Form of normal density function :-

→ Univariate normal density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \quad ; \quad x, \mu \in \mathbb{R}, \ \sigma^2 > 0$$

$$= \frac{1}{(2\pi)^{1/2}\, |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} \quad ; \quad \Sigma \text{ (+)ve definite}$$

→ Bivariate normal density function

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]}$$

$$f(\underline{x}) = \frac{1}{2\pi\, |\Sigma|^{1/2}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu})}$$

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \text{ and } \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

where $\rho = \dfrac{\sigma_{ij}}{\sigma_i \sigma_j}$

$$|\Sigma| = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2(1-\rho^2).$$

→ By the same analogy MVN density function may be written in the form

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

where, $K > 0$ is a constant to be determined such that the integral over the entire p-dimensional Euclidean space of $x_1, x_2, \cdots, x_p$ is unity.

Scanned with CamScanner

We assume that the matrix A is positive definite (if the quadratic form $\underline{x}' A \underline{x} \geq 0$) and symmetric.

Since A is positive definite

$\Rightarrow (\underline{x} - \underline{b})' A (\underline{x} - \underline{b}) \geq 0$

$\Rightarrow f(\underline{x}) \geq 0$ and is bounded above.

Since A is positive definite $\exists$ a non-singular matrix C ($|C| \neq 0$) s.t. $C'AC = I$

Now make a non-singular transformation

$\underline{x} - \underline{b} = C \underline{y} \Rightarrow \underline{y} = C^{-1} (\underline{x} - \underline{b})$

& $\underline{y} \in \mathbb{R}$

Jacobian of the transformation

$|J| = \frac{d(\underline{x} - \underline{b})}{d \underline{y}} = \frac{d \underline{x}}{d \underline{y}} = |C| \Rightarrow$ absolute value of $|C|$

$\Rightarrow$ density function of $\underline{y}$

$$g(\underline{y}) = K e^{-\frac{1}{2} \underline{y}' C' A C \underline{y}} |C|$$

$$= K e^{-\frac{1}{2} \underline{y}' \underline{y}} |C| \quad \text{—— (1)}$$

Integrating eq^n (1) over entire range

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} K |C| e^{-\frac{1}{2} \underline{y}' \underline{y}} \, dy_1 \cdots dy_p = 1$$

or $$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} K |C| e^{-\frac{1}{2} (y_1^2 + y_2^2 + \cdots + y_p^2)} \, dy_1 dy_2 \cdots dy_p = 1$$

or $$\cancel{\int_{-\infty}^{\infty}} K |C| \left[\left(\int_{-\infty}^{\infty} e^{-\frac{1}{2} y_1^2} dy_1 \right) \cdots \left(\int_{-\infty}^{\infty} e^{-\frac{1}{2} y_p^2} dy_p \right) \right] = 1$$

or, $K|C| (\sqrt{2\pi}) \cdots (\sqrt{2\pi}) = 1$; as $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = 1$

or, $\boxed{K = \frac{1}{|C| (2\pi)^{p/2}}}$

$$\Rightarrow g(\underline{y}) = \frac{|C| e^{-\frac{1}{2}\underline{y}'\underline{y}}}{|C| (2\pi)^{p/2}} = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_1^2}\right) \cdots \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_p^2}\right)$$

Thus, $y_1, y_2, \cdots y_p$ are independently normally distributed each with mean zero and variance unity.

Therefore, $E(\underline{X} - \underline{b}) = E(C\underline{Y}) = C\,E(\underline{Y}) = 0$

and $\boxed{\underline{b} = E(\underline{X}) = \underline{\mu}_x}$

Let Σ_y denotes the variance-covariance matrix of $\underline{Y}$. Since $y_1, y_2 \cdots y_p$ are independent $N(0,1)$, we have

$$\Sigma_y = E[\underline{Y} - 0][\underline{Y} - 0]' = E[\underline{Y}\underline{Y}'] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I$$

Hence, $I = E[\underline{Y}\underline{Y}'] = E[C^{-1}(\underline{X} - \underline{b})(\underline{X} - \underline{b})'C^{-1'}]$

$$= C^{-1}[E(\underline{X} - \underline{\mu})(\underline{X} - \underline{\mu})']C^{-1'}$$

$$= C^{-1}\Sigma_x C^{-1'}$$

Pre-multiplication by C and Post-multiplication by C'

$$CIC' = \Sigma_x \quad \text{or} \quad CC' = \Sigma_x$$

Also, we have

$$C'AC = I \quad \text{or} \quad A = C'^{-1} I C^{-1} = C'^{-1} C^{-1}$$

$$\text{or} \quad A^{-1} = CC'$$

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

Also, $|A^{-1}| = |CC'| = |C||C'| = |C|^2$

as $|C| = |C'|$

$\Rightarrow |C| = |\Sigma_x|^{1/2}$

Hence, $f(\underline{x}) = \dfrac{1}{(2\pi)^{p/2} |\Sigma_x|^{1/2}} \exp\left[\dfrac{-1}{2} (\underline{x} - \underline{\mu}_x)' \Sigma_x^{-1} (\underline{x} - \underline{\mu}_x)\right]$

Def^n. - A random vector $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$ taking values $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$ in E^p (Euclidean space of dimension p) is said to have a p- ~~di~~ variate normal distribution if its p.d.f. can be written as

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

where $\underline{\mu}_x = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$ and Σ_x is a positive definite symmetric matrix of order p.

Properties of MVN distribution :-

Theorem :- If the variance - covariance matrix of p-variate normal random vector $\underline{X} = (X_1, X_2, \cdots X_p)'$ is a diagonal matrix, then the components of $\underline{X}$ are independently normally distributed random variables.

Proof- We know that

$$f(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma_x|^{1/2}} \exp\left[-\frac{1}{2} (\underline{x} - \underline{\mu}_x)' \Sigma_x^{-1} (\underline{x} - \underline{\mu}_x)\right] \quad \text{—①}$$

Given, $\Sigma_x = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_p^2 \end{bmatrix}$

$$\Rightarrow (\underline{x} - \underline{\mu}_x)' \Sigma_x^{-1} (\underline{x} - \underline{\mu}_x) = \left[(x_1 - \mu_1) \cdots (x_p - \mu_p)\right] \begin{bmatrix} 1/\sigma_1^2 & \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 & \cdots & 1/\sigma_p^2 \end{bmatrix} \begin{bmatrix} (x_1 - \mu_1) \\ \vdots \\ (x_p - \mu_p) \end{bmatrix}$$

$$= \sum_{i=1}^{p} \left[\frac{x_i - \mu_i}{\sigma_i}\right]^2$$

and $|\Sigma_x| = \prod_{i=1}^{p} \sigma_i^2 \Rightarrow |\Sigma_x|^{1/2} = \prod_{i=1}^{p} \sigma_i$

Hence by eq^n. ①,

$$f(\underline{x}) = \frac{1}{(2\pi)^{p/2} \prod_{i=1}^{p} \sigma_i} \exp\left[-\frac{1}{2} \sum_{i=1}^{p} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right]$$

$$= \prod_{i=1}^{p} \frac{1}{(2\pi)^{1/2} \sigma_i} \exp\left[-\frac{1}{2} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right]$$

$$= f(x_1) f(x_2) \cdots f(x_p)$$

Therefore, $X_1, X_2, \cdots, X_p$ are independently normally distributed random variables.

Theorem:- If $\underline{x}$ (with p components) be distributed according to $N(\underline{\mu}, \Sigma)$. Then $\underline{Y} = C\underline{X}$ (non-singular) is distributed according to $N(C\underline{\mu}, C\Sigma C')$, for C non-singular.

Proof:- We have

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

Now $\underline{Y} = C\underline{X}$ or $\underline{X} = C^{-1}\underline{Y}$

$$\Rightarrow |J| = |C^{-1}|$$

Therefore, $$g(\underline{y}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} \underbrace{(C^{-1}\underline{y} - \underline{\mu})' \Sigma^{-1} (C^{-1}\underline{y} - \underline{\mu})}\right] |C^{-1}| \quad \text{---} (1)$$

$$\begin{aligned}(C^{-1}\underline{y} - \underline{\mu})' \Sigma^{-1} (C^{-1}\underline{y} - \underline{\mu}) &= (C^{-1}\underline{y} - C^{-1}C\underline{\mu})' \Sigma^{-1} (C^{-1}\underline{y} - C^{-1}C\underline{\mu}) \\ &= [C^{-1}(\underline{y} - C\underline{\mu})]' \Sigma^{-1} [C^{-1}(\underline{y} - C\underline{\mu})] \\ &= (\underline{y} - C\underline{\mu})' C^{-1'} \Sigma^{-1} C^{-1} (\underline{y} - C\underline{\mu}) \\ &= (\underline{y} - C\underline{\mu})' (C\Sigma C')^{-1} (\underline{y} - C\underline{\mu}) \quad \text{---} (2)\end{aligned}$$

Now $$|C^{-1}| = \frac{1}{|C|} = \sqrt{\frac{1}{|C|^2}} = \sqrt{\frac{1}{|C||C'|}} = \sqrt{\frac{|\Sigma|}{|C||\Sigma||C'|}}$$

$$= \frac{|\Sigma|^{1/2}}{|C\Sigma C'|^{1/2}} \quad ; \text{ since } |AB| = |A||B| \quad \text{---} (3)$$

Using (2) & (3), (1) can be rewritten as,

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

$$\Rightarrow \underline{y} \sim N(C\underline{\mu}, C\Sigma C')$$

Theorem - Let $\underline{x}' = (x_1, x_2, \cdots, x_p)$ have a joint normal distribution. A necessary and sufficient condition that a subset $\underline{x}^{(1)}$ of the components of $\underline{x}$ be independent of the subset $\underline{x}^{(2)}$ consisting of the remaining components of $\underline{x}$ is that the covariance between each component of $\underline{x}^{(1)}$ with a component of $\underline{x}^{(2)}$ is zero.

Proof - Let us partition.

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{pmatrix} \underline{x}^{(1)} \\ \underline{x}^{(2)} \end{pmatrix}, \text{ where } \underline{x}^{(1)} = \begin{bmatrix} x_1 \\ \vdots \\ x_q \end{bmatrix}_{q\times 1}, \text{ and } \underline{x}^{(2)} = \begin{bmatrix} x_{q+1} \\ \vdots \\ x_p \end{bmatrix}_{(p-q)\times 1}$$

The corresponding partition of $\underline{\mu}$ and Σ will be,

$$\underline{\mu} = \begin{bmatrix} \underline{\mu}^{(1)} \\ \underline{\mu}^{(2)} \end{bmatrix}, \text{ and } \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where,
$$\Sigma_{11} = E(\underline{x}^{(1)} - \underline{\mu}^{(1)})(\underline{x}^{(1)} - \underline{\mu}^{(1)})'$$
$$\Sigma_{22} = E(\underline{x}^{(2)} - \underline{\mu}^{(2)})(\underline{x}^{(2)} - \underline{\mu}^{(2)})'$$
$$\Sigma_{12} = E(\underline{x}^{(1)} - \underline{\mu}^{(1)})(\underline{x}^{(2)} - \underline{\mu}^{(2)})'$$
$$\Sigma_{21} = E(\underline{x}^{(2)} - \underline{\mu}^{(2)})(\underline{x}^{(1)} - \underline{\mu}^{(1)})' = \Sigma_{12}'$$

Necessary part,

Let $\underline{x}^{(1)}$ and $\underline{x}^{(2)}$ be independent

$\Rightarrow \Sigma_{12} = E(\underline{x}^{(1)} - \underline{\mu}^{(1)})(\underline{x}^{(2)} - \underline{\mu}^{(2)})' = 0$

$\Rightarrow$ The components are uncorrelated, i.e., covariances are 0.

Sufficient part,

From necessary part, we have $\Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}$

We have to prove that $\underline{x}^{(1)}$ & $\underline{x}^{(2)}$ are independent.

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

$$= \frac{1}{(2\pi)^{q/2} (2\pi)^{(p-q)/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} Q\right]$$

where $Q = (\underline{x}-\underline{\mu})' \Sigma^{-1} (\underline{x}-\underline{\mu})$

$$= \left[(\underline{x}^{(1)}-\underline{\mu}^{(1)})' \quad (\underline{x}^{(2)}-\underline{\mu}^{(2)})'\right] \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix} \begin{bmatrix} \underline{x}^{(1)} - \underline{\mu}^{(1)} \\ \underline{x}^{(2)} - \underline{\mu}^{(2)} \end{bmatrix}$$

$$= \left[(\underline{x}^{(1)}-\underline{\mu}^{(1)})' \Sigma_{11}^{-1} \quad (\underline{x}^{(2)}-\underline{\mu}^{(2)})' \Sigma_{22}^{-1}\right] \begin{bmatrix} \underline{x}^{(1)} - \underline{\mu}^{(1)} \\ \underline{x}^{(2)} - \underline{\mu}^{(2)} \end{bmatrix}$$

$$= (\underline{x}^{(1)}-\underline{\mu}^{(1)})' \Sigma_{11}^{-1} (\underline{x}^{(1)}-\underline{\mu}^{(1)}) + (\underline{x}^{(2)}-\underline{\mu}^{(2)})' \Sigma_{22}^{-1} (\underline{x}^{(2)}-\underline{\mu}^{(2)})$$

$$= Q_1 + Q_2$$

Also, $|\Sigma| = \begin{vmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{vmatrix} = \begin{vmatrix} \Sigma_{11} & 0 \\ 0 & I \end{vmatrix} \times \begin{vmatrix} I & 0 \\ 0 & \Sigma_{22} \end{vmatrix}$

$$= |\Sigma_{11}| |\Sigma_{22}|$$

$$\Rightarrow f(\underline{x}) = \frac{1}{(2\pi)^{q/2} (2\pi)^{(p-q)/2} |\Sigma_{11}|^{1/2} |\Sigma_{22}|^{1/2}} \exp\left[-\frac{1}{2}(Q_1+Q_2)\right]$$

$$= \frac{1}{(2\pi)^{q/2} |\Sigma_{11}|^{1/2}} \exp\left[-\frac{1}{2} Q_1\right] \times \frac{1}{(2\pi)^{(p-q)/2} |\Sigma_{22}|^{1/2}} \exp\left[-\frac{1}{2} Q_2\right]$$

$$= f(\underline{x}^{(1)}) \times f(\underline{x}^{(2)})$$

Therefore $\underline{x}^{(1)}$ & $\underline{x}^{(2)}$ are independent.

Further the marginal density of $\underline{X}^{(1)}$ is given by

$$h(\underline{x}^{(1)}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\underline{x}^{(1)}) \cdot f(\underline{x}^{(2)}) \, dx_{q+1} \cdots dx_p$$

$$= f(\underline{x}^{(1)}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\underline{x}^{(2)}) \, dx_{q+1} \cdots dx_p$$

$$= f(\underline{x}^{(1)})$$

Hence, the marginal distribution of $\underline{X}^{(1)}$ is $N_q(\underline{\mu}^{(1)}, \Sigma_{11})$ and the same of $\underline{X}^{(2)}$ is $N_{p-q}(\underline{\mu}^{(2)}, \Sigma_{22})$.

Scanned with CamScanner

Conditional distribution

The joint density function of $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ can be obtained by substituting $\underline{x}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underline{x}^{(2)}$ for $\underline{y}^{(1)}$, $\underline{x}^{(2)}$ for $\underline{y}^{(2)}$ and multiplying by the Jacobian of the transformation. The Jacobian of the transformation is $|J| = 1$, since $\underline{y} = C\underline{x}$ or $\underline{x} = C^{-1}\underline{y}$ and $\frac{d\underline{x}}{d\underline{y}} = |C^{-1}| = \frac{1}{|C|}$, where $|C| = \begin{vmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ \underline{0} & I \end{vmatrix} = 1$.

#

The resulting density of $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ is

$$f(\underline{x}^{(1)}, \underline{x}^{(2)}) = \frac{1}{(2\pi)^{q/2}|\Sigma_{11.2}|^{1/2}} \exp[-\frac{1}{2}(\underline{x}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underline{x}^{(2)} - \underline{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}\underline{\mu}^{(2)})'$$

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

$$\times \frac{1}{(2\pi)^{(p-q)/2}|\Sigma_{22}|^{1/2}} \exp[-\frac{1}{2}(\underline{x}^{(2)} - \underline{\mu}^{(2)})'\Sigma_{22}^{-1}(\underline{x}^{(2)} - \underline{\mu}^{(2)})]$$

$$= \frac{1}{(2\pi)^{q/2}|\Sigma_{11.2}|^{1/2}} \exp[-\frac{1}{2}\{\underline{x}^{(1)} - \underline{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underline{x}^{(2)} - \underline{\mu}^{(2)})\}'$$

$$\Sigma_{11.2}^{-1}\{\underline{x}^{(1)} - \underline{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underline{x}^{(2)} - \underline{\mu}^{(2)})\}]$$

$$\times \frac{1}{(2\pi)^{(p-q)/2}|\Sigma_{22}|^{1/2}} \exp[-\frac{1}{2}(\underline{x}^{(2)} - \underline{\mu}^{(2)})'\Sigma_{22}^{-1}(\underline{x}^{(2)} - \underline{\mu}^{(2)})].$$

Now

$$\frac{f(\underline{x}^{(1)}, \underline{x}^{(2)})}{f(\underline{x}^{(2)})} = \frac{1}{(2\pi)^{q/2}|\Sigma_{11.2}|^{1/2}} \exp[-\frac{1}{2}\{\underline{x}^{(1)} - \underline{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underline{x}^{(2)} - \underline{\mu}^{(2)})\}'$$

$$\Sigma_{11.2}^{-1}\{\underline{x}^{(1)} - \underline{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}(\underline{x}^{(2)} - \underline{\mu}^{(2)})\}]$$

$$= f(\underline{x}^{(1)} \mid \underline{x}^{(2)}).$$

Therefore, the density function $f(\underline{x}^{(1)} \mid \underline{x}^{(2)})$ is a $q-$variate normal with mean $\underline{\mu}^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\underline{x}^{(2)} - \underline{\mu}^{(2)}) = \underline{\mu}^{(1)*}$, and covariance matrix $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, i.e. $(\underline{X}^{(1)} \mid \underline{X}^{(2)} = \underline{x}^{(2)}) \sim N_q(\underline{\mu}^{(1)*}, \Sigma_{11.2})$.

#

Theorem: If $\underline{X}$ is distributed according to $N(\underline{\mu}, \Sigma)$, the marginal distribution of any set of components of $\underline{X}$ is multivariate normal with means, variances and covariances obtained by taking corresponding components of $\underline{\mu}$ and Σ, respectively.

Proof: We partition $\underline{X}$, $\underline{\mu}$ and Σ as

$$\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{pmatrix}, \text{ where } \underline{X}^{(1)} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_q \end{pmatrix}, \text{ and } \underline{X}^{(2)} = \begin{pmatrix} X_{q+1} \\ X_{q+2} \\ \vdots \\ X_p \end{pmatrix}$$

and the corresponding partition of $\underline{\mu}$ and Σ will be

$$\underline{\mu} = \begin{pmatrix} \underline{\mu}^{(1)} \\ \underline{\mu}^{(2)} \end{pmatrix}, \text{ and } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

We shall make a nonsingular linear transformation to sub vectors

$$\underline{Y}^{(1)} = \underline{X}^{(1)} + B\underline{X}^{(2)}, \text{ and } \quad \underline{Y}^{(2)} = \underline{X}^{(2)}.$$

where B is the matrix chosen such that $\underline{Y}^{(1)}$ and $\underline{Y}^{(2)}$ are uncorrelated. i.e. B must satisfy the equation

$$E(\underline{Y}^{(1)} - E\underline{Y}^{(1)})(\underline{Y}^{(2)} - E\underline{Y}^{(2)})' = 0$$

or $E(\underline{X}^{(1)} + B\underline{X}^{(2)} - \underline{\mu}^{(1)} - B\underline{\mu}^{(2)})(\underline{X}^{(2)} - \underline{\mu}^{(2)})' = 0$

or $E(\underline{X}^{(1)} - \underline{\mu}^{(1)})(\underline{X}^{(2)} - \underline{\mu}^{(2)})' + BE(\underline{X}^{(2)} - \underline{\mu}^{(2)})(\underline{X}^{(2)} - \underline{\mu}^{(2)})' = 0$

or $\Sigma_{12} + B\Sigma_{22} = 0.$

Thus,

$$B = -\Sigma_{12}\Sigma_{22}^{-1}, \text{ and } \quad \underline{Y}^{(1)} = \underline{X}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underline{X}^{(2)}.$$

Therefore, the transformation is

$$\begin{pmatrix} \underline{Y}^{(1)} \\ \underline{Y}^{(2)} \end{pmatrix} = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ \underline{0} & I \end{pmatrix} \begin{pmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{pmatrix}$$

i.e. $\underline{Y} = C\underline{X}$.

#

Since the transformation is nonsingular, therefore, the distribution $\underline{Y}$ is also $p-$variate normal with

$$E\underline{Y} = E\begin{pmatrix}\underline{Y}^{(1)}\\ \underline{Y}^{(2)}\end{pmatrix} = E\begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1}\\ \underline{0} & I\end{pmatrix}\begin{pmatrix}\underline{X}^{(1)}\\ \underline{X}^{(2)}\end{pmatrix} = E\begin{pmatrix}\underline{X}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underline{X}^{(2)}\\ \underline{X}^{(2)}\end{pmatrix}$$

$$= \begin{pmatrix}\underline{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underline{\mu}^{(2)}\\ \underline{\mu}^{(2)}\end{pmatrix}$$

and

$$\Sigma_{\underline{Y}} = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1}\\ \underline{0} & I\end{pmatrix}\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\begin{pmatrix} I & \underline{0}\\ -\Sigma_{12}\Sigma_{22}^{-1} & I\end{pmatrix}, \text{ since } \underline{Y} = C\,\underline{X}.$$

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

$$= \begin{pmatrix}\Sigma_{11.2} & \underline{0}\\ \Sigma_{21} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22} & \Sigma_{22}\end{pmatrix} = \begin{pmatrix}\Sigma_{11.2} & \underline{0}\\ \underline{0} & \Sigma_{22}\end{pmatrix}.$$

The joint density function of $\underline{Y}^{(1)}$ and $\underline{Y}^{(2)}$ is

$$f(\underline{y}) = \frac{1}{(2\pi)^{q/2}|\Sigma_{11.2}|^{1/2}}\exp[-\frac{1}{2}\{\underline{y}^{(1)} - (\underline{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underline{\mu}^{(2)})\}'$$

$$\Sigma_{11.2}^{-1}\{\underline{y}^{(1)} - (\underline{\mu}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\underline{\mu}^{(2)})\}]$$

$$\times\frac{1}{(2\pi)^{(p-q)/2}|\Sigma_{22}|^{1/2}}\exp[-\frac{1}{2}(\underline{y}^{(2)} - \underline{\mu}^{(2)})'\Sigma_{22}^{-1}(\underline{y}^{(2)} - \underline{\mu}^{(2)})]$$

Since $\underline{Y}^{(2)} = \underline{X}^{(2)}$ and $\underline{Y}^{(1)}$ are independently distributed, so

$f(\underline{y}) = f(\underline{y}^{(1)})\,f(\underline{x}^{(2)})$. Therefore, the marginal density function of $\underline{X}^{(2)} = \underline{Y}^{(2)}$ is given by the integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(\underline{y}^{(1)})\,f(\underline{x}^{(2)})\,d\underline{y}^{(1)} = f(\underline{x}^{(2)})\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(\underline{y}^{(1)})\,d\underline{y}^{(1)} = f(\underline{x}^{(2)})$$

Hence, the marginal distribution of $\underline{X}^{(2)}$ is $N_{p-q}(\underline{\mu}^{(2)},\ \Sigma_{22})$.

#

Theorem: If $\underline{X}$ is distributed according to $N_p(\underline{\mu}, \Sigma)$ and $\underline{Y} = D\underline{X}$, where D is $q \times p$ $(q < p)$ matrix of rank q, then $\underline{Y}$ is distributed according to $N_q(D\underline{\mu}, D\Sigma D')$.

#

Proof: The transformation is $\underline{Y} = D\underline{X}$, where $\underline{Y}$ has q components and D is $q \times p$ real matrix. The expected value of $\underline{Y}$ is

$E(\underline{Y}) = D\,E(\underline{X}) = D\underline{\mu}$, and the covariance matrix is

$$\Sigma_y = E[\underline{Y} - E(\underline{Y})]\,[\underline{Y} - E(\underline{Y})]' = E[D\underline{X} - D\,\underline{\mu}]\,[D\underline{X} - D\,\underline{\mu}]' = D\Sigma D'.$$

Since $R(D) = q$ the q rows of D are independent. We know that a set of q independent vectors can be extended to form a basis of the $p-$dimensional vector space by adding to it $p-q$ vectors.

Let $C = \begin{pmatrix} D_{q\times p} \\ E_{(p-q)\times p} \end{pmatrix}$, then C is nonsingular. Make the transformation

$\underline{Z} = C\underline{X}$. Since C nonsingular, therefore, $\underline{Z} \sim N_p(C\underline{\mu}, C\Sigma C')$, i.e.

$$\underline{Z} = \begin{pmatrix} D \\ E \end{pmatrix}\underline{X} = \begin{pmatrix} D\underline{X} \\ E\underline{X} \end{pmatrix}.$$

But, $D\underline{X}$ being the partition vector of $\underline{Z}$ has a marginal $q-$variate normal distribution, therefore,

$$\underline{Y} = D\underline{X} \sim N_q(D\underline{\mu}, D\Sigma D').$$

Note: This theorem tells us if $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$, then every linear combination of the components of $\underline{X}$ has a univariate normal distribution.

Proof: Let $Y = D_{1\times p}\,\underline{X} = (l_1, l_2, \ldots, l_p)\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} = \underline{l}'\,\underline{X}$, then $E(Y) = \underline{l}'\,\underline{\mu}$, and $\Sigma_y = \underline{l}'\,\Sigma\,\underline{l}$,

therefore,

$$Y = D_{1\times p}\,\underline{X} = \underline{l}'\,\underline{X} \sim N_1(\underline{l}'\,\underline{\mu}, \underline{l}'\,\Sigma\,\underline{l}).$$

#

Theorem: If every linear combination of the components of a vector $\underline{X}$ is normally distributed, then $\underline{X}$ has normal distribution.

Proof: Consider a vector $\underline{X}$ of $p-$components with density function $f(\underline{x})$ and characteristic function $\phi_{\underline{X}}(\underline{u}) = E[e^{i\underline{u}'\underline{X}}]$ and suppose the mean of $\underline{X}$ is $\underline{\mu}$ and the covariance matrix is Σ.

Since $\underline{u}'\underline{X}$ is normally distributed for every $\underline{u}$. Then the characteristic function of $\underline{u}'\underline{X}$ is

$$E\,e^{it(\underline{u}'\underline{X})} = e^{it\underline{u}'\underline{\mu} - \frac{1}{2}t^2\underline{u}'\Sigma\underline{u}}$$, taking $t=1$, this reduces to

$$E\,e^{i\underline{u}'\underline{X}} = e^{i\underline{u}'\underline{\mu} - \frac{1}{2}\underline{u}'\Sigma\underline{u}}.$$

Therefore, $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$.

#

Theorem: The moment generating function of a vector $\underline{X}$, which is distributed according to $N_p(\underline{\mu}, \Sigma)$ is $M_{\underline{X}}(\underline{t}) = e^{\underline{t}'\underline{\mu} + \frac{1}{2}\underline{t}'\Sigma\underline{t}}$.

Proof: Since Σ is a symmetric and positive definite, then there exits a non-singular matrix C such that

$$C'\Sigma^{-1}C = I \text{ and } \Sigma = CC'.$$

Make the nonsingular transformation

$$\underline{X} - \underline{\mu} = C\underline{Y}, \text{ then } \underline{Y} = C^{-1}(\underline{X} - \underline{\mu}) \text{ and } |J| = |C|.$$

Therefore, the density function of $\underline{Y}$ is

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

$$= \frac{1}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\underline{y}'C'\Sigma^{-1}C\underline{y})\right], \text{ since } |C| = |\Sigma|^{1/2}$$

$$= \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}\underline{y}'\underline{y}\right).$$

It shows that $Y_1, Y_2, \cdots, Y_p$ are independently normally distributed each with mean zero and variance one.

Now the moment generating function of $\underline{Y}$ is

$$M_{\underline{Y}}(\underline{u}) = E\,e^{\underline{u}'\underline{Y}} = E\,e^{(u_1Y_1 + \cdots + u_pY_p)} = E\,e^{(u_1Y_1)} \cdots E\,e^{(u_pY_p)}$$

$$= \prod_{i=1}^{p} E\,e^{u_iY_i} = e^{\frac{1}{2}\underline{u}'\underline{u}}, \text{ since } Y_i \sim N(0, 1).$$

Thus we can say

$$\phi_{\underline{X}}(\underline{t}) = E\,e^{\underline{t}'\underline{X}} = E\,e^{\underline{t}'(C\underline{Y} + \underline{\mu})} = e^{\underline{t}'\underline{\mu}}\,E[e^{\underline{t}'(C\underline{Y})}] = e^{\underline{t}'\underline{\mu}}\,E[e^{(C'\underline{t})'\underline{Y}}]$$

$$= e^{\underline{t}'\underline{\mu}}\,e^{\frac{1}{2}(C'\underline{t})'(C'\underline{t})} = e^{\underline{t}'\underline{\mu} + \frac{1}{2}\underline{t}'CC'\underline{t}} = e^{\underline{t}'\underline{\mu} + \frac{1}{2}\underline{t}'\Sigma\underline{t}}.$$

#

Characteristic function

The characteristic function of a random vector $\underline{X}$ is defined as

$$\phi_{\underline{X}}(\underline{t}) = E[e^{i\underline{t}'\underline{X}}], \text{ where } \underline{t} \text{ is a vector of reals, } i=\sqrt{-1}.$$

Theorem: Let $\underline{X} = (X_1, X_2, \cdots, X_p)'$ be normally distributed random vector with mean $\underline{\mu}$ and positive definite covariance matrix Σ, then the characteristic function of $\underline{X}$ is given by

$$\phi_{\underline{X}}(\underline{t}) = e^{i\underline{t}'\underline{\mu} - \frac{1}{2}\underline{t}'\Sigma\underline{t}}, \text{ where } \underline{t} = (t_1, t_2, \cdots, t_p)' \text{ is a real vector of order } p\times 1.$$

Proof: We have

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

Since Σ is a symmetric and positive definite, there exits a non-singular matrix C such that

$$C'\Sigma^{-1}C = I, \text{ and } \Sigma = CC'.$$

Let $\underline{X} - \underline{\mu} = C\underline{Y}$, so that $\underline{Y} = C^{-1}(\underline{X} - \underline{\mu})$ a nonsingular transformation and the Jacobian of the transformation is $|J| = |C|$, therefore, the density function of $\underline{Y}$ is

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

$$= \frac{1}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\underline{y}'C'\Sigma^{-1}C\,\underline{y})\right], \text{ since } |C| = |\Sigma|^{1/2}$$

$$= \frac{1}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2}\underline{y}'\underline{y}\right] = \left(\frac{1}{(2\pi)^{1/2}}\exp-\frac{1}{2}y_1^2\right)\cdots\left(\frac{1}{(2\pi)^{1/2}}\exp-\frac{1}{2}y_p^2\right).$$

It shows that $Y_1, Y_2, \cdots, Y_p$ are independently normally distributed each with mean zero and variance one.

Now the characteristic function of $\underline{Y}$ is

$$\phi_{\underline{Y}}(\underline{u}) = E[e^{i\underline{u}'\underline{Y}}] = E\, e^{i(u_1Y_1+\cdots+u_pY_p)} = E\, e^{iu_1Y_1}\cdots E\, e^{iu_pY_p}$$

Since $Y_1, Y_2, \cdots, Y_p$ are independent and distributed according to $N(0, 1)$,

$$\phi_{\underline{Y}}(\underline{u}) = \left(\exp-\frac{1}{2}u_1^2\right)\cdots\left(\exp-\frac{1}{2}u_p^2\right), \text{ since } X \sim N(\mu, \sigma^2), \text{ and } \phi_X(t) = e^{it\mu-\frac{1}{2}t^2\sigma^2}$$

$$= e^{-\frac{1}{2}(u_1^2+\cdots+u_p^2)} = e^{-\frac{1}{2}\underline{u}'\underline{u}}$$

#

Thus

$$\phi_{\underline{X}}(\underline{t}) = E[e^{i\underline{t}'\underline{X}}] = E[e^{i\underline{t}'(C\underline{Y}+\underline{\mu})}] = e^{i\underline{t}'\underline{\mu}}\,E[e^{i\underline{t}'(C\underline{Y})}]$$ #

$$= e^{i\underline{t}'\underline{\mu}}\,E\,e^{i(C'\underline{t})'\underline{Y}} = e^{i\underline{t}'\underline{\mu}}\;e^{-\frac{1}{2}(C'\underline{t})'(C'\underline{t})}$$

$$= e^{i\underline{t}'\underline{\mu}}\;e^{-\frac{1}{2}\underline{t}'CC'\underline{t}} = e^{i\underline{t}'\underline{\mu}-\frac{1}{2}\underline{t}'\Sigma\underline{t}}.$$ #

ESTIMATION OF PARAMETERS IN MULTIVARIATE NORMAL DISTRIBUTION

The normal distribution is completely specified if its mean vector $\underline{\mu}$ and covariance matrix Σ are known. In case of unknown parameters, the problems of their estimation arise. We estimate these parameters by the method of maximum likelihood estimation. For using this method we require a random sample of size n from the given $p-$variate normal population.

Let the random sample of size n from $N_p(\underline{\mu},\Sigma)$ be $\underline{x}_1,\underline{x}_2,\cdots,\underline{x}_\alpha,\cdots,\underline{x}_n$, where $n>p$ and $\underline{x}_\alpha$ is $p\times 1$ vector, $\alpha=1,2,\cdots,n$. In extended vector notation the data are as follows:

Suppose we have n individuals $1,2,\cdots,n$ and p characteristics $X_1,X_2,\cdots,X_p$ and each individuals studied.

Characteristic	Individual						Mean
	1	2	$\cdots$	α	$\cdots$	n	
X_1	x_{11}	x_{12}	$\cdots$	$x_{1\alpha}$	$\cdots$	x_{1n}	$\bar{x}_1$
X_2	x_{21}	x_{22}	$\cdots$	$x_{2\alpha}$	$\cdots$	x_{2n}	$\bar{x}_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_i	x_{i1}	x_{i2}	$\cdots$	$x_{i\alpha}$	$\cdots$	x_{in}	$\bar{x}_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_p	x_{p1}	x_{p2}	$\cdots$	$x_{p\alpha}$	$\cdots$	x_{pn}	$\bar{x}_p$
	$\underline{x}_1$	$\underline{x}_2$	$\cdots$	$\underline{x}_\alpha$	$\cdots$	$\underline{x}_n$	

Corresponding to $\alpha-$th individual there is a vector $\underline{x}_\alpha$ which is representing a point in $p-$dimensional space. So all these n point in E^n. Therefore,

$$\underline{\bar{x}}=\frac{1}{n}\sum_{\alpha=1}^{n}\underline{x}_\alpha=\begin{pmatrix}\frac{1}{n}\sum_{\alpha=1}^{n}x_{1\alpha}\\ \vdots\\ \frac{1}{n}\sum_{\alpha=1}^{n}x_{i\alpha}\\ \vdots\\ \frac{1}{n}\sum_{\alpha=1}^{n}x_{p\alpha}\end{pmatrix}=\begin{pmatrix}\bar{x}_1\\ \bar{x}_2\\ \vdots\\ \bar{x}_i\\ \vdots\\ \bar{x}_p\end{pmatrix},$$ is the sample mean vector, and sample variance covariance matrix is

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

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

Also

$$S=\frac{1}{n-1}\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{pmatrix}=\frac{A}{n-1}.$$

The matrix A is called sum of square and cross products of deviations about the mean or $(SS\,\&\,CP)$.

where

$$a_{ij}=\sum_{\alpha=1}^{n}(x_{i\alpha}-\overline{x}_i)(x_{j\alpha}-\overline{x}_j)\text{, for all } i \text{ and } j$$

$$=\sum_{\alpha=1}^{n}[x_{i\alpha}(x_{j\alpha}-\overline{x}_j)-\overline{x}_i(x_{j\alpha}-\overline{x}_j)]=\sum_{\alpha=1}^{n}x_{i\alpha}x_{j\alpha}-\overline{x}_j\sum_{\alpha=1}^{n}x_{i\alpha}-\overline{x}_i\sum_{\alpha=1}^{n}x_{j\alpha}+n\overline{x}_i\overline{x}_j$$

$$=\sum_{\alpha=1}^{n}x_{i\alpha}x_{j\alpha}-n\overline{x}_i\overline{x}_j.$$

Some results

1) Consider a quadratic form

$$Q=\underline{x}'A\underline{x}=\sum_{i,j}^{p}a_{ij}x_i x_j\text{, where } A \text{ is symmetric, then}$$

$$\frac{\partial Q}{\partial \underline{x}}=\begin{pmatrix} \frac{\partial Q}{\partial x_1} \\ \frac{\partial Q}{\partial x_2} \\ \vdots \\ \frac{\partial Q}{\partial x_p} \end{pmatrix}=2A\underline{x}.$$

In particular, for $p=2$

$$Q=(x_1 \quad x_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=(a_{11}x_1+a_{21}x_2 \quad a_{12}x_1+a_{22}x_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$=(a_{11}x_1^2+a_{21}x_1x_2+a_{12}x_1x_2+a_{22}x_2^2)\text{, then}$$

$$\frac{\partial Q}{\partial \underline{x}}=\begin{pmatrix} \frac{\partial Q}{\partial x_1} \\ \frac{\partial Q}{\partial x_2} \end{pmatrix}=\begin{pmatrix} 2a_{11}x_1+a_{21}x_2+a_{12}x_2 \\ a_{21}x_1+a_{12}x_1+2a_{22}x_2 \end{pmatrix}=2\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=2A\underline{x}.$$

2) If $B=\underline{y}'A\underline{x}=\underline{x}'A'\underline{y}$, then $\quad \frac{\partial B}{\partial \underline{y}}=A\underline{x}$, and $\frac{\partial B}{\partial \underline{x}}=A'\underline{y}$

3) If $Q = (\underline{x} - \underline{b})' A (\underline{x} - \underline{b}) = (\underline{b} - \underline{x})' A (\underline{b} - \underline{x})$, then

$$\frac{\partial Q}{\partial \underline{x}} = 2A(\underline{x} - \underline{b}), \text{ and } \frac{\partial Q}{\partial \underline{b}} = 2A(\underline{b} - \underline{x}).$$

4) A **sub matrix** of A is a rectangular array obtained from A by deleting rows and columns. A **minor** is the determinant of square sub matrix of A.

$$|A| = \sum_{j=1}^{p} a_{ij} A_{ij} = \sum_{j=1}^{p} a_{jk} A_{jk},$$

where A_{ij}, is $(-1)^{i+j}$ times the minor of a_{ij}, and the minor of an element a_{ij} is the determinant of the sub matrix of a square matrix A obtained by deleting the $i-$th row and $j-$th column.

If $|A| \neq 0$, there exist a unique matrix B such that $AB = I$, B is called the inverse of A and is denoted by A^{-1}.

Let a^{ij} be the element of A^{-1} in the $i-$th row and $j-$th column, then

$$a^{ij} = \frac{A_{ji}}{|A|} \text{ and } a_{ij} = \frac{A^{ji}}{|A^{-1}|}.$$

Maximum likelihood estimate of the mean vector

Let $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_\alpha, \cdots, \underline{x}_n$ be a random sample of size $n(> p)$ from $N_p(\underline{\mu}, \Sigma)$.

The likelihood function is

$$\phi = f(\underline{x}_1) f(\underline{x}_2) \cdots f(\underline{x}_n) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left[-\frac{1}{2} \sum_{\alpha=1}^{n} (\underline{x}_\alpha - \underline{\mu})' \Sigma^{-1} (\underline{x}_\alpha - \underline{\mu}) \right]$$

$$\log \phi = -\frac{np}{2} \log(2\pi) - \frac{n}{2} \log|\Sigma| - \frac{1}{2} \sum_{\alpha=1}^{n} (\underline{x}_\alpha - \underline{\mu})' \Sigma^{-1} (\underline{x}_\alpha - \underline{\mu}).$$

Differentiating with respect to $\underline{\mu}$ and equating to zero

$$\frac{\partial \log \phi}{\partial \underline{\mu}} = \underline{0} = -0 - 0 - \frac{1}{2} \sum_{\alpha=1}^{n} 2\Sigma^{-1} (\underline{\mu} - \underline{x}_\alpha) \quad \text{or} \quad \Sigma^{-1} \sum_{\alpha=1}^{n} (\underline{\mu} - \underline{x}_\alpha) = 0$$

or $\hat{\underline{\mu}} = \frac{1}{n} \sum_{\alpha=1}^{n} \underline{x}_\alpha = \overline{\underline{x}}$.

Maximum likelihood estimate of variance covariance matrix

Let $\Sigma^{-1} = (\sigma^{ij})$ and $\Sigma = (\sigma_{ij})$ then the $\left|\Sigma^{-1}\right| = \sigma^{i1}\Sigma^{i1} + \cdots + \sigma^{ip}\Sigma^{ip}$, where Σ^{ij} is the cofactor of σ^{ij} in Σ^{-1}, therefore, $\dfrac{\Sigma^{ij}}{\left|\Sigma^{-1}\right|} = (i\ j)^{th}$ element of $(\Sigma^{-1})^{-1} = (i\ j)^{th}$ element of $\Sigma = \sigma_{ij}$.

Now, the logarithm of the likelihood function is

$$\log\phi = -\frac{np}{2}\log(2\pi) + \frac{n}{2}\log\left|\Sigma^{-1}\right| - \frac{1}{2}\sum_{\alpha=1}^{n}(\underline{x}_\alpha - \underline{\mu})'\Sigma^{-1}(\underline{x}_\alpha - \underline{\mu})$$

$$= -\frac{np}{2}\log(2\pi) + \frac{n}{2}\log(\sigma^{i1}\Sigma^{i1} + \cdots + \sigma^{ij}\Sigma^{ij} + \cdots + \sigma^{ip}\Sigma^{ip})$$

$$-\frac{1}{2}\sum_{\alpha}\sum_{i,j}\sigma^{ij}(x_{i\alpha} - \mu_i)(x_{j\alpha} - \mu_j), \text{ since } \underline{x}' A\,\underline{x} = \sum_{i,j} a_{ij}\,x_i\,x_j$$

Differentiating with respect to σ^{ij} and equating to zero, we get

$$\frac{\partial\log\phi}{\partial\sigma^{ij}} = \underline{0} = \frac{n}{2}\frac{\Sigma^{ij}}{\left|\Sigma^{-1}\right|} - \frac{1}{2}\sum_{\alpha=1}^{n}(x_{i\alpha} - \mu_i)(x_{j\alpha} - \mu_j), \text{ because } \frac{\partial}{\partial x}\log f(x) = \frac{f'(x)}{f(x)}$$

or $\dfrac{n}{2}\sigma_{ij} = \dfrac{1}{2}\displaystyle\sum_{\alpha=1}^{n}(x_{i\alpha} - \mu_i)(x_{j\alpha} - \mu_j)$ or $\hat{\sigma}_{ij} = \dfrac{1}{n}\displaystyle\sum_{\alpha=1}^{n}(x_{i\alpha} - \hat{\mu}_i)(x_{j\alpha} - \hat{\mu}_j)$

Hence, $\hat{\Sigma} = \dfrac{A}{n}$.

Theorem: Given $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_\alpha, \cdots, \underline{x}_n$ be an independent random sample of size $n(> p)$ from $N_p(\underline{\mu}, \Sigma)$, then $\underline{\overline{x}} \sim N(\underline{\mu}, \Sigma / n)$.

Proof: $E(\underline{\overline{x}}) = E\left(\frac{1}{n}\sum_\alpha \underline{x}_\alpha\right) = \frac{1}{n}E(\underline{x}_1 + \cdots + \underline{x}_n) = \frac{1}{n}(n\underline{\mu}) = \underline{\mu}$ and

$$\Sigma_{\underline{\overline{x}}} = E(\underline{\overline{x}} - \underline{\mu})(\underline{\overline{x}} - \underline{\mu})' = E\left[\frac{1}{n}(\underline{x}_1 + \cdots + \underline{x}_n) - \underline{\mu}\right]\left[\frac{1}{n}(\underline{x}_1 + \cdots + \underline{x}_n) - \underline{\mu}\right]'$$

$$= \frac{1}{n^2}E(\underline{x}_1 + \cdots + \underline{x}_n - n\underline{\mu})(\underline{x}_1 + \cdots + \underline{x}_n - n\underline{\mu})'$$

$$= \frac{1}{n^2}[E(\underline{x}_1 - \underline{\mu})(\underline{x}_1 - \underline{\mu})' + \cdots + E(\underline{x}_n - \underline{\mu})(\underline{x}_n - \underline{\mu})' + 0], \text{ as } \underline{x}_\alpha\text{'s are independent.}$$

$$= \frac{1}{n^2}(n\Sigma) = \Sigma / n.$$

Thus,

$$\underline{\overline{x}} \sim N(\underline{\mu}, \Sigma / n).$$

Theorem: $\frac{A}{n-1}$ is an unbiased estimate of Σ.

Proof: We have

$$A = \sum_\alpha (\underline{x}_\alpha - \underline{\overline{x}})(\underline{x}_\alpha - \underline{\overline{x}})' = \sum_\alpha [(\underline{x}_\alpha - \underline{\mu}) - (\underline{\overline{x}} - \underline{\mu})][(\underline{x}_\alpha - \underline{\mu}) - (\underline{\overline{x}} - \underline{\mu})]'$$

$$= \sum_\alpha [(\underline{x}_\alpha - \underline{\mu})(\underline{x}_\alpha - \underline{\mu})' - (\underline{\overline{x}} - \underline{\mu})(\underline{x}_\alpha - \underline{\mu})' - (\underline{x}_\alpha - \underline{\mu})(\underline{\overline{x}} - \underline{\mu})' + (\underline{\overline{x}} - \underline{\mu})(\underline{\overline{x}} - \underline{\mu})']$$

$$= \sum_\alpha (\underline{x}_\alpha - \underline{\mu})(\underline{x}_\alpha - \underline{\mu})' - n(\underline{\overline{x}} - \underline{\mu})(\underline{\overline{x}} - \underline{\mu})'.$$

Taking expectation on both the sides

$$E(A) = \sum_\alpha E(\underline{x}_\alpha - \underline{\mu})(\underline{x}_\alpha - \underline{\mu})' - nE(\underline{\overline{x}} - \underline{\mu})(\underline{\overline{x}} - \underline{\mu})' = n\Sigma - n\Sigma / n = (n-1)\Sigma$$

or $E\left(\frac{A}{n-1}\right) = \Sigma$, this shows that the maximum likelihood estimate of Σ is biased, i.e. $E\left(\frac{A}{n}\right) = \frac{n-1}{n}\Sigma$.

Theorem: Given a random sample $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_\alpha, \cdots, \underline{x}_n$ from $N_p(\underline{\mu}, \Sigma)$, $\overline{\underline{x}} = \frac{1}{n}\sum_\alpha \underline{x}_\alpha$ and $A = \sum_\alpha (\underline{x}_\alpha - \overline{\underline{x}})(\underline{x}_\alpha - \overline{\underline{x}})'$. Thus $\overline{\underline{x}}$ and A are independently distributed.

Proof: Make an orthogonal transformation

$$\begin{aligned}
\underline{y}_1 &= c_{11}\underline{x}_1 + \cdots + c_{1n}\underline{x}_n \\
\vdots \quad & \quad \vdots \qquad\qquad \vdots \\
\underline{y}_{n-1} &= c_{n-11}\underline{x}_1 + \cdots + c_{n-1n}\underline{x}_n \\
\underline{y}_n &= \frac{1}{\sqrt{n}}\underline{x}_1 + \cdots + \frac{1}{\sqrt{n}}\underline{x}_n
\end{aligned}$$

#

or $\underline{Y} = C\,\underline{X}$, where $\underline{Y} = \begin{pmatrix} \underline{y}_1 \\ \vdots \\ \underline{y}_n \end{pmatrix}$, $\underline{X} = \begin{pmatrix} \underline{x}_1 \\ \vdots \\ \underline{x}_n \end{pmatrix}$, and

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ c_{n-11} & c_{n-12} & \cdots & c_{n-1n} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{pmatrix} \text{ is orthogonal.}$$

Since C is orthogonal

$$\sum_{k=1}^{n} c_{ik}\frac{1}{\sqrt{n}} = 0, \quad i = 1, 2, \cdots, n-1 \quad \Rightarrow \sum_{k=1}^{n} c_{ik} = 0$$

Also

$$\sum_{k=1}^{n} c_{ik} \times c_{jk} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \text{ for } i, j = 1, 2, \cdots, n.$$

Now consider,

$$\underline{y}_n = \frac{1}{\sqrt{n}}(\underline{x}_1 + \cdots + \underline{x}_n) = \sqrt{n}\,\overline{\underline{x}}$$

So that

$$E\,\underline{y}_n = E\sqrt{n}\,\overline{\underline{x}} = \sqrt{n}\,\underline{\mu}, \text{ and}$$

$$E\,\underline{y}_i = E\sum_{k=1}^{n} c_{ik}\,\underline{x}_k = \underline{\mu}\sum_{k=1}^{n} c_{ik} = \underline{0}, \text{ for } i = 1, 2, \cdots, n-1$$

#

$$\operatorname{Cov}(\underline{y}_i, \underline{y}_j) = E(\underline{y}_i - E\,\underline{y}_i)(\underline{y}_j - E\,\underline{y}_j)'$$

$$= E[c_{i1}(\underline{x}_1 - \underline{\mu}) + \cdots + c_{in}(\underline{x}_n - \underline{\mu})][c_{j1}(\underline{x}_1 - \underline{\mu}) + \cdots + c_{jn}(\underline{x}_n - \underline{\mu})]'$$

$$= c_{i1}\, c_{j1}\, E(\underline{x}_1 - \underline{\mu})(\underline{x}_1 - \underline{\mu})' + \cdots + c_{in}\, c_{jn}\, E(\underline{x}_n - \underline{\mu})(\underline{x}_n - \underline{\mu})' + 0$$

since $\underline{x}_1, \cdots, \underline{x}_n$ are independent

$$= (c_{i1}\, c_{j1} + \cdots + c_{in}\, c_{jn})\Sigma = \Sigma \sum_{k=1}^{n} c_{ik}\, c_{jk} = \begin{cases} 0, & if \quad i \neq j \\ \Sigma, & if \quad i = j \end{cases}$$

This implies $\underline{y}_1, \cdots, \underline{y}_n$ are independent.

Now

$$\sum_{i=1}^{n} \underline{y}_i \underline{y}_i' = \sum_{i=1}^{n} (c_{i1}\, \underline{x}_1 + \cdots + c_{in}\, \underline{x}_n)(c_{i1}\, \underline{x}_1 + \cdots + c_{in}\, \underline{x}_n)'$$

$$= \underline{x}_1 \underline{x}_1' \sum_{i=1}^{n} c_{i1}^2 + \cdots + \underline{x}_1 \underline{x}_n' \sum_{i=1}^{n} c_{i1}\, c_{in} + \cdots + \underline{x}_n \underline{x}_1' \sum_{i=1}^{n} c_{in}\, c_{i1} + \cdots + \underline{x}_n \underline{x}_n' \sum_{i=1}^{n} c_{in}^2$$

#

$$= \underline{x}_1 \underline{x}_1' + \cdots + \underline{x}_n \underline{x}_n' = \sum_{i=1}^{n} \underline{x}_i \underline{x}_i' .$$

Therefore,

$$\sum_{i=1}^{n} \underline{y}_i \underline{y}_i' - \underline{y}_n \underline{y}_n' = \sum_{i=1}^{n} \underline{x}_i \underline{x}_i' - \underline{y}_n \underline{y}_n'$$

or $$\sum_{i=1}^{n-1} \underline{y}_i \underline{y}_i' = \sum_{i=1}^{n} \underline{x}_i \underline{x}_i' - n \underline{\bar{x}}\, \underline{\bar{x}}' = A$$

#

#

#

#

#

#

#

#

#

#

Test for $\underline{\mu}$, when Σ is known

Given a random sample $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_\alpha, \cdots, \underline{x}_n$ from $N_p(\underline{\mu}, \Sigma)$. The hypothesis of interest is $H_0: \underline{\mu} = \underline{\mu}_0$, where $\underline{\mu}_0$ is a specified vector, then, under H_0, the test statistic is $n(\underline{\bar{x}} - \underline{\mu}_0)'\Sigma^{-1}(\underline{\bar{x}} - \underline{\mu}_0) \sim \chi_p^2$.

Proof: Let C be a non-singular matrix such that

$C'\Sigma^{*-1}C = I$, and $CC' = \Sigma^* = \dfrac{\Sigma}{n}$. Make the transformation

$(\underline{\bar{x}} - \underline{\mu}_0) = C\underline{y} \quad \Rightarrow \quad \underline{y} = C^{-1}(\underline{\bar{x}} - \underline{\mu}_0)$, and

$E\,\underline{y} = C^{-1}E(\underline{\bar{x}} - \underline{\mu}_0) = C^{-1}(\underline{\mu}_0 - \underline{\mu}_0) = \underline{0}$, under H_0.

$$\Sigma_{\underline{y}} = E(\underline{y} - E\,\underline{y})(\underline{y} - E\,\underline{y})' = C^{-1}E(\underline{\bar{x}} - \underline{\mu}_0)(\underline{\bar{x}} - \underline{\mu}_0)'C^{-1'}$$

$$= C^{-1}\Sigma^* C^{-1'} = (C'\Sigma^{*-1}C)^{-1} = I.$$

Therefore,

$\underline{y} \sim N_p(\underline{0}, I)$, i.e. $y_i \sim N(0,1)$, for all $i = 1, 2, \cdots, p$.

Now

$$n(\underline{\bar{x}} - \underline{\mu}_0)'\Sigma^{-1}(\underline{\bar{x}} - \underline{\mu}_0) = (\underline{\bar{x}} - \underline{\mu}_0)'(CC')^{-1}(\underline{\bar{x}} - \underline{\mu}_0)$$

$$= [C^{-1}(\underline{\bar{x}} - \underline{\mu}_0)]'[C^{-1}(\underline{\bar{x}} - \underline{\mu}_0)] = \underline{y}'\underline{y} = \sum_{i=1}^{p} y_i^2 \sim \chi_p^2.$$

Let $\chi_p^2(\alpha)$ be the number such that $\Pr[\chi_p^2 \geq \chi_p^2(\alpha)] = \alpha$, then for testing H_0, we use the critical region

$$n(\underline{\bar{x}} - \underline{\mu}_0)'\Sigma^{-1}(\underline{\bar{x}} - \underline{\mu}_0) \geq \chi_p^2(\alpha).$$

#

#

#

#

#

#

#

#

Two sample problem

Given $\underline{x}_1^{(1)}, \underline{x}_2^{(1)}, \cdots, \underline{x}_\alpha^{(1)}, \cdots, \underline{x}_{n_1}^{(1)}$ be a random sample from $N_p(\underline{\mu}^{(1)}, \Sigma)$ and $\underline{x}_1^{(2)}, \underline{x}_2^{(2)}, \cdots, \underline{x}_\alpha^{(2)}, \cdots, \underline{x}_{n_2}^{(2)}$ from $N_p(\underline{\mu}^{(2)}, \Sigma)$. $H_0: \underline{\mu}^{(1)} = \underline{\mu}^{(2)}$, then, in this case, the statistic is

$$\frac{n_1 n_2}{n_1 + n_2}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' \Sigma^{-1} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) \sim \chi_p^2 .$$

Proof: We know that

$\underline{\bar{x}}^{(1)} \sim N_p(\underline{\mu}^{(1)}, \Sigma / n_1)$, and $\underline{\bar{x}}^{(2)} \sim N_p(\underline{\mu}^{(2)}, \Sigma / n_2)$. Further, $\underline{\bar{x}}^{(1)}$ and $\underline{\bar{x}}^{(2)}$ are independent and

$$\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)} \sim N_p\left[\underline{\bar{\mu}}^{(1)} - \underline{\mu}^{(2)}, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\Sigma\right].$$

Make the transformation (nonsingular0

$$(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) = C\,\underline{y} \quad \Rightarrow \quad \underline{y} = C^{-1}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})$$

Since C is a non-singular matrix such that

$C'\Sigma^{*-1}C = I$, where $\Sigma^* = \dfrac{n_1 + n_2}{n_1 n_2}\Sigma = CC'$, and

$E\,\underline{y} = C^{-1}E(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) = \underline{0}$, under H_0.

#

$$\Sigma_{\underline{y}} = E(\underline{y} - E\underline{y})(\underline{y} - E\,\underline{y})' = C^{-1}E(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' C^{-1'}$$

$$= C^{-1}\Sigma^* C^{-1'} = (C'\Sigma^{*-1}C)^{-1} = I$$

Therefore,

$\underline{y} \sim N_p(\underline{0}, I)$, i.e. $y_i \sim N(0,1)$, for all $i = 1, 2, \cdots, p$.

Now

$$\frac{n_1 n_2}{n_1 + n_2}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' \Sigma^{-1} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) = (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})'(CC')^{-1}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})$$

$$= [C^{-1}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})]'[C^{-1}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})] = \underline{y}'\,\underline{y} = \sum_{i=1}^{p} y_i^2 \sim \chi_p^2 .$$

Let $\chi_p^2(\alpha)$ be the number such that $\Pr[\chi_p^2 \geq \chi_p^2(\alpha)] = \alpha$

Then for testing H_0, we use the critical region

$$\frac{n_1 n_2}{n_1 + n_2}(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' \Sigma^{-1} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) \geq \chi_p^2(\alpha) .$$

#

If $x_1, x_2, \cdots, x_n$ are independent observations from $N(\mu, \sigma^2)$, it is well known $(n-1)s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \sim \sigma^2 \chi^2_{n-1}$. The multivariate analogue of $(n-1)s^2$ is the matrix A and is called Wishart matrix. In other words the Wishart matrix is defined as the $p \times p$ symmetric matrix of sums of squares and cross products (of deviations about the mean) of the sample observations, from a $p-$variate nonsingular normal distribution. The distribution of A when the multivariate distribution is assumed normal is called Wishart distribution and is a generalization of χ^2 distribution in the univariate case.

By definition of A, we mean the joint distribution of the $\frac{p(p+1)}{2}$ distinct elements a_{ij}, $(i, j = 1, 2, \cdots, p;\ i \le j)$ of the symmetric matrix A.

#

Theorem: Let $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ be a random sample from $N_p(\underline{\mu}, I)$, $A = \sum_{\alpha=1}^{n}(\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' = \sum_{\alpha=1}^{n-1}\underline{Z}_\alpha \underline{Z}'_\alpha$, where $\underline{Z}_\alpha$ are independent, each distributed according to $N_p(\underline{0}, I)$. Then the density of $A = \sum_{\alpha=1}^{n-1}\underline{Z}_\alpha \underline{Z}'_\alpha$ is

$$\frac{|A|^{(\upsilon-p-1)/2}\exp(-\frac{1}{2}tr\, A)}{2^{\upsilon p/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}, \text{ where } \upsilon = n-1.$$

#

This is the form of wishart distribution when $\Sigma = I$ and is denoted by $W_p(\upsilon, I)$

#

Theorem: Suppose the $p-$component vectors $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ $(n > p)$ are independent, each distributed according to $N_p(\underline{\mu}, \Sigma)$, then the density of $A = \sum_{\alpha=1}^{n}(\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' = \sum_{\alpha=1}^{n-1}\underline{Z}_\alpha \underline{Z}'_\alpha$, where, $\underline{Z}_\alpha \sim N_p(\underline{0}, \Sigma)$ is

$$\frac{|A|^{(\upsilon-p-1)/2}\exp(-\frac{1}{2}tr\, A\Sigma^{-1})}{2^{\upsilon p/2}\pi^{p(p-1)/4}|\Sigma|^{\upsilon/2}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}, \text{ where } \upsilon = n-1.$$

#

$A \sim W_p(\upsilon, \Sigma)$ #

#

#

Theorem: Let $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ $(n \geq p+1)$ are distributed independently, each according to $N_p(\underline{\mu}, \Sigma)$. Then the distribution of $S = \frac{1}{n-1}\sum_{\alpha=1}^{n}(\underline{x}_\alpha - \underline{\overline{x}})(\underline{x}_\alpha - \underline{\overline{x}})'$ is $W_p(\upsilon, \Sigma/\upsilon)$, where $\upsilon = n-1$.

#

$$f(S) = \frac{|S|^{(\upsilon-p-1)/2}\exp\left[-\frac{1}{2}tr\, S\left(\frac{\Sigma}{n-1}\right)^{-1}\right]}{2^{\upsilon p/2}\pi^{p(p-1)/4}\left|\frac{\Sigma}{n-1}\right|^{\upsilon/2}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}.$$

#

Characteristic function

Consider

$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, the elements of the matrix are $A_{11}, A_{22}, 2A_{12}$, because $A_{12} = A_{21}$

Introduce a real matrix $\Theta = (\theta_{ij})$, with $\theta_{ij} = \theta_{ji}$

$\Theta = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix}$ then, $A\Theta = \begin{pmatrix} A_{11}\theta_{11} + A_{12}\theta_{21} & A_{11}\theta_{12} + A_{12}\theta_{22} \\ A_{21}\theta_{11} + A_{22}\theta_{21} & A_{21}\theta_{12} + A_{22}\theta_{22} \end{pmatrix}$

$tr\, A\Theta = A_{11}\theta_{11} + A_{12}\theta_{21} + A_{21}\theta_{12} + A_{22}\theta_{22} = A_{11}\theta_{11} + A_{22}\theta_{22} + 2A_{12}\theta_{12}$

#

We know that the characteristic function of a vector $\underline{x}' = (x_1, x_2, \cdots, x_p)$ is defined as

$Ee^{i\underline{t}'\underline{x}}$, where $\underline{t}'\underline{x} = t_1x_1 + t_2x_2 + \cdots + t_px_p$

#

Theorem: If $\underline{Z}_1, \underline{Z}_2, \cdots, \underline{Z}_{n-1}$ are independent, each with distribution $N_p(\underline{0}, \Sigma)$, the characteristic function of $A_{11}, A_{22}, \cdots, A_{pp}, 2A_{12}, \cdots, 2A_{p-1p}$, where $A = (A_{ij}) = \sum_{\alpha=1}^{n-1}\underline{Z}_\alpha \underline{Z}_\alpha'$ is given by

$\phi_A(\Theta) = Ee^{itr\, A\Theta} = |I - 2i\Theta\Sigma|^{-\upsilon/2}$, where $n-1 = \upsilon$.

#

By definition,

$$\phi_A(\Theta) = Ee^{itr\,\Theta A} = \int \frac{e^{itr\,\Theta A}|A|^{(\upsilon-p-1)/2}\exp(-\frac{1}{2}tr\, A\Sigma^{-1})}{2^{\upsilon p/2}\pi^{p(p-1)/4}|\Sigma|^{\upsilon/2}\prod_{i=1}^{p}\Gamma(\upsilon-i+1)/2}dA$$

$$= \frac{K}{|\Sigma|^{\upsilon/2}}\int |A|^{(\upsilon-p-1)/2} e^{-\frac{1}{2}tr(\Sigma^{-1}-2i\Theta)A}\, dA$$

#

$$= \frac{K\left|\Sigma^{-1} - 2i\Theta\right|^{\upsilon/2}}{\left|\Sigma\right|^{\upsilon/2}\left|\Sigma^{-1} - 2i\Theta\right|^{\upsilon/2}} \int \left| A \right|^{(\upsilon-p-1)/2} e^{-\frac{1}{2}tr(\Sigma^{-1}-2i\Theta)A} dA$$

$$= \frac{1}{\left|\Sigma\right|^{\upsilon/2}\left|\Sigma^{-1} - 2i\Theta\right|^{\upsilon/2}}, \text{ since } K\left|\Sigma^{-1}\right|^{\upsilon/2} \int \left| A \right|^{(\upsilon-p-1)/2} e^{-\frac{1}{2}tr A\Sigma^{-1}} dA = 1$$

$$= \frac{1}{\left|I - 2i\Theta\Sigma\right|^{\upsilon/2}} = \left|I - 2i\Theta\Sigma\right|^{-\upsilon/2}.$$

This shows that, if $A \sim W_p(\upsilon, \Sigma)$, and then the characteristic function of A is $\left|I - 2i\Theta\Sigma\right|^{-\upsilon/2}$.

#

#

Properties of Wishart distribution

Theorem: Suppose $A_i \quad (i=1,2)$ are distributed independently according to $W_p(\upsilon_i, \Sigma)$ respectively, then $A_1 + A_2 \sim W_p(\upsilon_1 + \upsilon_2, \Sigma)$.

Proof: We know that the characteristic function of A_1, if $A_1 \sim W_p(\upsilon_1, \Sigma)$, is

$$\phi_{A_1}(\Theta) = \left| I - 2i\,\Theta\Sigma \right|^{-\upsilon_1/2}.$$

#

Similarly, the characteristic function of A_2 will be

$$\phi_{A_2}(\Theta) = \left| I - 2i\,\Theta\Sigma \right|^{-\upsilon_2/2}$$

Since A_1 and A_2 are independently distributed, so

$$\phi_{A_1+A_2}(\Theta) = \phi_{A_1}(\Theta)\phi_{A_2}(\Theta) = \left| I - 2i\,\Theta\Sigma \right|^{-(\upsilon_1+\upsilon_2)/2}$$

But this is the characteristic function of $W_p(\upsilon_1 + \upsilon_2, \Sigma)$, therefore,

$$A_1 + A_2 \sim W_p(\upsilon_1 + \upsilon_2, \Sigma).$$

#

Theorem: If $A \sim W_p(n-1, \Sigma)$, then the distribution of $\underline{l}' A \underline{l} \sim (\underline{l}'\Sigma \underline{l})\chi^2_{n-1}$, where $\underline{l}$ is a known vector.

Proof: Given $A \sim W_p(n-1, \Sigma)$, then $A = \sum_{\alpha=1}^{n-1} \underline{Z}_\alpha \underline{Z}_\alpha'$, where $\underline{Z}_\alpha \sim N_p(\underline{0}, \Sigma)$, and

$$\underline{l}' A \underline{l} = \sum_{\alpha=1}^{n-1} \underline{l}' \underline{Z}_\alpha \underline{Z}_\alpha' \underline{l} = \sum_{\alpha=1}^{n-1} (\underline{l}' \underline{Z}_\alpha)(\underline{l}' \underline{Z}_\alpha)' = \sum_{\alpha=1}^{n-1} V_\alpha^2, \text{ where } V_\alpha = \underline{l}' \underline{Z}_\alpha \text{ is } N(0, \underline{l}'\Sigma \underline{l}).$$

Therefore,

$$\underline{l}' A \underline{l} \sim (\underline{l}'\Sigma \underline{l})\chi^2_{n-1}.$$

Theorem: If $A \sim W_p(n-1, \Sigma)$, and if a is a positive constant, then $aA \sim W_p(n-1, a\Sigma)$.

Proof: Since $A \sim W_p(n-1, \Sigma)$, there are $n-1$ independent $p-$component random vectors $\underline{Z}_1, \cdots, \underline{Z}_{n-1}$ each distributed as $N_p(\underline{0}, \Sigma)$, such that

$A = \sum_{\alpha=1}^{n-1} \underline{Z}_\alpha \underline{Z}_\alpha'$. Evidently, we have

$aA = \sum_{\alpha=1}^{n-1} (\sqrt{a}\,\underline{Z}_\alpha)(\sqrt{a}\,\underline{Z}_\alpha')$ and $\sqrt{a}\,\underline{Z}_1, \cdots, \sqrt{a}\,\underline{Z}_{n-1}$ are independently identically distributed as normal $N_p(\underline{0}, a\Sigma)$. Thus

$$aA \sim W_p(n-1, a\Sigma).$$

#

Theorem: Let A and Σ be partitioned into q and $p-q$ rows and columns as

$$A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},\ \Sigma=\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

If A is distributed according to $W_p(\upsilon, \Sigma)$, then A_{11} is distributed according to $W_q(\upsilon, \Sigma_{11})$ (the marginal distribution of some sets of elements of A).

Proof: Given $A \sim W_p(\upsilon, \Sigma)$, where

$$A=\sum_{\alpha=1}^{n-1} \underline{Z}_\alpha \underline{Z}_\alpha{}',\ \text{and } \underline{Z}_\alpha \text{ are independent, each with distribution } N_p(\underline{0}, \Sigma).$$

Partition $\underline{Z}_\alpha$ into sub vectors of q and $p-q$ components as

$$\underline{Z}_\alpha=\begin{pmatrix} \underline{Z}_\alpha^{(1)} \\ \underline{Z}_\alpha^{(2)} \end{pmatrix}.$$

#

Then $\underline{Z}_\alpha^{(1)}$ are independent, each with distribution $N_q(\underline{0}, \Sigma_{11})$, because $\underline{Z}_\alpha^{(1)}$ and $\underline{Z}_\alpha^{(2)}$ are independent, so that $\Sigma_{12}=\Sigma_{21}=0$, and

$$A_{11}=\sum_{\alpha=1}^{n-1} \underline{Z}_\alpha^{(1)} \underline{Z}_\alpha^{(1)}{}', \quad \Rightarrow A_{11} \sim W_q(\upsilon, \Sigma_{11}).$$

Similarly,

$$A_{22} \sim W_{p-q}(\upsilon, \Sigma_{22}).$$

#

Theorem: Suppose A is distributed according to $W_p(\upsilon, \Sigma)$ and let $A^*=BAB'$, where B is a matrix of order $q \times p$, then $A^* \sim W_q(\upsilon, \Phi)$, where $\Phi=B\Sigma B'$.

Proof: We have

$$A=\sum_{\alpha=1}^{n-1} \underline{Z}_\alpha \underline{Z}_\alpha{}',\ \text{where } \underline{Z}_\alpha \sim N_p(\underline{0}, \Sigma)$$

$$BAB'=\sum_{\alpha=1}^{n-1} B\underline{Z}_\alpha \underline{Z}_\alpha{}' B'=\sum_{\alpha=1}^{n-1}(B\underline{Z}_\alpha)(B\underline{Z}_\alpha)'$$

$$=\sum_{\alpha=1}^{n-1} \underline{V}_\alpha \underline{V}_\alpha{}',\ \text{where } \underline{V}_\alpha=B\underline{Z}_\alpha \sim N_q(\underline{0}, B\Sigma B').$$

Therefore,

$$A^*=BAB'=\sum_{\alpha=1}^{n-1} \underline{V}_\alpha \underline{V}_\alpha{}' \sim W_q(\upsilon, B\Sigma B').$$

#

#

Multiple Correlation coefficient

18 June 2021 15:11

MULTIPLE AND PARTIAL CORRELATIONS

Result:

$X_1^* = E(X_1 | X_2 = x_2) = a + bx_2$ is the regression line of X_1 on X_2, b is the regression coefficient. It follows that

$$\int x_1 f(x_1 | x_2) dx_1 = a + bx_2 \tag{5.1}$$

On multiplying both the sides of equation (5.1) by $f(x_2)$, and integrating with respect to x_2, we get

$$\int\int x_1 f(x_1 | x_2) f(x_2) dx_1 dx_2 = a \int f(x_2) dx_2 + b \int x_2 f(x_2) dx_2$$

or $\int\int x_1 [f(x_1, x_2) dx_2] dx_1 = a + b\mu_2$ or $\int x_1 f(x_1) dx_1 = a + b\mu_2$

or $\mu_1 = a + b\mu_2$ (5.2)

On multiplying both the sides of equation (5.1) by $x_2 f(x_2)$, and integrating with respect to x_2, we get

$$\int\int x_1 x_2 f(x_1, x_2) dx_2 dx_1 = a \int x_2 f(x_2) dx_2 + b \int x_2^2 f(x_2) dx_2$$

or $E(X_1, X_2) = a\mu_2 + bE(X_2^2)$

or $\sigma_{12} + \mu_1\mu_2 = a\mu_2 + b(\sigma_2^2 + \mu_2^2)$ (5.3)

By solving equations (5.2) and (5.3), we get

$$\begin{array}{r} \mu_1\mu_2 = a\mu_2 + b\mu_2^2 \\ \sigma_{12} + \mu_1\mu_2 = a\mu_2 + b(\sigma_2^2 + \mu_2^2) \\ \hline \sigma_{12} = b[(\sigma_2^2 + \mu_2^2) - \mu_2^2] \end{array}$$

$b = \dfrac{\sigma_{12}}{\sigma_2^2} = \rho \dfrac{\sigma_1}{\sigma_2}$. After substituting the value of b in equation (5.1), we get

$a = \mu_1 - \rho \dfrac{\sigma_1}{\sigma_2} \mu_2$, therefore,

$$X_1^* = E(X_1 | X_2 = x_2) = \mu_1 - \rho \frac{\sigma_1}{\sigma_2} \mu_2 + \rho \frac{\sigma_1}{\sigma_2} x_2 = \mu_1 + \underbrace{\rho \frac{\sigma_1}{\sigma_2}}_{\text{regression coeff.}} (x_2 - \mu_2).$$

Multiple correlation

If X_1 is the first component of $\underline{X}$ and $\underline{X}^{(2)}$ the vector of remaining $(p-1)$ components. We first express X_1 as a linear combination of $\underline{X}^{(2)}$ defined by the relation

$X_1^* = \mu_1 + \underline{\beta}'(\underline{X}^{(2)} - \underline{\mu}^{(2)})$, the coefficient vector $\underline{\beta}$ is determined by minimizing

$$U = E[X_1 - X_1^*]^2 = E[X_1 - \mu_1 - \underline{\beta}'(\underline{X}^{(2)} - \underline{\mu}^{(2)})]^2.$$

Differentiating with respect to $\underline{\beta}$ and equating to zero

$$-2E(\underline{X}^{(2)} - \underline{\mu}^{(2)})[(X_1 - \mu_1) - \underline{\beta}'(\underline{X}^{(2)} - \underline{\mu}^{(2)})] = 0$$

or $E(\underline{X}^{(2)} - \underline{\mu}^{(2)})(X_1 - \mu_1) - E(\underline{X}^{(2)} - \underline{\mu}^{(2)})(\underline{X}^{(2)} - \underline{\mu}^{(2)})'\underline{\beta} = 0$

or $\underline{\sigma}_{12} = \Sigma_{22}\underline{\beta}$ and

$$\hat{\underline{\beta}}' = \underline{\sigma}_{12}'\Sigma_{22}^{-1}.$$

Therefore, the best linear function of X_1 in terms of $\underline{X}^{(2)}$ is

$$X_1^* = \mu_1 + \underline{\sigma}_{12}'\Sigma_{22}^{-1}(\underline{X}^{(2)} - \underline{\mu}^{(2)}) = \hat{X}_1.$$

The correlation coefficient between X_1 and its best linear estimate in terms of $\underline{X}^{(2)}$ is called **Multiple correlation** between X_1 and $X_2, \cdots, X_p$. This is denoted by

$$\rho_{1(2,3,\cdots,p)} = \frac{Cov(X_1, \hat{X}_1)}{\sqrt{V(X_1)\,V(\hat{X}_1)}}, \text{ where } \quad V(X_1) = E[X_1 - E(X_1)]^2 = \sigma_{11}.$$

$$V(\hat{X}_1) = E[\hat{X}_1 - E(\hat{X}_1)][\hat{X}_1 - E(\hat{X}_1)]'$$

$$= E[\mu_1 + \underline{\sigma}_{12}'\Sigma_{22}^{-1}(\underline{X}^{(2)} - \underline{\mu}^{(2)}) - \mu_1][\mu_1 + \underline{\sigma}_{12}'\Sigma_{22}^{-1}(\underline{X}^{(2)} - \underline{\mu}^{(2)}) - \mu_1]'$$

since $E(\hat{X}_1) = \mu_1 + \underline{\sigma}_{12}'\Sigma_{22}^{-1}E(\underline{X}^{(2)} - \underline{\mu}^{(2)}) = \mu_1$

$$= \underline{\sigma}_{12}'\Sigma_{22}^{-1}E(\underline{X}^{(2)} - \underline{\mu}^{(2)})(\underline{X}^{(2)} - \underline{\mu}^{(2)})'\Sigma_{22}^{-1}\underline{\sigma}_{12}$$

$$= \underline{\sigma}_{12}'\Sigma_{22}^{-1}\underline{\sigma}_{12}.$$

$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} X_1 \\ \underline{X}^{(2)}_{(p-1)\times 1} \end{bmatrix}$

$\underline{\beta}_{(p-1)\times 1}$

$\Rightarrow$ $\underline{\beta}'(\underline{X}^{(2)} - \underline{\mu}^{(2)})$ is a scalar.

$\Rightarrow$ $\underline{\beta}'[\underline{X}^{(2)} - \underline{\mu}^{(2)}] = [\underline{X}^{(2)} - \underline{\mu}^{(2)}]'\underline{\beta}$

and Σ_{22} being a symmetric matrix

$\Sigma_{22} = \Sigma_{22}'$

$$Cov(X_1,\hat{X}_1)=E[X_1-E(X_1)][\hat{X}_1-E(\hat{X}_1)]'$$

$$=E(X_1-\mu_1)[\mu_1+\underline{\sigma}_{12}{}'\Sigma_{22}^{-1}(\underline{X}^{(2)}-\underline{\mu}^{(2)})-\mu_1]'$$

$$=E(X_1-\mu_1)(\underline{X}^{(2)}-\underline{\mu}^{(2)})'\Sigma_{22}^{-1}\underline{\sigma}_{12}=\underline{\sigma}_{12}{}'\Sigma_{22}^{-1}\underline{\sigma}_{12}.$$

Hence,

$$\rho_{1(2,3,\cdots,p)}=\frac{\underline{\sigma}_{12}{}'\Sigma_{22}^{-1}\underline{\sigma}_{12}}{\sqrt{\sigma_{11}(\underline{\sigma}_{12}{}'\Sigma_{22}^{-1}\underline{\sigma}_{12})}}=\sqrt{\frac{\underline{\sigma}_{12}{}'\Sigma_{22}^{-1}\underline{\sigma}_{12}}{\sigma_{11}}}=\sqrt{\frac{\underline{\beta}'\Sigma_{22}\underline{\beta}}{\sigma_{11}}},$$

where $\underline{\sigma}_{12}$ and Σ_{22} are defined as $\begin{pmatrix}\sigma_{11} & \underline{\sigma}_{12}{}' \\ \underline{\sigma}_{21} & \Sigma_{22}\end{pmatrix}$.

$$\hat{\underline{\beta}}' = \sigma_{12}'\Sigma_{22}^{-1} \Rightarrow \hat{\underline{\beta}} = \Sigma_{22}^{-1}\sigma_{12}$$

$$\underline{\sigma}_{12}'\Sigma_{22}^{-1}\sigma_{12} = \underbrace{\sigma_{12}'\Sigma_{22}^{-1}}\Sigma_{22}\Sigma_{22}^{-1}\sigma_{12}$$

$$= \hat{\underline{\beta}}'\Sigma_{22}\hat{\underline{\beta}}$$

Estimation of multiple correlation coefficient

The multiple correlation in the population is

$$\rho_{1(2,\cdots,p)}=\sqrt{\frac{\underline{\sigma}_{12}{}'\Sigma_{22}^{-1}\underline{\sigma}_{12}}{\sigma_{11}}}=\sqrt{\frac{\underline{\beta}'\Sigma_{22}\underline{\beta}}{\sigma_{11}}}.$$

Given $\underline{x}_\alpha$ $(\alpha=1,\cdots,n)$, $n>p$. We estimate Σ by $\hat{\Sigma}=\frac{A}{n}=\frac{n-1}{n}S$,

where, $A=\sum_\alpha(\underline{x}_\alpha-\underline{\bar{x}})(\underline{x}_\alpha-\underline{\bar{x}})'$.

Now A is partitioned as follows

$$\frac{A}{n}=\begin{pmatrix}\frac{a_{11}}{n} & \frac{\underline{a}_{12}{}'}{n} \\ \frac{\underline{a}_{12}}{n} & \frac{A_{22}}{n}\end{pmatrix}$$, and the estimate of $\underline{\beta}$ is $\hat{\underline{\beta}}'=\underline{\sigma}_{12}{}'\Sigma_{22}^{-1}=\frac{\underline{a}_{12}{}'}{n}\left(\frac{A_{22}}{n}\right)^{-1}=\underline{a}_{12}{}'A_{22}^{-1}$.

Let A (→ square) $= \begin{bmatrix}A_{11} & A_{12} \\ A_{21} & A_{22}\end{bmatrix}$

$$\begin{bmatrix}I & -A_{12}A_{22}^{-1} \\ 0 & I\end{bmatrix}\begin{bmatrix}A_{11} & A_{12} \\ A_{21} & A_{22}\end{bmatrix}\begin{bmatrix}I & 0 \\ -A_{22}^{-1}A_{21} & I\end{bmatrix}$$

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

Using the above estimate, the sample multiple correlation coefficient of X_1 on $X_2,\cdots,X_p$ is

$$R_{1(2,\cdots,p)}=\sqrt{\frac{\hat{\underline{\sigma}}_{12}{}'\hat{\Sigma}_{22}^{-1}\underline{\sigma}_{12}}{\hat{\sigma}_{11}}}=\sqrt{\frac{\underline{a}_{12}{}'A_{22}^{-1}\underline{a}_{12}}{a_{11}}}$$

and

$$1-R^2_{1(2,\cdots,p)}=\frac{a_{11}-\underline{a}_{12}{}'A_{22}^{-1}\underline{a}_{12}}{a_{11}}=\frac{\left|a_{11}-\underline{a}_{12}{}'A_{22}^{-1}\underline{a}_{12}\right|\left|A_{22}\right|}{a_{11}\left|A_{22}\right|}=\frac{|A|}{a_{11}|A_{22}|}.$$

Taking determinants on both sides,

$$|A| = |A_{11}-A_{12}A_{22}^{-1}A_{21}||A_{22}|$$

$$= |A_{11\cdot 2}||A_{22}|$$

Distribution of sample multiple correlation coefficient in null case

The sample multiple correlation coefficient between X_1 and $\underline{X}^{(2)}$ is defined by relation

$$R^2 = \frac{\underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11}} \quad \text{and} \quad 1 - R^2 = \frac{a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11}},$$

where $R^2 = R^2_{1(2,3,\cdots,p)}$ and $A = \begin{pmatrix} a_{11} & \underline{a}_{12}' \\ \underline{a}_{12} & A_{22} \end{pmatrix}$.

Therefore,

$$\frac{R^2}{1-R^2} = \frac{\underline{a}_{12}' A_{22}^{-1} \underline{a}_{12}}{a_{11.2}}.$$

We know that, if A is partitioned as

$$A = \begin{array}{cc} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} & \begin{array}{l} q \\ p-q \end{array} \end{array}$$ and $A \sim W_p(n-1, \Sigma)$, then, $A_{11} \sim W_q(n-1, \Sigma_{11})$ and

$$A_{11} - A_{12} A_{22}^{-1} A_{21} \sim W_q(n-1-(p-q), \Sigma_{11.2}).$$

Thus, in our case

$$a_{11} \sim W_1(n-1, \sigma_{11}), \; \Rightarrow \; \frac{a_{11}}{\sigma_{11}} \sim \chi^2_{n-1}.$$

In null case $\rho_{1(2,3,\cdots,p)} = 0$

$$\Sigma_{11.2} = \sigma_{11} - \underline{\sigma}_{12}' \Sigma_{22}^{-1} \underline{\sigma}_{21} = \sigma_{11}, \text{ since } \underline{\sigma}_{12}' = 0, \text{ so that}$$

$$a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{21} \sim W_1(n-1-(p-1), \sigma_{11})$$

$$\Rightarrow \frac{a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{21}}{\sigma_{11}} \sim \chi^2_{n-p}.$$

Consider

$$\frac{a_{11}}{\sigma_{11}} = \frac{a_{11} - \underline{a}_{12}' A_{22}^{-1} \underline{a}_{21}}{\sigma_{11}} + \frac{\underline{a}_{12}' A_{22}^{-1} \underline{a}_{21}}{\sigma_{11}}$$

or $Q = Q_1 + Q_2$, (say),

where $Q \sim \chi^2_{n-1}$, and $Q_1 \sim \chi^2_{n-p}$.

From Fisher Cochran theorem Q_2 is independently distributed as $\chi^2_{n-1-(n-p)}$, i.e. $Q_2 \sim \chi^2_{p-1}$ and is independent of Q_1, hence,

$$F = \frac{R^2}{1-R^2} \times \frac{n-p}{p-1} = \frac{\underline{a}_{12}' A_{22}^{-1} \underline{a}_{12} / \sigma_{11}}{a_{11.2} / \sigma_{11}} \times \frac{n-p}{p-1} = \frac{\chi^2_{p-1} / p-1}{\chi^2_{n-p} / n-p} \sim F_{p-1,\, n-p}.$$

The distribution of the statistic F is

$$df(F) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} F^{\frac{\nu_1}{2}-1}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)\left(1+\frac{\nu_1}{\nu_2}F\right)^{(\nu_1+\nu_2)/2}} dF,$$

where $\nu_1 = p-1$, $\nu_2 = n-p$.

In this put $F = \frac{R^2}{1-R^2}\frac{\nu_2}{\nu_1}$, then $dF = \frac{dR^2}{(1-R^2)^2}\frac{\nu_2}{\nu_1}$

$$df(R^2) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2}\left(\frac{R^2}{1-R^2}\frac{\nu_2}{\nu_1}\right)^{\frac{\nu_1}{2}-1}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)\left(1+\frac{R^2}{1-R^2}\right)^{(\nu_1+\nu_2)/2}} \frac{\nu_2}{\nu_1} \frac{dR^2}{(1-R^2)^2}$$

$$= \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}-\frac{\nu_1}{2}+1-1}\left(\frac{R^2}{1-R^2}\right)^{\frac{\nu_1}{2}-1}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)\left(\frac{1}{1-R^2}\right)^{(\nu_1+\nu_2)/2}} \frac{dR^2}{(1-R^2)^2}$$

$$= \frac{1}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} (R^2)^{\frac{\nu_1}{2}-1} (1-R^2)^{\frac{\nu_1+\nu_2}{2}-\frac{\nu_1}{2}+1-2} dR^2$$

$$= \frac{1}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} (R^2)^{\frac{\nu_1}{2}-1} (1-R^2)^{\frac{\nu_2}{2}-1} dR^2.$$ Put $dR^2 = 2R\, dR$, thus the distribution of R.

$$df(R) = \frac{2R^{(\nu_1-1)} (1-R^2)^{\frac{\nu_2}{2}-1}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} dR = \frac{2R^{p-2} (1-R^2)^{\frac{n-p}{2}-1}}{B\left(\frac{p-1}{2}, \frac{n-p}{2}\right)} dR, \qquad 0 < R < 1.$$

Generalized variance

The multivariate analogue of the variance σ^2 of a univariate distribution is the covariance matrix Σ, and the determinant of covariance matrix is termed as generalized variance of the multivariate distribution. Similarly, the generalized variance of a sample $\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n$ is defined as

$$|S| = \left| \frac{1}{n-1} \sum_{\alpha=1}^{n} (\underline{x}_\alpha - \underline{\overline{x}})(\underline{x}_\alpha - \underline{\overline{x}})' \right|.$$

Distribution of the generalized variance

By definition, the distribution of $|S|$ is the same as the distribution of $\frac{1}{(n-1)^p}|A|$, where $A = \sum_{\alpha=1}^{n-1} \underline{Z}_\alpha \underline{Z}_\alpha{}'$, with $\underline{Z}_\alpha \sim N_p(\underline{0}, \Sigma)$. Since Σ is symmetric and positive definite there exist a nonsingular matrix C such that $C\Sigma C' = I$.

#

Make the transformation

$\underline{Z}_\alpha^* = C\,\underline{Z}_\alpha$, then, $\underline{Z}_\alpha^* \sim N_p(\underline{0}, I)$, are independent, and

$$\sum_{\alpha=1}^{n-1} \underline{Z}_\alpha^* \underline{Z}_\alpha^{*\prime} = C\left(\sum_\alpha \underline{Z}_\alpha \underline{Z}_\alpha{}' \right) C' = CAC' = B \text{ (say)}$$

We have

$$|CAC'| = |B| \quad \text{or} \quad |A| = \frac{1}{|C|^2}|B|. \text{ But } C\Sigma C' = I, \text{ then}$$

$$|C\Sigma C'| = 1, \quad \text{or} \quad |C|^2 = \frac{1}{|\Sigma|} \quad \Rightarrow |A| = |B||\Sigma|.$$

Let

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pp} \end{pmatrix}$$

and $B_{ii} = \begin{pmatrix} b_{ii} & \underline{b}'_{(i)} \\ \underline{b}_{(i)} & B_{i+1\,i+1} \end{pmatrix}$, where $\underline{b}'_{(i)} = (b_{i\,i+1}, b_{i\,i+2}, \cdots, b_{i\,p})$.

So that

$B_{11} = B$ and $B_{pp} = b_{pp}$, and

#

$$b_{ii.i+1,\cdots,p} = b_{ii} - \underline{b}'_{(i)}\, B^{-1}_{i+1i+1}\, \underline{b}_{(i)}$$

$$= \frac{|B_{ii}|}{\left|B_{i+1i+1}\right|}, \text{ as } A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$ and assume that A_{22} is square and nonsingular, then

$$|A| = \left|A_{11} - A_{12}\, A_{22}^{-1}\, A_{21}\right| |A_{22}| = |A_{11.2}||A_{22}|.$$

Also

$$|B| = |B_{11}| = \frac{|B_{11}|}{|B_{22}|}\frac{|B_{22}|}{|B_{33}|}\cdots\frac{\left|B_{p-1\,p-1}\right|}{\left|B_{pp}\right|}\left|B_{pp}\right|$$

$$= (b_{11.2,\cdots,p})(b_{22.3,\cdots,p})\cdots(b_{p-1\,p-1.\,p})\,b_{pp}.$$

Since

$$B = \sum_{\alpha=1}^{n-1} \underline{Z}^*_\alpha \underline{Z}^{*'}_\alpha, \text{ with } \underline{Z}^*_\alpha \sim N_p(\underline{0}, I)$$

$$\Rightarrow b_{11.2,\cdots,p} = \sum_{\alpha=1}^{n-1-(p-1)} V_\alpha^2, \text{ where } V_\alpha \sim N(0, 1).$$

#

Thus,

$b_{11.2,\cdots,p}$ has χ^2 – distribution with $(n-p)$ degree of freedom.

Since $\underline{Z}^*_\alpha{}'s$ are independently distributed

$\Rightarrow (b_{11.2,\cdots,p})(b_{22.3,\cdots,p})\cdots(b_{p-1\,p-1.\,p})\,b_{pp}$ are independently distributed

$\Rightarrow b_{ii.i+1,\cdots,p}$ has χ^2 – distribution with $n-1-(p-i)$ degrees of freedom.

Therefore,

$|B|$ is distributed as $\chi^2_{n-1-(p-1)} \cdot \chi^2_{n-1-(p-2)} \cdot \cdots \cdot \chi^2_{n-2} \cdot \chi^2_{n-1}$.

Since $|S| = \dfrac{1}{(n-1)^p}|A| = \dfrac{1}{(n-1)^p}|\Sigma||B|$, then,

$|S|$ is distributed as $\dfrac{1}{(n-1)^p}|\Sigma|\ \chi^2_{n-1} \cdot \chi^2_{n-2} \cdots \chi^2_{n-1-(p-2)} \cdot \chi^2_{n-p}$.

For $p=1$

$$|S| = \frac{|\Sigma|}{n-1}\chi^2_{n-1} \quad \Rightarrow \quad \frac{(n-1)\,s_{11}}{\sigma_{11}} \sim \chi^2_{n-1} \quad \text{ or } \quad \frac{a_{11}}{\sigma^2} \sim \chi^2_{n-1}.$$

#

#

DISCRIMINANT ANALYSIS

The problem of discriminant analysis deals with assigning an individual to one of several categories on the basis of measurements on a $p-$component vector of variable $\underline{x}$ on that individual. For example, we take certain measurements on the skull of an animal and want to know whether it was male or female, a patient is to be classified as diabetic or non diabetic on the basis of certain tests such as blood, urine, blood pressure etc., a salesman is to be classified as successful or unsuccessful on different psychological tests.

Procedure of classification into one of the two populations with known probability distribution

Let R denote the entire $p-$dimensional space in which the point of observation $\underline{x}$ falls. We then have to divided R into two, say, R_1 and R_2 so that

If $\underline{x}$ falls in R_1, then assign the individual to population π_1

If $\underline{x}$ falls in R_2, assign the individual to population π_2

Obviously, with any such procedure an error of misclassification is inevitable (unavoidably) i.e. the rule may assign an individual to π_2, when he really belongs to π_1 and vice versa. A rule should control this error of discrimination.

Let $f_1(\underline{x})$ and $f_2(\underline{x})$ are the probability density function of $\underline{x}$ in the two populations π_1 and π_2. Let

$q_1 =$ a priori probability that an individual comes from π_1

$q_2 =$ a priori probability that an individual comes from π_2

$\Pr(1|2) = \Pr($ an individual belongs to π_2 is misclassified to $\pi_1)$

$\Pr(2|1) = \Pr($ an individual belongs to π_1 is misclassified to $\pi_2)$.

Obviously,

$$\Pr(1|2) = \int_{R_1} f_2(\underline{x})\,d\underline{x}, \text{ and } \Pr(2|1) = \int_{R_2} f_1(\underline{x})\,d\underline{x}.$$

Since the probability of drawing an observation from π_1 is q_1 and from π_2 is q_2, we have

$\Pr($ drawing an observation from π_1 and is misclassified as from $\pi_2) = q_1 \Pr(2|1)$

$\Pr($ drawing an observation from π_2 and is misclassified as from $\pi_1) = q_2 \Pr(1|2)$

Then the total chance of misclassification, say ϕ, is

$$\phi = q_1 \Pr(2|1) + q_2 \Pr(1|2) \tag{7.1}$$

We choose regions R_1 and R_2 such that equation (7.1) is minimized. The procedure that minimize (7.1) for a given q_1 and q_2 is called a Bayes procedure. Consider,

$$\begin{aligned}\phi &= q_1 \Pr(2|1) + q_2 \Pr(1|2) = q_1 \int_{R_2} f_1(\underline{x})\,d\underline{x} + q_2 \int_{R_1} f_2(\underline{x})\,d\underline{x} \\ &= q_1 \int_{R_2} f_1(\underline{x})\,d\underline{x} + q_1 \int_{R_1} f_1(\underline{x})\,d\underline{x} - q_1 \int_{R_1} f_1(\underline{x})\,d\underline{x} + q_2 \int_{R_1} f_2(\underline{x})\,d\underline{x} \\ &= q_1 \int_{R} f_1(\underline{x})\,d\underline{x} + \int_{R_1} [q_2\, f_2(\underline{x}) - q_1\, f_1(\underline{x})]\,d\underline{x}\end{aligned}$$

$\Rightarrow$ ϕ is minimize, when $q_2 f_2(\underline{x}) \le q_1 f_1(\underline{x})$, so we divide R such as

$$R_1 = \{\underline{x} | q_2 f_2(\underline{x}) \le q_1 f_1(\underline{x})\} = \left\{\underline{x} \Big| \frac{q_1 f_1(\underline{x})}{q_2 f_2(\underline{x})} \ge 1\right\} = \left\{\underline{x} \Big| \frac{f_1(\underline{x})}{f_2(\underline{x})} \ge \frac{q_2}{q_1}\right\}.$$

Similarly,

$$R_2 = \left\{\underline{x} \Big| \frac{q_1 f_1(\underline{x})}{q_2 f_2(\underline{x})} < 1\right\} = \left\{\underline{x} \Big| \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{q_2}{q_1}\right\}.$$

Further, if the cost of misclassification is given, $C(2|1)$ cost of misclassification to π_2 when it actually comes from π_1, $C(1|2)$ the cost of misclassification to π_1 when it actually comes from π_2. Example, if a potentially good candidate for admission to a medical school is rejected, the nation will suffer a shortage in medical persons, but, on the contrary, if a bad candidate is admitted, he may not be able to complete the course successfully and money, resources equipment used by him will be a waste. Total expected cost from misclassification

$$E(C) = C(2|1)\Pr(2|1)q_1 + C(1|2)\Pr(1|2)q_2 \text{ and}$$

classification rule will be

$$R_1 = \left\{\underline{x} \Big| \frac{f_1(\underline{x})}{f_2(\underline{x})} \ge \frac{q_2 C(1|2)}{q_1 C(2|1)}\right\}, \text{ and } \quad R_2 = \left\{\underline{x} \Big| \frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{q_2 C(1|2)}{q_1 C(2|1)}\right\}.$$

Classification into one of two known multivariate normal populations

Let $f_1(\underline{x})$ = density function of $N_p(\underline{\mu}^{(1)}, \Sigma)$ and $f_2(\underline{x})$ = density function of $N_p(\underline{\mu}^{(2)}, \Sigma)$ and the region R_1 of the classification into the population first is given by

$$R_1 = \left\{\underline{x} \Big| \frac{f_1(\underline{x})}{f_2(\underline{x})} \ge k\right\}, \text{ where } k = \frac{q_2 C(1|2)}{q_1 C(2|1)}. \text{ Consider}$$

$$\ln \frac{f_1(\underline{x})}{f_2(\underline{x})} \ge \ln k \tag{7.2}$$

The left hand side of (7.2) can be expanded as

$$\ln f_1(\underline{x}) - \ln f_2(\underline{x}) = -\frac{1}{2}\{(\underline{x} - \underline{\mu}^{(1)})'\Sigma^{-1}(\underline{x} - \underline{\mu}^{(1)}) - (\underline{x} - \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{x} - \underline{\mu}^{(2)})\}$$

$$= -\frac{1}{2}\{(\underline{x}'\Sigma^{-1}\underline{x} - \underline{x}'\Sigma^{-1}\underline{\mu}^{(1)} - \underline{\mu}^{(1)'}\Sigma^{-1}\underline{x} + \underline{\mu}^{(1)'}\Sigma^{-1}\underline{\mu}^{(1)})$$

$$-(\underline{x}'\Sigma^{-1}\underline{x} - \underline{x}'\Sigma^{-1}\underline{\mu}^{(2)} - \underline{\mu}^{(2)'}\Sigma^{-1}\underline{x} + \underline{\mu}^{(2)'}\Sigma^{-1}\underline{\mu}^{(2)})\}$$

$$= -\frac{1}{2}\{-2\underline{x}'\Sigma^{-1}\underline{\mu}^{(1)} + 2\underline{x}'\Sigma^{-1}\underline{\mu}^{(2)} + \underline{\mu}^{(1)'}\Sigma^{-1}\underline{\mu}^{(1)} - \underline{\mu}^{(2)'}\Sigma^{-1}\underline{\mu}^{(2)}\}$$

$$= \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) - \frac{1}{2}(\underline{\mu}^{(1)'}\Sigma^{-1}\underline{\mu}^{(1)} - \underline{\mu}^{(2)'}\Sigma^{-1}\underline{\mu}^{(2)}).$$

Thus,

$$R_1 = \{\underline{x} | \underline{x}'\Sigma^{-1}(\mu^{(1)} - \mu^{(2)}) - \frac{1}{2}(\mu^{(1)'}\Sigma^{-1}\mu^{(1)} - \mu^{(2)'}\Sigma^{-1}\mu^{(2)}) \ge \ln k\}.$$

In particular, when

$C(1|2) = C(2|1)$, and $q_1 = q_2 = 1/2$, then $k = 1$, $\ln k = 0$.

Then the region of classification into population first is

$$R_1 = \{\underline{x} | \ \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) - \frac{1}{2}(\underline{\mu}^{(1)'}\Sigma^{-1}\underline{\mu}^{(1)} - \underline{\mu}^{(2)'}\Sigma^{-1}\underline{\mu}^{(2)}) \geq 0\}$$

$$= \{\underline{x} | \ \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) \geq \frac{1}{2}(\underline{\mu}^{(1)'}\Sigma^{-1}\underline{\mu}^{(1)} - \underline{\mu}^{(2)'}\Sigma^{-1}\underline{\mu}^{(2)})\}$$

$$= \{\underline{x} | \ \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) \geq \frac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})\}.$$

The term $\underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})$ is known as the Fisher's discriminant function.

Similarly, the region of classification into population second is

$$R_2 = \{\underline{x} | \ \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) < \frac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})\}$$

The regions are computed easily as follows:

Consider

$\Sigma^{-1}\underline{d} = \underline{\delta}$, then solve, $\Sigma\underline{\delta} = \underline{d}$ (by Doolittle method), where $\underline{d} = \underline{\mu}^{(1)} - \underline{\mu}^{(2)}$.

Probability of misclassification (Two known $p-$variate normal population)

If the observation is from π_1, then

$\Pr(2|1) = \int_{R_2} f_1(\underline{x})\,d\underline{x} = \int_{R_2} f(\underline{x}; \underline{\mu}^{(1)}, \Sigma)\,d\underline{x}$, where

$R_2 = \{\underline{x} | \ \underline{x}'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) < \frac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})\}$. Put

$U = \underline{x}'\underline{\delta}$, where $\underline{\delta} = \Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})$, and $h = \frac{1}{2}(\underline{\mu}^{(1)} + \underline{\mu}^{(2)})\underline{\delta}$

$\Rightarrow R_2 = \{\underline{x} \mid U < h\}$.

Since $\underline{x} \sim N_p(\underline{\mu}^{(1)}, \Sigma)$, then $U = \underline{x}'\underline{\delta}$ is the univariate normal with following parameters

$E(U) = \underline{\mu}^{(1)'}\underline{\delta}$, and $\mathrm{Var}(U) = \underline{\delta}'\Sigma\underline{\delta} = (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})'\Sigma^{-1}(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) = \Delta^2$.

Therefore,

$$\Pr(2|1) = \int_{-\infty}^{h} \frac{1}{\Delta\sqrt{2\pi}} \exp\left\{-\frac{1}{2\Delta^2}(U - \underline{\mu}^{(1)'}\underline{\delta})^2\right\} du$$

Make a transformation

$$\frac{U - \underline{\mu}^{(1)'}\underline{\delta}}{\Delta} = y, \quad \Rightarrow du = \Delta\,dy.$$

Thus,

$$\Pr(2|1) = \int_{-\infty}^{(h-\underline{\mu}^{(1)'}\underline{\delta})/\Delta} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\underline{\mu}^{(2)}-\underline{\mu}^{(1)})'\underline{\delta}/2\Delta} e^{-y^2/2}\, dy .$$

Similarly,

$\Pr(1|2) = \int_{R_1} f_2(\underline{x})\, d\underline{x} = \int_{R_1} f(\underline{x}; \underline{\mu}^{(2)}, \Sigma)\, d\underline{x}$, where $R_1 = \{\underline{x} \mid U \geq h\}$, $E(U) = \underline{\mu}^{(2)'}\underline{\delta}$ and $\mathrm{Var}(U) = \Delta^2$. Put

$\dfrac{U - \underline{\mu}^{(2)'}\underline{\delta}}{\Delta} = y, \quad \Rightarrow \quad du = \Delta\, dy$. Therefore,

$$\Pr(1|2) = \frac{1}{\sqrt{2\pi}} \int_{(h-\underline{\mu}^{(2)'}\underline{\delta})/\Delta}^{\infty} e^{-y^2/2}\, dy = \frac{1}{\sqrt{2\pi}} \int_{(\underline{\mu}^{(1)}-\underline{\mu}^{(2)})'\underline{\delta}/2\Delta}^{\infty} e^{-y^2/2}\, dy .$$

Sample discriminant function

Suppose that we have a sample $\underline{x}_1^{(1)}, \cdots, \underline{x}_{n_1}^{(1)}$ from $N_p(\underline{\mu}^{(1)}, \Sigma)$ and a sample $\underline{x}_1^{(2)}, \cdots, \underline{x}_{n_2}^{(2)}$ from $N_p(\underline{\mu}^{(2)}, \Sigma)$, and the unbiased estimate of $\underline{\mu}^{(1)}$ is

$\overline{\underline{x}}^{(1)} = \dfrac{1}{n_1} \sum_{\alpha=1}^{n_1} \underline{x}_\alpha^{(1)}$, and $\underline{\mu}^{(2)}$ is $\overline{\underline{x}}^{(2)} = \dfrac{1}{n_2} \sum_{\alpha=1}^{n_2} \underline{x}_\alpha^{(2)}$, and of Σ is S defined by

$$S = \frac{1}{n_1 + n_2 - 2} \left[\sum_{\alpha=1}^{n_1} (\underline{x}_\alpha^{(1)} - \overline{\underline{x}}^{(1)})(\underline{x}_\alpha^{(1)} - \overline{\underline{x}}^{(1)})' + \sum_{\alpha=1}^{n_1} (\underline{x}_\alpha^{(2)} - \overline{\underline{x}}^{(2)})(\underline{x}_\alpha^{(2)} - \overline{\underline{x}}^{(2)})' \right].$$

Substitute these estimates for the parameters in the function $\underline{x}'\underline{\delta}$, Fisher's discriminant function becomes

$$\underline{x}' S^{-1} (\overline{\underline{x}}^{(1)} - \overline{\underline{x}}^{(2)}).$$

This is known as sample discriminant function. The classification procedure now becomes

i) Compute $\underline{x}' S^{-1} (\overline{\underline{x}}^{(1)} - \overline{\underline{x}}^{(2)}) = \underline{x}'\underline{\delta}$

ii) Compute $\dfrac{1}{2} (\overline{\underline{x}}^{(1)} + \overline{\underline{x}}^{(2)})' S^{-1} (\overline{\underline{x}}^{(1)} - \overline{\underline{x}}^{(2)}) = \dfrac{1}{2} (\overline{\underline{x}}^{(1)} + \overline{\underline{x}}^{(2)})'\underline{\delta}$

iii) Assign the individual with measurements $\underline{x}$ to population first or population second, according as $\underline{x}'\underline{\delta} - \dfrac{1}{2} (\overline{\underline{x}}^{(1)} + \overline{\underline{x}}^{(2)})'\underline{\delta}$ is ≥ 0 or < 0.

Classification into one of several populations

Consider the m populations say $\pi_1, \cdots, \pi_m$ with the priori probabilities $q_1, \cdots, q_m$ and the density functions $f_1(\underline{x}), \cdots, f_m(\underline{x})$. We wish to divide the $p-$dimensional space R, in which the point of observation $\underline{x}$ falls, into m mutually exclusive and exhaustive regions $R_1, \cdots, R_m$.

If $\underline{x}$ falls in R_i, then assign the individual to population π_i.

Let

$$\Pr(j|i) = \Pr(\text{an individual belongs to } \pi_i \text{ is misclassified to } \pi_j) = \int_{R_j} f_i(\underline{x})\, d\underline{x}$$

and

$$C(j|i) = \text{Cost of misclassifying an observation from } \pi_i \text{ as coming from } \pi_j.$$

Since the probability that an observation comes from π_i is q_i.

$$\Pr(\text{an observation belongs to } \pi_i \text{ and is classified as coming from } \pi_j)$$

$$= \Pr(\text{an observation comes from } \pi_i) \times \Pr(\text{misclassified it as coming from } \pi_j)$$

$$= q_i \Pr(j|i).$$

The total expected cost from misclassification

$$\sum_{i=1}^{m}\left[\sum_{j=1}^{m} C(j|i)\, q_i \Pr(j|i)\right], \quad i \neq j.$$

We would like to choose regions $R_1, \cdots, R_m$ to make this expected cost minimum.

It can be seen that the classification rule comes out to be

Assign $\underline{x}$ to π_k if

$$\sum_{\substack{i=1 \\ i\neq k}}^{m} q_i f_i(\underline{x}) C(k|i) < \sum_{\substack{i=1 \\ i\neq j}}^{m} q_i f_i(\underline{x}) C(j|i), \; j = 1, 2, \cdots, m; \quad j \neq k$$

Let $\underline{X} \sim N_p(\underline{\mu}^{(i)}, \Sigma)$, $i = 1, 2, \cdots, m$, let the cost of misclassification be equal. Then the rule is assign $\underline{x}$ to π_k if

$$\sum_{\substack{i=1 \\ i\neq k}}^{m} q_i f_i(\underline{x}) < \sum_{\substack{i=1 \\ i\neq j}}^{m} q_i f_i(\underline{x}), \; j = 1, 2, \cdots, m; \quad j \neq k$$

$$\sum_{\substack{i=1 \\ i\neq k}}^{m} q_i f_i(\underline{x}) + q_k f_k(\underline{x}) - q_k f_k(\underline{x}) < \sum_{\substack{i=1 \\ i\neq j}}^{m} q_i f_i(\underline{x}) + q_j f_j(\underline{x}) - q_j f_j(\underline{x})$$

$$\Rightarrow \sum_{\substack{i=1 \\ i\neq k}}^{m} q_i f_i(\underline{x}) + q_k f_k(\underline{x}) = \sum_{i=1}^{m} q_i f_i(\underline{x}) = \sum_{\substack{i=1 \\ i\neq j}}^{m} q_i f_i(\underline{x}) + q_j f_j(\underline{x})$$

$$\Rightarrow q_k f_k(\underline{x}) > q_j f_j(\underline{x})$$

$$\Rightarrow \frac{f_k(\underline{x})}{f_j(\underline{x})} > \frac{q_j}{q_k}.$$

or assign $\underline{x}$ to π_j if $\dfrac{f_j(\underline{x})}{f_k(\underline{x})} > \dfrac{q_k}{q_j}$

or assign $\underline{x}$ to π_j if $U_{jk}(\underline{x}) = \ln \dfrac{f_j(\underline{x})}{f_k(\underline{x})} > \ln \dfrac{q_k}{q_j}$

$$\ln \frac{f_j(\underline{x})}{f_k(\underline{x})} = \ln \frac{\exp[-1/2(\underline{x}-\underline{\mu}^{(j)})'\Sigma^{-1}(\underline{x}-\underline{\mu}^{(j)})]}{\exp[-1/2(\underline{x}-\underline{\mu}^{(k)})'\Sigma^{-1}(\underline{x}-\underline{\mu}^{(k)})]}$$

$$= \left[\underline{x} - \frac{1}{2}(\underline{\mu}^{(j)} + \underline{\mu}^{(k)})\right]' \Sigma^{-1}(\underline{\mu}^{(j)} - \underline{\mu}^{(k)}).$$

Note that

$$U_{kj}(\underline{x}) = \left[\underline{x} - \frac{1}{2}(\underline{\mu}^{(k)} + \underline{\mu}^{(j)})\right]' \Sigma^{-1}(\underline{\mu}^{(k)} - \underline{\mu}^{(j)})$$

Therefore,

$$U_{jk}(\underline{x}) = -U_{kj}(\underline{x}).$$

There are ${}^{m}C_2$ combinations of $U_{jk}(\underline{x})$ for m categories

For $m = 3$, combinations are $U_{12}\ U_{13}\ U_{23}$.

$$\left.\begin{aligned} \ln \frac{f_1(\underline{x})}{f_2(\underline{x})} &= U_{12} > \ln \frac{q_2}{q_1} \\ \ln \frac{f_1(\underline{x})}{f_3(\underline{x})} &= U_{13} > \ln \frac{q_3}{q_1} \end{aligned}\right\} \text{assign } \underline{x} \text{ to } \pi_1$$

$$\left.\begin{aligned} \ln \frac{f_2(\underline{x})}{f_1(\underline{x})} &= U_{21} > \ln \frac{q_1}{q_2} \\ \ln \frac{f_2(\underline{x})}{f_3(\underline{x})} &= U_{23} > \ln \frac{q_3}{q_2} \end{aligned}\right\} \text{assign } \underline{x} \text{ to } \pi_2$$

$$\left.\begin{aligned} \ln \frac{f_3(\underline{x})}{f_1(\underline{x})} &= U_{31} > \ln \frac{q_1}{q_3} \\ \ln \frac{f_3(\underline{x})}{f_2(\underline{x})} &= U_{32} > \ln \frac{q_2}{q_3} \end{aligned}\right\} \text{assign } \underline{x} \text{ to } \pi_3.$$

Sample discriminant function

$\underline{X} \sim N_p(\underline{\mu}^{(i)}, \Sigma)$, $\underline{\mu}^{(i)}$, Σ are unknown then replace $\underline{\mu}^{(i)}$ by $\underline{\overline{x}}^{(i)}$ and Σ by S, where

$$S = \frac{1}{\sum_{i=1}^{m} n_i - m} \sum_{i=1}^{m} \sum_{\alpha=1}^{n_i} (\underline{x}_{i\alpha} \underline{x}'_{i\alpha} - n_i \underline{\overline{x}}^{(i)} \underline{\overline{x}}^{(i)\prime}) \quad \text{or} \quad \left(\sum_{i=1}^{m} n_i - m\right) S = A_1 + A_2 + \cdots + A_m.$$

Substitute these estimates for the parameters in the function, the function becomes

$$V_{jk}(\underline{x}) = \left[\underline{x} - \frac{1}{2}(\underline{\bar{x}}^{(j)} + \underline{\bar{x}}^{(k)})\right]' S^{-1}(\underline{\bar{x}}^{(j)} - \underline{\bar{x}}^{(k)}).$$

HOTELLING'S$-T^2$

If X is univariate normal with mean μ and standard deviation σ, then $U = \dfrac{\sqrt{n}\,(\bar{x}-\mu)}{\sigma} \sim N(0,1)$, and $V = \dfrac{1}{\sigma^2}\sum_i (x_i - \bar{x})^2 = \dfrac{(n-1)\,s^2}{\sigma^2} \sim \chi^2_{n-1}$, where s^2 is the sample variance from a sample of size n. If U and V are independently distributed, then **Student's**$-t$ is defined as

$$t = \frac{U}{\sqrt{V/n-1}} = \frac{\sqrt{n}\,(\bar{x}-\mu)/\sigma}{\sqrt{(n-1)\,s^2/(n-1)\sigma^2}} = \frac{\sqrt{n}\,(\bar{x}-\mu)}{s} \sim t_{n-1}.$$

The multivariate analogue of Student's$-t$ is Hotelling's T^2.

If $\underline{x}_\alpha$ $(\alpha = 1, 2, \cdots, n)$ is an independent sample of size n from $N_p(\underline{\mu}, \Sigma)$ and, if $\underline{\bar{x}}$ is the sample mean vector, S the matrix of variance covariance, then the Hotelling's$-T^2$ is defined by the relation

$$T^2 = n\,(\underline{\bar{x}} - \underline{\mu})'\,S^{-1}\,(\underline{\bar{x}} - \underline{\mu}).$$

$$\frac{T^2}{n-1}\frac{n-p}{p} = \frac{\chi^2_p(\lambda^2)/p}{\chi^2_{n-p}/(n-p)} \sim F_{p,n-p}(\lambda^2).$$

Properties of Hotelling's $-T^2$

T^2 – Statistic as a function of likelihood ratio criterion

Let $\underline{x}_\alpha$ $(\alpha = 1, 2, \cdots, n > p)$ be a random sample of size n from $N_p(\underline{\mu}, \Sigma)$. The likelihood function is

$$L(\underline{\mu}, \Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} \exp\left[-\frac{1}{2}\sum_\alpha (\underline{x}_\alpha - \underline{\mu})' \Sigma^{-1} (\underline{x}_\alpha - \underline{\mu})\right]$$

and the likelihood ratio criterion

$$\lambda = \frac{\max L_\omega (\underline{\mu}, \Sigma)}{\max L_\Omega (\underline{\mu}, \Sigma)} = \frac{\max L_0}{\max L}.$$

In the parameter space Ω, the maximum of L occurs when the parameters $\underline{\mu}$ and Σ are estimated by their maximum likelihood estimators i.e. $\hat{\underline{\mu}} = \underline{\bar{x}}$, and $\hat{\Sigma} = A/n$. In the space ω, we have $\underline{\mu} = \underline{\mu}_0$, and $\hat{\Sigma} = \frac{1}{n}\sum_\alpha (\underline{x}_\alpha - \underline{\mu}_0)(\underline{x}_\alpha - \underline{\mu}_0)'$, therefore,

$$\max L_\Omega = \frac{1}{(2\pi)^{np/2} \left|\frac{A}{n}\right|^{n/2}} \exp\left[-\frac{1}{2}\sum_\alpha (\underline{x}_\alpha - \underline{\bar{x}})' \left(\frac{A}{n}\right)^{-1} (\underline{x}_\alpha - \underline{\bar{x}})\right]$$

$$= \frac{n^{np/2}}{(2\pi)^{np/2} |A|^{n/2}} \exp\left(-\frac{1}{2} np\right).$$

Similarly,

$$\max L_\omega = \frac{1}{(2\pi)^{np/2} \left|\frac{1}{n}\sum_\alpha (\underline{x}_\alpha - \underline{\mu}_0)(\underline{x}_\alpha - \underline{\mu}_0)'\right|^{n/2}}$$

$$\exp\left[-\frac{1}{2} tr\, n \left\{\sum_\alpha (\underline{x}_\alpha - \underline{\mu}_0)(\underline{x}_\alpha - \underline{\mu}_0)'\right\}^{-1} \left\{\sum_\alpha (\underline{x}_\alpha - \underline{\mu}_0)(\underline{x}_\alpha - \underline{\mu}_0)'\right\}\right]$$

$$= \frac{n^{np/2}}{(2\pi)^{np/2} \left|\sum_\alpha (\underline{x}_\alpha - \underline{\mu}_0)(\underline{x}_\alpha - \underline{\mu}_0)'\right|^{n/2}} \exp\left(-\frac{1}{2} np\right).$$

Consider

$$\sum_\alpha (\underline{x}_\alpha - \underline{\mu}_0)(\underline{x}_\alpha - \underline{\mu}_0)' = \sum_\alpha [(\underline{x}_\alpha - \underline{\bar{x}}) + (\underline{\bar{x}} - \underline{\mu}_0)][(\underline{x}_\alpha - \underline{\bar{x}}) + (\underline{\bar{x}} - \underline{\mu}_0)]'$$

$$= A + n(\underline{\bar{x}} - \underline{\mu}_0)(\underline{\bar{x}} - \underline{\mu}_0)'.$$

Hence,

$$\max L_{\omega} = \frac{n^{np/2}}{(2\pi)^{np/2}\left|A + n(\underline{\bar{x}} - \underline{\mu}_0)(\underline{\bar{x}} - \underline{\mu}_0)'\right|^{n/2}} \exp\left(-\frac{1}{2}np\right)$$

Thus, the likelihood ratio criterion is

$$\lambda = \frac{|A|^{n/2}}{\left|A + n(\underline{\bar{x}} - \underline{\mu}_0)(\underline{\bar{x}} - \underline{\mu}_0)'\right|^{n/2}}$$

or $\lambda^{2/n} = \dfrac{|A|}{\begin{vmatrix} 1 & -\sqrt{n}\,(\underline{\bar{x}} - \underline{\mu}_0)' \\ \sqrt{n}\,(\underline{\bar{x}} - \underline{\mu}_0) & A \end{vmatrix}} = \dfrac{|A|}{|A|\left|1 + n(\underline{\bar{x}} - \underline{\mu}_0)'A^{-1}(\underline{\bar{x}} - \underline{\mu}_0)\right|}$,

since $|\Sigma| = |\Sigma_{22}|\left|\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right|$.

$$= \frac{1}{1 + n(\underline{\bar{x}} - \underline{\mu}_0)'A^{-1}(\underline{\bar{x}} - \underline{\mu}_0)}$$

$$= \frac{1}{1 + \frac{n}{n-1}(\underline{\bar{x}} - \underline{\mu}_0)'S^{-1}(\underline{\bar{x}} - \underline{\mu}_0)} = \frac{1}{1 + \frac{T^2}{n-1}},$$

where, $T^2 = n\,(\underline{\bar{x}} - \underline{\mu}_0)'S^{-1}(\underline{\bar{x}} - \underline{\mu}_0) = n(n-1)(\underline{\bar{x}} - \underline{\mu}_0)'A^{-1}(\underline{\bar{x}} - \underline{\mu}_0)$.

The likelihood ratio test is defined by the critical region $\lambda \le \lambda_0$, where, λ_0 is so chosen so as to have level α, i.e. $\Pr[\lambda \le \lambda_0 \mid H_0] = \alpha$.

Thus

$$\lambda^{2/n} \le \lambda_0^{2/n}, \text{ or } \frac{1}{1 + T^2/(n-1)} \le \lambda_0^{2/n}, \quad \text{or } 1 + \frac{T^2}{n-1} \ge \lambda_0^{-n/n},$$

or $T^2 \ge (n-1)(\lambda_0^{-2/n} - 1) = T_0^2$, (say)

$\Rightarrow T^2 \ge T_0^2$.

Therefore, $\Pr[T^2 \ge T_0^2 \mid H_0] = \alpha$.

Invariance property of T^2

Let $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$, then $T_x^2 = n\,(\underline{\bar{x}} - \underline{\mu}_{0x})'S_x^{-1}(\underline{\bar{x}} - \underline{\mu}_{0x})$,

where

$$S_x = \frac{1}{n-1}\sum_{\alpha}(\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' = \frac{1}{n-1}(\underline{x}_1\underline{x}_1' + \cdots + \underline{x}_n\underline{x}_n' - n\underline{\bar{x}}\,\underline{\bar{x}}')$$

Make a non-singular transformation

$$\underline{Y} = C\,\underline{X}, \;\; \Rightarrow \;\; \underline{y}_\alpha = C\,\underline{x}_\alpha$$

Now

$$S_y = \frac{1}{n-1}\sum_\alpha (\underline{y}_\alpha - \underline{\bar{y}})(\underline{y}_\alpha - \underline{\bar{y}})' = \frac{1}{n-1}(\underline{y}_1\underline{y}_1' + \cdots + \underline{y}_n\underline{y}_n' - n\,\underline{\bar{y}}\,\underline{\bar{y}}')$$

$$= \frac{1}{n-1}(C\,\underline{x}_1\underline{x}_1'C' + \cdots + C\,\underline{x}_n\underline{x}_n'C' - nC\,\underline{\bar{x}}\,\underline{\bar{x}}'C')$$

$$= C\left[\frac{1}{n-1}(\underline{x}_1\underline{x}_1' + \cdots + \underline{x}_n\underline{x}_n' - n\,\underline{\bar{x}}\,\underline{\bar{x}}')\right]C'$$

$$= C\,S_x\,C'.$$

By definition

$$T_y^2 = n\,(\underline{\bar{y}} - \underline{\mu}_{0y})'S_y^{-1}(\underline{\bar{y}} - \underline{\mu}_{0y}) = n\,(C\,\underline{\bar{x}} - C\,\underline{\mu}_{0x})'(C\,S_x\,C')^{-1}(C\,\underline{\bar{x}} - C\,\underline{\mu}_{0x})$$

$$= n\,(\underline{\bar{x}} - \underline{\mu}_{0x})'C'C'^{-1}S_x^{-1}C^{-1}C\,(\underline{\bar{x}} - \underline{\mu}_{0x})$$

$$= n\,(\underline{\bar{x}} - \underline{\mu}_{0x})'S_x^{-1}(\underline{\bar{x}} - \underline{\mu}_{0x}) = T_x^2.$$

Uses of T^2 – statistic

One sample problem

Let $\underline{x}_1, \cdots, \underline{x}_n$ be a random sample from $N_p(\underline{\mu}, \Sigma)$, when the variance covariance matrix Σ is unknown. Suppose we are required to test

$H_0: \underline{\mu} = \underline{\mu}_0$ (specific mean vector).

Let

$\underline{Y} = \sqrt{n}\,(\underline{\bar{x}} - \underline{\mu}_0)$, then, $E\underline{Y} = \underline{0}$, under H_0, and $\Sigma_{\underline{Y}} = \Sigma$. Thus, $\underline{Y} \sim N_p(\underline{0}, \Sigma)$ and

$$(n-1)S = \sum_{\alpha=1}^{n}(\underline{x}_\alpha - \underline{\bar{x}})(\underline{x}_\alpha - \underline{\bar{x}})' = A = \sum_{\alpha=1}^{n-1}\underline{Z}_\alpha\underline{Z}_\alpha', \text{ with } \underline{Z}_\alpha \sim N_p(\underline{0}, \Sigma).$$

Therefore, by definition

$$T^2 = \underline{Y}'S^{-1}\underline{Y}, \text{ and } \frac{T^2}{n-1}\frac{n-p}{p} \sim F_{p,n-p}.$$

Thus adopting a significance level of size α, the null hypothesis is rejected if

$$T^2 \ge T_0^2, \text{ where, } T_0^2 = \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha).$$

Two sample problem

Let $\underline{x}_1^{(i)}, \cdots, \underline{x}_n^{(i)}$ be a random sample from $N_p(\underline{\mu}^{(i)}, \Sigma)$, $i = 1, 2$, where the variance covariance matrices are assumed equal but unknown. The hypothesis of interest is $H_0: \underline{\mu}^{(1)} = \underline{\mu}^{(2)}$.

Let

$$\underline{\bar{x}}^{(i)} = \frac{1}{n_i} \sum_{\alpha=1}^{n_i} \underline{x}_\alpha^{(i)} \text{ be the sample mean vector, and}$$

$$\underline{\bar{x}}^{(i)} \sim N_p(\underline{\mu}^{(i)}, \Sigma / n_i) \text{, then, } \underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)} \sim N_p\left[\underline{0}, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\Sigma\right] \text{, under } H_0.$$

$$\Rightarrow \sqrt{\frac{n_1 n_2}{n_1 + n_2}} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) \sim N_p(\underline{0}, \Sigma).$$

Let

$$\underline{Y} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) \text{, then,}$$

$$E\underline{Y} = \underline{0} \text{, under } H_0 \text{, and } \quad \Sigma_{\underline{Y}} = \frac{n_1 n_2}{n_1 + n_2} E(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' = \Sigma.$$

Therefore, $\underline{Y} \sim N_p(\underline{0}, \Sigma)$.

Let

$$S^{(i)} = \frac{1}{n_i - 1} \sum_{\alpha=1}^{n_i} (\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})(\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})' \quad \text{and}$$

$$S = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^{2} \sum_{\alpha=1}^{n_i} (\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})(\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})'$$

$$= \frac{(n_1 - 1) S^{(1)} + (n_2 - 1) S^{(2)}}{n_1 + n_2 - 2} \text{ be the pooled sample variance covariance matrix.}$$

or $(n_1 + n_2 - 2) S = (n_1 - 1) S^{(1)} + (n_2 - 1) S^{(2)}$

$$= A^{(1)} + A^{(2)} = \sum_{\alpha=1}^{n_1 + n_2 - 2} \underline{Z}_\alpha \underline{Z}_\alpha' \text{, with } \underline{Z}_\alpha \sim N_p(\underline{0}, \Sigma).$$

Hence, $(n_1 + n_2 - 2) S$ is distributed as $\sum_{\alpha=1}^{n_1 + n_2 - 2} \underline{Z}_\alpha \underline{Z}_\alpha'$.

Therefore, by definition

$$T^2 = \underline{Y}' S^{-1} \underline{Y} = \frac{n_1 n_2}{n_1 + n_2} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' S^{-1} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})$$

and

$$\frac{T^2}{n_1 + n_2 - 2} \frac{n_1 + n_2 - 2 - (p-1)}{p} \sim F_{p, n_1 + n_2 - p - 1}.$$

Thus adopting a significance level of size α, the null hypothesis is rejected if $T^2 \geq T_0^2$, where, $T_0^2 = \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p, n_1 + n_2 - p - 1}(\alpha)$.

Distribution of Mahalanobis's D^2

The quantity $(\underline{\mu}^{(1)} - \underline{\mu}^{(2)})' \Sigma^{-1} (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})$ is denoted by Δ^2 and was proposed by Mahalanobis as a measure of the distance between the two populations, $N_p(\underline{\mu}^{(1)}, \Sigma)$, and $N_p(\underline{\mu}^{(2)}, \Sigma)$. If the parameters are replaced by their unbiased estimates, is denoted by D^2, which is given by

$$D^2 = (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})' S^{-1} (\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})$$ and is known as Mahalanobis's D^2,

where

$$S = \frac{(n_1 - 1) S^{(1)} + (n_2 - 1) S^{(2)}}{n_1 + n_2 - 2} = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^{2} \sum_{\alpha=1}^{n_i} (\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})(\underline{x}_\alpha^{(i)} - \underline{\bar{x}}^{(i)})',$$

where $\underline{\bar{x}}^{(i)} = \frac{1}{n_i} \sum_{\alpha=1}^{n_i} \underline{x}_\alpha^{(i)}$, $i = 1, 2$.

It is obvious that

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} D^2,$$

i.e. two-sample T^2 and D^2 are almost the same, except for the constant $k^2 = \frac{n_1 n_2}{n_1 + n_2}$.

Let

$\underline{Y} = k(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)})$, then expected value of $\underline{Y}$ is $E(\underline{Y}) = k(\underline{\mu}^{(1)} - \underline{\mu}^{(2)}) = \underline{\delta}$

and the variance covariance matrix of $\underline{Y}$ is

$$\Sigma_{\underline{Y}} = k^2 E[(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) - (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})][(\underline{\bar{x}}^{(1)} - \underline{\bar{x}}^{(2)}) - (\underline{\mu}^{(1)} - \underline{\mu}^{(2)})]'$$

$$= k^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right) \Sigma = \Sigma, \text{ because } \frac{n_1 + n_2}{n_1 n_2} = \frac{1}{k^2}.$$

Therefore,

$$\underline{Y} \sim N_p(\underline{\delta}, \Sigma)\text{, then,}\quad k^2 D^2 = \underline{Y}' S^{-1} \underline{Y}.$$

Since Σ is positive definite matrix there exist a nonsingular matrix C such that

$$C\Sigma C' = I \;\Rightarrow\; CC' = \Sigma^{-1}.$$

Define,

$\underline{Y}^* = C\underline{Y}$, $S^* = C\,S\,C'$, and $\underline{\delta}^* = C\,\underline{\delta}$, then,

$k^2 D^2 = \underline{Y}^{*'} S^{*-1} \underline{Y}^*$, and the expected value of $\underline{Y}^*$ is

$E(\underline{Y}^*) = C\,E(\underline{Y}) = C\underline{\delta} = \underline{\delta}^*$, and the variance covariance matrix of $\underline{Y}^*$ is

$$\Sigma_{\underline{Y}^*} = C\,E[\underline{Y} - E(\underline{Y})][(\underline{Y} - E(\underline{Y})]' C' = C\,\Sigma\,C' = I.$$

Thus,

$$\underline{Y}^* \sim N_p(\underline{\delta}^*, I), \;\Rightarrow\; \underline{Y}^{*'} \underline{Y}^* \sim \chi_p^2(\underline{\delta}^{*'} \underline{\delta}^*),$$

where

$$\underline{\delta}^{*'} \underline{\delta}^* = \underline{\delta}' C' C \underline{\delta} = \underline{\delta}' \Sigma^{-1} \underline{\delta} = \lambda^2.$$

Let

$$(n_1 + n_2 - 2)\,S = \sum_{\alpha=1}^{n_1+n_2-2} \underline{Z}_\alpha \underline{Z}_\alpha'\text{, where } \underline{Z}_\alpha \sim N_p(\underline{0}, \Sigma)$$

$$\Rightarrow\; (n_1 + n_2 - 2)\,S^* = \sum_{\alpha=1}^{n_1+n_2-2} (C\underline{Z}_\alpha)(C\underline{Z}_\alpha)'\text{, where } C\underline{Z}_\alpha \sim N_p(\underline{0}, I).$$

Therefore,

$$k^2 D^2 = \underline{Y}^* S^{*-1} \underline{Y}^* = (n_1 + n_2 - 2)\frac{\chi_p^2(\lambda^2)}{\chi^2_{n_1+n_2-2-(p-1)}}$$

$$\Rightarrow\; \frac{n_1 n_2}{n_1 + n_2} \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)\,p} D^2 = \frac{\chi_p^2(\lambda^2)/p}{\chi^2_{n_1+n_2-2-(p-1)}/n_1 + n_2 - p - 1} \sim F_{p, n_1+n_2-p-1}(\lambda^2).$$

If $\underline{\mu}^{(1)} = \underline{\mu}^{(2)}$, then the $F-$ distribution is central.

Date: ______ Page: ______

<u>Theorem</u> If $A \sim \sum_{\alpha=1}^{n} Z_\alpha Z_\alpha'$, where Z_i's are independent each with distribution $N_p(\underline{0}, \Sigma)$ then $A_{11\cdot 2} \sim \sum_{\alpha=1}^{n-(p-q)} U_\alpha U_\alpha'$ where U_α's are independent and distributed as $N_q(\underline{0}, \Sigma_{11\cdot 2})$; where

$$A_{11\cdot 2} = A_{11} - A_{12} A_{22}^{-1} A_{21}$$

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{array}{l} q \text{ rows} \\ p-q \text{ rows} \end{array}$$

If $A \sim W_p(n, \Sigma)$ then $A_{11\cdot 2} \sim W_p(n-(p-q), \Sigma_{11\cdot 2})$.

<u>Proof</u>- $$A = \sum_{\alpha=1}^{n} Z_\alpha Z_\alpha'$$

Let $$Z_\alpha = \begin{pmatrix} Z_\alpha^{(1)} \\ Z_\alpha^{(2)} \end{pmatrix} \begin{array}{l} q \\ p-q \end{array}$$

Then $$A = \sum_{\alpha=1}^{n} \begin{pmatrix} Z_\alpha^{(1)} \\ Z_\alpha^{(2)} \end{pmatrix} \begin{pmatrix} Z_\alpha^{(1)'} & Z_\alpha^{(2)'} \end{pmatrix}$$

$$= \begin{bmatrix} WW' & WY' \\ YW' & YY' \end{bmatrix} \quad \text{—(1)}$$

where $W = [Z_1^{(1)} \; Z_2^{(1)} \; \dots \; Z_n^{(1)}]_{q \times n}$

$Y = [Z_1^{(2)} \; Z_2^{(2)} \; \dots \; Z_n^{(2)}]_{(p-q) \times n}$

Let F be a non-singular matrix s.t.

$FYY'F' = I_{p-q}$ —(2)

& $FY = G_{2\,(p-q)\times n}$ —(3)

Therefore $G_2 G_2' = I_{p-q}$ —(4)

$\Rightarrow$ Rows of G_2 are linearly independent.

Date: ______ Page: ______

$\exists$ a matrix $G_{1\,(n-(p-q))\times n}$ s.t. $G_{n\times n} = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix}$ is an orthogonal matrix.

Consider the transformation $U_{q\times n} = W G'$ — (5)

$U = [U_1 \;\; U_2 \; \cdots \; U_n]$

Let $G = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix} \Rightarrow G' = [g_1' \;\; g_2' \; \cdots \; g_n']$

$\therefore \; U_1 = W g_1', \quad U_2 = W g_2', \; \cdots$

$\Rightarrow \mathrm{Cov}(U_1, U_2) = E[U_1 U_2'] = E[W g_1' g_2 W'] = 0$

($\because g_1' g_2 = 0$ as G is orthogonal)

Thus U_α; $\alpha = 1, 2, \ldots, n$ are independent.

Since U_α's are linear transformation of normal vectors so have normal distribution.

The conditional mean of $U | Y = y$ is

$$E[U | Y = y] = E[WG' | Y = y] = E[W | Y = y] G'$$

($\because$ for given $Y = y$; G is constant)

We know that if $X = \begin{bmatrix} X^{(1)} \\ X^{(2)} \end{bmatrix} \sim N_p(\mu, \Sigma)$ then

$$E[X^{(1)} | X^{(2)} = x^{(2)}] = \mu^{(1)} + \Sigma_{12} \Sigma_{22}^{-1} (x^{(2)} - \mu^{(2)})$$

or $E(Z_\alpha^{(1)} | Z_\alpha^{(2)} = z_\alpha^{(2)}) = \Sigma_{12} \Sigma_{22}^{-1} z_\alpha^{(2)}$

Date: ______ Page: ______

$\therefore E[U|Y=y] = \Sigma_{12}\Sigma_{22}^{-1} y G'$

$= \underset{q\times(p-q)}{\beta} \underset{(p-q)\times n}{y} \underset{n\times n}{G'}$

By eq^n ③,

$FY = G_2 \Rightarrow Y = F^{-1}G_2$

$\therefore E[U|Y=y] = \beta F^{-1} G_2 G'$

$= \beta F^{-1} G_2 (G_1' \; G_2')$

$= (\beta F^{-1} \underset{(p-q)\times n}{G_2 G_1'} \quad \beta F^{-1} \underset{n\times(n-(p-q))}{G_2 G_2'})$

$\because G$ is orthogonal $\Rightarrow G_2 G_1' = 0$

& $G_2 G_2' = I_{p-q}$

$\Rightarrow E[U|Y=y] = (\underset{q\times(n-(p-q))}{0} \quad \underset{q\times(p-q)}{\beta F^{-1}})$ — ⑥

Thus, all $U_1, U_2, \dots U_{n-(p-q)}$ vectors have mean 0 and $U_{n-(p-q)+1}, \dots, U_n$ have mean βF^{-1} which is non-zero.

Cov. matrix of $U_\alpha | Y=y$ is $\Sigma_{11\cdot2}$. This doesn't depend on Y.

$\therefore U_\alpha \sim N_q(0, \Sigma_{11\cdot2})$ for $\alpha = 1, 2, \dots, n-(p-q)$

and for $\alpha = n-(p-q)+1, \dots, n$ $U_\alpha \sim N_q(\beta F^{-1}, \Sigma_{11\cdot2})$

Now $A_{11} = WW' = UGG'U' = UU' = \sum_{\alpha=1}^{n} U_\alpha U_\alpha'$

$(\because U = WG' \Rightarrow UG = WG'G \Rightarrow UG = W)$

and $A_{12}A_{22}^{-1}A_{21} = WY'(YY')^{-1}YW'$

$= UG(F^{-1}G_2)'(YY')^{-1}F^{-1}G_2G'U'$

$= UGG_2'F^{-1'}(YY')^{-1}F^{-1}G_2G'U'$

Date: ________ Page: ______

$$A_{12} A_{22}^{-1} A_{21} = U G G_2' (F Y Y' F')^{-1} G_2 G' U'$$

$$= U \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} G_2' \, I_{p-q} \, G_2 (G_1' \; G_2') U'$$

$$= U \begin{pmatrix} \underline{0} \\ I_{p-q} \end{pmatrix} (\underline{0} \quad I_{p-q}) U'$$

$$= U \begin{pmatrix} \underline{0} & \underline{0} \\ \underline{0} & I_{p-q} \end{pmatrix} U'$$

$$= \sum_{\alpha = n-(p-q)+1}^{n} U_\alpha U_\alpha'$$

& $$A_{11\cdot 2} = A_{11} - A_{12} A_{22}^{-1} A_{21}$$

$$= \sum_{\alpha=1}^{n} U_\alpha U_\alpha' - \sum_{\alpha=n-(p-q)+1}^{n} U_\alpha U_\alpha'$$

$$= \sum_{\alpha=1}^{n-(p-q)} U_\alpha U_\alpha'$$

where $U_\alpha \sim N_q(\underline{0}, \Sigma_{11\cdot 2})$ for $\alpha = 1, 2, \ldots n-(p-q)$

$\Rightarrow A_{11\cdot 2} \sim W_q (n-(p-q), \Sigma_{11\cdot 2})$

Here $A_{11\cdot 2}$ & $A_{12} A_{22}^{-1} A_{21}$ are independent.

PRINCIPAL COMPONENTS

1 Introduction

A principal component analysis is concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables. Its general objectives are (1) data reduction and (2) interpretation.
Although p components are required to reproduce the total system variability, often much of this variability can be accounted for by a small number k of the principal components. If so, there is (almost) as much information in the k components as there is in the original p variables. The$k$ principal components can then replace the initial p variables, and the original data set, consisting of n measurements on p variables, is reduced to a data set consisting of n measurements on k principal components.
An analysis of principal components often reveals relationships that were not previously suspected and thereby allows interpretations that would not ordinarily result.

2 Population Principal Components

Algebraically, principal components are particular linear combinations of the p random variablesX_1, $X_2, \ldots, X_p$. Geometrically, these linear combinations represent the selection of a new coordinate system obtained by rotating the original system with $X_1, X_2, \ldots, X_p$ as the coordinate axes. The new axes represent the directions with maximum variability and provide a simpler and more parsimonious description of the covariance structure.
As we shall see, principal components depend solely on the covariance matrix $\mathbf{\Sigma}$ (or the correlation matrix $\boldsymbol{\rho}$) of $X_1, X_2, \ldots, X_p$. Their development does not require a multivariate normal assumption. Let the random vector $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$ have the covariance matrix $\mathbf{\Sigma}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ Consider the linear combinations

$$\begin{aligned}
Y_1 &= \mathbf{a}_1'\mathbf{X} = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_P \\
Y_2 &= \mathbf{a}_2'\mathbf{X} = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\
&\vdots \qquad\qquad\qquad \vdots \\
Y_p &= \mathbf{a}_p'\mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pP}X_P
\end{aligned} \tag{2.1}$$

Then, we obtain

$$\operatorname{Var}(Y_i) = \mathbf{a}_i'\mathbf{\Sigma}\mathbf{a}_i \quad i = 1, 2, \ldots, p \tag{2.2}$$

$$\operatorname{Cov}(Y_i, Y_k) = \mathbf{a}_i'\mathbf{\Sigma}\mathbf{a}_k \quad i, k = 1, 2, \ldots, p \tag{2.3}$$

The principal components are those *uncorrelated* linear combinations $Y_1, Y_2, \ldots, Y_p$ whose variances in (2.2) are as large as possible.
The first principal component is the linear combination with maximum variance. That is, it maximizes Var $(Y_1) = \mathbf{a}_1'\boldsymbol{\Sigma}\mathbf{a}_1$. It is clear that $\mathrm{Var}\,(Y_1) = \mathbf{a}_1'\boldsymbol{\Sigma}\mathbf{a}_1$ can be increased by multiplying any a $_1$ by some constant. To eliminate this indeterminacy, it is convenient to restrict attention to coefficient vectors of unit length. We therefore define

$$\begin{aligned} \text{First principal component} &= \text{linear combination } a_1'\mathbf{X} \text{ that maximizes} \\ & \quad \mathrm{Var}\,(\mathbf{a}_1'\mathbf{X}) \ \text{ subject to } \mathbf{a}_1'\mathbf{a}_1 = 1 \\ \text{Second principal component} &= \text{linear combination a } _2'\mathbf{X} \text{ that maximizes} \\ & \quad \mathrm{Var}\,(\mathbf{a}_2'\mathbf{X}) \ \text{ subject to } \mathbf{a}_2'\mathbf{a}_2 = 1 \text{ and} \\ & \quad \mathrm{Cov}\,(\mathbf{a}_1'\mathbf{X}, \mathbf{a}_2'\mathbf{X}) = 0 \end{aligned}$$

At the ith step.

$$\begin{aligned} i\text{th principal component} \ &= \ \text{linear combination } a_i'\mathbf{X} \text{ that maximizes} \\ & \quad \mathrm{Var}\,(\mathbf{a}_i'\mathbf{X}) \ \text{ subject to a } _i'\mathbf{a}_i = 1 \text{ and} \\ & \quad \mathrm{Cov}\,(\mathbf{a}_i'\mathbf{X}, \mathbf{a}_k'\mathbf{X}) = 0 \quad \text{ for } \quad k < i \end{aligned}$$

Result 2.1. Let $\boldsymbol{\Sigma}$ be the covariance matrix associated with the random vector $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$. Let $\boldsymbol{\Sigma}$ have the eigenvalue-eigenvector pairs $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Then the *ith principal component* is given by

$$Y_i = \mathbf{e}_i'\mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X_p, \quad i = 1, 2, \ldots, p \tag{2.4}$$

With these choices,

$$\begin{aligned} \mathrm{Var}\,(Y_i) &= \mathbf{e}_i'\boldsymbol{\Sigma}\mathbf{e}_i = \lambda_i \quad i = 1, 2, \ldots, p \\ \mathrm{Cov}\,(Y_i, Y_k) &= \mathbf{e}_i'\boldsymbol{\Sigma}\mathbf{e}_k = 0 \quad i \neq k \end{aligned} \tag{2.5}$$

If some λ_i are equal, the choices of the corresponding coefficient vectors, $\mathbf{e}_i$, and hence Y_i, are not unique.

Result 2.2. Let $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$ have covariance matrix $\boldsymbol{\Sigma}$, with eigenvalueeigenvector pairs $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Let $Y_1 = \mathbf{e}_1'\mathbf{X}, Y_2 = \mathbf{e}_2'\mathbf{X}, \ldots, Y_p = \mathbf{e}_p'\mathbf{X}$ be the principal components. Then

$$\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp} = \sum_{i=1}^{p} \mathrm{Var}\,(X_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} \mathrm{Var}\,(Y_i)$$

Result 2.2 says that

$$\begin{aligned}\text{Total population variance } &= \sigma_{11} + \sigma_{22} + \cdots + \sigma_{pP} \\ &= \lambda_1 + \lambda_2 + \cdots + \lambda_p \end{aligned} \tag{2.6}$$

and consequently, the proportion of total variance due to (explained by) the k th principal component is

$$\left(\begin{array}{c}\text{Proportion of total} \\ \text{population variance} \\ \text{due to } k \text{ th principal} \\ \text{component}\end{array}\right) = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p} \quad k = 1, 2, \ldots, p \tag{2.7}$$

If most (for instance, 80 to 90%) of the total population variance, for large p, can be attributed to the first one, two, or three components, then these components can "replace" the original p variables without much loss of information.

Each component of the coefficient vector $\mathbf{e}_i' = [e_{i1}, \ldots, e_{ik}, \ldots, e_{ip}]$ also merits inspection. The magnitude of e_{ik} measures the importance of the k th variable to the ith principal component, irrespective of the other variables. In particular, e_{ik} is proportional to the correlation coefficient between Y_i and X_k.

Result 2.3. If $Y_1 = \mathbf{e}_1'\mathbf{X}, Y_2 = \mathbf{e}_2'\mathbf{X}, \ldots, Y_p = \mathbf{e}_p'\mathbf{X}$ are the principal components obtained from the covariance matrix $\mathbf{\Sigma}$, then

$$\rho_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \quad i, k = 1, 2, \ldots, p \tag{2.8}$$

are the correlation coefficients between the components Y_i and the variables X_k. Here $(\lambda_1, \mathbf{e}_1)$, $(\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ are the eigenvalue-eigenvector pairs for $\mathbf{\Sigma}$.

Although the correlations of the variables with the principal components often help to interpret the components, they measure only the univariate contribution of an individual X to a component Y. That is, they do not indicate the importance of an X to a component Y in the presence of the other X' s.

Example 2.1. (Calculating the population principal components) Suppose the random variables X_1, X_2 and X_3 have the covariance matrix

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

It may be verified that the eigcnvalue-cigenvector pairs are

$$\lambda_1 = 5.83, \mathrm{e}_1' = [.383, -924, 0]$$

$$\lambda_2 = 2.00, \mathrm{e}_2' = [0, 0, 1]$$

$$\lambda_3 = 0.17, \mathrm{e}_3' = [.924, 383, 0]$$

Therefore, the principal components become

$$
\begin{aligned}
Y_1 &= e_1'\mathbf{X} = .383X_1 - 924X_2 \\
Y_2 &= e_2'\mathbf{X} = X_3 \\
Y_3 &= e_j'\mathbf{X} = .924X_1 + 383X_2
\end{aligned}
$$

The variable X_3 is one of the principal components, because it is uncorrelated with the other two variables. Equation (8-5) can be demonstrated from first principles. For example,

$$
\begin{aligned}
\text{Var}\,(Y_1) &= \text{Var}\,(.383X_1 - .924X_2) \\
&= (.383)^2\,\text{Var}\,(X_1) + (-924)^2\,\text{Var}\,(X_2) \\
&+ 2(.383)(-924)\,\text{Cov}\,(X_1, X_2) \\
&= .147(1) + .854(5) - .708(-2) \\
&= 5.83 = \lambda_1 \\
\text{Cov}\,(Y_1, Y_2) &= \text{Cov}\,(.383X_1 - .924X_2, X_3) \\
&= .383\,\text{Cov}\,(X_1, X_3) - .924\,\text{Cov}\,(X_2, X_3) \\
&= .383(0) - .924(0) = 0
\end{aligned}
$$

It is also readily apparent that

$$\sigma_{11} + \sigma_{22} + \sigma_{33} = 1 + 5 + 2 = \lambda_1 + \lambda_2 + \lambda_3 = 5.83 + 2.00 + .17$$

validating Equation (2.6) for this example. The proportion of total variance accounted for by the first principal component is $\lambda_1/\,(\lambda_1 + \lambda_2 + \lambda_3) = 5.83/8 = .73$. Further, the first two components account for a proportion $(5.83+2)/8 = .98$ of the population variance. In this case, the components Y_1 and Y_2 could replace the original three variables with little loss of information.
Next, using (2.8), we obtain

$$
\begin{aligned}
\rho_{Y_1,X_1} &= \frac{e_{11}\sqrt{\lambda_1}}{\sqrt{\sigma_{11}}} = \frac{.383\sqrt{5.83}}{\sqrt{1}} = .925 \\
\rho_{Y_1,X_2} &= \frac{e_{12}\sqrt{\lambda_1}}{\sqrt{\sigma_{22}}} = \frac{-.924\sqrt{5.83}}{\sqrt{5}} = -.998
\end{aligned}
$$

Notice here that the variable X_2, with coefficient $-.924$, receives the greatest weight in the component Y_1. It also has the largest correlation (in absolute value) with Y_1. The correlation of X_1, with Y_1, .925, is almost as large as that for X_2, indicating that the variables are about equally important to the first principal component. The relative sizes of the coefficients of X_1 and X_2 suggest, however, that X_2 contributes more to the determination of Y_1 than does X_1. since, in this case, both coefficients are reasonably large and they have opposite signs, we would argue that both variables aid in

the interpretation of Y_1.
Finally,

$$\rho_{Y_2,x_1} = \rho_{Y_2,X_2} = 0 \quad \text{and} \quad \rho_{Y_2,X_3} = \frac{\sqrt{\lambda_2}}{\sqrt{\sigma_{33}}} = \frac{\sqrt{2}}{\sqrt{2}} = 1 \quad \text{(as it should)}$$

The remaining correlations can be neglected, since the third component is unimportant.

Lesson 13: Canonical Correlation Analysis

Overview

Canonical correlation analysis is a method for exploring the relationships between two multivariate sets of variables (vectors), all measured on the same individual.

Consider, as an example, variables related to exercise and health. On one hand, you have variables associated with exercise, observations such as the climbing rate on a stair stepper, how fast you can run a certain distance, the amount of weight lifted on bench press, the number of push-ups per minute, etc. On the other hand, you have variables that attempt to measure overall health, such as blood pressure, cholesterol levels, glucose levels, body mass index, etc. Two types of variables are measured and the relationships between the exercise variables and the health variables are of interest.

As a second example consider variables measured on environmental health and environmental toxins. A number of environmental health variables such as frequencies of sensitive species, species diversity, total biomass, productivity of the environment, etc. may be measured and a second set of variables on environmental toxins are measured, such as the concentrations of heavy metals, pesticides, dioxin, etc.

For a third example consider a group of sales representatives, on whom we have recorded several sales performance variables along with several measures of intellectual and creative aptitude. We may wish to explore the relationships between the sales performance variables and the aptitude variables.

One approach to studying relationships between the two sets of variables is to use canonical correlation analysis which describes the relationship between the first set of variables and the second set of variables. We do not necessarily think of one set of variables as independent and the other as dependent, though that may potentially be another approach.

Objectives

Upon completion of this lesson, you should be able to:

- Carry out a canonical correlation analysis using SAS (Minitab does not have this functionality);
- Assess how many canonical variate pairs should be considered;
- Interpret canonical variate scores;
- Describe the relationships between variables in the first set with variables in the second set.

Next »

13.1 - Setting the Stage for Canonical Correlation Analysis

What motivates canonical correlation analysis?

It is possible to create pairwise scatter plots with variables in the first set (e.g., exercise variables), and variables in the second set (e.g., health variables). But if the dimension of the first set is *p* and that of the second set is *q*, there will be *pq* such scatter plots, it may be difficult, if not impossible, to look at all of these graphs together and interpret the results.

Similarly, you could compute all correlations between variables from the first set (e.g., exercise variables), and variables in the second set (e.g., health variables), however interpretation is difficult when *pq* is large.

Canonical Correlation Analysis allows us to summarize the relationships into a lesser number of statistics while preserving the main facets of the relationships. In a way, the motivation for canonical correlation is very similar to principal component analysis. It is another dimension reduction technique.

Canonical Variates

Let's begin with the notation:

We have two sets of variables X and Y.

Suppose we have *p* variables in set 1: $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$

and suppose we have *q* variables in set 2: $\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{pmatrix}$

We select X and Y based on the number of variables that exist in each set so that $p \le q$. This is done for computational convenience.

We look at linear combinations of the data, similar to principal components analysis. We define a set of linear combinations named *U* and *V*. *U* corresponds to the linear combinations from the first set of variables, *X*, and *V* corresponds to the second set of variables, *Y*. Each member of *U* is paired with a member of *V*. For example, U_1 below is a linear combination of the *p X* variables and V_1 is the corresponding linear combination of the *q Y* variables. Similarly, U_2 is a linear combination of the *p X* variables, and V_2 is the corresponding linear combination of the *q Y* variables. And, so on....

$$\begin{aligned} U_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ U_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\vdots \\ U_p &= a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \\ \\ V_1 &= b_{11}Y_1 + b_{12}Y_2 + \dots + b_{1q}Y_q \\ V_2 &= b_{21}Y_1 + b_{22}Y_2 + \dots + b_{2q}Y_q \\ &\vdots \\ V_p &= b_{p1}Y_1 + b_{p2}Y_2 + \dots + b_{pq}Y_q \end{aligned}$$

Thus define

$$(U_i, V_i)$$

as the i^{th} *canonical variate pair*. (U_1, V_1) is the first canonical variate pair, similarly (U_2, V_2) would be the second canonical variate pair and so on. With $p \le q$ there are p canonical covariate pairs.

We hope to find linear combinations that maximize the correlations between the members of each canonical variate pair.

We compute the variance of U_i variables with the following expression:

$$\text{var}(U_i) = \sum_{k=1}^{p}\sum_{l=1}^{p} a_{ik}a_{il}cov(X_k, X_l)$$

The coeffcients a^{i1} through a^{ip} that appear in the double sum are the same coefficients that appear in the definition of U_i. The covariances between the k^{th} and l^{th} *X*-variables are multiplied by the corresponding coefficients a^{ik} and a^{il} for the variate U_i.

Similar calculations can be made for the variance of V_j as shown below:

$$\text{var}(V_j) = \sum_{k=1}^{p}\sum_{l=1}^{q} b_{jk}b_{jl}\text{cov}(Y_k, Y_l)$$

The covariance between U_i and V_j is:

$$\text{cov}(U_i, V_j) = \sum_{k=1}^{p}\sum_{l=1}^{q} a_{ik}b_{jl}\text{cov}(X_k, Y_l)$$

The correlation between U_i and V_j is calculated using the usual formula. We take the covariance between the two variables and divide it by the square root of the product of the variances:

$$\frac{\text{cov}(U_i, V_j)}{\sqrt{\text{var}(U_i)\text{var}(V_j)}}$$

The *canonical correlation* is a specific type of correlation. The canonical correlation for the i^{th} canonical variate pair is simply the correlation between U_i and V_i:

$$\rho_i^* = \frac{\text{cov}(U_i, V_i)}{\sqrt{\text{var}(U_i)\text{var}(V_i)}}$$

This is the quantity to maximize. We want to find linear combinations of the *X*s and linear combinations of the *Y*s that maximize the above correlation.

Canonical Variates Defined

Let us look at each of the *p* canonical variates pair one by one.

First canonical variate pair: (U_1, V_1):

The coefficients $a_{11}, a_{12}, \dots, a_{1p}$ and $b_{11}, b_{12}, \dots, b_{1q}$ are selected to maximize the canonical correlation ρ_1^* of the first canonical variate pair. This is subject to the constraint that variances of the two canonical variates in that pair are equal to one.

$$\text{var}(U_1) = \text{var}(V_1) = 1$$

This is required to obtain unique values for the coefficients.

Second canonical variate pair: (U_2, V_2)

Similarly we want to find the coefficients $a_{21}, a_{22}, \dots, a_{2p}$ and $b_{21}, b_{22}, \dots, b_{2q}$ that maximize the canonical correlation ρ_2^* of the second canonical variate pair, (U_2, V_2). Again, we will maximize this canonical correlation subject to the constraints that the variances of the individual canonical variates are both equal to one. Furthermore, we require the additional constraints that (U_1, U_2), and (V_1, V_2) are uncorrelated. In addition, the combinations (U_1, V_2) and (U_2, V_1) must be uncorrelated. In summary, our constraints are:

$$\text{var}(U_2) = \text{var}(V_2) = 1,$$

$$\text{cov}(U_1, U_2) = \text{cov}(V_1, V_2) = 0,$$

$$\text{cov}(U_1, V_2) = \text{cov}(U_2, V_1) = 0.$$

Basically, we require that all of the remaining correlations equal zero.

This procedure is repeated for each pair of canonical variates. In general, ...

i^{th} **canonical variate pair:** (U_i, V_i)

We want to find the coefficients $a_{i1}, a_{i2}, \dots, a_{ip}$ and $b_{i1}, b_{i2}, \dots, b_{iq}$ that maximize the canonical correlation ρ_i^* subject to the constraints that

$$\text{var}(U_i) = \text{var}(V_i) = 1,$$

$$\text{cov}(U_1, U_i) = \text{cov}(V_1, V_i) = 0,$$

$$\text{cov}(U_2, U_i) = \text{cov}(V_2, V_i) = 0,$$

$$\vdots$$

$$\text{cov}(U_{i-1}, U_i) = \text{cov}(V_{i-1}, V_i) = 0,$$

$$\text{cov}(U_1, V_i) = \text{cov}(U_i, V_1) = 0,$$

$$\text{cov}(U_2, V_i) = \text{cov}(U_i, V_2) = 0,$$

$$\vdots$$

$$\text{cov}(U_{i-1}, V_i) = \text{cov}(U_i, V_{i-1}) = 0.$$

Again, requiring all of the remaining correlations to be equal to zero.

Next, let's see how this is carried out in SAS...

« Previous Next »