

\* Multivariate Data :-

- Let there be a sample of size 'n' drawn randomly from a pop<sup>n</sup> and from each sample object the values of 'p' variable are observed
- Let us assume that each variable is measured on 'k' occasions and the value of the  $j^{\text{th}}$  variable measured on  $k^{\text{th}}$  occasions from  $i^{\text{th}}$  object be

$$x_{rj} = (i=1, 2, \dots, n), j=1, 2, \dots, p, l=1, 2, \dots, k)$$

Object	Ocassion	Values of the variable
1	1	$x_{111}, x_{121}, \dots, x_{1j1}, \dots, x_{1pk}$
	2	
	1	
	$\vdots$	
	$k$	$x_{11k}, x_{12k}, \dots, x_{1jk}, \dots, x_{1pk}$
2	1	
	2	
	1	
	$\vdots$	
	$k$	
$n$	1	$x_{n11}, x_{n21}, \dots, x_{nj1}, \dots, x_{np1}$
	2	$x_{n12}, x_{n22}, \dots, x_{nj2}, \dots, x_{np2}$
	1	
	$\vdots$	
	$k$	$x_{n1k}, x_{n2k}, \dots, x_{njk}, \dots, x_{npk}$

The above elements  $x_{rj}$  values in 'nk' rows and 'p' columns is known as super matrix [rattel (1952)]

- In practice most of the multivariate data are recorded on only one occasion  $K = 1$  and the observed data are arranged in 'n' rows and 'p' cols.

$$\underline{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{bmatrix}_{n \times p}$$

The  $\underline{X}_i$  matrix of order  $(n \times p)$  can also be written as

$$\underline{X} = (\underline{x}_{ij}) = \begin{bmatrix} \underline{x}'_1 \\ \underline{x}'_2 \\ \vdots \\ \vdots \\ \underline{x}'_n \end{bmatrix} = \begin{bmatrix} \underline{x}_{(1)} & \underline{x}_{(2)} & \dots & \underline{x}_{(p)} \end{bmatrix}$$

where

$$\underline{x}'_i = [x_{i1} \ x_{i2} \ \dots \ x_{ip}]$$

$$\underline{x}'_j = [x_{1j} \ x_{2j} \ \dots \ x_{nj}]$$

- Here  $\underline{x}'_i$  is a vector of values of p-variables observed from  $i^{\text{th}}$  object ( $i = 1, 2, \dots, n$ )
- $\underline{x}'_j$  is a vector of values of  $j^{\text{th}}$  variable observed from  $n$  objects ( $j = 1, 2, \dots, p$ )
- The collected information is usually represented by a data vector  $\underline{x} = [\underline{x}_1 \ \underline{x}_2 \ \dots \ \underline{x}_p]$

Ex:-

Socioeconomic status ( $x=0$ )

	$x_1$ qual	$x_2$ lab	$x_3$ M	$x_4$ F
Object	No. of new born children	No. of dead children under ages		

\* Data Summary :-

\* Mean Vector :- The Mean of  $j^{\text{th}}$  variable of 'n' objects is

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad j=1, 2, \dots, p$$

Mean of all  $p$ -variables is represented by  $\bar{x}$ .  
where  $\bar{x} = [\bar{x}_1 \ \bar{x}_2 \ \dots \ \bar{x}_p]$

In matrix notation,  $\bar{x} = n^{-1} \mathbf{x}' \mathbf{1}$ ,  
where,  $\mathbf{1} = [1 \ 1 \ \dots \ 1]^T$

-  $\bar{x}$  is called Centroid (Mean)

- The Mean vector represented above are called Sample Mean vector

- Let  $x_j$  be the  $j^{\text{th}}$  element of data vector  $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_p]$  ( $j=1, 2, \dots, p$ ) is a continuous r.v. with mean

$$\mu_j = E(x_j) \quad \& \quad \text{mean } \mu = E(\mathbf{x})$$

Variance  $\sigma_{jj}^2 = E[(x_j - \mu_j)^2]$  so that random vector

$\mathbf{x}$  follows Multivariate with mean vector  $E(\mathbf{x}) = \mu = [\mu_1 \ \mu_2 \ \dots \ \mu_p]^T$  and

$$\text{Var-Cov Matrix } \Sigma = E[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T]$$

$$\text{Cov}(x_i, x_j) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}_{p \times p}$$

Dispersion → Matrix.

- Here  $\sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)]$   $i \neq j = 1, 2, \dots, p$  are called Covariances of  $i^{\text{th}}$  &  $j^{\text{th}}$  variable and  $\underline{\mu} \leftarrow \text{pop}^n$  mean vector.

17/25

\* Sample Vari-Cov Matrix:

$$\text{S} = n(S_{ij}) = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{bmatrix}_{p \times p}$$

$$S_{jj} = \frac{1}{n} \sum_{j=1}^n (x_{j1} - \bar{x}_j)^2, \quad S_{ij} = \frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_i)(x_{j1} - \bar{x}_j)$$

$$S = n^{-1} \mathbf{x}' \mathbf{x} - \bar{\mathbf{x}} \bar{\mathbf{x}}'$$

\* Data Transformation:

- Multivariate Data are large in volume and this are analysed to infer about the parameter vector or pop<sup>n</sup> (x).
- The variables under study are measured in different units and this differentials in unit show variations in the amount of variances of variables. Hence, Data Transformation is needed

which provides variables of unique variances

The transformation is taken in

- (1) Linear Combination of variables.
- (2) Scaling transformation
- (3) Mahalanobis transformation
- (4) Orthogonal transformation
- (5) Principal Component transformation
- (6) Lower Triangular transformation.

### (1) Linear Combination of variables

- let there be 'p' variables  $x_i$  in the data vector  $\underline{x}$ , where  $E(\underline{x}) = \underline{u}$ ,  $\text{Var}(\underline{x}) = S$
- Assume ( $\underline{x}$ ) sample Var-Cov Matrix is ('S')
- let  $\underline{a} = [a_1 \ a_2 \ \dots \ a_p]$  be a vector of known elements. Then  $y_i = \underline{a}' \underline{x}_{ij}$  is a linear combination of variables.

$$\text{Sample Mean} = \bar{y} = \underline{a}' \bar{\underline{x}}$$

$$\text{Sample Variance} = \underline{a}' S \underline{a}$$

- The  $i$ th value of ( $\underline{y}$ ,  $S$ )

$$y_i = a_1 x_{i1} + a_2 x_{i2} + \dots + a_p x_{ip}; \quad i=1, 2, \dots, n$$

- The above transformation of variables is 1-dimensional. In practice, multi-dimensional transformation has wide use so Multi-dim. transformation is given by

$$\underline{y} = \underline{x} \underline{A}' + \underline{b}'$$

where  $A$  is a  $(q \times p)$  matrix of known elements

$\underline{b}$  is a col<sup>n</sup> vector of 'q' known elements.  
 $\underline{y} = A\underline{x} + \underline{b}$

### (2) Scaling transformation

- This transformation changes the variable to a new one of unique variance.
- let  $D = \text{diag}(\Sigma_j)$ , where  $\Sigma_j = \sqrt{\Sigma_{jj}}$ ;  $j=1, 2, \dots, p$
- The random vector  $\underline{x}$  transforms to a new random vector  $\underline{y}$  by scaling transformation if  

$$\underline{y} = D^{-1}(\underline{x} - \underline{\mu})$$

$$E(\underline{y}) = D^{-1} E[\underline{x} - \underline{\mu}] = D^{-1} [E(\underline{x}) - \underline{\mu}] = 0$$

$$V(\underline{y}) = D^{-1} E[(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})^T] D^{-1} = D^{-1} \Sigma D^{-1}$$

### (3) Mahalanobis transformation

- If  $\Sigma > 0$ , then  $\Sigma^{-1}$  has a unique symmetric positive definite square root  $\Sigma^{-1/2}$ .

- Mahalanobis transformation is defined by  

$$\underline{y} = \Sigma^{-1/2} (\underline{x} - \bar{\underline{x}})$$
 where

$$E(\underline{y}) = 0 \quad \text{and} \quad V(\underline{y}) = I$$

so that the transformation eliminates the correlation between the variables and standardises the variance of each variable.

### (4) Orthogonal transformation

- let  $\Gamma$  be an orthogonal matrix of order  $p \times p$

such that  $\Gamma'\Gamma = \Gamma\Gamma' = I$

Then the transformation of a random vector  $\underline{x}$  to a new vector  $\underline{y} = \Gamma\underline{x}$  is called Orthogonal Transformation.

$$E(\underline{y}) = \Gamma \mu$$

$$V(\underline{y}) = \Gamma \Sigma \Gamma'$$

### (5) Principal Component Transformation

Let  $\lambda_1 > \lambda_2 \dots > \lambda_p > 0$  be the eigen values of the Cov Matrix  $\Sigma$  where  $V_{ar}(\underline{x}) = \Sigma, \underline{\Gamma}_1, \underline{\Gamma}_2, \dots, \underline{\Gamma}_p$  be the corresponding eigen vectors. Then  $\Gamma = [\underline{\Gamma}_1 \underline{\Gamma}_2 \dots \underline{\Gamma}_p]$  is a matrix formed with eigen vectors of  $\Sigma$ . If  $\lambda_i$ 's are distinct & greater than 0,  $\Gamma$  will be an orthogonal matrix. In such a case  $\Sigma$  can be written as  $\Sigma = \Gamma \Lambda \Gamma'$

where  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$   
then  $\underline{y} = \Gamma'(\underline{x} - \mu)$  is Principal Component transformation of  $\underline{x}$  and  $\Sigma$ .

$$E(\underline{y}) = 0, V_{ar}(\underline{y}) = \Gamma' \Sigma \Gamma = \Lambda$$

### (6) Lower Triangular Transformation

For any matrix  $\Sigma$  one can determine a lower triangular matrix  $C$

$$C = \begin{bmatrix} C_{11} & 0 & 0 & \cdots & 0 \\ C_{21} & C_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{p1} & C_{p2} & C_{p3} & \cdots & C_{pp} \end{bmatrix}_{p \times p}$$

such that

$$\Sigma = CC'$$

Then the transformation of  $\underline{x}$  to

$y = C^{-1}(x - \mu)$  is called Lower Triangular Transformation.

### \* Multivariate Normal dist:

- let  $X$  be a  $p$ -dimensional random vector such that  $x = [x_1, x_2, \dots, x_p]$ , the dist of  $x$  is called Multivariate Normal dist if the joint density funct of  $x$  is

$$f(x) = f(x_1, x_2, \dots, x_p)$$

$$= [2\pi\Sigma]^{1/2} \exp \left[ -\frac{1}{2} (x - \mu)' \Sigma (x - \mu) \right]$$

$$= \frac{|V|^{1/2}}{(2\pi)^{p/2}} \exp \left[ -\frac{1}{2} (x - \mu)' V (x - \mu) \right]$$

H.W

7/7b

where  $V = (v_{ij}) = \Sigma^{-1}$  is a positive definite matrix.

Here  $x \sim N_p(\mu, \Sigma)$ ;  $-\infty < x_j < \infty$

$-\infty < x_j < \infty$ ;  $j = 1, 2, \dots, p$

- The function  $f(x)$  is a density funct since  $V$  is +ve so  $V^{1/2}$  is positive. Also, any power of  $(2\pi)^{p/2}$  exponent is positive. Now, we need to show

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_p) dx_1 dx_2 \cdots dx_p = 1$$

let us make a transformation  $z = x - \mu$  such that  $z_j = x_j - \mu_j$ ,  $j = 1, 2, \dots, p$ .

The transformation of  $x$  to  $z$  gives Jacobian as unity since  $\frac{dx_j}{dz_k} = 0$  &  $\frac{dx_j}{dz_j} = 1$

$$\frac{dx_j}{dz_k} \quad (j \neq k)$$

$$\therefore f(z) = \frac{1}{(2\pi)^{1/2}} \frac{|v|^{1/2}}{e^{\frac{-1}{2}z'Vz}}$$

Now

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(z) dz$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{|v|^{1/2}}{(2\pi)^{1/2}} e^{\frac{-1}{2}z'Vz} dz_1 \dots dz_p$$

Here the transformation of  $x$  to  $z$  is one-to-one.

- H.W - Consider an orthogonal transformation of  $z$  to  $T$  by  $z = PT$  where  $P$  is an orthogonal matrix.

717/25

- The pdf of a univariate normal dist' is of the form

$$f(x) = K e^{-\frac{1}{2}\alpha(x-\mu)^2}$$

$$= K e^{-\frac{1}{2}\alpha(x-\mu)\alpha(x-\mu)} \text{ if } x \in \mathbb{R}$$

where  $\mu \in \mathbb{R}$ ,  $\alpha > 0$ ,  $K > 0$

generalising concept, the pdf of a multivariate normal is taken as

$$f(\underline{x}) = K e^{-\frac{1}{2}(\underline{x}-\underline{\mu})' A (\underline{x}-\underline{\mu})}$$

$$= K e^{-\frac{1}{2}(\underline{x}-\underline{\mu})' P^{-1} (\underline{x}-\underline{\mu})} \text{ if } \underline{x} \in \mathbb{R}^p, \underline{\mu} \in \mathbb{R}^p$$

A is P.d. matrix &  $K > 0$   
(Positive definite).

- Our objective is to find the constant  $\underline{\mu}$  and  $A$  in terms of the moments of a Multivariate Normal dist'.

- Let  $\underline{x}$  be a Multivariate Normal Variate with the pdf  $f(\underline{x}) = K e^{-\frac{1}{2}(\underline{x}-\underline{\mu})' A (\underline{x}-\underline{\mu})}$ ,  $\underline{x} \in \mathbb{R}^p$ .

Now

$$\int_{\mathbb{R}^p} f(\underline{x}) d\underline{x} = 1$$

$$\Rightarrow K \int_{\mathbb{R}^p} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})' A (\underline{x}-\underline{\mu})} d\underline{x} = 1$$

(

$$\text{let } \underline{y} = A'(\underline{x} - \underline{\mu})$$

$A$  is P.d. Matrix  $\exists$  a non-singular matrix

$$P \Rightarrow PP' = A \text{ or } P'AP = I_p$$

$$\Rightarrow K \int_{\mathbb{R}^p} e^{-\frac{1}{2}\underline{y}'\underline{y}} \frac{1}{|J|} d\underline{y} = 1$$

$$( |J| = \left| J \begin{pmatrix} \underline{y} \\ \underline{x} \end{pmatrix} \right| = |P'| = \sqrt{|PP'|} = \sqrt{|A|} )$$

$$\Rightarrow \frac{K}{\sqrt{|A|}} \int_{\mathbb{R}^p} e^{-\frac{1}{2}\underline{y}'\underline{y}} d\underline{y} = 1$$

$$\Rightarrow \frac{K}{\sqrt{|A|}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2} \sum y_i^2} dy_1 dy_2 \cdots dy_p = 1$$

$$\Rightarrow \frac{K}{\sqrt{|A|}} \prod_{i=1}^p \left\{ \int_{-\infty}^{\infty} e^{-\frac{1}{2} y_i^2} dy_i \right\} = 1$$

$$\Rightarrow \frac{K}{\sqrt{|A|}} \cdot (\sqrt{2\pi})^p = 1 \quad \left( \because z \sim N(0, 1) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1 \right)$$

$$\Rightarrow K = \frac{\sqrt{|A|}}{(\sqrt{2\pi})^p}$$

Now the pdf of  $\underline{y}$  is

$$g(\underline{y}) = \frac{1}{(\sqrt{2\pi})^P} e^{-\frac{1}{2}\underline{y}^T \underline{y}}, \quad \underline{y} \in \mathbb{R}^P$$

$$\begin{aligned} E(y_i) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ y_i \cdot \frac{1}{(\sqrt{2\pi})^P} e^{-\frac{1}{2} \sum_{j=1}^P y_j^2} dy_1 dy_2 \cdots dy_P \right\} \\ &= \left( \int_{-\infty}^{\infty} y_i \left( \frac{1}{\sqrt{2\pi}} \right)^{-\frac{1}{2} y_i^2} dy_i \right) \prod_{j \neq i} \left\{ \int_{-\infty}^{\infty} \frac{e^{-\frac{1}{2} y_j^2}}{\sqrt{2\pi}} dy_j \right\} \end{aligned}$$

$$E(y_i) = 0$$

$$\begin{aligned} \text{Cov}(y_i, y_j) &= E(y_i y_j) - E(y_i) E(y_j) \\ &= E(y_i y_j) \quad (\because E(y_i) = 0) \end{aligned}$$

$$\Rightarrow \text{Cov}(y_i, y_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} y_i y_j \cdot \left( \frac{1}{(\sqrt{2\pi})^P} e^{-\frac{1}{2} \sum_{k=1}^P y_k^2} \right) dy_1 dy_2 \cdots dy_P$$

$$= \int_{-\infty}^{\infty} \left( y_i^2 \cdot \left( \frac{1}{\sqrt{2\pi}} \right)^{-\frac{1}{2} y_i^2} \right) \left( \prod_{j \neq i} \left\{ \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{2\pi}} \right)^{-\frac{1}{2} y_j^2} dy_j \right\} \right) \quad i=j$$

$$= \left\{ \left( \int_{-\infty}^{\infty} y_i \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y_i^2} dy_i \right) \left( \int_{-\infty}^{\infty} y_j \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} y_j^2} dy_j \right) \right\}_{k \neq i \neq j} \quad i \neq j$$

$$= \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases} \quad (i=j \text{ means variance})$$

To find  $E(\underline{y})$  if  $i \neq j$  then  $\Sigma i \neq j$  means sum mean

$$\therefore E(\underline{y}) = 0 \quad \& \quad \text{Disp}(\underline{y}) = I_p$$

Since  $E(\underline{y}) = 0$

$$\underline{y} = P'(\underline{x} - \underline{1})$$

$$E(\underline{y}) = E[P'(\underline{x} - \underline{1})] = P'[E(\underline{x}) - \underline{1}] = 0$$

$$\Rightarrow E(\underline{x}) - \underline{1} = 0$$

$$\Rightarrow E(\underline{x}) = \underline{1} = \underline{\mu} \quad (\text{say})$$

$$D(\underline{y}) = J_P$$

$$D[P'(\underline{x} - \underline{1})] = J_P$$

$$\Rightarrow P'D(\underline{x})P = J_P$$

$$\Rightarrow D(\underline{x}) = (P')^{-1}P^{-1}$$

$$= (PP')^{-1} = A^{-1}$$

$$\Rightarrow A = \Sigma^{-\frac{1}{2}} \quad \text{where } \Sigma = D(\underline{x})$$

$$\text{Hence } K = \frac{1}{(\sqrt{2\pi})^P |\Sigma|} = (\sqrt{2\pi})^{-P} |\Sigma|^{-\frac{1}{2}}$$

Defn: A random vector  $\underline{x}(px1)$  is said to have a multivariate Normal dist<sup>n</sup> with mean vector  $E(\underline{x}) = \underline{\mu}$  & dispersion matrix  $P_x = \Sigma$  if the pdf of  $\underline{x}$  is given by

$$f_{\underline{x}}(\underline{x}) = \frac{1}{(2\pi)^{P/2} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})}; \underline{x} \in \mathbb{R}^P$$

we can write  $\underline{x} \sim N_p(\underline{\mu}, \Sigma)$  (density funct<sup>n</sup> of p-variate Normal)

\* Remarks:

- (±) The pdf of  $f_{\underline{x}}(\underline{x})$  is maximum if the exponent is minimum i.e.  $(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})$  is minimum.

Note that  $(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) > 0$  if  $\underline{x} \neq \underline{\mu}$   
 = 0 if  $\underline{x} = \underline{\mu}$   
 since  $\Sigma^{-1}$  is positive definite.

Hence  $\underline{x} = \underline{\mu}$  is the mode of the dist.

$$\text{Let } g(\underline{x}) = (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})$$

$$\text{then } \frac{\partial g(\underline{x})}{\partial \underline{x}} = 0 \text{ at } \underline{x} = \underline{\mu}$$

(2) Note that the exponent

$$\begin{aligned} g(\underline{x}) &= (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \\ &= \underline{x}' \Sigma^{-1} \underline{x} - \underline{x}' \Sigma^{-1} \underline{\mu} - \underline{\mu}' \Sigma^{-1} \underline{x} + \underline{\mu}' \Sigma^{-1} \underline{\mu} \\ &= \underline{x}' \Sigma^{-1} \underline{x} - 2 \underline{x}' \Sigma^{-1} \underline{\mu} + \underline{\mu}' \Sigma^{-1} \underline{\mu} \end{aligned}$$

$$\text{Let } g^*(\underline{x}) = \sum_{i=1}^p \sum_{j=1}^p a_{ij} x_i x_j + \sum_{i=1}^p b_i x_i + c$$

be the exponent of a multivariate dist' pdf.

$$\text{Note that } g^*(\underline{x}) = \underline{x}' A \underline{x} + \underline{x}' \underline{b} + c$$

Comparing we get

$$A = \Sigma^{-1} \Rightarrow \Sigma = A^{-1}$$

Thm (1): If  $\underline{x}_{px1} \sim N_p(\underline{\mu}, \Sigma)$  then for a non-singular matrix  $P_{pxp}$

$$P' \underline{x} \sim N_p(P' \underline{\mu}, P' \Sigma P)$$

Proof: The pdf of  $\underline{x}$  is  $\frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2} (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})}$ ,  $\underline{x} \in \mathbb{R}^p$

$$\text{let } \underline{y} = P' \underline{x} \Rightarrow \underline{x} = (P')^{-1} \underline{y}$$

$$|\underline{J}| = \left| \frac{\partial \underline{y}}{\partial \underline{x}} \right| = ||\underline{P}'|| = \sqrt{|\underline{P}'\underline{P}|}$$

The pdf of  $\underline{y}$  is

$$\begin{aligned} g(\underline{y}) &= \frac{1}{(\sqrt{2\pi})^p \sqrt{|\underline{\Sigma}|}} e^{-\frac{1}{2} \{ (\underline{P}')^{-1} \underline{y} - \underline{\mu} \}' \underline{\Sigma}^{-1} \{ (\underline{P}')' \underline{y} - \underline{\mu} \}} \\ &= \frac{1}{(\sqrt{2\pi})^p \sqrt{|\underline{P}'\underline{\Sigma}\underline{P}|}} e^{-\frac{1}{2} (\underline{y} - \underline{P}'\underline{\mu})' \underline{P}'\underline{\Sigma}^{-1} (\underline{P}')^{-1} (\underline{y} - \underline{P}'\underline{\mu})} \\ &= \frac{1}{(\sqrt{2\pi})^p \sqrt{|\underline{P}'\underline{\Sigma}\underline{P}|}} e^{-\frac{1}{2} (\underline{y} - \underline{P}'\underline{\mu})' (\underline{P}'\underline{\Sigma}^{-1}\underline{P})^{-1} (\underline{y} - \underline{P}'\underline{\mu})} \\ &\therefore \underline{y} \sim N_p(\underline{P}'\underline{\mu}, \underline{P}'\underline{\Sigma}\underline{P}) \end{aligned}$$

\* Mgf :-

The mgf of a  $p$ -dimensional r.v.  $\underline{x}$  is defined as  $M_{\underline{x}}(\underline{t}) = E[e^{\underline{t}'\underline{x}}]$  provided it exists, where  $\underline{t}$  belongs to a region containing the origin as an interior point.

m(2) :- If  $\underline{x} \sim N_p(\underline{\mu}, \underline{\Sigma})$  then the mgf of  $\underline{x}$  is  $M_{\underline{x}}(\underline{t}) = e^{\underline{t}'\underline{\mu} + \frac{1}{2}\underline{t}'\underline{\Sigma}\underline{t}}$  where  $\underline{t}$  belongs to a region containing the origin as an interior point.

Proof:-

$$\begin{aligned} M_{\underline{x}}(\underline{t}) &= E[e^{\underline{t}'\underline{x}}] \\ &= \int e^{\underline{t}'\underline{x}} f(\underline{x}) d\underline{x} \\ &= \int_{\mathbb{R}^p} e^{\underline{t}'\underline{x}} \cdot \frac{1}{\sqrt{(2\pi)^p \sqrt{|\underline{\Sigma}|}}} e^{-\frac{1}{2} (\underline{x} - \underline{\mu})' \underline{\Sigma}^{-1} (\underline{x} - \underline{\mu})} d\underline{x} \end{aligned}$$

$$= \frac{1}{(\sqrt{2\pi})^p \sqrt{|\Sigma|}} \int_{R^p} e^{-\frac{1}{2} \left\{ \underline{x}' \Sigma^{-1} \underline{x} - \underline{x}' \Sigma^{-1} \underline{\mu} - \underline{\mu}' \Sigma^{-1} \underline{x} + \underline{\mu}' \Sigma^{-1} \underline{\mu} \right\}} dx$$

( $\because \underline{x}'$  and  $\underline{x}$  are symmetric.)

$$= \frac{1}{(\sqrt{2\pi})^p \sqrt{|\Sigma|}} \int_{R^p} e^{-\frac{1}{2} \left\{ \underline{x}' \Sigma^{-1} \underline{x} - 2\underline{x}' \Sigma^{-1} \underline{\mu} + \underline{\mu}' \Sigma^{-1} \underline{\mu} - 2\underline{t}' \Sigma^{-1} \underline{x} \right\}} dx$$

$$= \frac{1}{(\sqrt{2\pi})^p \sqrt{|\Sigma|}} \int_{R^p} e^{-\frac{1}{2} \left\{ \underline{x}' \Sigma^{-1} \underline{x} - 2\underline{x}' (\Sigma^{-1} \underline{\mu} + \underline{t}') + \underline{\mu}' \Sigma^{-1} \underline{\mu} \right\}} dx$$

$$= \frac{1}{(\sqrt{2\pi})^p \sqrt{|\Sigma|}} \int_{R^p} e^{-\frac{1}{2} \left\{ \underline{x}' \Sigma^{-1} \underline{x} - 2\underline{x}' (\Sigma^{-1} (\underline{\mu} + \Sigma \underline{t})) + \underline{\mu}' \Sigma^{-1} \underline{\mu} \right\}} dx$$

$$= \frac{1}{(\sqrt{2\pi})^p \sqrt{|\Sigma|}} \int_{R^p} e^{-\frac{1}{2} (\underline{x} - \underline{\mu} - \Sigma \underline{t})' \Sigma^{-1} (\underline{x} - \underline{\mu} - \Sigma \underline{t}) + \underline{t}' \underline{\mu} - \underline{t}' \Sigma \underline{t}} dx$$

$$(\because (\underline{x}' \Sigma^{-1} \underline{x} - 2\underline{x}' \Sigma^{-1} (\underline{\mu} + \Sigma \underline{t}) + \underline{\mu}' \Sigma^{-1} \underline{\mu})$$

$$= \underline{x}' \Sigma^{-1} \underline{x} - 2\underline{x}' \Sigma^{-1} (\underline{\mu} + \Sigma \underline{t}) + (\underline{\mu} + \Sigma \underline{t})' \Sigma^{-1} (\underline{\mu} + \Sigma \underline{t}) +$$

$$= (\underline{x} - \underline{\mu} - \Sigma \underline{t})' \Sigma^{-1} (\underline{x} - \underline{\mu} - \Sigma \underline{t}) - \underline{\mu}' \Sigma^{-1} \underline{t} - \underline{t}' \Sigma \Sigma^{-1} \underline{\mu}$$

$$= (\underline{x} - \underline{\mu} - \Sigma \underline{t})' \Sigma^{-1} (\underline{x} - \underline{\mu} - \Sigma \underline{t}) - 2\underline{t}' \underline{\mu} - \underline{t}' \Sigma \underline{t}$$

$$= e^{-\frac{1}{2} \underline{t}' \underline{\mu} + \frac{1}{2} \underline{t}' \Sigma \underline{t}} \int_{R^p} \frac{1}{e^{-\frac{1}{2} (\underline{x} - \underline{\mu} - \Sigma \underline{t})' \Sigma^{-1} (\underline{x} - \underline{\mu} - \Sigma \underline{t})}} dx$$

= 1 ( $\because$  P.d.f.)

$$= e^{-\frac{1}{2} \underline{t}' \underline{\mu} + \frac{1}{2} \underline{t}' \Sigma \underline{t}}$$

Ex :- Prove Thm (2) using mgf technique

$$M_x(t) = E[e^{\underline{t}' \underline{x}}]$$

$$= E[e^{\underline{t}' P \underline{x}}] = E[e^{(\underline{P} \underline{t})' \underline{x}}]$$

$$\text{let } \underline{u} = \underline{P} \underline{t}$$

$$M_x(t) = E[e^{\underline{u}' \underline{x}}]$$

$$= e^{\underline{u}' \underline{u} + \frac{1}{2} \underline{u}' \Sigma \underline{u}}$$

$$= e^{\underline{t}' P \underline{u}} = e^{\frac{1}{2} \underline{t}' P' \Sigma P \underline{t}}$$

$$= e^{\underline{t}' (\underline{P} \underline{u}) + \frac{1}{2} \underline{t}' (\underline{P}' \Sigma \underline{P}) \underline{t}}$$

which is mgf of  $N_p(\underline{P} \underline{u}, \underline{P}' \Sigma \underline{P})$

Theorem (3) :- Let  $\underline{X}_{px1} \sim N_p(\underline{u}, \Sigma)$  &  $\underline{Y}_{q,px1} \sim N_q(B \underline{u}, B \Sigma B')$  be a matrix of rank  $q$ , then

$$\underline{Y}_{q,px1} = B \underline{X} \sim N_q(B \underline{u}, B \Sigma B')$$

Proof :-  $M_y(t) = E[e^{\underline{t}' \underline{y}}] = E[e^{\underline{t}' B \underline{x}}]$

$$= E[e^{(\underline{B}' \underline{t})' \underline{x}}]$$

$$\text{let } \underline{u} = \underline{B}' \underline{t}$$

$$= E[e^{\underline{u}' \underline{x}}]$$

$$= e^{\underline{u}' \underline{u} + \frac{1}{2} \underline{u}' \Sigma \underline{u}}$$

$$= e^{(\underline{B}' \underline{t})' \underline{u} + \frac{1}{2} (\underline{B}' \underline{t})' \Sigma (\underline{B}' \underline{t})}$$

$$= e^{\underline{t}' (\underline{B} \underline{u}) + \frac{1}{2} \underline{t}' (\underline{B}' \Sigma \underline{B}) \underline{t}}$$

which is mgf of  $N_q(B \underline{u}, B \Sigma B')$

$\therefore$  By uniqueness of mgf

$$\underline{y} \sim N_q(B \underline{u}, B \Sigma B')$$

Ex :- Prove Thm (1) using mgf technique.

$$M_x(t) = E \left[ e^{\underline{t}' \underline{x}} \right] = E \left[ e^{\underline{t}' P' \underline{x}} \right] = E \left[ e^{(\underline{P} \underline{t})' \underline{x}} \right]$$

$$\text{let } \underline{u} = \underline{P} \underline{t}$$

$$\begin{aligned} M_x(t) &= E \left[ e^{\underline{u}' \underline{x}} \right] \\ &= e^{\underline{u}' \underline{\mu} + \frac{1}{2} \underline{u}' \Sigma \underline{u}} \\ &= e^{\underline{t}' P' \underline{\mu} + \frac{1}{2} \underline{t}' P' \Sigma P \underline{t}} \\ &= e^{\underline{t}' (\underline{P}' \underline{\mu}) + \frac{1}{2} \underline{t}' (\underline{\Sigma} \underline{P}) \underline{t}} \end{aligned}$$

which is mgf of  $N_p(\underline{P}' \underline{\mu}, \underline{\Sigma} \underline{P})$

Theorem (3) :- Let  $\underline{x}_{px1} \sim N_p(\underline{\mu}, \Sigma)$  &  $B_{q \times p} (q \leq p)$  be a matrix of rank q, then

$$\underline{y}_{q \times 1} = B \underline{x} \sim N_q(B \underline{\mu}, B \Sigma B')$$

Proof :-  $M_y(t) = E \left[ e^{\underline{t}' \underline{y}} \right] = E \left[ e^{\underline{t}' B \underline{x}} \right] = E \left[ e^{(\underline{B}' \underline{t})' \underline{x}} \right]$

$$\text{let } \underline{u} = \underline{B}' \underline{t}$$

$$\begin{aligned} &= E \left[ e^{\underline{u}' \underline{x}} \right] \\ &= e^{\underline{u}' \underline{\mu} + \frac{1}{2} \underline{u}' \Sigma \underline{u}} \\ &= e^{(\underline{B}' \underline{t})' \underline{\mu} + \frac{1}{2} (\underline{B}' \underline{t})' \Sigma (\underline{B}' \underline{t})} \\ &= e^{\underline{t}' (B \underline{\mu}) + \frac{1}{2} \underline{t}' (B \Sigma B') \underline{t}} \end{aligned}$$

which is mgf of  $N_q(B \underline{\mu}, B \Sigma B')$

$\therefore$  By uniqueness of mgf

$$\underline{y} \sim N_q(B \underline{\mu}, B \Sigma B')$$

Theore

Proof

Corolla

Proof



Theorem (4): If  $\underline{x}_{pxi} \sim N_p(\underline{\mu}, \Sigma)$ , then we can write  
 $\underline{x} = \underline{\mu} + P\underline{y}$  where  $PP' = \Sigma$  &  $\underline{y} \sim N_p(0, I_p)$

Proof: Pdf of  $\underline{x}$  is

$$f(\underline{x}) = \frac{1}{(\sqrt{2\pi})^p \sqrt{|\Sigma|}} e^{-\frac{1}{2} (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})}, \quad \underline{x} \in \mathbb{R}^p$$

Since  $\Sigma$  is positive-definite  $\exists$  a non-singular  
 $P \ni PP' = \Sigma$

$$\text{let } \underline{y} = \frac{\underline{x} - \underline{\mu}}{P} \Rightarrow \underline{y} = P^{-1}(\underline{x} - \underline{\mu})$$

$$\text{Note that } |J| = \left| \begin{array}{c} \partial \underline{y} \\ \hline \partial \underline{x} \end{array} \right| = \frac{|P^{-1}|}{|P^{-1}'| |P^{-1}|} = \sqrt{|P^{-1}'| |P^{-1}|} = \sqrt{|(P P')^{-1}|} = \sqrt{|\Sigma^{-1}|}$$

$$\begin{aligned} g(\underline{y}) &= \frac{1}{(\sqrt{2\pi})^p} \cdot \frac{1}{\sqrt{|\Sigma|}} e^{-\frac{1}{2} (\underline{P}\underline{y})' (P P')^{-1} (\underline{P}\underline{y})} \cdot \sqrt{|\Sigma|} \\ &= \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2} \underline{y}' \underline{y}} \end{aligned}$$

$$\Rightarrow \underline{y} \sim N_p(0, I_p) \text{ where } \underline{x} = \underline{\mu} + P\underline{y}$$

&  $I = P'P$

Corollary: If  $\underline{x}_{pxi} \sim N_p(\underline{\mu}, \Sigma)$  then  
 $(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \sim \chi_p^2$

Proof: If  $\underline{x} \sim N_p(\underline{\mu}, \Sigma)$  then we have

$$\begin{aligned} \underline{x} - \underline{\mu} &= P\underline{y} \text{ where } PP' = \Sigma \\ &\text{& } \underline{y} \sim N_p(0, I_p) \end{aligned}$$

PDF of  $\underline{y}$

$$g(\underline{y}) = \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2} \underline{y}' \underline{y}} = \frac{1}{(\sqrt{2\pi})^p} e^{-\frac{1}{2} \sum_{i=1}^p y_i^2}$$

Now

$$\begin{aligned} & (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \\ &= (\underline{y}' P \Sigma^{-1} P \underline{y}) \\ &= \underline{y}' (P' P) \Sigma^{-1} \underline{y} \\ &= \underline{y}' \underline{y} = \sum_{i=1}^p y_i^2 \quad \left( \text{Quadratic form of } \chi^2 \text{ dist}^2 \right) \\ &= \chi^2(p) \text{ df.} \end{aligned}$$

Let  $\underline{x} \sim N_p(\underline{\mu}, \Sigma)$

Consider the following partition

$$\underline{x}_{p \times 1} = \begin{pmatrix} \underline{x}_{(1)p \times 1} \\ \vdots \\ \underline{x}_{(2)} \end{pmatrix}, \quad \underline{\mu}_{p \times 1} = \begin{pmatrix} \underline{\mu}_{(1)} \\ \vdots \\ \underline{\mu}_{(2)} \end{pmatrix},$$

$$\text{Disp}(\underline{x}) = \Sigma_{p \times p} = \begin{bmatrix} \Sigma_{11}(p \times p_1) & \Sigma_{12} \\ \Sigma_{21}[(\Sigma_{12})'] & \Sigma_{22} \end{bmatrix}$$

$$= \begin{bmatrix} \text{Cov}(\underline{x}_{(1)}, \underline{x}_{(1)}) & \text{Cov}(\underline{x}_{(1)}, \underline{x}_{(2)}) \\ \text{Cov}(\underline{x}_{(2)}, \underline{x}_{(1)}) & \text{Cov}(\underline{x}_{(2)}, \underline{x}_{(2)}) \end{bmatrix}$$

1417125

Theorem (5): If  $\underline{x} \sim N_p(\underline{\mu}, \Sigma)$ , then the necessary and sufficient condition for  $\underline{x}_{(1)}$  &  $\underline{x}_{(2)}$  to be independent is that  $\Sigma_{12} = 0$ .

Proof

Proof - Let  $\Sigma_{12} = \text{Cov}(\underline{x}_{(1)}, \underline{x}_{(2)}) = 0$

Then the pdf of  $\underline{x}_{(p+1)}$  is  $N_p(\underline{\mu}, \Sigma)$

$$= \frac{1}{(2\pi)^{P/2} \sqrt{|\Sigma|}} e^{-\frac{1}{2} (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})}, \quad \underline{x} \in \mathbb{R}.$$

$$\text{Now } \Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \quad (\because \Sigma_{12} = 0)$$

$$\text{Hence } |\Sigma| = |\Sigma_{11}| \cdot |\Sigma_{22}| \quad \&$$

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix}$$

Now the exponent becomes

$$\begin{aligned} & (\underline{x} - \underline{\mu}') \Sigma^{-1} (\underline{x} - \underline{\mu}) \\ &= \begin{pmatrix} \underline{x}_{(1)} - \underline{\mu}_{(1)} \\ \underline{x}_{(2)} - \underline{\mu}_{(2)} \end{pmatrix}' \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} \underline{x}_{(1)} - \underline{\mu}_{(1)} \\ \underline{x}_{(2)} - \underline{\mu}_{(2)} \end{pmatrix} \\ &= (\underline{x}_{(1)} - \underline{\mu}_{(1)})' \Sigma_{11}^{-1} (\underline{x}_{(1)} - \underline{\mu}_{(1)}) + (\underline{x}_{(2)} - \underline{\mu}_{(2)})' \Sigma_{22}^{-1} (\underline{x}_{(2)} - \underline{\mu}_{(2)}) \end{aligned}$$

Hence  $f(\underline{x} | \underline{\mu}, \Sigma)$

$$= \frac{1}{(\sqrt{2\pi})^{P_1} \sqrt{|\Sigma_{11}|}} e^{-\frac{1}{2} (\underline{x}_{(1)} - \underline{\mu}_{(1)})' \Sigma_{11}^{-1} (\underline{x}_{(1)} - \underline{\mu}_{(1)})}$$

$$\times \frac{1}{(\sqrt{2\pi})^{P-P_1} \sqrt{|\Sigma_{22}|}} e^{-\frac{1}{2} (\underline{x}_{(2)} - \underline{\mu}_{(2)})' \Sigma_{22}^{-1} (\underline{x}_{(2)} - \underline{\mu}_{(2)})}$$

Hence  $\underline{x}_{(1)}$  and  $\underline{x}_{(2)}$  are independent

$$\underline{x}_{(1)} \sim N_p(\underline{\mu}_{(1)}, \Sigma_{11})$$

&

$$\underline{x}_{(2)} \sim N_{p-p_1}(\underline{\mu}_{(2)}, \Sigma_{22})$$

Conversely,

let  $\underline{x}_{(1)}$  and  $\underline{x}_{(2)}$  are independent

$$\text{Cov}(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)] \quad i = 1, 2, \dots, p_1 \\ j = p_1 + 1, \dots, p$$

$$= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j) f(x_1, \dots, x_p) dx_1 dx_2 \dots dx_p$$

$$= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_i - \mu_i) f_1(x_1, x_2, \dots, x_{p_1}) dx_1 dx_2 \dots dx_{p_1}$$

$$\times \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_j - \mu_j) f_2(x_{p_1+1}, x_{p_1+2}, \dots, x_p) dx_{p_1+1} \dots dx_p$$

$$\leq \{E(x_i) - \mu_i\} \{E(x_j) - \mu_j\}$$

$$= 0 \quad (\because E(x_i) = \mu_i \text{ & } E(x_j) = \mu_j)$$

$$\boxed{\sigma_{ij} = 0}$$

$$i = 1, \dots, p_1$$

$$j = p_1 + 1, p_1 + 2, \dots, p$$

$$\Sigma_{12} = \Sigma_{21} = 0$$

Hence the proof.

Theorem (6): If  $\underline{x}_{px1} \sim N_p(\underline{\mu}, \Sigma)$ , then any sub-vector is also a multivariate normal with mean vector and dispersion matrix obtained by taking the corresponding components  $\underline{\mu}$  &  $\Sigma$ . In particular  $\underline{x}_{(2)} \sim N_{p-p_1}(\underline{\mu}_{(2)}, \Sigma_{22})$

Proof: From previous Theorem,

If  $\underline{x} \sim N_p(\underline{\mu}, \Sigma)$  then

$\underline{x}_{(1)} \sim N_{p_1}(\underline{\mu}_{(1)}, \Sigma_{11})$  and

$\underline{x}_{(2)} \sim N_{p-p_1}(\underline{\mu}_{(2)}, \Sigma_{22})$  if  $\Sigma_{12} = 0$

Consider the transformation

$$\underline{y}_{(1)} = \underline{x}_{(1)} + M \underline{x}_{(2)}$$

$\underline{y}_{(2)} = \underline{x}_{(2)}$  and  $M$  is  $\exists \text{ Cov}(\underline{y}_{(1)}, \underline{y}_{(2)}) = 0$

$$\Rightarrow E[\underline{y}_{(1)} - E(\underline{y}_{(1)})][\underline{y}_{(2)} - E(\underline{y}_{(2)})]' = 0$$

$$\Rightarrow E[\underline{x}_{(1)} + M \underline{x}_{(2)} - E(\underline{x}_{(1)} + M \underline{x}_{(2)})][\underline{x}_{(2)} - E(\underline{x}_{(2)})]' = 0$$

$$\Rightarrow E[\underline{x}_{(1)} + M \underline{x}_{(2)} - \underline{\mu}_{(1)} - M \underline{\mu}_{(2)}][\underline{x}_{(2)} - \underline{\mu}_{(2)}]' = 0$$

$$\Rightarrow E[(\underline{x}_{(1)} - \underline{\mu}_{(1)}) + M(\underline{x}_{(2)} - \underline{\mu}_{(2)})][\underline{x}_{(2)} - \underline{\mu}_{(2)}]' = 0$$

$$\Rightarrow E[\underline{x}_{(1)} - \underline{\mu}_{(1)}][\underline{x}_{(2)} - \underline{\mu}_{(2)}]' + M E[\underline{x}_{(2)} - \underline{\mu}_{(2)}][\underline{x}_{(2)} - \underline{\mu}_{(2)}]' = 0$$

$$\Rightarrow \text{Cov}(\underline{x}_{(1)}, \underline{x}_{(2)}) + M \text{Cov}(\underline{x}_{(2)}, \underline{x}_{(2)}) = 0$$

$$\Rightarrow \Sigma_{12} + M \Sigma_{22} = 0$$

$$\Rightarrow M = -\Sigma_{12} \Sigma_{22}^{-1}$$

Hence the transformation becomes

$$\underline{y}_{(1)} = \underline{x}_{(1)} - \Sigma_{12} \Sigma_{22}^{-1} \underline{x}_{(2)} \text{ and } \underline{y}_{(2)} = \underline{x}_{(2)}$$

$$\Rightarrow \underline{Y} = \begin{pmatrix} \underline{Y}_{(1)} \\ \underline{Y}_{(2)} \end{pmatrix} = \begin{pmatrix} I_{P_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{P-P_1} \end{pmatrix} \begin{pmatrix} \underline{\Sigma}_{(1)} \\ \underline{X}_{(2)} \end{pmatrix}$$

$\Rightarrow \underline{Y} = P\underline{X}$  (say) where

$$P = \begin{pmatrix} I_{P_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{P-P_1} \end{pmatrix} \text{ is non-singular matrix}$$

Hence by theorem

$$\underline{Y}_{px1} = P\underline{X} \sim N_p(P\underline{\mu}, P\Sigma P')$$

Note that

$$P\underline{\mu} = \begin{pmatrix} I_P & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{P-P_1} \end{pmatrix} \begin{pmatrix} \underline{\mu}_{(1)} \\ \underline{\mu}_{(2)} \end{pmatrix} = \begin{pmatrix} \underline{\mu}_{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underline{\mu}_{(2)} \\ \underline{\mu}_{(2)} \end{pmatrix}$$

$$P\Sigma P' = \begin{pmatrix} I_{P_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{P-P_1} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I_{P_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{P-P_1} \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}$$

Hence

$$\begin{pmatrix} \underline{Y}_{(1)} \\ \underline{Y}_{(2)} \end{pmatrix} \sim N_p \left( \begin{pmatrix} \underline{\mu}_{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underline{\mu}_{(2)} \\ \underline{\mu}_{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \right)$$

Here  $\text{Cov}(\underline{Y}_{(1)}, \underline{Y}_{(2)}) = 0$

Hence we have,

$$\underline{Y}_{(2)} = \underline{x}_{(2)} \sim N_{p-p_1}(\underline{\mu}_{(2)}, \Sigma_{22})$$

15/7/25

Ex: (i) If  $\underline{x} \sim N_3(\underline{\mu}, \Sigma)$ , find the dist' of

$$\begin{bmatrix} x_1 - x_2 \\ x_2 - x_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$A\underline{x} \sim N(A\underline{\mu}, A\Sigma A')$$

$$A\underline{\mu} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_2 - \mu_3 \end{bmatrix}$$

$$A\Sigma A' = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} - \sigma_{21} & \sigma_{12} - \sigma_{22} & \sigma_{13} - \sigma_{23} \\ \sigma_{21} - \sigma_{31} & \sigma_{22} - \sigma_{32} & \sigma_{23} - \sigma_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} - \sigma_{21} - \sigma_{12} + \sigma_{22} & \sigma_{12} - \sigma_{22} - \sigma_{13} + \sigma_{23} \\ \sigma_{21} - \sigma_{31} - \sigma_{22} + \sigma_{32} & \sigma_{22} - \sigma_{32} - \sigma_{23} + \sigma_{33} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{12} - \sigma_{22} - \sigma_{13} + \sigma_{23} \\ \sigma_{21} - \sigma_{31} - \sigma_{22} + \sigma_{32} & \sigma_{22} - 2\sigma_{23} + \sigma_{33} \end{bmatrix}$$

$\therefore \sigma_{12} = \sigma_{21} \text{ & } \sigma_{23} = \sigma_{32}$

Ex: (2) If  $\underline{x} \sim N_5 (\underline{\mu}, \Sigma)$  find the dist' of  $\begin{bmatrix} x_2 \\ x_4 \end{bmatrix}$   
we set

$$\underline{x}_{(1)} = \begin{bmatrix} x_2 \\ x_4 \end{bmatrix}, \underline{\mu}_{(1)} = \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \Sigma_{11} = \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix}$$

Now

$$\underline{x} = \begin{array}{c|c} x_2 & \xleftarrow{\quad} \underline{x}_{(1)} \\ \hline x_4 & \\ \hline \cdots & \\ \hline x_1 & \\ \hline x_3 & \xleftarrow{\quad} \underline{x}_{(2)} \\ \hline x_5 & \end{array} \quad \underline{\mu} = \begin{bmatrix} \mu_2 \\ \mu_4 \\ \vdots \\ \mu_1 \\ \mu_3 \\ \mu_5 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_{22} & \sigma_{24} & \sigma_{21} & \sigma_{23} & \sigma_{25} \\ \sigma_{24} & \sigma_{44} & \sigma_{14} & \sigma_{34} & \sigma_{45} \\ \hline \cdots & (2 \times 2) & \cdots & \cdots & (2 \times 3) \\ \sigma_{12} & \sigma_{14} & \sigma_{11} & \sigma_{13} & \sigma_{15} \\ \sigma_{32} & \sigma_{34} & \sigma_{13} & \sigma_{33} & \sigma_{35} \\ \hline \sigma_{52} & \sigma_{45} & \sigma_{15} & \sigma_{35} & \sigma_{55} \\ \hline (3 \times 2) & & (3 \times 2) & & (3 \times 3) \end{bmatrix}$$

$\xrightarrow{\Sigma_{11}}$        $\xrightarrow{\Sigma_{12}}$   
 $\xrightarrow{\Sigma_{21} \text{ or } \Sigma_{12}}$        $\xrightarrow{\Sigma_{22}}$

$$\underline{x}_{(1)} \sim N_2 \left( \begin{bmatrix} \mu_2 \\ \mu_4 \end{bmatrix}, \begin{bmatrix} \sigma_{22} & \sigma_{24} \\ \sigma_{24} & \sigma_{44} \end{bmatrix} \right)$$

Ex: (3) Let  $\underline{x} \sim N_3 (\underline{\mu}, \Sigma)$

with

$$\Sigma = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$\xrightarrow{\Sigma_{11}}$        $\xrightarrow{\Sigma_{12}}$   
 $\xrightarrow{\Sigma_{21} \text{ or } \Sigma_{12}}$        $\xrightarrow{\Sigma_{22}}$

Are  $x_1$  and  $x_2$  independent? What about  $(x_1, x_2) \& x_3$ ?

$\rightarrow \text{Cov}(x_1, x_2) \neq 0 \Rightarrow x_1 \text{ and } x_2 \text{ are not indep.}$

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_3 \end{bmatrix}$$

Since

$$\text{Cov}((x_1, x_2), x_3) = \Sigma_{12} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ so we can conclude}$$

that  $x_3$  is indep of  $x_1$  as well as  $x_2$ .

Ex:- (4) :- Let  $\underline{x} \sim N_3(\underline{\mu}, \Sigma)$  with  $\underline{\mu} = \begin{bmatrix} -3 \\ 1 \\ 4 \end{bmatrix}$  and

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

which of the following r.v.'s are indep? Explain.

(i)  $x_1$  and  $x_2 \rightarrow$  Not indep.

(ii)  $x_2$  and  $x_3 \rightarrow$  Indep

(iii)  $(x_1, x_2)$  and  $x_3 \rightarrow$  Independent.

(iv)  $\frac{(x_1 + x_2)}{2}$  and  $x_3$ .

(v) Also find the dist' of  $3x_1 - 2x_2 + x_3$

$\rightarrow$  (iv) let  $y = \frac{x_1 + x_2}{2}$

$$\text{Var}(y, x_3) = \text{Var}(y) + \text{Var}(x_3) + 2\text{Cov}(y, x_3)$$

$$= \text{Var}\left(\frac{x_1 + x_2}{2}\right) + 2 + 0$$

$$= \frac{1}{4} \text{Var}(x_1 + x_2) + 2$$

$$= \frac{1}{4} [\text{Var}(x_1) + \text{Var}(x_2) + 2\text{Cov}(x_1, x_2)] + 2$$

$$= \frac{1}{4} [4 + 5 - 4] + 2 = \frac{1}{2} + 2 = \boxed{\frac{5}{2}}$$

(v)  $3x_1 - 2x_2 + x_3 \sim N(-7, 55)$

Mean  $\rightarrow 3(-3) - 2(1) + 4$

$$-9 - 2 + 4 = -7$$

$$\text{Var}(3x_1 - 2x_2 + x_3)$$

$$= 9\text{Var}(x_1) + 4\text{Var}(x_2) + \text{Var}(x_3)$$

$$- 12\text{Cov}(x_1, x_2) - 4\text{Cov}(x_2, x_3) + 6\text{Cov}(x_1, x_3)$$

$$= 9(1) + 4(5) + 2 - 12(-2) - 4(0) + 6(0)$$

$$= 9 + 20 + 2 + 24$$

$$= 55 )$$

\* Conditional Distribution:

Theorem (7): If  $\underline{x} \sim N_p(\underline{\mu}, \Sigma)$ , then the conditional dist'n of  $\underline{x}_{(1)}$  given  $\underline{x}_{(2)} = \underline{x}_{(2)}$  is given by

$$N_p(\underline{\mu}_{(1)} + \Sigma_{12} \Sigma_{22}^{-1} (\underline{x}_{(2)} - \underline{\mu}_{(2)}), \Sigma_{11.2})$$

where

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

Proof :- Consider the following transformation

$$\underline{y}_{(1)} = \underline{x}_{(1)} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

$$\underline{y}_{(2)} = \underline{x}_{(2)}$$

Then,

$$\begin{pmatrix} \underline{y}_{(1)} \\ \underline{y}_{(2)} \end{pmatrix} \sim N_p \left[ \begin{pmatrix} \underline{\mu}_{(1)} - \Sigma_{12} \Sigma_{22}^{-1} \underline{\mu}_{(2)} \\ \underline{\mu}_{(2)} \end{pmatrix}, \begin{pmatrix} \Sigma_{11.2} & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \right]$$

Hence the pdf of  $\begin{pmatrix} \underline{y}_{(1)} \\ \underline{y}_{(2)} \end{pmatrix}$  is

$$f(\underline{y}_{(1)}, \underline{y}_{(2)}) = f(\underline{y}_{(1)} | \underline{\mu}_{(1)} - \Sigma_{12} \Sigma_{22}^{-1} \underline{\mu}_{(2)}, \Sigma_{11.2}) \cdot f(\underline{y}_{(2)} | \underline{\mu}_{(2)}, \Sigma_{22})$$

(\*)

$\therefore \underline{y}_{(1)} \sim N_p (\underline{\mu}_{(1)} - \Sigma_{12} \Sigma_{22}^{-1} \underline{\mu}_{(2)}, \Sigma_{11.2})$   
and

$\underline{y}_{(2)} \sim N_{p-p_1} (\underline{\mu}_{(2)}, \Sigma_{22})$   
independently as  $\text{Cov}(\underline{y}_{(1)}, \underline{y}_{(2)}) = 0$

The pdf of  $\begin{pmatrix} \underline{x}_{(1)} \\ \underline{x}_{(2)} \end{pmatrix}$  will be obtained from (\*) by

replacing  $\underline{y}_{(1)}$  by  $\underline{x}_{(1)} - \Sigma_{12} \Sigma_{22}^{-1} \underline{x}_{(2)}$  and  $\underline{y}_{(2)}$  by  $\underline{x}_{(2)}$

Since the jacobian of the transformation is unity

Hence the pdf of  $\begin{pmatrix} \underline{x}_{(1)} \\ \underline{x}_{(2)} \end{pmatrix}$  is

$$f(\underline{x}_{(1)}, \underline{x}_{(2)}) = f(\underline{x}_{(1)} - \Sigma_{12} \Sigma_{22}^{-1} \underline{x}_{(2)} | \underline{\mu}_{(1)} - \Sigma_{12} \Sigma_{22}^{-1} \underline{\mu}_{(2)}, \Sigma_{11.2})$$

$$\times f(\underline{x}_{(2)} | \underline{\mu}_{(2)}, \Sigma_{22})$$

Hence the conditional Pdf of  $\underline{x}_{(1)}$  given  $\underline{x}_{(2)} = \underline{x}_{(2)}$   
is

$$f(\underline{x}_{(1)} | \underline{x}_{(2)}) = \frac{f(\underline{x}_{(1)}, \underline{x}_{(2)})}{f(\underline{x}_{(2)})}$$

$$= \frac{1}{(\sqrt{2\pi})^{p_1} \sqrt{|\Sigma_{11.2}|}} e^{-\frac{1}{2} [(\underline{z} - \underline{\mu})' \Sigma_{11.2}^{-1} (\underline{z} - \underline{\mu})]}$$

$$= \frac{1}{(\sqrt{2\pi})^{p_1} \sqrt{|\Sigma_{11.2}|}} e^{\frac{1}{2} \left[ \left( (\underline{x}_{(1)} - \Sigma_{12} \Sigma_{22}^{-1} \underline{x}_{(2)}) - (\underline{\mu}_{(1)} - \Sigma_{12} \Sigma_{22}^{-1} \underline{\mu}_{(2)}) \right)' \Sigma_{11.2}^{-1} \right]}$$

$$\times \left\{ (\underline{x}_{(1)} - \Sigma_{12} \Sigma_{22}^{-1} \underline{x}_{(2)}) - (\underline{\mu}_{(1)} - \Sigma_{12} \Sigma_{22}^{-1} \underline{\mu}_{(2)}) \right\}'$$

$$= \frac{1}{e^{\frac{1}{2}} \left( \sum_{11-2} \right)} \left[ \left( \underline{x}_{(1)} - \underline{u}_{(1)} - \sum_{12} \sum_{22}^{-1} (\underline{x}_{(2)} - \underline{u}_{(2)}) \right)^T \right]$$

$$\left( \frac{1}{2\pi} \right)^{\frac{1}{2}} \sqrt{\sum_{11-2}} \times \left[ \underline{x}_{(1)} - \underline{u}_{(1)} - \sum_{12} \sum_{22}^{-1} (\underline{x}_{(2)} - \underline{u}_{(2)}) \right]$$

$$\frac{\underline{x}_{(1)}}{\underline{x}_{(2)} = \underline{x}_{(2)}} \sim N_p \left( \underline{u}_{(1)} + \sum_{12} \sum_{22}^{-1} (\underline{x}_{(2)} - \underline{u}_{(2)}) \rightarrow \sum_{11-2} \right)$$

\* Remark :-

22/7/25

(±)

(+)  $E[\underline{x}_{(1)} | \underline{x}_{(2)} = \underline{x}_{(2)}] = \underline{u}_{(1)} + \sum_{12} \sum_{22}^{-1} (\underline{x}_{(2)} - \underline{u}_{(2)})$  is a linear function of  $\underline{x}_{(2)}$ .

Hence the regression of  $\underline{x}_{(1)}$  and  $\underline{x}_{(2)}$  is linear.

Also, the dispersion of  $\underline{x}_{(1)}$  and given  $\underline{x}_{(2)} = \underline{x}_{(2)}$  is  $\sum_{11-2} = \sum_{11} - \sum_{12} \sum_{22}^{-1} \sum_{21}$  is indep. of  $\underline{x}_{(2)}$

Hence, the conditional dist' of  $\underline{x}_{(1)} | \underline{x}_{(2)} = \underline{x}_{(2)}$  is Homoscedastic.

Soln.

Ex :- (5) :-  $\underline{x} = (x_1, x_2, x_3) \sim N_3(\underline{0}, \Sigma)$  where

$$\Sigma = \begin{bmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{bmatrix}$$

Show that for any  $c > 0$

$$P[(x_1^2 + c)g^2 - 2(x_1 x_2 + x_2 x_3)g + x_1^2 + x_2^2 + x_3^2 - c \leq 0] = \int_0^c \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \cdot y^{1/2} dy$$

→ Since  $\underline{x} \sim N_3(\underline{0}, \Sigma)$

$$(\underline{x} - \underline{0})^T \Sigma^{-1} (\underline{x} - \underline{0}) \sim \chi^2_3$$

$$\underline{x}^T \Sigma^{-1} \underline{x} \sim \chi^2_3$$

$$\left(\frac{\underline{x} - \underline{u}}{\sigma}\right)^2 \sim \chi^2_{(1)}$$

$$(\underline{x} - \underline{u})' \Sigma^{-1} (\underline{x} - \underline{u}) \sim \chi^2_{(3)}$$

PAGE NO.: 11  
DATE: / /

$$\Rightarrow P \left[ \underline{x}' \Sigma^{-1} \underline{x} \leq c \right] = \int_0^c \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \chi^2_{(2)}(y) dy$$

$$= \int_0^c \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} y dy$$

(Chi-square dist)  
Here  $n=3$  &  
 $\chi^2 = y$

22/7/25

(2) Suppose  $(\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p)' \sim N_p(\underline{u}, \Sigma)$

where  $\underline{u} = (1, 1, \dots, 1)', \Sigma = \begin{bmatrix} 1 & 1 & \dots & 1 & 1 \\ 1 & 2 & \dots & 2 & 2 \\ \vdots & & & & \\ 1 & 2 & 3 & \dots & 3 \end{bmatrix}$

then Show that

$$Q = (x_2 - x_1)^2 + (x_3 - x_2)^2 + \dots + (x_p - x_{p-1})^2 \text{ has a } \chi^2 \text{-dist}.$$

Sol<sup>n</sup>: let  $y_i = x_i - x_{i-1}, i = 1, 2, \dots, p$

$$\therefore \underline{y}_{(p-1)} x_1 = \begin{pmatrix} y_2 \\ y_3 \\ \vdots \\ y_p \end{pmatrix}$$

$$= \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

$(p-1)x_p$

$$= B_{(p-1)x_p} \underline{x}_{p+1} \text{ (say)}$$

Note that  $g(B) = (p-1)$

Hence  $\underline{y} = B \underline{x} \sim N_{p-1}(B\underline{u}, B\Sigma B')$

$$\therefore E(\underline{y}) = E(B \underline{x}) = B E(\underline{x}) = B \underline{u},$$

$$D(\underline{y}) = D(B \underline{x}) = B D(\underline{x}) B' = B \Sigma B'$$



$$E(B\bar{x}) = E[B(x_i - \bar{x}_{i-1})] = B[E(x_i) - E(\bar{x}_{i-1})] \\ = B[\mu - \mu] \\ = [0]$$

$$B^T B' = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 2 & \dots & 2 \\ \vdots & & & & \\ 1 & 2 & 3 & \dots & 3 \\ 2 & 3 & \dots & (p-1) & p \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & 1 \end{bmatrix} (p-1) \times (p-1)$$

$\underline{y} \sim N_{p-1}(\underline{0}, I_{p-1})$   
 $\Rightarrow y_2, y_3, \dots, y_p$  are iid  $N_1(0, 1)$

$$\textcircled{1} = y_2^2 + y_3^2 + \dots + y_p^2 \\ = \sum_{i=2}^p y_i^2 \sim \chi^2_{(p-1) \text{ df}}$$

Theorem (8) :-  $\underline{x}_{p+1} \sim N_p(\underline{\mu}, \Sigma)$  iff  
 $\underline{l}' \underline{x} \sim N_1(l'\underline{\mu}, l'\Sigma l)$ ,  $\Leftrightarrow l_{p+1} \in \mathbb{R}$

Proof :- If : Let  $\underline{l}' \underline{x} \sim N_1(l'\underline{\mu}, l'\Sigma l)$

Then the mgf of  $\underline{l}' \underline{x}$  is  $E[e^{t(\underline{l}' \underline{x})}]$

$$= e^{t(l'\underline{\mu}) + \frac{1}{2} t^2 (l'\Sigma l)}$$

$$\text{putting } t = 1 \\ = E[e^{\underline{l}' \underline{x}}] = e^{\underline{l}' \underline{\mu} + \frac{1}{2} \underline{l}' \Sigma \underline{l}} + l$$

$=$  mgf of  $p$ -variate Normal  
 ie.  $N_p(\underline{\mu}, \Sigma)$

By uniqueness property of mgf  
 $\underline{x} \sim N_p(\underline{\mu}, \Sigma)$

Only if : let  $\underline{x} \sim N_p(\underline{\mu}, \Sigma)$

Consider a vector  $k\underline{t}$  where  $k$  is a constant

The mgf of  $\underline{x}$  is  $E[e^{(k\underline{t}')\underline{x}}] = e^{(\underline{k}\underline{t})'\underline{\mu} + \frac{1}{2}(\underline{k}\underline{t})'\Sigma(\underline{k}\underline{t})}$ ,  $\forall t \in R^p$

$$\Leftrightarrow E[e^{k(\underline{t}'\underline{x})}] = e^{k(\underline{t}'\underline{\mu}) + \frac{1}{2}k^2(\underline{t}'\Sigma\underline{t})}$$

$$\Leftrightarrow E[e^{k(\underline{t}'\underline{x})}] = e^{k(\underline{t}'\underline{\mu}) + \frac{1}{2}k^2(\underline{t}'\Sigma\underline{t})}$$

$$\Leftrightarrow \text{mgf of } \underline{t}'\underline{x} = e^{k(\underline{t}'\underline{\mu}) + \frac{1}{2}k^2(\underline{t}'\Sigma\underline{t})} \quad \forall t \in R^p$$

$$= \text{mgf of } N_1(\underline{t}'\underline{\mu}, \underline{t}'\Sigma\underline{t})$$

Hence

$$\underline{t}'\underline{x} \sim N(\underline{t}'\underline{\mu}, \underline{t}'\Sigma\underline{t}) \quad \forall t \in R^p$$

22/7/25

\* MLE of the parameters of Multivariate Normal Dist<sup>n</sup>

Theorem (a) : If  $\underline{x}_n$ ,  $n = 1, 2, \dots, N$  be a random sample of size  $n$  from a  $N_p(\underline{\mu}, \Sigma)$  Then the MLE of its parameters  $\underline{\mu}$  and  $\Sigma$  are given by

$$\hat{\underline{\mu}} = \bar{\underline{x}} \quad \text{and} \quad \hat{\Sigma} = \frac{\underline{A}}{N} \quad \text{where}$$

$$\underline{A} = \sum_{n=1}^N (\underline{x}_n - \bar{\underline{x}})(\underline{x}_n - \bar{\underline{x}})'$$

Proof :- The density function of the  $n^{\text{th}}$  obs<sup>n</sup> is given

by

$$f(\underline{x}_n) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (\underline{x}_n - \underline{\mu})' \Sigma^{-1} (\underline{x}_n - \underline{\mu}) \right]$$

Then the likelihood funct<sup>n</sup> corresponding to (1)

is given by

$$L(\underline{\mu}, \Sigma) = \prod_{n=1}^N f(\underline{x}_n)$$

$$= \prod_{n=1}^N \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (\underline{x}_n - \underline{\mu})' \Sigma^{-1} (\underline{x}_n - \underline{\mu}) \right]$$

$$= \frac{1}{(2\pi)^{Np/2}} \frac{1}{|\Sigma|^{N/2}} \exp \left[ -\frac{1}{2} \sum_{n=1}^N (\underline{x}_n - \underline{\mu})' \Sigma^{-1} (\underline{x}_n - \underline{\mu}) \right] \quad (2)$$

Taking log on both sides of (2), we get

$$\ln L(\underline{\mu}, \Sigma) = -\frac{Np}{2} \ln(2\pi) - \frac{N}{2} \ln |\Sigma| - \frac{1}{2} \sum_{n=1}^N (\underline{x}_n - \underline{\mu})' \Sigma^{-1} (\underline{x}_n - \underline{\mu}) \quad (3)$$

(Consider,

$$\sum_{n=1}^N (\underline{x}_n - \underline{\mu})' \Sigma^{-1} (\underline{x}_n - \underline{\mu})$$

$$= \sum_{n=1}^N (\underline{x}_n - \bar{\underline{x}} + \bar{\underline{x}} - \underline{\mu})' \Sigma^{-1} (\underline{x}_n - \bar{\underline{x}} + \bar{\underline{x}} - \underline{\mu}) \text{ where } \bar{\underline{x}} = \frac{1}{N} \sum_{n=1}^N \underline{x}_n$$

$$= \sum_{n=1}^N (\underline{x}_n - \bar{\underline{x}})' \Sigma^{-1} (\underline{x}_n - \bar{\underline{x}}) + \sum_{n=1}^N (\underline{x}_n - \bar{\underline{x}})' \Sigma^{-1} (\underline{x}_n - \underline{\mu}) \quad \rightarrow = 0$$

$$+ \sum_{n=1}^N (\bar{\underline{x}} - \underline{\mu})' \Sigma^{-1} (\underline{x}_n - \bar{\underline{x}}) + \sum_{n=1}^N (\bar{\underline{x}} - \underline{\mu})' \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu}) \quad \rightarrow = 0 \quad (\text{Property of AM})$$

$$\text{Since } \sum_{n=1}^N (\underline{x}_n - \bar{\underline{x}})' \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu}) = 0 \text{ & } \sum_{n=1}^N (\bar{\underline{x}} - \underline{\mu})' \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu}) = 0$$

Also the term

$(\underline{x}_n - \bar{\underline{x}})' \Sigma^{-1} (\underline{x}_n - \bar{\underline{x}})$  is a scalar term and the trace of scalar is the same scalar.

$$= \sum_{n=1}^N \text{trace} (\underline{x}_n - \bar{\underline{x}}) (\underline{x}_n - \bar{\underline{x}})' \Sigma^{-1} + N (\bar{\underline{x}} - \underline{\mu})' \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu})$$

(As  $\text{trace}(AB) = \text{trace}(BA)$ )

$$= \text{trace} \left[ \sum_{n=1}^N (\underline{x}_n - \bar{\underline{x}})(\underline{x}_n - \bar{\underline{x}})' \right] \Sigma^{-1} + N(\bar{\underline{x}} - \underline{\mu})' \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu}) \quad (4)$$

Using (4) in (3),

$$\ln L(\underline{\mu}, \Sigma) = -\frac{NP}{2} \ln(2\pi) - \frac{N}{2} \ln |\Sigma| - \frac{1}{2} \text{trace} \left[ \sum_{n=1}^N (\underline{x}_n - \bar{\underline{x}})(\underline{x}_n - \bar{\underline{x}})' \right] \Sigma^{-1} - \frac{N}{2} (\bar{\underline{x}} - \underline{\mu})' \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu}) \quad (5)$$

$$\frac{\partial \ln L(\underline{\mu}, \Sigma)}{\partial \underline{\mu}} = 0 \quad \left( \ln |\Sigma| = \ln \frac{1}{|\Sigma^{-1}|} \right)$$

$$\Rightarrow -\frac{N}{2} \frac{\partial}{\partial \underline{\mu}} (\bar{\underline{x}} - \underline{\mu})' \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu}) = 0$$

$$\Rightarrow -\frac{N}{2} \left[ (\bar{\underline{x}} - \underline{\mu})' \Sigma^{-1} (-1) + (-1) \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu}) \right] = 0$$

$$\Rightarrow -\frac{N}{2} \left[ 2 \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu}) (-1) \right] = 0$$

$$\Rightarrow \bar{\underline{x}} - \underline{\mu} = 0$$

$$\Rightarrow \underline{\hat{\mu}} = \bar{\underline{x}} \quad \left( \begin{array}{l} \therefore \Sigma^{-1} (\bar{\underline{x}} - \underline{\mu}) = 0 \\ \text{But } \Sigma^{-1} \neq 0 \Rightarrow \bar{\underline{x}} - \underline{\mu} = 0 \end{array} \right)$$

Using (5)

$$\frac{\partial \ln L(\underline{\mu}, \Sigma)}{\partial \Sigma^{-1}} = 0$$

$$\Rightarrow \frac{N}{2} (\Sigma^{-1})^{-1} - \frac{1}{2} \left( \sum_{n=1}^N (\underline{x}_n - \bar{\underline{x}})(\underline{x}_n - \bar{\underline{x}})' \right) = 0 \quad (\because \underline{\hat{\mu}} = \bar{\underline{x}})$$

$$\Rightarrow \hat{\Sigma} = \frac{1}{N} \left[ \sum_{n=1}^N (\underline{x}_n - \bar{\underline{x}})(\underline{x}_n - \bar{\underline{x}})' \right]$$

H.W

Take

$$\underline{\mu} = \underline{\mu}_0$$

## \* Elliptically Contoured Dist = (ECD):

- ECD are a class of multivariate prob. dist<sup>n</sup>s that generalize the multi-variate Normal dist<sup>n</sup>. They are characterized by the property that they contours of constant prob. density are ellipsoids.
- A key feature of ECD is that they retain many of the desirable properties of Normal dist<sup>n</sup> such as Linearity Marginal and Conditional dist<sup>n</sup> while also allowing for heavier or lighter tails.
- A random vector  $\underline{x}$  is said to have ECD with parameters  $\underline{\mu}$ ,  $\Sigma$  and  $g$  if its pdf can be expressed as

$$f(\underline{x}) = k |\Sigma|^{-1/2} g \left[ (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \right]$$

where

$\underline{\mu}$   $\leftarrow$  location parameter

$\Sigma \leftarrow$  positive definite shape matrix

$g \leftarrow$  generator funct<sup>n</sup> that determines the specific dist<sup>n</sup> within the class.

$k \leftarrow$  normalization constant.

Proof

- A special case of ECD is spherically symmetric dist<sup>n</sup> where the density function depends on the distance from the center

$$g(\underline{x}) = h(\|\underline{x}\|) \text{ for some function } h.$$

↳ Norm of Matrix.

- If  $\underline{x}$  has an ECD and  $\underline{y} = T\underline{x} + \underline{b}$ , where  $T$  is a non-singular matrix, then  $\underline{y}$  also has an ECD.

- The most well-known example is Multivariate Normal dist' where

$$g(\underline{u}) = \exp\left(-\frac{\underline{u}^2}{2}\right)$$

\* Applications :-

- Financial Modelling
- Robust Statistics
- Spatial Statistics
- Machine Learning

28/7/25

(1) Let  $\underline{x} \sim N_p(\underline{0}, \sigma^2 I_p)$  and  $P_1 (m \times p)$  is a matrix  
 $\Rightarrow P_1 P_1' = I_m$  then show that  $\underline{y} = P_1 \underline{x} \sim N_m(\underline{0}, \sigma^2 I_m)$   
 is independently distributed with  $\frac{1}{\sigma^2} (\underline{x}' \underline{x} - \underline{y}' \underline{y}) \sim \chi^2_{p-m}$

Proof :- Note that  $P_1 P_1' = I_m$ , here  $P_1 (m \times p)$  is semi-orthogonal matrix. Hence we can find  $P_2 (p-m \times p)$   $\Rightarrow P_2 P_2' = I_{p-m}$   
 where  $P(p \times p) = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix}$

$$\text{Let } \underline{z} = P \underline{x} = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \underline{x} = \begin{pmatrix} P_1 \underline{x} \\ P_2 \underline{x} \end{pmatrix} = \begin{pmatrix} \underline{y} \\ \underline{w} \end{pmatrix}$$

$$\text{where } \underline{w} = P_2 \underline{x}$$

Since  $P$  is non-singular

$$\underline{z} = P \underline{x} \sim N_p(\underline{0}, \sigma^2 I_p)$$

$$\Rightarrow \begin{pmatrix} \underline{y} \\ \underline{w} \end{pmatrix} \sim N_p(\underline{0}, \sigma^2 I_p)$$

$$\Rightarrow \begin{pmatrix} \underline{y} \\ \underline{w} \end{pmatrix} \sim N_m(\underline{0}, \sigma^2 I_m) \text{ and } \begin{pmatrix} \underline{y} \\ \underline{w} \end{pmatrix} \sim N_{(p-m)}(\underline{0}, \sigma^2 I_{p-m}) \text{ independently}$$

$$\text{Now, } \underline{\underline{z'z}} = \underline{\underline{x'P'Px}} \\ = \underline{\underline{x'x}}$$

$$\Rightarrow \underline{\underline{x'x}} = \underline{\underline{z'z}} = \left( \frac{\underline{\underline{y'}}}{\underline{\underline{w}}} \right)' \left( \frac{\underline{\underline{y}}}{\underline{\underline{w}}} \right) \\ = \left( \frac{\underline{\underline{y'}}}{\underline{\underline{w}}} \right) \left( \frac{\underline{\underline{y}}}{\underline{\underline{w}}} \right) \\ = \underline{\underline{y'y}} + \underline{\underline{w'w}}$$

$$\Rightarrow \underline{\underline{w'w}} = \underline{\underline{x'x}} - \underline{\underline{y'y}}$$

Since  $\underline{\underline{w}} \sim N_{p-m}(\underline{0}, \sigma^2 I_{p-m})$

$$\Rightarrow \frac{\underline{\underline{w'w}}}{\sigma^2} \sim \chi^2_{(p-m)}$$

$$\Rightarrow \frac{1}{\sigma^2} (\underline{\underline{x'x}} - \underline{\underline{y'y}}) \sim \chi^2_{(p-m)}$$

which is independently dist with

$$\underline{\underline{y}} \sim N_m(\underline{0}, \sigma^2 I_m)$$

$$-\underline{x} - \underline{x} -$$

# UNIT - 2

PAGE NO.:  
DATE: / /

## WISHART DISTRIBUTION :-

Defn :- A  $p \times p$  random matrix  $W$  is called a Wishart matrix based on ' $m$ ' d.f. if it can be expressed as

$$W = \sum_{\alpha=1}^m \underline{Y}_{\alpha} \underline{Y}_{\alpha}' \quad \text{& follows } p\text{-variate Normal distn.}$$

where  $\underline{Y}_1, \underline{Y}_2, \dots, \underline{Y}_m$  are some ' $m$ ' indep.  $N_p(\mathbf{0}, \Sigma)$  random vectors and we say that

$W \sim \text{Wishart distn}$  on ' $m$ ' d.f. with parameters  $\Sigma$  and we write

$$W \sim W_p(\Sigma, m)$$

$\hookrightarrow$  (p-variate)

- In particular in the univariate case when  $p=1$  in which  $\Sigma = \sigma^2$  (say) is the common variance of the ' $m$ ' normal popn ' $w$ ' on ' $m$ ' d.f. is S.S. for ' $m$ ' indep.  $N(0, \sigma^2)$  variates

$$\begin{aligned} W &= \sum_{\alpha=1}^m \underline{Y}_{\alpha} \underline{Y}_{\alpha}' = \underline{Y}_1 \underline{Y}_1' + \underline{Y}_2 \underline{Y}_2' + \dots + \underline{Y}_m \underline{Y}_m' \\ &= \sum_{i=1}^m Y_i^2 \quad (\text{taking } p=1) \quad | \quad X \sim N(\mu, \sigma^2) \\ \Rightarrow \frac{\sum_{i=1}^m Y_i^2}{\sigma^2} &= \frac{W}{\sigma^2} \sim \chi^2_{(m) \text{ d.f.}} \quad | \quad Z = X - \mu \sim N(0, 1) \\ &\quad \sigma^2 \quad | \quad Z^2 \sim \chi^2_{(1)} \\ &\quad | \quad \Sigma Z^2 \sim \chi^2_{(m)} \end{aligned}$$

Note :- This is why multivariate Wishart distn is generally regarded as Multivariate generalization of the  $\chi^2$  distn.

$$\left( \underline{Y} \sim N(\mathbf{0}, \Sigma) \rightarrow \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{\underline{Y}}{\sigma} \right)^2 \right] \right)$$

29/7/25

PAGE NO.:

DATE: / /

### \* Basic Properties of Wishart dist<sup>n</sup>:

Theorem

(1) :- If  $w \sim W_p(\Sigma, m)$  and if  $c$  is a positive constant then  $(cw) \sim W_p(c\Sigma, m)$

Proof :-

Since  $w \sim W_p(\Sigma, m)$ , there are  $m$ -indep.  $p$ -component random vectors  $\underline{y}_1, \underline{y}_2, \dots, \underline{y}_m$  each distributed as  $N_p(\underline{0}, \Sigma)$  such that  $w = \sum_{\alpha=1}^m \underline{y}_{\alpha} \underline{y}_{\alpha}'$

Evidently we have  $(cw) = c \sum_{\alpha=1}^m \underline{y}_{\alpha} \underline{y}_{\alpha}'$

$$= \sum_{\alpha=1}^m (\sqrt{c} \underline{y}_{\alpha}) (\sqrt{c} \underline{y}_{\alpha}')$$

$\sqrt{c} \underline{y}_1, \sqrt{c} \underline{y}_2, \dots, \sqrt{c} \underline{y}_m$  are iid as  $N(\underline{0}, c\Sigma)$ .   
 $\therefore \underline{y}_{\alpha} \sim N_p(\underline{0}, \Sigma)$   
 $E(\sqrt{c} \underline{y}_{\alpha}) = \sqrt{c} \cdot \underline{0} = \underline{0}$   
 $D(\sqrt{c} \underline{y}_{\alpha}) = (\sqrt{c})^2 D(\underline{y}_{\alpha}) = c\Sigma$

Theorem (2) :- If  $w \sim W_p(\Sigma, m)$  and if  $A$  is  $g \times p$  matrix of rank  $R$ , then  $A w A' \sim W_p(A\Sigma A', m)$

Proof :-

As  $w \sim W_p(\Sigma, m)$  and if  $A$  is  $g \times p$  matrix of rank ' $r$ '. Also  $w = \sum_{\alpha=1}^m \underline{y}_{\alpha} \underline{y}_{\alpha}'$ , there are ' $m$ ' indep.  $N_p(\underline{0}, \Sigma)$

random vectors  $\underline{y}_1, \underline{y}_2, \dots, \underline{y}_m$ .

$$\text{Let } z_{\alpha} = A \underline{y}_{\alpha} \quad (\alpha = 1, 2, \dots, m)$$

then  $z_{\alpha}$ 's are indep.  $N_g(\underline{0}, A\Sigma A')$

$$(\because E(z_{\alpha}) = E(A \underline{y}_{\alpha}) = \underline{0})$$

$$D(z_{\alpha}) = D(A \underline{y}_{\alpha}) = A D(\underline{y}_{\alpha}) A' = A \Sigma A'$$



so that  $n \times n$  random matrix

$$\sum_{\alpha=1}^m Z_{\alpha} Z_{\alpha}' = \sum_{\alpha=1}^m (A \underline{y}_{\alpha})(A \underline{y}_{\alpha})'$$

$$= \sum_{\alpha=1}^m A(\underline{y}_{\alpha} \underline{y}_{\alpha}') A'$$

$$= A \left( \sum_{\alpha=1}^m \underline{y}_{\alpha} \underline{y}_{\alpha}' \right) A'$$

$$= A \Sigma A'$$

$\Sigma Z_{\alpha} Z_{\alpha}'$  is a  $n \times n$  random matrix on 'm' d.f.  
i.e.  $\Sigma Z_{\alpha} Z_{\alpha}' \sim W_n (\Sigma A A', m)$

$$(\because D(A \Sigma A') = A D(\Sigma) A' = A \Sigma A')$$

Theorem (3) = Let  $W \sim W_p(\Sigma, m)$  and let  $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ ,  $W = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}$  where  $\Sigma_{11}$  and  $w_{11}$  are  $k \times k$  then

(i)  $w_{11} \sim W_k (\Sigma_{11}, m)$  and  $w_{22} \sim W_{p-k} (\Sigma_{22}, m)$

(ii)  $w_{11}$  and  $w_{22}$  are indep. iff  $\Sigma_{12} = 0$

Proof = (i) Since  $W \sim W_p(\Sigma, m)$

$W = \sum_{\alpha=1}^m \underline{y}_{\alpha} \underline{y}_{\alpha}'$   $\underline{y}_{\alpha}'$ 's are some indep  $N_p(0, I)$  random vectors

Now partitioning  $\underline{y}_{\alpha}$  as

$$\underline{y}_{\alpha}' = (\underline{y}_{\alpha}^{(1)'}, \underline{y}_{\alpha}^{(2)'}) \quad \alpha = 1, 2, \dots, m$$

where each  $\underline{y}_{\alpha}^{(1)}$  has 'k' components and  $\underline{y}_{\alpha}^{(2)}$  has  $(p-k)$  components.

$\underline{y}_{\alpha}^{(1)}$  are iid as  $N_k(0, \Sigma_1)$  and

$\underline{y}_{\alpha}^{(2)}$  are iid as  $N_{p-k}(0, \Sigma_2)$

Moreover we have

$$w_{11} = \sum_{\alpha=1}^m y_{\alpha}^{(1)} y_{\alpha}^{(1)'} \quad \text{and} \quad w_{22} = \sum_{\alpha=1}^m y_{\alpha}^{(2)} y_{\alpha}^{(2)'}$$

Hence  $w_{11} \sim w_p(\Sigma_{11}, m)$  and

$$w_{22} \sim w_{p-k}(\Sigma_{22}, m)$$

(ii) Further more for each  $\alpha$ ,  $y_{\alpha}^{(1)}$  and  $y_{\alpha}^{(2)}$  are independent iff  $\Sigma_{12} = 0$

Hence (ii) follows from the independence of  $y$ 's

(Corollary)

Corollary :- Let  $w \sim w_p(\Sigma, m)$  then

(i) Every principal sub-matrix of  $w$  has a Wishart dist' on ' $m$ ' d.f.

(ii) For a given  $K \leq p$  the principal submatrix  $w_{11}, w_{22}, \dots, w_{kk}$  of  $w$  are mutually indep. iff  $\Sigma_{ij} = 0 \forall i \neq j$  where  $\Sigma_{ij}$ 's are the sub matrix of the properly partitioned  $\Sigma$ .

(iii) For each  $i = 1, 2, \dots, p \rightarrow \frac{w_{ii}}{\sigma_{ii}} \sim \chi^2_{(n)}$  where

$w_{ii}$ 's and  $\sigma_{ii}$ 's are  $(i, i)$ th element of  $w$  and  $\Sigma$  respectively.

Theorem (4) :- If  $w_1 \sim w_p(\Sigma, m_1)$  and  $w_2 \sim w_p(\Sigma, m_2)$  independently then  $w_1 + w_2 \sim w_p(\Sigma, m_1 + m_2)$

Proof :- Since  $w_1 \sim w_p(\Sigma, m_1)$  and  $w_2 \sim w_p(\Sigma, m_2)$  independently, there are  $m_1 + m_2$  indep  $N_p(0, \Sigma)$  random vector  $y_1, y_2, \dots, y_{m_1+m_2}$  such that

$$w_1 = \sum_{\alpha=1}^{m_1} y_{\alpha} y_{\alpha}' \quad \text{and} \quad w_2 = \sum_{\alpha=m_1+1}^{m_1+m_2} y_{\alpha} y_{\alpha}'$$

(consequently,

$$W_1 + W_2 = \sum_{\alpha=1}^{m_1} \underline{y}_\alpha \underline{y}'_\alpha + \sum_{\alpha=m_1+1}^{m_2} \underline{y}_\alpha \underline{y}'_\alpha \\ = \sum_{\alpha=1}^{m_1+m_2} \underline{y}_\alpha \underline{y}'_\alpha$$

Hence  $W_1 + W_2 \sim W_p(\Sigma, m_1 + m_2)$

are

Corollary = IF  $w_1, w_2, \dots, w_k$  are independently distributed as  $W_p(\Sigma, m_1), W_p(\Sigma, m_2), \dots, W_p(\Sigma, m_k)$  resp. then

$$w_1 + w_2 + \dots + w_k \sim W_p(\Sigma, m_1 + m_2 + \dots + m_k)$$

\* Remarks :

(1) Thm (i) is a reproductive property of the Wishart dist". As we have seen that two independent Wishart matrices with same  $\Sigma$  parameter can combine to give birth to a third Wishart Matrix with a same parameter and d.f. equal to the sum of the d.f. of the parent matrices.

(2) Wishart dist" can be regarded as Multivariate Generalization of chi-square dist"

(3) From the def" of Wishart Matrix  $W = \sum_{\alpha=1}^m \underline{y}_\alpha \underline{y}'_\alpha$

$$E(W) = E\left(\sum_{\alpha=1}^m \underline{y}_\alpha \underline{y}'_\alpha\right) \quad \underline{y}_\alpha \sim N_p(0, \Sigma)$$

$$= \sum_{\alpha=1}^m [E(\underline{y}_\alpha \underline{y}'_\alpha)]$$

$$= [E(\underline{y}_1 \underline{y}'_1) + E(\underline{y}_2 \underline{y}'_2) + \dots + E(\underline{y}_m \underline{y}'_m)]$$

$$= V(\underline{y}_1) + V(\underline{y}_2) + \dots + V(\underline{y}_m) \quad \left( \because V(\underline{y}_\alpha) = E(\underline{y}_\alpha \underline{y}'_\alpha) - E(\underline{y}_\alpha)^2 \right)$$

$$= \boxed{m \sum} \quad \Rightarrow E(\underline{y}_\alpha \underline{y}'_\alpha) = V(\underline{y}_\alpha) \quad (\because E(\underline{y}_\alpha) = 0) \quad (i)$$

(4) Assumption  $m \geq p$  is invariably made in the application of Wishart dist<sup>n</sup> in order to have a non-singular dist<sup>n</sup>.

4/8/25

\* Characteristic function of a Wishart dist<sup>n</sup>:

Theorem (i) :- The characteristic funct<sup>n</sup> of a Wishart  $W_p(\Sigma, m)$  dist<sup>n</sup>

$$\Phi(t) = \left( \frac{|\Sigma^{-1}|}{|\Sigma^{-1} - 2tI|} \right)^{m/2} \quad \text{where } T \text{ is a}$$

Symmetric matrix of real no.s.

Proof: Suppose  $w \sim W_p(\Sigma, m)$  then there are 'm' indep. normal  $N_p(\underline{0}, I)$  random vectors  $\underline{y}_1, \underline{y}_2, \dots, \underline{y}_m \rightarrow$

$$w = \sum_{\alpha=1}^m \underline{y}_\alpha \underline{y}'_\alpha$$

Now Since  $\Sigma^{-1}$  is positive definite for a given real symmetric matrix  $T$ , we find a non-singular matrix  $P \rightarrow P^{-1} \Sigma^{-1} P = I$

$$P' T P = \text{diag}(d_1, d_2, \dots, d_p) = D \text{ (say)}$$

where  $d_1, d_2, \dots, d_p$  are the roots of the determinants eqns.

$$|T - d \Sigma^{-1}| = 0$$

$$\text{If } z_\alpha = P^{-1} \underline{y}_\alpha, \quad \alpha = 1, 2, \dots, m$$

we see that

(i)  $z_1, z_2, \dots, z_m$  are distributed independently as  $N_p(0, I)$ .

$$E(z_\alpha) = E(P^{-1}y_\alpha) = P^{-1}E(y_\alpha) = 0$$

$$\begin{aligned} D(z_\alpha) &= D(P^{-1}y_\alpha) = P^{-1}E(y_\alpha)y_\alpha^T P^{-1} \rightarrow \text{symmetric} \\ &= P^{-1}P^{-1} = I \end{aligned}$$

(ii) trace  $TW = \text{trace} \left( T \sum_{\alpha=1}^m y_\alpha y_\alpha' \right)$

Since  $\text{Tr } AB = \text{Tr } BA$

$\text{trace } a = a$  for all scalars.

$$= \sum_{\alpha=1}^m y_\alpha' T y_\alpha = \sum_{\alpha=1}^m y_\alpha' P' T P y_\alpha$$

$$= \sum_{\alpha=1}^m z_\alpha' P' T P z_\alpha \quad \left( \because (Pz_\alpha)' T (Pz_\alpha) = z_\alpha' P' T P z_\alpha \right)$$

$$= \sum_{\alpha=1}^m z_\alpha' D z_\alpha$$

$$= \sum_{\alpha=1}^m \sum_{j=1}^p d_j z_{j\alpha}^2$$

$$\left( \because \sum_{\alpha=1}^m [d_1^2 z_{1\alpha}^2 + d_2^2 z_{2\alpha}^2 + \dots + d_p^2 z_{p\alpha}^2] \right)$$

where  $z_{j\alpha}$  denotes the  $j^{th}$  component of  $z_\alpha$   
 $(j = 1, 2, \dots, p, \alpha = 1, \dots, m)$

The C.F. of  $W$ , which is the joint C.F. of elements of  $W$ , is

$$E[e^{it\pi T W}] = E[e^{it\pi T \sum_{\alpha=1}^m y_\alpha y_\alpha'}]$$

$$= E[e^{\frac{i}{2} \sum j d_j z_{j\alpha}^2}]$$

$$= \prod_{\alpha} \prod_j E \left[ e^{i d_j z_{j\alpha}} \right]^2 \quad \left( \because E[e^{x+y}] = E[e^x] \cdot E[e^y] \right)$$

$$= E[e^x]^2 E[e^y]^2$$

$\therefore$  since  $z_j$ 's are indep. and so

Since each  $z_{j\alpha} \sim N(0, 1)$  and hence each  $z_{j\alpha}^2 \sim \chi^2_1$

$$= \prod_{\alpha} \prod_j (\pm - 2 i d_j)^{-1/2} \quad (\text{Using C.F. of } \chi^2 \text{ dist})$$

$$= \prod_{\alpha} \left| \pm - 2 \sum_j d_j \right|^{-1/2}$$

$$= \prod_{\alpha} \left| \pm - 2 S D \right|^{-1/2}$$

$$= \left| P' \Sigma^{-1} P - 2 i P' T P \right|^{-m/2}$$

$$= \left[ |P'| |P| \left| \Sigma^{-1} - 2 i T \right| \right]^{-m/2}$$

$$= \left[ \frac{|\Sigma|}{|\Sigma^{-1} - 2 i T|} \right]^{-m/2}$$

$$= \left[ \frac{|\Sigma|}{|\Sigma^{-1}|} \right]^{m/2}$$

Hence the proof.

5/8/25

\*

Gram Matrix :-

- A real symmetric matrix  $A$  is called a gram matrix if it is expressible as  $A = BB'$  for some matrix  $B$  in which case  $B$  is called a grammian component of  $A$ .

Theorem: Suppose  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  are independently distributed as  $N_p(\underline{B}\underline{a}_1, \Sigma), \dots, N_p(\underline{B}\underline{a}_n, \Sigma)$  respectively where  $B$  is some  $P \times K$  matrix  $n \geq p+k$  and  $A = (\underline{a}_1, \underline{a}_2, \dots, \underline{a}_n)$  is a given  $K \times n$  matrix of rank  $K$ . Also let  $x = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$ ,  $H = AA'$ ,  $G = xA'H^{-1}$  then

$$x^*x' - GHG' \sim w_p(I, n-k)$$

Proof: Since  $A$  is full row rank, the gram matrix  $H$  is p.d. and hence we can find a non-singular matrix  $P \rightarrow$

$$\begin{aligned} I &= PHP' \\ &= P(AA')P' \\ &= (PA)(PA)' \end{aligned}$$

Showing that the matrix  $\varPhi = PA$  is some orthogonal matrix, we can find an  $(n-k) \times n$  matrix  $R$  of rank  $(n-k) \rightarrow (\varPhi' | R')$  is orthogonal in which case  $\varPhi R' = 0$ .

Evidently, since  $P$  is non-singular we have  $AR' = 0$ .

Consider the orthogonal transformation

$$y = x(\varPhi' | R') = (x\varPhi' | xR')$$

$$\text{let } x\varPhi' = \underline{y}^{(1)} = (\underline{y}_1, \underline{y}_2, \dots, \underline{y}_k)$$

$$xR' = \underline{y}^{(2)} = (\underline{y}_{k+1}, \underline{y}_{k+2}, \dots, \underline{y}_n)$$

$$E(\underline{y}^{(1)}) = E(x\varPhi') = E(x)\varPhi'$$

$$\text{Since } \underline{x}_1 \sim N_p(\underline{B}\underline{a}_1, \Sigma)$$

$$\text{Also } x = (\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n)$$

$$\begin{aligned}
 E(\underline{y}^{(1)}) &= E(x_1, x_2, \dots, x_n) \varphi' \\
 &= E(x_1) E(x_2) \dots E(x_n) \varphi' \\
 &= (B \underline{a}_1) (B \underline{a}_2) \dots (B \underline{a}_n) \varphi' \\
 &= B(\underline{a}_1, \underline{a}_2, \dots, \underline{a}_n) \varphi' \\
 &= BA\varphi'
 \end{aligned}$$

Since  $\varphi = PA \Rightarrow A = P^{-1}\varphi$ .

$$E(\underline{y}^{(1)}) = B P^{-1} \varphi \varphi'$$

$$\Rightarrow E(\underline{y}^{(1)}) = B P^{-1} \varphi \varphi' \quad (\text{since } \varphi' \varphi = I)$$

$$\begin{aligned}
 E(\underline{y}^{(2)}) &= E(x R') = E(x) R' \\
 &= B A R' \quad (\text{since } A P = I) \\
 &= 0 \quad (\text{since } A R' = 0)
 \end{aligned}$$

$$\begin{aligned}
 D(\underline{y}) &= D[x \times (\varphi' | R')] \\
 &= E[(x - E(x))(\varphi' | R')(\varphi' | R')^T (x - E(x))'] \\
 &= E[(x - E(x))(x - E(x))'] \\
 &= D(x) \\
 &= D(x_1, x_2, \dots, x_n) \\
 &= \text{diag}(\Sigma, \Sigma, \dots, \Sigma) \quad (\because \text{Since } x_i \text{'s are independent})
 \end{aligned}$$

$$D(x_\alpha) = \Sigma \quad \forall \alpha = 1, 2, \dots, n$$

Note that  $D(\underline{y})$  is  $P_n \times P_n$

$$\therefore D(\underline{y}_\alpha) = \Sigma \quad \forall \alpha = 1, 2, \dots, n$$

It follows that  $y_{k+1}, y_{k+2}, \dots, y_n$  are independently distributed as  $N_p(\underline{0}, \Sigma)$  each

Finally we have,

$$\underline{x}^* \underline{x}' = \underline{G} \underline{H} \underline{G}'$$

$$= \underline{\psi} \begin{pmatrix} \underline{\varphi}' & \underline{R}' \end{pmatrix}^{-1} \begin{pmatrix} \underline{\varphi} \\ \hline \underline{R} \end{pmatrix} \underline{\psi}' - \underline{X} \underline{A}' \underline{H}' \underbrace{\underline{H} \underline{H}' \underline{A} \underline{X}'}_{\underline{I}}$$

$$= \underline{\psi} \underline{\psi}' - \underline{X} \underline{A}' \underline{H}' \underline{A} \underline{X}' \quad (\because \underline{\varphi}' \text{ and } \underline{R}' \text{ are orthogonal})$$

$$= \underline{\psi} \underline{\psi}' - \underline{X} \underline{\varphi}' (\underline{P}'')' \underline{H}' \underline{P}^{-1} \underline{\varphi} \underline{X}'$$

$$= \underline{\psi} \underline{\psi}' - \underline{X} \underline{\varphi}' (\underline{P} \underline{H} \underline{P}')^{-1} \underline{\varphi} \underline{X}' \quad (\text{Since } \underline{A} = \underline{P}^{-1} \underline{\varphi})$$

Since  $\underline{P} \underline{H} \underline{P}' = \underline{I}$

$$\begin{aligned} \Rightarrow &= \underline{\psi} \underline{\psi}' - \underline{X} \underline{\varphi}' \underline{\varphi} \underline{X}' \\ &= \underline{\psi} \underline{\psi}' - \underline{\psi}^{(1)} \underline{\psi}^{(1)'} \quad (\text{Since } \underline{X} \underline{\varphi}' = \underline{\psi}^{(1)}) \\ &= \underline{\psi}^{(1)} \underline{\psi}^{(1)'} + \underline{\psi}^{(2)} \underline{\psi}^{(2)'} - \underline{\psi}^{(1)} \underline{\psi}^{(1)'} \\ &= \underline{\psi}^{(2)} \underline{\psi}^{(2)'} \sim w_p(\Sigma, n-k) \end{aligned}$$

Remark:  $\underline{G} \underline{H} \underline{G}' = \underline{\psi}^{(1)} \underline{\psi}^{(1)'} = \sum_{\alpha=1}^k \underline{\psi}_{\alpha}^{(1)} \underline{\psi}_{\alpha}^{(1)'}$  and columns of

$\underline{\psi}_{\alpha}^{(1)}$  and  $\underline{\psi}_{\alpha}^{(1)'}$  are independently distributed as  $N_p(\underline{0}, \underline{I}), \dots, N_p(\underline{0}_k, \Sigma)$  respectively where  $v_{\alpha}' = B \underline{f}_{\alpha} \rightarrow \underline{f}_{\alpha}$  being the  $\alpha^{\text{th}}$  column of  $\underline{P}'$ ,  $\alpha = 1, 2, \dots, k$ .

The dist<sup>n</sup> of such a matrix  $\underline{\psi}^{(1)} \underline{\psi}^{(1)'}$  is called a non-central Wishart dist<sup>n</sup> on  $k$  degrees of freedom with non-centrality parameter.

$$\delta^2 = \sum_{\alpha=1}^k v_{\alpha}' \Sigma^{-1} v_{\alpha}$$

and it is denoted as

$$\underline{G} \underline{H} \underline{G}' \sim w_{p, \delta}(\Sigma, k)$$

Corollary (1) :-  $\underline{x} \underline{x}' - G H G'$  and  $G$  are independent.

→ This follows as col<sup>m</sup> of  $\underline{y}_\alpha^{(1)}$ 's are independently distributed.

Corollary (2) :- If  $B = 0$  then  $G H G' = \underline{y}_\alpha^{(1)} \underline{y}_\alpha^{(1)'} \sim W_p(\Sigma, k)$

Theorem :- Suppose  $W \sim W_p(\Sigma, m)$  and let  $W \in \Sigma$  be partitioned respectively as  $(w_{11} \ w_{12})$  &  $(\Sigma_{11} \ \Sigma_{12})$   
 $w_{21} \ w_{22}$   $\Sigma_{21} \ \Sigma_{22}$

where  $w_{11}$  and  $\Sigma_{11}$  are  $k \times k$  then for a given (fixed)  $w_{22}$  the random matrix

$$w_{11 \cdot 2} = w_{11} + w_{12} \Sigma_{22}^{-1} w_{21} \sim W_k(\Sigma_{11 \cdot 2}, m-p+k)$$

$$\text{where } \Sigma_{11 \cdot 2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$$

Proof :- By definition of Wishart matrix, we have

$$W = \sum_{\alpha=1}^m \underline{y}_\alpha \underline{y}_\alpha'$$
 for some 'm' indep.  $N_p(\underline{0}, \Sigma)$

random vectors  $\underline{y}_1, \underline{y}_2, \dots, \underline{y}_m$

Now partitioning each  $\underline{y}_\alpha$  as

$$\underline{y}_\alpha = (\underline{y}_\alpha^{(1)}, \underline{y}_\alpha^{(2)}), \quad \alpha = 1, 2, \dots, m$$

where each  $\underline{y}_\alpha^{(1)}$  has 'k' and each  $\underline{y}_\alpha^{(2)}$  has remaining components.

So,

$$w_{11} = \sum_{\alpha=1}^m \underline{y}_\alpha^{(1)} \underline{y}_\alpha^{(1)'} \quad \text{and} \quad w_{12} = \sum_{\alpha=1}^m \underline{y}_\alpha^{(1)} \underline{y}_\alpha^{(2)'} = w_{21}$$

$$w_{22} = \sum_{\alpha=1}^m \underline{y}_\alpha^{(2)} \underline{y}_\alpha^{(2)'}$$

Not

Evidently holding  $w_{22}$  fixed means holding  $y_\alpha^{(2)} = \underline{y}_\alpha^{(2)}$  (say) fixed;  $\alpha = 1, 2, \dots, m$

Now for  $\alpha = 1, 2, \dots, m$ , the conditional dist<sup>n</sup> of  $y_\alpha^{(1)}$  holding  $\underline{y}_\alpha^{(2)} = \underline{y}_\alpha^{(2)}$  fixed  
 i.e.  $N_k(\Sigma_{12} \Sigma_{22}^{-1} \underline{y}_\alpha^{(2)}, \Sigma_{11-2})$

Moreover, since  $\underline{y}_\alpha$  are independent, so are  $\underline{y}_\alpha^{(1)}$   
 disregarding of what conditions the corresponding  $\underline{y}_\alpha^{(2)}$  are subjected to.

The 'n' conditional dist<sup>n</sup> of  $\underline{y}_\alpha^{(1)}$  satisfy all the conditions of theorem (2) with  $n=m$

$$x = (\underline{y}_1^{(1)}, \underline{y}_2^{(1)}, \dots, \underline{y}_m^{(1)})$$

$$B = \Sigma_{12} \Sigma_{22}^{-1}$$

$$A = (-\underline{y}_1^{(2)}, -\underline{y}_2^{(2)}, \dots, -\underline{y}_n^{(2)}), \Sigma = \Sigma_{11-2}$$

so that we have,

$$H = \sum_{\alpha=1}^m \underline{y}_\alpha^{(2)} \underline{y}_\alpha'^{(2)} = w_{22}$$

$$G = -\underline{y}^{(1)} A' w_{22}^{-1} = w_{12} w_{22}^{-1}$$

$$xx' = \underline{y}^{(1)} \underline{y}^{(1)'} = w_{11}$$

$$GHG' = w_{12} w_{22}^{-1} w_{21}$$

$$xx' - GHG' = w_{11} - w_{12} w_{22}^{-1} w_{21} = w_{11-2}$$

Hence the proof.

NOTE : Conditional dist<sup>n</sup> of a Wishart sub-matrix given another Wishart sub-matrix is again Wishart.

Def<sup>n</sup> :- Let us call a Wishart  $W_p(\Sigma, m)$  dist<sup>n</sup>, a standard Wishart dist<sup>n</sup> and it is denoted by  $\Phi_{p,m}$ .

12/8/25

\* Derivation of Pdf of Wishart dist<sup>n</sup> :-

Theorem :- If  $W \sim W_p(\Sigma, m)$  then its joint pdf is

$$f(W) = |\Sigma|^{-m/2} |W|^{(m-p-1)/2} e^{\frac{-1}{2} \text{tr}(W \Sigma^{-1})} \frac{m^{m/2}}{2^{mp/2}} \prod_{i=1}^p \Gamma\left(\frac{m-i+1}{2}\right) \quad (1)$$

Proof :- As  $W$  is Wishart  $W_p(\Sigma, m)$  it is expressible as  $W = \sum_{\alpha=1}^m Y_{\alpha} Y_{\alpha}'$  where  $Y_{\alpha}$ 's are indep  $N_p(\underline{0}, \Sigma)$  random vectors, moreover  $\Sigma$  is p.d. we can find a n.s. lower triangular Matrix  $T$   $\ni$

$$T \Sigma T' = I \quad (2)$$

Then the random vectors

$Z = T Y_{\alpha} \quad (3)$ ,  $\alpha = 1, 2, \dots, m$  are where  $Y_{\alpha}$  independently and identically distributed as  $N_p(\underline{0}, \Sigma)$  So that the  $p \times p$  random matrix

$$\begin{aligned} \Phi_{p,m} &= TW\Gamma' = T \Sigma Y_{\alpha} Y_{\alpha}' T' \\ &= \Sigma (T Y_{\alpha})(T Y_{\alpha})' \\ &= \sum_{\alpha=1}^m Z_{\alpha} Z_{\alpha}' \quad (4) \end{aligned}$$

This has a standard Wishart dist<sup>n</sup> on m d.f.

$$E[\Phi] = E\left[\sum_{\alpha=1}^m z_\alpha z'_\alpha\right] = 0$$

$$\left( \because E\left[\sum_{\alpha=1}^m E(z_\alpha z'_\alpha)\right] = 0 \right)$$

$$D(\Phi) = D\left[\sum_{\alpha=1}^m z_\alpha z'_\alpha\right] = D\left[\sum_{\alpha=1}^m T y_\alpha y'_\alpha T'\right]$$

$$= T D(\sum y_\alpha y'_\alpha) T'$$

$$= T \Sigma T' \quad \left( \because D(\sum y_\alpha y'_\alpha) = \Sigma \right)$$

as  $\Sigma = \sum y_\alpha y'_\alpha \sim W_p(\Sigma, m)$

Hence  $\Phi_{pm} \sim W_p(\Sigma, I)$

Let  $\underline{u}_i'$  denote the  $i$ th row of  $p \times m$  matrix

$$Z = (z_1, z_2, \dots, z_m), \quad i=1, 2, \dots, p$$

Since  $\underline{u}_i \neq 0$  & a random orthogonal matrix  $P$   
 $\Rightarrow \underline{u}_i' P = (0, w_i), \quad w_i = \sqrt{\underline{u}_i' \underline{u}_i}, \quad \left. \quad \right\} \quad (5)$   
 $\underline{v}_i' P = (\underline{v}_i', w_i), \quad i = 1, 2, \dots, p$

then  $(p-1) \times (p-1)$  gram matrix  $\underline{v} \underline{v}'$  where  
 $\underline{v}' = (\underline{v}_1, \underline{v}_2, \dots, \underline{v}_p)$  is standard Wishart  
 $\Phi_{(p-1)(m-1)}$  matrix and

$$(a) |\Phi_{pm}| = w_i^2 \cdot |\Phi_{(p-1)(m-1)}| \quad (6)$$

Also since 'p' is orthogonal we have

$$\sum_{i=1}^p \underline{u}_i' \underline{u}_i = \sum_{i=1}^p w_i^2 + \sum_{i=1}^p \underline{v}_i' \bullet \underline{v}_i \quad (7)$$

$$\Rightarrow \sum_{j=1}^P w_j^2 = \text{tr} \sum_{\alpha=1}^m z_\alpha z_\alpha' - \text{tr} VV' \quad (\text{since } \sum u_i' u_i = \text{tr} z z') \\ = \text{tr } \Phi_{pm} = \text{tr } \Phi_{(p-1)(m-1)} \quad (8)$$

Moreover by our transformation (5)

(i) the r.v.'s  $w_1, w_2, \dots, w_p$  are independently  $N(0, 1)$  variates.

(ii)  $w_i^2 \sim \chi^2_{(m)}$  (since  $u_i \sim N_m(0, I)$ )

(iii)  $w_1, w_2, \dots, w_p$  and  $\Phi_{(p-1)(m-1)}$  are mutually independent.

Hence the joint pdf of  $w_1, w_2, \dots, w_p$  and

$\Phi_{(p-1)(m-1)}$  is

$$= \frac{1}{2^{m/2} \Gamma(m/2)} (w_1^2)^{(m-1)/2} e^{-w_1^2/2} \times \frac{1}{(2\pi)^{(p-1)/2}} e^{-\Phi_{(p-1)(m-1)}/2}$$

$$\times g_1\left(\frac{\Phi_{(p-1)(m-1)}}{2}\right) \quad (9)$$

where  $g_1\left(\frac{\Phi_{(p-1)(m-1)}}{2}\right)$  denotes the pdf of

$$\Phi_{(p-1)(m-1)}$$

Besides since  $V P$  is orthogonal by (5) we have

$$(10) \quad \Phi_{pm} = ZZ' = (Z_p)(Z_p)' \quad (10)$$

and hence  $\Phi_{ij}$  denotes the  $(i, j)^{\text{th}}$  element of  $\Phi_{pm}$ .

(ii) we have,

$$\Phi_{11} = w_1^2, \quad \Phi_{ii} = w_i w_i, \quad \text{whenever } i=j \text{ and } \\ = v_i v_j, \quad \text{whenever } i+j \geq 2$$

Now consider the inverse transformation from  $w_1^2, w_2, \dots, w_p, \Phi_{(p-1)(m-1)}$  to  $\Phi_{pm}$ . To evaluate the jacobian of the transformation we work out the partial differential table as shown below.

	$\Phi_{11}$	$\Phi_{12}$	$\Phi_{13} - \Phi_{1p}$	$\Phi_{22} - \dots - \Phi_{2p}$	$w_1^2 \rightarrow \Phi_{11}$
$w_1^2$	1	0	-	0	$w_1, w_2 \rightarrow \Phi_{12}$
$w_2$	0	$w_1$	-	0	$w_1, w_p \rightarrow \Phi_{1p}$
$w_3$	0	0	$w_1$	0	$w_{22} = v_2 v'_2 \rightarrow \Phi_{22}$
$w_p$	0	0	0	$w_1$	$w_{23} = v_2 v'_3 \rightarrow \Phi_{23}$
$v'_2 v_2$	0	0	0	-	$w_{2p} = v_2 v'_p \rightarrow \Phi_{2p}$
$v'_p v_p$	0	0	0	-	$w_{pp} = v_p v'_p \Rightarrow \Phi_{pp}$

Clearly, the jacobian of the transformation is

$$-(p-1) \cdot \frac{1}{|\mathcal{J}|}$$

The pdf of  $\Phi_{pm}$  is then

$$g(\Phi_{pm}) = \frac{\Phi_{11}^{(m-p-1)/2}}{2 \cdot m!_2 \sqrt{m!_2} (2\pi)^{(p-1)/2}} \left[ e^{-\frac{1}{2} (\text{tr } \Phi_{pm} - \text{tr } \Phi_{(p-1)(m-1)})} \cdot g_1(\Phi_{p-1 \times (m-1)}) \right] \quad (12)$$

(since  $\frac{1}{|\mathcal{J}|} = w_1^{-(p-1)}$  and since  $w_1 = \sqrt{\Phi_{11}} \Phi_{(m-p-1)/2}$ )

$$\Rightarrow \frac{1}{|\mathcal{J}|} = (\Phi_{11})^{\frac{-(p-1)}{2}} \xrightarrow{\text{since (9),}} (\omega_1^2)^{m/2-1} = (\Phi_{11})^{m/2-1}$$

But following (c) we have

$$\Phi_{ii} = \omega_i^2 = \frac{|\Phi_{pm}|}{|\Phi_{(p-1)\times(m-1)}|} \quad (13)$$

Hence substituting for  $\Phi_{ii}$  in (12)

$$\begin{aligned} & |\Phi_{pm}| g\left(\frac{\Phi_{pm}}{|\Phi_{(p-1)\times(m-1)}|}\right) e^{-\frac{t\pi\Phi_{pm}/2}{2}} \\ &= \frac{|\Phi_{(p-1)\times(m-1)}|}{2 \Gamma(m/2) (2\pi)^{(p-1)/2}} g_1\left(\frac{|\Phi_{(p-1)\times(m-1)}|}{|\Phi_{(p-2)\times(m-2)}|} e^{\frac{(m-p-1)/2}{2}}\right) \end{aligned} \quad (14)$$

$$(\therefore (\Phi_{ii})^{fm-p-1})_2 = \left( \frac{|\Phi_{pm}|}{|\Phi_{(p-2)\times(m-2)}|} e^{\frac{(m-p-1)/2}{2}} \right)$$

Similarly it can be shown that

$$\begin{aligned} & |\Phi_{(p-2)\times(m-2)}| g\left(\frac{\Phi_{(p-2)\times(m-2)}}{|\Phi_{(p-3)\times(m-3)}|} e^{\frac{t\pi\Phi_{(p-2)\times(m-2)}/2}{2}}\right) \\ &= \frac{|\Phi_{(p-2)\times(m-2)}|}{2 \Gamma(m-1/2) (2\pi)^{(p-2)/2}} \cdot g_2\left(\frac{|\Phi_{(p-2)\times(m-2)}|}{|\Phi_{(p-3)\times(m-3)}|} e^{\frac{(m-p-1)/2}{2}}\right) \end{aligned}$$

where  $g_2$  is the pdf of  $\Phi_{(p-3), m-3}$ .

Continuing in this way we get

$$(15) \quad \frac{1}{2} \left( \frac{|\Phi_{pm}|}{|\Phi_{(p-1)\times(m-1)}|} \right)^2 g\left(\frac{\Phi_{pm}}{|\Phi_{(p-1)\times(m-1)}|} e^{\frac{t\pi\Phi_{pm}/2}{2}}\right)$$

$$= \frac{1}{2} \left( \frac{m/2 + (m-1)/2 + \dots + (n-p-1)/2}{(2\pi)^{2/2}} \right)^2 \frac{\frac{(n-1)/2 + (p-2)/2 + \dots + 1/2}{(2\pi)^{2/2}}}{\frac{(m-1)/2 + (m-2)/2 + \dots + (n-p-1)/2}{(2\pi)^{2/2}}} \frac{(m-1)!! (n-p-1)!!}{(m/2)!! ((n-p)/2)!!}$$

which gives

$$g(\Phi_{pm}) = \frac{|\Phi_{pm}|}{\frac{mp/2}{2} \pi^{\frac{p(p-1)/4}{2}} \prod_{i=1}^p \Gamma\left(\frac{m-i-1}{2}\right)}$$

~~18/25~~

- Finally let us consider the inverse transformation from  $T_p$  to  $w$ . Recalling that  $\Phi_{pm} = TW\Gamma'$  and  $T$  is lower triangular and denoting by  $t_{ij}$  and  $w_{ij}$  the  $(i, j)^{th}$  element of  $T$  and  $w$  respectively we have

$$\Phi_{ij} = \sum_{s < i} \sum_{s < j} t_{is} w_{is} t_{sj} \quad (17)$$

Now the partial differential table for the said transformation is as follows :-

	$\phi_{11}$	$\phi_{12}$	$\dots$	$\phi_{1p}$	$\phi_{22}$	$\dots$	$\phi_{2p}$
$w_{11}$	$t_{11}^2$	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$w_{12}$	-	$t_{11}t_{12}$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
!	:	0	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$w_{1p}$	0	0	$\dots$	$t_{11}t_{pp}$	$\dots$	$\dots$	$\dots$
$w_{22}$	0	0	$\dots$	$\dots$	$t_{22}^2$	$\dots$	$\dots$
!	:	1	$\dots$	$\dots$	:	$\dots$	$\dots$
!	:	1	$\dots$	$\dots$	:	$\dots$	$\dots$
$w_{2p}$	0	0	$\dots$	$\dots$	$\dots$	$\dots$	$t_{pp}^2$

The required Jacobian transformation is

$$J = (t_{11}, t_{12}, \dots, t_{pp})^{p+1}$$

$$= \left| \sum \right|^2 \quad (18) \quad (\text{by (2)})$$

Hence by (16) the pdf of  $w$  is

$$\left( |T|^2 |w| \right)^{(m-p-1)/2} e^{-\operatorname{tr} w T' \Sigma^{-1} T} \frac{-(p+1)}{2}$$

$$\frac{mp}{2^{\frac{m}{2}} \pi^{\frac{p(p-1)}{4}}} \prod_{i=1}^p \left| \left( \frac{m-i+1}{2} \right) \right|$$

Since  $T' T = \Sigma^{-1} |T|^2 = |\Sigma|^{\frac{1}{2}}$  and  $\operatorname{tr} w T' T = \operatorname{tr} w \Sigma^{-1}$

$$= |\Sigma| (w) \frac{(m-p-1)}{2} e^{-\frac{1}{2} \operatorname{tr} w \Sigma^{-1}}$$

$$\frac{mp}{2^{\frac{m}{2}} \pi^{\frac{p(p-1)}{4}}} \prod_{i=1}^p \left| \left( \frac{m-i+1}{2} \right) \right|$$

Final result obtained without involving unitary soft word  
conversion or any intermediate.

$$\begin{aligned} & q_1 = \phi, q_2 = \phi, q_3 = \phi, q_4 = \phi \\ & q_5 = \phi, q_6 = \phi, q_7 = \phi, q_8 = \phi \\ & q_9 = \phi, q_{10} = \phi, q_{11} = \phi, q_{12} = \phi \\ & q_{13} = \phi, q_{14} = \phi, q_{15} = \phi, q_{16} = \phi \\ & q_{17} = \phi, q_{18} = \phi, q_{19} = \phi, q_{20} = \phi \end{aligned}$$

Final result obtained without involving soft word conversion or any intermediate.

Final result obtained without involving soft word conversion or any intermediate.

Final result obtained without involving soft word conversion or any intermediate.

# UNIT - 3

PAGE NO.:  
DATE: / /

## \* Discrimination and classification :-

- Discrimination analysis and classification are multivariate technique concerned with separating different or distinct set of object ( $or obs^n$ ) and with allocating new object ( $or obs^n$ ) to a previously defined groups.
- Discrimination analysis is rather exploratory in nature as a separatory procedure. It is often employed on a basis in order to investigate observed differences when causal relationship are not well understood.
- Classification procedure are less exploratory in the same that they need to well define rules which can be used for assigning objects. Classification ordinarily requires more problem structure than discrimination.

## \* Goal of discrimination :-

- (a) To describe either graphically (in  $n$  there are fewer dimensions) or algebraically the differential features of  $n$  objects ( $obs^n$ ) from several known collections ( $pop^n$ ).
- (b) Try to find discrimination whose numerical values are such that collection are separated as much as possible.

## \* Goal of classification :-

- (a) To sort objects ( $obs^n$ ) into two or more labelled classes, the emphasis is on deriving a rule that can be used to optimally assign a new object to the labelled classes.

## \* Separation and classification for two pop :-

- We may be interested in finding out:-

- (1) Separating two classes of objects or
- (2) Assigning a new object to one of the two classes (or both).

Let the classes be labelled as  $J_1$  and  $J_2$

The objects are ordinarily separated or classified

on the basis of the measurements for e.g.:-

$$\text{associated w.r.t. } \underline{x} = [x_1, x_2, \dots, x_p]$$

- The observed value of  $x$  (matrix) differ to some extent from one class to another.

- Let the measured variables  $x$  being the pop<sup>n</sup> for 1<sup>st</sup> class as  $J_1$  and those from 2<sup>nd</sup> class be  $J_2$ . This two pop<sup>n</sup> can be described by pdf  $f_1(x)$  and  $f_2(x)$  and consequently we can talk of assigning obs<sup>n</sup> to pop<sup>n</sup> or objects to classes interchangeably.

19/8/25

PAGE NO.:

DATE: 11

Classes:

Pop<sup>n</sup> J<sub>1</sub> and J<sub>2</sub>

Measured variables

(1) Successful or unsuccessful (Scores, H<sub>avg</sub>, grade (fail or graduate

(2) Male and females - Anthropological measurements take circumference and

(3) Good and Poor Income, age, family

Credit risks Size, Cibil's score, etc.

We shall concentrate on classification for two

Allocation or classification rules are usually developed from learning samples.

Measured characteristic randomly selected object

Known to come from two populations are examined and difference

- Essentially set of all possible sample

outcome is divided into two regions R<sub>1</sub> and R<sub>2</sub>

→ If a new obs<sup>n</sup> falls in R<sub>1</sub> it is allocated

to J<sub>1</sub> and if it falls in R<sub>2</sub> we allocated

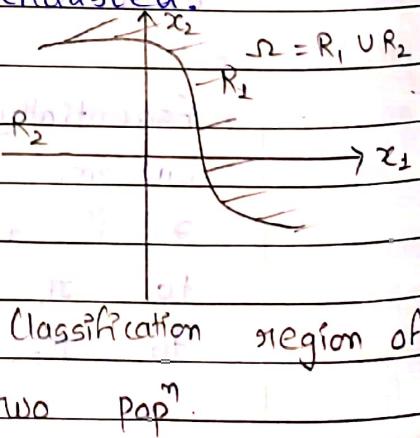
to J<sub>2</sub>.



Scanned with OKEN Scanner

- Thus one set of observed values favours  $\pi_1$ , the other set of value favour  $\pi_2$ .
- A good classification procedure should result in few misclassification. In other words, the chances or probability of misclassification should be small.
- Let  $f_1(\underline{x})$  and  $f_2(\underline{x})$  be the pdf associated with  $p \times 1$  random vector variable  $\underline{x}$  for the pop<sup>n</sup>  $\pi_1$  and  $\pi_2$  respectively.
- An object with associated measurement  $\underline{x}$  must be assigned to either  $\pi_1$  or  $\pi_2$ .
- Let  $\Omega$  be the sample space i.e. the collection of all possible obs  $\underline{x}$ .
- Let  $R_1$  be that set of  $\underline{x}$  values for which we classify objects as  $\pi_1$  and  $R_2 = \Omega - R_1$  be the remaining  $\underline{x}$  values for which we classify objects as  $\pi_2$ .
- Since every object must be assigned to one and only one of two pop<sup>n</sup>, the set  $R_1$  and  $R_2$  are mutually exclusive and exhausted.

The conditional prob.  $P(z|z)$  of classifying an object as  $\pi_2$  when in fact it is from  $\pi_1$



$$P(2|z) = P(x \in R_2 | \pi_1)$$

$$= \int_{R_2} f(x) dx$$

$$R_2 = \Omega - R_1$$

Similarly,

$$P(z|2) = P(x \in R_1 | \pi_2)$$

$$= \int_{R_1} f_2(x) dx$$

- Let  $P_1$  be the prior probability of  $\pi_1$ , and  $P_2$  be the prior probability of  $\pi_2$ , where  $P_1 + P_2 = 1$

- The overall probability of correctly or incorrectly classify object can be derive as the product of prior and conditional classification probability.

$\rightarrow P(\text{Correctly classification in } \pi_1)$

$$= P(\text{obs. comes from } \pi_1 \text{ and is correctly classified as } \pi_1)$$

$$= P(x \in R_1 | \pi_1) \cdot P(\pi_1)$$

$$= P(\bar{z}|z) \cdot P_1$$

$\rightarrow P(\text{Misclassified as } \pi_1)$

$$= P(\text{obs. comes from } \pi_2 \text{ and is misclassified as } \pi_1)$$

$$= P(x \in R_1 | \pi_2) \cdot P(\pi_2)$$

$$= P(\bar{z}|z) \cdot P_2$$

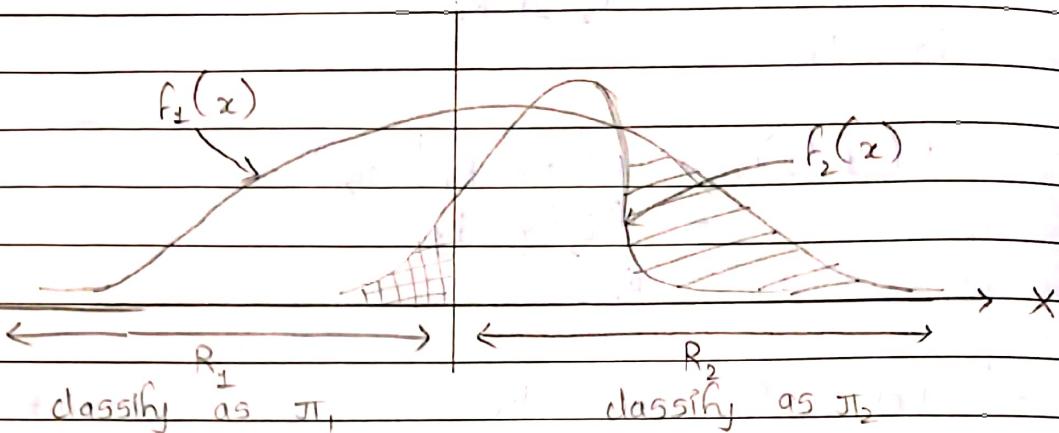
$\rightarrow P(\text{Correctly classified as } \pi_2)$

$$= P(x \in R_2 | \pi_2) \cdot P(\pi_2)$$

$$= P(z|\bar{z}) \cdot P_2$$

$$\rightarrow P(\text{misclassified as } \pi_2) = P(x \in R_2 | \pi_2) \cdot P(\pi_1)$$

$$= P(\bar{2} | \pm) \cdot P_1$$



- Classification skills are often evaluated in terms of their misclassification probability but it ignores misclassification cost.
- The cost of misclassification can be defined by a cost matrix.

		Classify as	
		$\pi_1$	$\pi_2$
True pop	$\pi_1$	0	$c(2 \pm)$
	$\pi_2$	$c(\pm 2)$	0

Expected cost of misclassification ECM  
 $= c(2|\pm)P(2|\pm)P_1 + c(\pm|2)P(\pm|2) \cdot P_2$

- A reasonable classification rule should have minimum ECM.

Result: The regions  $R_1$  and  $R_2$  that minimise the ECM are defined by the values  $x$  for which the following inequalities hold

$$R_1 : \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \left( \frac{c(\pm 1_2)}{c(2|\pm)} \right) \cdot \left( \frac{P_2}{P_1} \right)$$

density ratio  $\geq$  (cost ratio) (prior prob. ratio)

$$R_2 : \frac{f_1(\underline{x})}{f_2(\underline{x})} \leq \left( \frac{c(\pm 1_2)}{c(2|\pm)} \right) \left( \frac{P_2}{P_1} \right)$$

\* Special cases for minimum expected cost regions.

(a)  $P_2 = \pm$  (equal prior prob.)

$$R_1 : \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq c(\pm 1_2), R_2 : \frac{f_1(\underline{x})}{f_2(\underline{x})} \leq c(\pm 1_2)$$

(b)  $c(\pm 1_2) = \pm$  (equal misclassification cost)

$$R_1 : \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \frac{P_2}{P_1}, R_2 : \frac{f_1(\underline{x})}{f_2(\underline{x})} \leq \frac{P_2}{P_1}$$

(c)  $\frac{P_2}{P_1} = \pm = c(\pm 1_2)$

$$R_1 : \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \pm, R_2 : \frac{f_1(\underline{x})}{f_2(\underline{x})} \leq \pm$$

- One might ignore the cost of misclassification and choose  $R_1$  and  $R_2$  to minimise the total Prob. of misclassification.

Total Prob. of misclassification = TPM

=  $P(\text{misclassifying a } \pi_1 \text{ obs} \text{ or misclassifying a } \pi_2 \text{ obs})$

$$= P(\text{obs comes from } \pi_1 \text{ and is misclassified}) + \\ P(\text{obs. comes from } \pi_2 \text{ and is misclassified}) \\ = P_1 \int_{R_2} f_1(x) dx + P_2 \int_{R_1} f_2(x) dx$$

- This Problem is equivalent to minimise the expected cost of misclassification when the cost of misclassification are equal which is same as decision (b).
- Problem of allocation :- We could allocate a new obs<sup>n</sup>  $\underline{x}_0$  to the pop<sup>n</sup> with the largest 'Posterior Prob.' (Classification new Bayes rule) :-

By Bayes rule, the posterior prob. are

$$P(\pi_1 / \underline{x}_0) = \frac{P(\pi_1 \text{ occurs and observe } \underline{x}_0)}{P(\text{observing } \underline{x}_0)}$$

$$= P(\text{observe } \underline{x}_0 / \pi_1) \cdot P(\pi_1)$$

$$= P(\text{observe } \underline{x}_0 / \pi_1) \cdot P(\pi_1) + P(\text{observe } \underline{x}_0 / \pi_2) P(\pi_2)$$

$$= P_1 f_1(\underline{x}_0)$$

$$\frac{P_1 f_1(\underline{x}_0)}{P_1 f_1(\underline{x}_0) + P_2 f_2(\underline{x}_0)}$$

Similarly,

$$P(\pi_2 / \underline{x}_0) = 1 - P(\pi_1 / \underline{x}_0)$$

$$= P_2 f_2(\underline{x}_0)$$

$$\frac{P_2 f_2(\underline{x}_0)}{P_1 f_1(\underline{x}_0) + P_2 f_2(\underline{x}_0)}$$

- Sample  $\underline{x}_0$  will be classified in that class which has higher posterior prob.

- This is equivalent to decision rule (b).

\* Classification with two multivariate normal pop<sup>n</sup>

→ When  $\Sigma_1 = \Sigma_2 = \Sigma$  (known)

Assume the joint densities of  $\underline{x} = [x_1, x_2, \dots, x_p]$  for the two pop<sup>n</sup>  $\pi_1$  and  $\pi_2$  are given by

$$f_i(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (\underline{x} - \underline{\mu}_i)' \Sigma^{-1} (\underline{x} - \underline{\mu}_i) \right] \text{ for } i=1,2$$

we assume pop<sup>n</sup> parameters  $\underline{\mu}_1, \underline{\mu}_2$  &  $\Sigma$  are known

$$R_1 : \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \left( \frac{c(1/2)}{c(2/1)} \right) \left( \frac{p_2}{p_1} \right)$$

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} = \exp \left[ -\frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma^{-1} (\underline{x} - \underline{\mu}_1) + \frac{1}{2} (\underline{x} - \underline{\mu}_2)' \Sigma^{-1} (\underline{x} - \underline{\mu}_2) \right] \geq \left( \frac{c(1/2)}{c(2/1)} \right) \left( \frac{p_2}{p_1} \right)$$

$$R_2 : \frac{f_1(\underline{x})}{f_2(\underline{x})} = \exp \left[ -\frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma^{-1} (\underline{x} - \underline{\mu}_1) + \frac{1}{2} (\underline{x} - \underline{\mu}_2)' \Sigma^{-1} (\underline{x} - \underline{\mu}_2) \right] < \left( \frac{c(1/2)}{c(2/1)} \right) \left( \frac{p_2}{p_1} \right)$$

Given the regions  $R_1$  and  $R_2$  above we can construct the following classification rule

- taking log on both sides and simplifying the LHS, we get

$$\begin{aligned}
 & -\frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma^{-1} (\underline{x} - \underline{\mu}_1) + \frac{1}{2} (\underline{x} - \underline{\mu}_2)' \Sigma^{-1} (\underline{x} - \underline{\mu}_2) \\
 &= -\frac{1}{2} \left[ (\underline{x} - \underline{\mu}_1)' \Sigma^{-1} (\underline{x} - \underline{\mu}_1) - (\underline{x} - \underline{\mu}_2)' \Sigma^{-1} (\underline{x} - \underline{\mu}_2) \right] \\
 &= -\frac{1}{2} \left[ \underline{x}' \Sigma^{-1} \underline{x} - \underline{x}' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_1' \Sigma^{-1} \underline{x} + \underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1 - \underline{x}' \Sigma^{-1} \underline{x} \right. \\
 &\quad \left. + \underline{x}' \Sigma^{-1} \underline{\mu}_2 + \underline{\mu}_2' \Sigma^{-1} \underline{x} - \underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2 \right] \\
 &= -\frac{1}{2} \left[ -2 \underline{\mu}_1' \Sigma^{-1} \underline{x} + 2 \underline{\mu}_2' \Sigma^{-1} \underline{x} + \underline{\mu}_1' \Sigma^{-1} \underline{\mu}_1 - \underline{\mu}_2' \Sigma^{-1} \underline{\mu}_2 \right] \\
 &= (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 + \underline{\mu}_2)
 \end{aligned}$$

and consequently

$$\begin{aligned}
 R_1 : & (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 + \underline{\mu}_2) \\
 & \geq \ln \left[ \frac{c(1|2)}{c(2|1)} \left( \frac{P_2}{P_1} \right) \right] \\
 R_2 : & (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{x} - \frac{1}{2} (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 + \underline{\mu}_2) \\
 & \leq \ln \left[ \frac{c(1|2)}{c(2|1)} \left( \frac{P_2}{P_1} \right) \right]
 \end{aligned} \tag{3}$$

- Above is the minimum ECM (Expected Cost Minimization) Classification Rule. In most practical situations, the popn. quantities  $\underline{\mu}_1, \underline{\mu}_2$  and  $\Sigma$  are unknown. So, the rule must be modified.

- When  $\underline{\mu}_1, \underline{\mu}_2$  and  $\Sigma$  are unknown

Suppose we have  $n_1$  obs<sup>ns</sup> of multivariate  
g.v.  $\underline{x}' = [\underline{x}_1, \underline{x}_2, \dots, \underline{x}_p]$  from  $\pi_1$  and  $n_2$   
measurements of this quantity from  $\pi_2$  with  
 $n_1 + n_2 - 2 \geq p$

The resp. data matrices are

$$\underline{x}_1(p \times n_1) = [\underline{x}_{11}, \underline{x}_{12}, \dots, \underline{x}_{1n_1}]$$

$$\underline{x}_2(p \times n_2) = [\underline{x}_{21}, \underline{x}_{22}, \dots, \underline{x}_{2n_2}]$$

From these data matrices the sample Mean  
vector and Cov. Matrices are determined by

$$\bar{\underline{x}}_1(p \times 1) = \frac{1}{n_1} \sum_{j=1}^{n_1} \underline{x}_{1j}$$

$$S_1(p \times p) = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (\underline{x}_{1j} - \bar{\underline{x}}_1)(\underline{x}_{1j} - \bar{\underline{x}}_1)'$$

$$\bar{\underline{x}}_2(p \times 1) = \frac{1}{n_2} \sum_{j=1}^{n_2} \underline{x}_{2j}$$

$$S_2(p \times p) = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (\underline{x}_{2j} - \bar{\underline{x}}_2)(\underline{x}_{2j} - \bar{\underline{x}}_2)'$$

Since it is assumed that parent pop<sup>n</sup> have  
the same covariance matrices for ex. I,  
the same covariance matrices  $S_1$  and  $S_2$  are  
combined to derive a single unbiased estimate  
of  $\Sigma$ .

$$S_{\text{pool}} = \left[ \frac{n_1 - 1}{(n_1 - 1) + (n_2 - 1)} \right] S_1 + \left[ \frac{n_2 - 1}{(n_1 - 1) + (n_2 - 1)} \right] S_2$$

$$= \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

is an unbiased estimate of  $\Sigma$  if the data matrices  $\underline{x}_1$  and  $\underline{x}_2$  contain random samples from the pop<sup>n</sup>  $\pi_1$  and  $\pi_2$  respectively.

- Substituting  $\bar{\underline{x}}_1$  for  $\mu_1$ ,  $\bar{\underline{x}}_2$  for  $\mu_2$  and  $S_{\text{pool}}$  for  $\Sigma$  gives sample classification rule.

- The estimated minimum rule for two normal pop<sup>n</sup> is to allocate  $\underline{x}_0$  to  $\pi_1$  if

$$(\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' S_{\text{pool}}^{-1} \underline{x}_0 - \frac{1}{2} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' S_{\text{pool}}^{-1} (\bar{\underline{x}}_1 + \bar{\underline{x}}_2) > 0$$

$$\Rightarrow \ln \left[ \frac{c(1/2) P_2}{c(21_1) P_1} \right] > 0 \quad (4)$$

Allocate  $\underline{x}_0$  (to  $\pi_2$ ) O.W.

Remark

\* particular case :-

$$\text{If in above rule } \frac{c(1/2) P_2}{c(21_1) P_1} = 1$$

$\Rightarrow \ln(1) = 0$  & the estimated min. ECM rule for two normal pop<sup>n</sup> amounts to comparing scalar variable

$$Y_1 = (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' S_{\text{pool}}^{-1} \underline{x}_0 = \hat{\underline{l}}' \underline{x}$$

evaluated at  $\underline{x}_0$  with the no.

$$\hat{m} = \frac{1}{2} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' S_{\text{pool}}^{-1} (\bar{\underline{x}}_1 + \bar{\underline{x}}_2)$$

$$= \frac{1}{2} (\bar{y}_1 + \bar{y}_2)$$

where

$$\bar{y}_1 = (\bar{x}_1 - \bar{x}_2)' S_{pool}^{-1} \bar{x}_1 = \hat{1}' \bar{x}_1$$

and

$$\bar{y}_2 = (\bar{x}_1 - \bar{x}_2)' S_{pool}^{-1} \bar{x}_2 = \hat{0}' \bar{x}_2$$

That is, the estimated minimum ECM rule for two normal pop<sup>n</sup> amounts to creating two univariate pop<sup>n</sup> for the  $y_i$  values by taking an appropriate linear combination of the obs<sup>n</sup> from the pop<sup>n</sup>  $\pi_1$  and  $\pi_2$  and then assigning a new obs<sup>n</sup>  $\bar{x}_0$  to  $\pi_1$  and  $\pi_2$  i.e.

$$\text{Assign } \bar{x}_0 \text{ to } \pi_1 \text{ if } y_0 \geq m \\ \bar{x}_0 \text{ to } \pi_2 \text{ if } y_0 < m \quad (5)$$

Remark → Once the parameter estimates are used for unknown pop<sup>n</sup> quantities, there is no assurance that the resulting rule will minimize the ECM in a particular application because optimal rule was derived assuming  $f_1(\underline{x})$  and  $f_2(\underline{x})$  are completely known

case →

$\Sigma_1 \neq \Sigma_2$  (pop<sup>n</sup> parameters known)

$$R_1 : \frac{f_1(\underline{x})}{f_2(\underline{x})} \geq \frac{c(1/2)}{c(2/1)} \left( \frac{p_2}{p_1} \right)$$

$$f_1(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma_1|^{1/2}} \exp \left[ -\frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) \right]$$

$$f_2(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma_2|^{1/2}} \exp \left[ -\frac{1}{2} (\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2) \right]$$

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} = \frac{|\Sigma_2|^{1/2}}{|\Sigma_1|^{1/2}} \exp \left[ -\frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) + \frac{1}{2} (\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2) \right]$$

taking log

$$\Rightarrow -\frac{1}{2} (\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) + \frac{1}{2} (\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2) + \ln \left( \frac{|\Sigma_2|}{|\Sigma_1|} \right)^{1/2}$$

$$\Rightarrow -\frac{1}{2} \left[ (\underline{x} - \underline{\mu}_1)' \Sigma_1^{-1} (\underline{x} - \underline{\mu}_1) - (\underline{x} - \underline{\mu}_2)' \Sigma_2^{-1} (\underline{x} - \underline{\mu}_2) \right] - \frac{1}{2} \ln \left( \frac{|\Sigma_2|}{|\Sigma_1|} \right)$$

$$\Rightarrow -\frac{1}{2} \left[ \underline{x}' \Sigma_1^{-1} \underline{x} - \underline{x}' \Sigma_1^{-1} \underline{\mu}_1 - \underline{\mu}_1' \Sigma_1^{-1} \underline{x} + \underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1 \right.$$

$$\quad \left. - \underline{x}' \Sigma_2^{-1} \underline{x} + \underline{x}' \Sigma_2^{-1} \underline{\mu}_2 + \underline{\mu}_2' \Sigma_2^{-1} \underline{x} - \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2 \right]$$

$$- \frac{1}{2} \ln \left( \frac{|\Sigma_1|}{|\Sigma_2|} \right)$$

$$\Rightarrow -\frac{1}{2} \left[ \underline{x}' (\Sigma_1^{-1} - \Sigma_2^{-1}) \underline{x} + 2 \underline{\mu}_1' \Sigma_1^{-1} \underline{x} + 2 \underline{\mu}_2' \Sigma_2^{-1} \underline{x} - \underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1 \right.$$

$$\quad \left. - \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2 \right] - \frac{1}{2} \ln \left( \frac{|\Sigma_1|}{|\Sigma_2|} \right)$$

$$\Rightarrow -\frac{1}{2} \left[ \underline{x}' (\Sigma_1^{-1} - \Sigma_2^{-1}) \underline{x} + (\underline{\mu}_1' \Sigma_1^{-1} - \underline{\mu}_2' \Sigma_2^{-1}) \underline{x} - \frac{1}{2} \ln \left( \frac{|\Sigma_1|}{|\Sigma_2|} \right) \right.$$

$$\quad \left. - \frac{1}{2} (\underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1 - \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2) \right]$$

$$\Rightarrow R_1 : -\frac{1}{2} \underline{x}' (\Sigma_1^{-1} - \Sigma_2^{-1}) \underline{x} + (\underline{\mu}_1' \Sigma_1^{-1} - \underline{\mu}_2' \Sigma_2^{-1}) \underline{x} - K$$

$$\geq \ln \left( \frac{c_{1/2} P_2}{c_{2/1} P_1} \right)$$

$$\text{where } K = \frac{1}{2} \ln \left( \frac{|\Sigma_1|}{|\Sigma_2|} \right) + \frac{1}{2} (\underline{\mu}_1' \Sigma_1^{-1} \underline{\mu}_1 - \underline{\mu}_2' \Sigma_2^{-1} \underline{\mu}_2)$$

- The classification regions are defined by the quadratic function of  $\underline{x}$  when the ~~over~~  
 $\Sigma_1 = \Sigma_2 = \Sigma$  the quadratic ~~term~~ term

$$-\frac{1}{2} \underline{x}' (\Sigma_1^{-1} - \Sigma_2^{-1}) \underline{x} \rightarrow \text{disappears.}$$

- Allocate  $\underline{x}_0$  to  $\pi_1$  if  
 $-\frac{1}{2} \underline{x}_0' (\Sigma_1^{-1} - \Sigma_2^{-1}) \underline{x}_0 + (\underline{\mu}_1' \Sigma_1^{-1} - \underline{\mu}_2' \Sigma_2^{-1}) \underline{x}_0 - K$

$$\ln \left[ \frac{C(1|2)}{C(2|1)} P_2 \right] \geq \ln \left[ \frac{C(1|2)}{C(2|1)} P_1 \right]$$

O.w. allocate  $\underline{x}_0$  to  $\pi_2$

#### \* Fisher's Discriminant Function:

→ Fisher's idea was to transform the multivariate obs<sup>n</sup>  $\underline{x}$  to the univariate obs<sup>n</sup>  $y$  such that  $y$ 's derived from pop<sup>n</sup>  $\pi_1$  and  $\pi_2$  were separated as much as possible.

→ Fisher suggested taking linear combination of  $\underline{x}$  to create  $y$  (because they are simple functions of  $\underline{x}$  to be handled easily)

→ Fisher's approach does not assume that pop<sup>n</sup> are normal, it assumes that pop<sup>n</sup> covariances are equal because a pooled estimate of the common covariance matrix is used.

→ A fixed linear combination of the  $\underline{x}$ 's takes

the values  $y_{11}, y_{12}, \dots, y_{1n_1}$  for the obs<sup>n</sup> from the 1<sup>st</sup> pop<sup>n</sup> and the values  $y_{21}, y_{22}, \dots, y_{2n_2}$  for the obs<sup>n</sup> from the 2<sup>nd</sup> pop<sup>n</sup>.

- The separation of this two sets of univariate  $y$ 's is assessed in terms of the difference bet<sup>n</sup>  $\bar{y}_1$  and  $\bar{y}_2$  expressed in standard deviation (s.d) units.

$$\text{Separation} = \frac{|\bar{y}_1 - \bar{y}_2|}{S_y}$$

where

$$S_y^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_1 + n_2 - 2}$$

is the pooled estimate of the variance.

- The objective is to select the linear combination of  $\underline{x}$  to achieve maximum separation of the sample means  $\bar{y}_1$  and  $\bar{y}_2$ .

Result: The linear combination

$$Y = \underline{b}' \underline{x} = (\bar{y}_1 - \bar{y}_2)' S_{\text{pool}}^{-1} \underline{x}$$

maximizes the ratio

$$\left( \text{Squared distance between bet' sample means of } Y \right) = \frac{(\bar{y}_1 - \bar{y}_2)^2}{S_y^2}$$

Sample variance of  $y$

$$= (\hat{\underline{l}}' \bar{\underline{x}}_1 - \hat{\underline{l}}' \bar{\underline{x}}_2)^2$$

$$= \hat{\underline{l}}' S_{\text{pool}}^{-1} \hat{\underline{l}}$$

$$= (\hat{\underline{l}}' \underline{d})^2 \quad (*)$$

$$= \frac{(\hat{\underline{l}}' \underline{S}_{\text{pool}}^{-1} \hat{\underline{l}})^2}{(\hat{\underline{l}}' \underline{S}_{\text{pool}}^{-1} \hat{\underline{l}})}$$

over all possible coeff. vector  $\hat{l}$ ,  $d = (\bar{x}_1 - \bar{x}_2)$

The max of the ratio is  $D^2 = (\bar{x}_1 - \bar{x}_2)' S_{\text{pool}}^{-1} (\bar{x}_1 - \bar{x}_2)$

For proof we use maximization lemma,

Let  $B_{p \times p}$  be positive definite and  $d_{px1}$  be a given vector then for any arbitrary non-zero vector  $x_{px1}$   $\max_{x \neq 0} \frac{(x'd)^2}{(x'Bx)} = d'B^{-1}d$  with the maximum attained when  $x_{px1} = C B^{-1}d$  for any constant  $C \neq 0$

The max of the ratio  $\max(*)$  is given by applying above lemma directly.

Thus getting  $d = (\bar{x}_1 - \bar{x}_2)$

$$\max_{\hat{l}} \frac{(\hat{l}'d)^2}{(\hat{l}'S_{\text{pool}}\hat{l})} = d'S_{\text{pool}}^{-1}d$$

where  $D^2$  is the sample squared distance bet' the two means

$$S_y^2 \text{ in } (*) \text{ can be calculated as:}$$

$$S_y^2 = \sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2 + \sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2$$

with

$$y_{1j} = \hat{l}' \bar{x}_{1j}$$

$$y_{2j} = \hat{l}' \bar{x}_{2j}$$

- An allocation rule based on Fisher's Discriminant function is

Allocate  $\underline{x}_0$  to  $JT_1$  if  $y_0 = (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' \bar{S}_{pool}^{-1} \underline{x}_0 \geq \hat{m} = \frac{1}{2} (\bar{\underline{x}}_1 - \bar{\underline{x}}_2)' \bar{S}_{pool}^{-1} (\bar{\underline{x}}_1 + \bar{\underline{x}}_2)$

O.w. Allocate  $\underline{x}_0$  to  $JT_2$ .

### \* Canonical Correlation Analysis (CCA)

- CCA used to identify and quantify the association between two set of variables.
- CCA focuses on the correlation bet' a linear combination of the variables in one set and a linear combination of the variables in the other set.
- The idea is
  - to determine the pair of linear combination having the largest correlation.

we determine the pairs of linear combinations having the largest correlation among all pairs uncorrelated with the initially selected pair.

Then the process is continuous.

The pairs of linear combinations are called Canonical variables and their correlations are called Canonical correlation. The canonical correlation measure the strength of association bet' two set of variables.

- The maximization aspect of the technique represent an attempt to concentrate a high-dimensional relationship between two sets of variables into a few pairs of Canonical variables.

\* Canonical variates and canonical correlations :-

- We shall measure the association bet<sup>n</sup> two groups of variables.
- The 1<sup>st</sup> group of 'p' variates variables is represented by ( $p \times 1$ ) random vectors  $\underline{x}^{(1)}$  and second group of 'q' variables is represented by ( $q \times 1$ ) random vectors  $\underline{x}^{(2)}$ .
- $\underline{x}^{(1)}$  represent the smaller set  $q \leq p$
- For the random vectors  $\underline{x}^{(1)}$  and  $\underline{x}^{(2)}$  let
 
$$E[\underline{x}^{(1)}] = \underline{\mu}^{(1)}$$

$$E[\underline{x}^{(2)}] = \underline{\mu}^{(2)}$$

$$\text{Var} - \text{Cov}(\underline{x}^{(1)}) = \Sigma_{11}$$

$$\text{Var}(\text{Cov}(\underline{x}^{(2)})) = \Sigma_{22}$$

$$\text{Cov}(\underline{x}^{(1)}, \underline{x}^{(2)}) = \Sigma_{12} = \Sigma_{21}$$

$$\text{let } \underline{x}_{(p+q) \times 1} = \begin{bmatrix} \underline{x}^{(1)} \\ \vdots \\ \underline{x}^{(2)} \end{bmatrix} = \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \vdots^{(1)} \\ x_p^{(1)} \\ \hline x_1^{(2)} \\ \vdots^{(2)} \\ x_q^{(2)} \end{bmatrix}$$

has mean vector

$$\underline{u}_{(p+q) \times 1} = E(\underline{x}) = \begin{bmatrix} E(\underline{x}^{(1)}) \\ E(\underline{x}^{(2)}) \end{bmatrix} = \begin{bmatrix} \underline{\mu}^{(1)} \\ \underline{\mu}^{(2)} \end{bmatrix}$$

and Cov Matrix

$$\Sigma = E[(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})']$$

$$= \begin{bmatrix} E(\underline{x}^{(1)} - \underline{\mu}^{(1)})(\underline{x}^{(1)} - \underline{\mu}^{(1)})' & E(\underline{x}^{(1)} - \underline{\mu}^{(1)})(\underline{x}^{(2)} - \underline{\mu}^{(2)})' \\ \vdots & \vdots \\ E(\underline{x}^{(2)} - \underline{\mu}^{(2)})(\underline{x}^{(1)} - \underline{\mu}^{(1)})' & E(\underline{x}^{(2)} - \underline{\mu}^{(2)})(\underline{x}^{(2)} - \underline{\mu}^{(2)})' \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_{11}(p \times p) & \Sigma_{12}(p \times q) \\ \vdots & \vdots \\ \Sigma_{21}(q \times p) & \Sigma_{22}(q \times q) \end{bmatrix}$$

- The main task of CCA is to summarise the association between  $\underline{x}^{(1)}$  and  $\underline{x}^{(2)}$  sets in terms of  $q$  carefully chosen covariances or correlation rather than  $pq$  covariances in  $\Sigma_2$ .

- Linear Combination provide simple summary measures of a set of variables.

$$u = \underline{a}' \underline{x}^{(1)}$$

$$v = \underline{b}' \underline{x}^{(2)}$$
 for some pair of coefficient vectors  $\underline{a}$  and  $\underline{b}$ .

$$\text{Var}(u) = \text{Var}(\underline{a}' \underline{x}^{(1)}) = \underline{a}' \Sigma_{11} \underline{a}$$

$$\text{Var}(v) = \text{Var}(\underline{b}' \underline{x}^{(2)}) = \underline{b}' \Sigma_{22} \underline{b}$$

$$\text{Cov}(u, v) = \text{Cov}(\underline{a}' \underline{x}^{(1)}, \underline{b}' \underline{x}^{(2)}) = \underline{a}' \text{Cov}(\underline{x}^{(1)}, \underline{x}^{(2)}) \underline{b}$$

$$= \underline{a}' \Sigma_{12} \underline{b}$$

$$\text{Corr}(u, v) = \frac{\text{Cov}(u, v)}{\sqrt{\text{Var}(u)\text{Var}(v)}} = \frac{\underline{a}' \Sigma_{12} \underline{b}^T}{\sqrt{(\underline{a}' \Sigma_{11} \underline{a}) (\underline{b}' \Sigma_{22} \underline{b})}}$$

is as large as possible. (\*)

We define :-

- The first pair of canonical variables are the pair of linear combinations  $u_1, v_1$  having unit variances which maximises the correlation  $(*)$ .
- The second pair of canonical variables are the linear combination  $u_2, v_2$  having unit variances which maximise the correlation among all choices which are uncorrelated with the first pair of canonical variables.
- At the  $k^{th}$  step, the  $k^{th}$  pair of canonical variables are the linear combination  $u_k, v_k$  having unique variances which maximise the correlation  $(*)$  among all the choices which are uncorrelated with the previous  $(k-1)$  canonical variable pairs.
- The correlation bet<sup>n</sup>  $k^{th}$  pair of canonical variables is called  $k^{th}$  canonical correlation.