoid: $p \geq 2$

The above idea may be generalized to the case of a random vector of order $p \times 1$, $p \geq 2$. Let the variables $x_1, \dots, x_p$ have a jt distn with mean vector $\underset{\sim}{\mu}$ & dispersion matrix $\Sigma = (\sigma_{ij})$ which is pd.

Let $\underset{\sim}{U}$ be a r.v. distd uniformly over a closed region, say, the region $S = \{\underset{\sim}{u} : (\underset{\sim}{u} - \underset{\sim}{\mu})' A (\underset{\sim}{u} - \underset{\sim}{\mu}) < 1\}$, where A is pd, such that $\underset{\sim}{U}$ has the same mean $\underset{\sim}{\mu}$ & dispersion matrix Σ as that of $\underset{\sim}{X}^{p\times 1}$.

The pdf of $\underset{\sim}{U}$ is

$$f(\underset{\sim}{u}) = \begin{cases} K & \text{if } \underset{\sim}{u} \in S \\ 0 & \text{, o.w.} \end{cases}$$

; where K is a constant such that

$$\int_S f(\underset{\sim}{u})\, d\underset{\sim}{u} = 1$$

Since A is pd, $\exists$ a n.s. $P \ni PP' = A$

Consider the transformation,

$$\underset{\sim}{v} = P'(\underset{\sim}{y} - \underset{\sim}{\mu})$$

$$|\text{Jacobian}| = 1/\|P'\| = \frac{1}{\sqrt{|PP'|}} = \frac{1}{\sqrt{|A|}}$$

Now, $1 = \int_S f(\underset{\sim}{y})\,d\underset{\sim}{y}$

$$= \frac{K}{\sqrt{|A|}} \int_{\underset{\sim}{v}'\underset{\sim}{v} < 1} d\underset{\sim}{v}$$

$$= \frac{K}{\sqrt{|A|}} \cdot \frac{\pi^{p/2}}{\Gamma(p/2+1)}$$

$$\Rightarrow \quad K = \sqrt{|A|}\,\frac{\Gamma(p/2+1)}{\pi^{p/2}}$$

Hence the pdf of $\underset{\sim}{v}$ is

$$g(\underset{\sim}{v}) = \begin{cases} \dfrac{\Gamma(p/2+1)}{\pi^{p/2}}, & \text{if } \underset{\sim}{v}'\underset{\sim}{v} < 1. \\ 0, & \text{o.w.} \end{cases}$$

Note that

$$E(v_i) = 0 \quad \forall i$$

$$\& \quad E(v_i^2) = \int_{\underset{\sim}{v}'\underset{\sim}{v} < 1} v_i^2\, g(\underset{\sim}{v})\,d\underset{\sim}{v}$$

$$= \frac{\Gamma(p/2+1)}{\pi^{p/2}} \cdot \frac{\pi^{p/2}}{2\,\Gamma(p/2+2)}$$

$$= \frac{1}{p+2} \quad \forall i.$$

& $cov(v_i, v_j) = 0 \quad \forall\ i \neq j$

Thus, $Disp(\underset{\sim}{v}) = \frac{1}{p+2} I_p$ & $E(\underset{\sim}{v}) = \underset{\sim}{0}$

$\therefore \quad E(\underset{\sim}{u}) = \underset{\sim}{\mu}$

& $Disp(\underset{\sim}{u}) = (P')^{-1} \frac{1}{p+2} I_p (P)^{-1}$

$$= \frac{1}{p+2} (PP')^{-1}$$

$$= \frac{1}{p+2} A^{-1}$$

By construction,

$$Disp(\underset{\sim}{x}) = Disp(\underset{\sim}{u})$$

$$\Rightarrow \quad A = \frac{\Sigma^{-1}}{p+2}.$$

Hence the region S is

$$\{ \underset{\sim}{u} : (\underset{\sim}{u} - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{u} - \underset{\sim}{\mu}) < p+2 \}$$

that is $\underset{\sim}{u}$ is uniformly dist^{td} over the region $S = \{ \underset{\sim}{u} : (\underset{\sim}{u} - \underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{u} - \underset{\sim}{\mu}) < p+2 \}$ having same mean $\underset{\sim}{\mu}$ & dispersion matrix Σ as of $\underset{\sim}{x}^{p\times 1}$. Hence the region S will serve as a geometrical representation of the mode of concentration of the dist^{n} of $\underset{\sim}{x}$ about the mean $\underset{\sim}{\mu}$.

If the ellipsoid of concentration of a random vector $\underset{\sim}{y}^{p\times 1}$ with the same mean $\underset{\sim}{\mu}$ as $\underset{\sim}{x}^{p\times 1}$, is enclosed entirely within the ellipsoid of concentration of $\underset{\sim}{x}^{p\times 1}$, then $\underset{\sim}{y}^{p\times 1}$ has greater concentration (or smaller dispersion) than that of $\underset{\sim}{x}^{p\times 1}$.

Note that

$\underset{\sim}{y}^{p\times 1}$ has greater concentration than $\underset{\sim}{x}^{p\times 1}$ if $S \supset S'$ where

$$S = \{ \underset{\sim}{y} : (\underset{\sim}{y}-\underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{y}-\underset{\sim}{\mu}) < p+2 \}$$

$$S' = \{ \underset{\sim}{y} : (\underset{\sim}{y}-\underset{\sim}{\mu})' \Sigma^{*-1} (\underset{\sim}{y}-\underset{\sim}{\mu}) < p+2 \},$$

$\Sigma^* = \text{disp}(\underset{\sim}{Y})$.

ie If $(\underset{\sim}{y}-\underset{\sim}{\mu}) \Sigma^{*-1} (\underset{\sim}{y}-\underset{\sim}{\mu}) - (\underset{\sim}{y}-\underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{y}-\underset{\sim}{\mu})$

$\geq 0 \ \forall \ \underset{\sim}{y}$

and strictly more then 0 for at least one $\underset{\sim}{y}$.

ie if $(\underset{\sim}{y}-\underset{\sim}{\mu})' (\Sigma^{*-1} - \Sigma^{-1}) (\underset{\sim}{y}-\underset{\sim}{\mu}) \geq 0 \ \forall \ \underset{\sim}{y}$

& $\exists \ \underset{\sim}{y} \ni$ it is > 0.

ie if $(\Sigma^{*-1} - \Sigma^{-1})$ is nnd $\Leftrightarrow (\Sigma - \Sigma^*)$ is nnd

Alternatively, we can compare the areas of S & S'. The smaller the area, the greater the concentration.

Problem:

(10) Suppose the jt pdf of $(X_1, \cdots, X_n)'$ is

$$f(x_1, \cdots, x_n) = \begin{cases} \text{constant} & \text{if } x_1^2 + \cdots + x_n^2 < r^2 \\ 0 & \text{o.w} \end{cases}$$

Find the correlation coeff betn X_1 & X_2. Are X_1 & X_2 indept? Justify

<u>Hint:</u> Let $x_1 = R\cos\theta_1$

$x_2 = R\sin\theta_1\cos\theta_2$

$x_3 = R\sin\theta_1\sin\theta_2\cos\theta_3$

$\vdots$

$x_n = R\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-1},$

$0 \le \theta_i < \pi,$
$i = 1(1)\overline{n-2}$
$0 \le \theta_{n-1} < 2\pi$
$0 < R < r.$

$$|J| = R^{n-1}(\sin\theta_1)^{n-2}(\sin\theta_2)^{n-3}\cdots\sin\theta_{n-2}$$

Now,

$$E(x_1) = \int_{\Sigma x_i^2 < r^2} x_1 f(x_1,\dots,x_n)\,dx_1\cdots dx_n$$

$$= K\int_{\theta_1=0}^{\pi}\cdots\int_{\theta_{n-2}=0}^{\pi}\int_{\theta_{n-1}=0}^{2\pi}\int_0^{r}(R\cos\theta_1)R^{n-1}(\sin\theta_1)^{n-2}\cdots\sin\theta_{n-2}\,d\theta_1\,d\theta_2\cdots d\theta_{n-1}\,dR$$

$$= K\left\{\int_0^{\pi}\cos\theta_1(\sin\theta_1)^{n-2}d\theta_1\right\}\left\{\int_0^{\pi}(\sin\theta_2)^{n-3}d\theta_2\right\}\cdots\left\{\int_0^{2\pi}d\theta_{n-1}\right\}\left\{\int_0^{r}R^n dR\right\}$$

$= 0$

[Let $g(\theta_1) = \cos\theta_1(\sin\theta_1)^{n-2}$

$g(\pi-\theta_1) = -\cos\theta_1(\sin\theta_1)^{n-2}$

$\therefore \int_0^{\pi}\cos\theta_1(\sin\theta_1)^{n-2}d\theta_1 = 0$]

Here $E(X_i) = 0$, since the pdf or distn is symmetric w.r.t the r.v.s X_i's.

Now,

$$\mathrm{cov}(X_1, X_2) = E(X_1 X_2)$$

$$= \int_{\sum x_i^2 < r^2} x_1 x_2 f(x_1, \cdots, x_n)\, dx_1 \cdots dx_n$$

$$= K\left\{\int_0^{\pi} \cos\theta_1 (\sin\theta_1)^{n-1} d\theta_1\right\}\left\{\int_0^{\pi} \cos\theta_2 (\sin\theta_2)^{n-3} d\theta_2\right\}$$

$$\cdots \left\{\int_0^{2\pi} d\theta_{n-1}\right\}\left\{\int_0^{R} R^{n+1} dR\right\} = 0$$

Hence, $\rho(X_1, X_2) = 0$.

But variables are related & they are not indept.

To integrate $\int_{S_K} y_1^{\nu_1 - 1} y_2^{\nu_2 - 1} \cdots y_K^{\nu_K - 1} (1 - y_1 - \cdots - y_K)^{\nu_{K+1} - 1} dy_1 \cdots dy_K$,

use transformatn:

$y_1 = x_1 - x_1 x_2 = x_1(1 - x_2)$

$y_2 = x_1 x_2 (1 - x_3)$

$\vdots$

$y_{n-1} = x_1 x_2 \cdots x_{n-1}(1 - x_n)$

$y_n = x_1 x_2 \cdots x_n$

$x_1 = \sum_1^n y_i$

$x_1 x_2 = \sum_2^n y_i$

$x_1 x_2 x_3 = \sum_3^n y_i$

$x_1 x_2 \cdots x_n = y_n$

$$|J| = x_1^{K-1} x_2^{K-2} \cdots x_{K-2}^{2} x_{K-1}$$

Dirichlet disn :

A random vector $\underset{\sim}{X}^{k \times 1} = (X_1, \ldots, X_k)'$ is said to have a Dirichlet disn if its pdf is

$$f(x_1, \ldots, x_k) = \begin{cases} \dfrac{\Gamma(\nu_1+\nu_2+\cdots+\nu_{k+1})}{\Gamma(\nu_1)\Gamma(\nu_2)\cdots\Gamma(\nu_{k+1})} \cdot x_1^{\nu_1-1} \cdots x_k^{\nu_k-1}(1-x_1-\cdots-x_k)^{\nu_{k+1}-1} \\ \quad \text{if } \underset{\sim}{x} \in S_k = \{\underset{\sim}{x} : 0<x_i<1, i=1(1)k \ \& \ \sum_1^k x_i < 1\} \\ 0, \text{ outside } S_k. \end{cases}$$

[Check this is a pdf, see Malik-Arora]

where $\nu_i > 0, \forall i = 1(1)\overline{k+1}$ are the parameters.

Then, we write

$$(X_1, \ldots, X_k) \sim D(\nu_1, \ldots, \nu_k ; \nu_{k+1})$$

Moments :

$$\mu'_{r_1, r_2, \ldots, r_k} = E(X_1^{r_1} X_2^{r_2} \cdots X_k^{r_k})$$

$$= \int_{S_k} x_1^{r_1} x_2^{r_2} \cdots x_k^{r_k} f(x_1, \ldots, x_k)\, dx_1 \cdots dx_k$$

$$= \frac{\Gamma(\nu_1+\cdots+\nu_{k+1})}{\Gamma(\nu_1)\cdots\Gamma(\nu_{k+1})} \int_{S_k} x_1^{\nu_1+r_1-1} \cdots x_k^{\nu_k+r_k-1}(1-x_1-\cdots-x_k)^{\nu_{k+1}-1} dx_1 \cdots dx_k$$

$$= \frac{\Gamma(\nu_1+\cdots+\nu_{k+1})}{\Gamma(\nu_1)\cdots\Gamma(\nu_{k+1})} \cdot \frac{\Gamma(\nu_1+r_1)\cdots\Gamma(\nu_k+r_k)\Gamma(\nu_{k+1})}{\Gamma(\nu_1+\cdots+\nu_{k+1}+r_1+\cdots+r_k)}$$

$$= \frac{\Gamma(\nu_1+\cdots+\nu_{k+1})}{\Gamma(\nu_1)\cdots\Gamma(\nu_k)} \cdot \frac{\Gamma(\nu_1+r_1)\cdots\Gamma(\nu_k+r_k)}{\Gamma(\nu_1+\cdots+\nu_{k+1}+r_1+\cdots+r_k)}$$

MT

$$E(X_i) = \frac{\nu_i}{\nu_1 + \cdots + \nu_{K+1}}, \quad i = 1(1)K$$

$$\left[\because E(X_i) = \frac{\Gamma(\nu_1 + \cdots + \nu_{K+1})\,\Gamma(\nu_i + 1)}{\Gamma(\nu_i)\,\Gamma(\nu_1 + \cdots + \nu_{K+1} + 1)}\right.$$

& $V(X_i)$

$$= E(X_i^2) - E^2(X_i)$$

$$= \frac{(\nu_i + 1)\nu_i}{(\nu_1 + \cdots + \nu_{K+1} + 1)(\nu_1 + \cdots + \nu_{K+1})} - \left(\frac{\nu_i}{\nu_1 + \cdots + \nu_{K+1}}\right)^2$$

$$\left[\because E(X_i^2) = \frac{\Gamma(\nu_1 + \cdots + \nu_{K+1})\,\Gamma(\nu_i + 2)}{\Gamma(\nu_i)\,\Gamma(\nu_1 + \cdots + \nu_{K+1} + 2)}\right]$$

$$= \frac{\nu_i(\nu_1 + \cdots + \nu_{K+1} - \nu_i)}{(\nu_1 + \cdots + \nu_{K+1})^2(\nu_1 + \cdots + \nu_{K+1} + 1)}$$

$$\text{cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$$

$$= \frac{\nu_i \nu_j}{(\nu_1 + \cdots + \nu_{K+1})(\nu_1 + \cdots + \nu_{K+1} + 1)} - \frac{\nu_i \nu_j}{(\nu_1 + \cdots + \nu_{K+1})^2}$$

$$= -\frac{\nu_i \nu_j}{(\nu_1 + \cdots + \nu_{K+1})^2(\nu_1 + \cdots + \nu_{K+1} + 1)}$$

Marginal distn:

Theorem 15: If X is a bounded r.v, then its cdf $F(\cdot)$, ie its distn is uniquely determined by the sequence $\{\mu_r'\}$ of moments.

proof: Since X is bounded, there exists two finite numbers a & b such that

$$P(a < x < b) = 1$$

Let $M = \max\{|a|, |b|\}$, then $P(|X| < M) = 1$.

Now, $|\mu_r'| = \left|\int_a^b x^r \, dF(x)\right|$

$$\leq \int_a^b |x|^r \, dF(x)$$

$$\leq M^r \int_a^b dF(x)$$

$$= M^r, \quad r = 0, 1, 2, \cdots.$$

Note that

$$\left|\sum_{r=0}^{\infty} \frac{\mu_r' t^r}{r!}\right| \leq \sum_{r=0}^{\infty} |\mu_r'| \frac{|t^r|}{r!} \leq \sum_{r=0}^{\infty} \frac{M^r |t|^r}{r!} = e^{M|t|},$$

which is finite for all t.

Hence the mgf of X exists & the cdf $F(\cdot)$ is uniquely determined by the sequence $\{\mu_r'\}$ of moments.

The multivariate version of the theorem is also true.

Note that,

$$\mu'_{r_1, \cdots, r_{K_1}, 0, 0, \cdots, 0}$$

$$= E(X_1^{r_1} \cdots X_{K_1}^{r_{K_1}})$$

$$= \frac{\Gamma(\nu_1 + \cdots + \nu_{K+1})}{\Gamma(\nu_1) \cdots \Gamma(\nu_{K_1})} \frac{\Gamma(\nu_1 + r_1) \cdots \Gamma(\nu_{K_1} + r_{K_1})}{\Gamma(\nu_1 + \cdots + \nu_{K+1} + r_1 + \cdots + r_{K_1})}, \quad K_1 < K$$

Thus, we see that the general moment $\mu'_{r_1,\ldots,r_{k_1}}$ of the marginal dis^{tn} of $(X_1,\ldots,X_{k_1})$, $k_1<k$, of k-variate Dirichlet dis^{tn} $D(\nu_1,\ldots,\nu_k;\nu_{k+1})$, has the value which is the general moment of $D(\nu_1,\nu_2,\ldots,\nu_{k_1};\nu_{k_1+1}+\cdots+\nu_{k+1})$, a k_1-variate Dirichlet dis^{tn}.

Since X_i's are bounded, ie.

$$P(0<X_i<1]=1,$$

the dis^{tn} of $(X_1,\ldots,X_{k_1})$, $k_1<k$, is uniquely determined by its moments.

Hence,

$$(X_1,\ldots,X_{k_1}) \sim D(\nu_1,\ldots,\nu_{k_1};\nu_{k_1+1}+\cdots+\nu_{k+1})$$

Conditional dis^{tn} & the regression of X_k on $(X_1,\ldots,X_{k-1})$:

$$f(x_k|x_1,\ldots,x_{k-1})$$

$$=\frac{f(x_1,x_2,\ldots,x_k)}{f_1(x_1,\ldots,x_{k-1})}$$

$$=\frac{\dfrac{\Gamma(\nu_1+\cdots+\nu_{k+1})}{\Gamma(\nu_1)\Gamma(\nu_2)\cdots\Gamma(\nu_{k+1})}\cdot x_1^{\nu_1-1}\cdots x_k^{\nu_k-1}(1-x_1-\cdots-x_k)^{\nu_{k+1}-1}}{\dfrac{\Gamma(\nu_1+\cdots+\nu_{k+1})}{\Gamma(\nu_1)\cdots\Gamma(\nu_{k-1})\Gamma(\nu_k+\nu_{k+1})}x_1^{\nu_1-1}\cdots x_{k-1}^{\nu_{k-1}-1}(1-x_1-\cdots-x_{k-1})^{\nu_k+\nu_{k+1}-1}}$$

$$= \frac{\Gamma(\nu_K+\nu_{K+1})}{\Gamma(\nu_K)\Gamma(\nu_{K+1})} \left(\frac{x_K}{1-x_1-\cdots-x_{K-1}}\right)^{\nu_K-1} \left(1-\frac{x_K}{1-x_1-\cdots-x_{K-1}}\right)^{\nu_{K+1}-1} \cdot \left(\frac{1}{1-x_1-\cdots-x_{K-1}}\right)$$

, if $0<x_K<(1-x_1-\cdots-x_{K-1})$

$\therefore$ The conditional distn.

$\frac{X_K}{1-X_1-\cdots-X_{K-1}} \Big| (X_1,\cdots,X_{K-1})$ is $B(\nu_K,\nu_{K+1})$ distn

[Transformn: $y=\frac{X_K}{1-\sum_1^{K-1}X_i}$]

[$\underset{\sim}{X} \rightarrow (X_1, X_2, \cdots, X_{K-1}, Y)$]

Note that

$$E(X_K \mid X_1=x_1,\cdots,X_{K-1}=x_{K-1})$$

$$= \int_0^{1-\sum_1^{K-1}x_i} x_K f(x_K \mid x_1,\cdots,x_{K-1})\, dx_K$$

$$= \int_0^{1-\sum_1^{K-1}x_i} x_K \frac{1}{B(\nu_K,\nu_{K+1})} \left(\frac{x_K}{1-\sum_1^{K-1}x_i}\right)^{\nu_K-1} \left(1-\frac{x_K}{1-\sum_1^{K-1}x_i}\right)^{\nu_{K+1}-1} \cdot \frac{1}{1-\sum_1^{K-1}x_i}\, dx_K$$

$\left[y=\frac{x_K}{1-\sum_1^{K-1}x_i}\right]$

$$= \int_0^1 \frac{y^{\nu_K}(1-y)^{\nu_{K+1}-1}\left(1-\sum_1^{K-1}x_i\right)}{B(\nu_K,\nu_{K+1})}\, dy$$

$$= \frac{B(\nu_K+1,\nu_{K+1})\left(1-\sum_1^{K-1}x_i\right)}{B(\nu_K,\nu_{K+1})}$$

$$= \frac{\nu_K}{\nu_K+\nu_{K+1}}\left(1-\sum_1^{K-1}x_i\right)$$

Hence the regression $E(X_K \mid X_1,\cdots,X_{K-1})$ is linear.

[Alt. Since $\frac{X_K}{1-\sum_1^{K-1} X_i} \sim B(\nu_K, \nu_{K+1})$

$$\Rightarrow E\left(\frac{X_K}{1-\sum_1^{K-1} x_i} \mid X_1 = x_1, \ldots, X_{K-1} = x_{K-1}\right) = \frac{\nu_K}{\nu_K + \nu_{K+1}}$$

$$\Rightarrow E(X_K \mid X_1 = x_1, \ldots, X_{K-1} = x_{K-1}) = \frac{\nu_K}{\nu_K + \nu_{K+1}}\left(1 - \sum_1^{K-1} x_i\right)$$]

Result:

If $(X_1, \ldots, X_K)$ is a random vector having $D(\nu_1, \ldots, \nu_K; \nu_{K+1})$ dis^n, then $(X_1 + \cdots + X_K)$ has the Beta dis^n $B(\nu_1 + \cdots + \nu_K; \nu_{K+1})$.

proof: $E\left[\left\{1 - \sum_1^K X_i\right\}^r\right]$

$$= \int_{S_K} \left\{1 - \sum_1^K x_i\right\}^r f(\underset{\sim}{x})\, d\underset{\sim}{x}$$

$$= \frac{\Gamma(\nu_1 + \cdots + \nu_{K+1})}{\Gamma(\nu_1) \cdots \Gamma(\nu_{K+1})} \cdot \frac{\Gamma(\nu_1) \cdots \Gamma(\nu_K)\Gamma(\nu_{K+1} + r)}{\Gamma(\nu_1 + \cdots + \nu_K + \nu_{K+1} + r)}$$

$$= \frac{B(\nu_{K+1} + r, \nu_1 + \cdots + \nu_K)}{B(\nu_{K+1}, \nu_1 + \cdots + \nu_K)},$$ which is the rth order moment of $B(\nu_{K+1}, \nu_1 + \cdots + \nu_K)$

Hence $\left\{1 - \sum_1^K X_i\right\} \sim B(\nu_{K+1}, \nu_1 + \cdots + \nu_K)$

[since $\left\{1 - \sum_1^K X_i\right\}$ is a bounded r.v., the dis^n is uniquely determined by its moments]

$$\Leftrightarrow \sum_1^K X_i \sim B(\nu_1 + \cdots + \nu_K, \nu_{K+1}).$$

————