Techniques and Applications of Multivariate Analysis

6. Discriminant Analysis (DCA)

Contents

6.1 Understanding DCA

[Definition] **Discrimination & Classification**

Multivariate techniques concerned with separating distinct sets of observations and with allocating new observations to previously defined groups.

[Goals]

function of variables

- **Discrimination** : Finding a good discriminant to accurately predict group membership of observations.
- **Classification** : Deriving a rule for allocating observations to the labeled groups

.

- The goals of two techniques frequently overlap, and the distinction bewteen **separation** and **allocation** becomes unclear.

- R.A. Fisher(1938) introduced the term discrimination in the first modern treatment of separative problem.

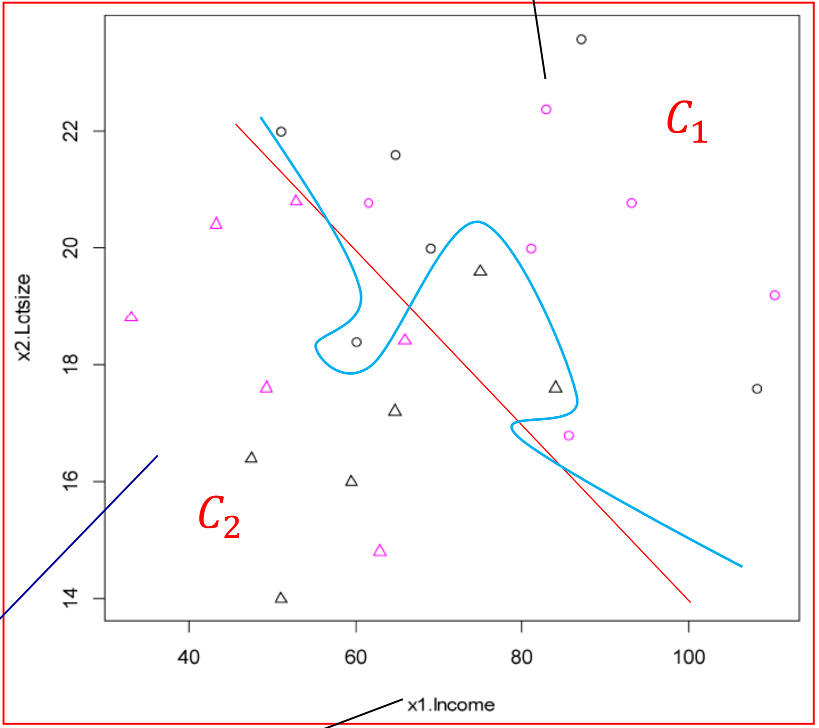
6.2 DCA with Two Clusters

Discriminating 12 owners from 12 non-owners of riding mowers [Table 6.3.1]

Identify the best sales prospects for an intensive sales campaign

Owners tends to have larger incomes and bigger lots than nonowners

C_1 : Owners Income ($1000)	C_1 : Owners Lot size ($1000ft^2$)	C_2 : Non-owners Income	C_2 : Non-owners Lot size
60.0	18.4	75.0	19.6
85.5	16.8	52.8	20.8
64.8	21.6	64.8	17.2
61.5	20.8	43.2	20.4
87.0	23.6	84.0	17.6
110.1	19.2	49.2	17.6
108.0	17.6	59.4	16.0
82.8	22.4	66.0	18.4
69.0	20.0	47.4	16.4
93.0	20.8	33.0	18.8
51.0	22.0	51.0	14.0
81.0	20.0	63.0	14.8

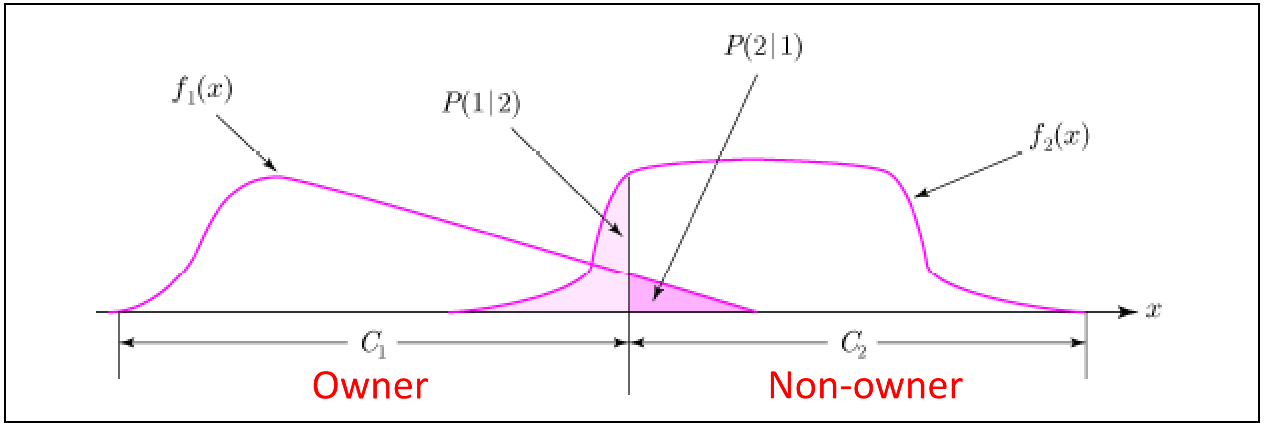


- Two C1 and C2 are determined by the red solid line
- Allocate observations to C1 and C2, respectively.
- 1 owner would be incorrectly classified as C2 and, conversely, 2 nonowners as C1.

Income seems to be a better discriminator than lot size

Misclassification Error Rate = 100(1 + 2)/(12+12)% = 12.5%

6.2 DCA with Two Clusters



[Table 6.2.1] The conditional probability of misclassification about two clusters in case $p=1$: $P(1|2)$, $P(2|1)$

$f_1(\boldsymbol{x})$ *and* $f_2(\boldsymbol{x})$: *pdfs associated with the p x 1 vector of random variables for the populations* C_1 *and* C_2*, respectively*

$$P(observation \ comes \ from \ C_1 \ and \ is \ misclassified \ as \ C_2)$$
$$= P(2|1) = P(X \in C_2 | C_1) = \int_{C_2 = \Omega - C_1} f_1(\boldsymbol{x}) d\boldsymbol{x} = 1 - \int_{C_1} f_1(\boldsymbol{x}) d\boldsymbol{x}$$

$$\Omega = C_1 \cup C_2$$

$$P(observation \ comes \ from \ C_2 \ and \ is \ misclassified \ as \ C_1)$$
$$= P(1|2) = P(X \in C_1 | C_2) = \int_{C_1} f_2(\boldsymbol{x}) d\boldsymbol{x} = 1 - \int_{C_2} f_2(\boldsymbol{x}) d\boldsymbol{x}$$

6.2 DCA with Two Clusters

Prior probability x **Conditional classification probability**

Overall probabilities of correctly or incorrectly classifying observations

p_1 : prior probability of C_1 $\Rightarrow p_1 + p_2 = 1 \Leftarrow$ p_2 : prior probability of C_2

$P(\text{observation is correctly classified as } C_1)$
$= P(X \in C_1 | C_1) P(C_1) = P(1|1) p_1$

$P(\text{observation is correctly classified as } C_2)$
$= P(X \in C_2 | C_2) P(C_2) = P(2|2) p_2$

$P(\text{observation is misclassified as } C_1)$
$= P(X \in C_1 | C_2) P(C_2) = P(1|2) p_2$

$P(\text{observation is misclassified as } C_2)$
$= P(X \in C_2 | C_1) P(C_1) = P(2|1) p_1$

Expected Cost of Misclassification (**ECM**) = $c(2|1)P(2|1)p_1 + c(1|2)P(1|2)p_2$

Classification Rule: making the ECM as small as possible

Cost matrix: usually costs are unknown

Misclassification Cost		Classify as: C_1	Classify as: C_2
True population	C_1	0	$c(2\|1)$
	C_2	$c(1\|2)$	0

Deriving the Minimum ECM Rule

$$
\begin{aligned}
ECM &= c(2|1)P(2|1)p_1 + c(1|2)P(1|2)p_2 \\
&= c(2|1)p_1 \int_{C_2} f_1(\boldsymbol{x})d\boldsymbol{x} + c(1|2)p_2 \int_{C_1} f_2(\boldsymbol{x})d\boldsymbol{x} \\
&= c(2|1)p_1 \left[1 - \int_{C_1} f_1(\boldsymbol{x})d\boldsymbol{x}\right] + c(1|2)p_2 \int_{C_1} f_2(\boldsymbol{x})d\boldsymbol{x} \\
&= c(2|1)p_1 + \int_{C_1} \left[c(1|2)p_2 f_2(\boldsymbol{x}) - c(2|1)p_1 f_1(\boldsymbol{x})\right] d\boldsymbol{x}
\end{aligned}
$$

$$
MinECM \Leftrightarrow c(1|2)p_2 f_2(\boldsymbol{x}) - c(2|1)p_1 f_1(\boldsymbol{x}) < 0
$$

$$
\Leftrightarrow c(1|2)p_2 f_2(\boldsymbol{x}) < c(2|1)p_1 f_1(\boldsymbol{x})
$$

$$
\Leftrightarrow \frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} > \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)
$$

6.2 DCA with Two Clusters

[Table 6.2.1] ECM Classification Rule = Minimum ECM Rule

A classification rule that minimizes the ECM can be applied to the observed $\boldsymbol{x}_o = (x_{o1}, x_{o2}, \cdots, x_{op})^t$: New observation

Density Ratio   Cost Ratio   Prior Prob. Ratio

$$\frac{f_1(\boldsymbol{x}_o)}{f_2(\boldsymbol{x}_o)} > \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right) \quad \Rightarrow \quad \boldsymbol{x}_o \text{ is assigned to cluster } C_1.$$

$$\frac{f_1(\boldsymbol{x}_o)}{f_2(\boldsymbol{x}_o)} \leq \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right) \quad \Rightarrow \quad \boldsymbol{x}_o \text{ is assigned to cluster } C_2.$$

6.2 DCA with Two Clusters

[Table 6.2.2] Special cases of ECM Classification Rules

For observed $\boldsymbol{x}_o = (x_{o1}, x_{o2}, \cdots, x_{op})^t$,

- $p_2/p_1 = 1$: $\dfrac{f_1(\boldsymbol{x}_o)}{f_2(\boldsymbol{x}_o)} > \dfrac{c(1|2)}{c(2|1)}$ $\Rightarrow$ Allocate $\boldsymbol{x}_o$ to C_1

- $c(1|2)/c(2|1) = 1$: $\dfrac{f_1(\boldsymbol{x}_o)}{f_2(\boldsymbol{x}_o)} > \dfrac{p_2}{p_1}$ $\Rightarrow$ Allocate $\boldsymbol{x}_o$ to C_1

- $p_1/p_2 = 1 \wedge c(1|2)/c(2|1) = 1$: $\dfrac{f_1(\boldsymbol{x}_o)}{f_2(\boldsymbol{x}_o)} > 1$ $\Rightarrow$ Allocate $\boldsymbol{x}_o$ to C_1

6.3 DCA of Two Multivariate Normal Clusters

$$\boldsymbol{x} = (x_1,\ x_2,\ ...,\ x_p)^t \sim N_p(\boldsymbol{\mu}_k,\ \Sigma_k),\ k=1,2 \quad \text{with} \quad f_k(\boldsymbol{x}) = \frac{1}{(\sqrt{2\pi})^p\,|\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^t \Sigma_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)\right],\ k=1,\ 2$$

- Case 1: Equal Covariance Matrices $\Sigma_1 = \Sigma_2 = \Sigma$

$$C_1:\quad \exp\left[-\frac{1}{2}(\boldsymbol{x}_o-\boldsymbol{\mu}_1)^t\Sigma^{-1}(\boldsymbol{x}_o-\boldsymbol{\mu}_1)+\frac{1}{2}(\boldsymbol{x}_o-\boldsymbol{\mu}_2)^t\Sigma^{-1}(\boldsymbol{x}_o-\boldsymbol{\mu}_2)\right] > \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$\Leftrightarrow\quad (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\Sigma^{-1}\boldsymbol{x}_o - \frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\Sigma^{-1}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)] > \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

$$L(\boldsymbol{x}_o) = (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\Sigma^{-1}\boldsymbol{x}_o \text{ and } \beta_0 = \frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^t\Sigma^{-1}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)$$

$$L(\boldsymbol{x}_o) > \beta_0 + \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \Rightarrow \boldsymbol{x}_o \text{ is assigned to cluster } C_1$$

Substitute $\overline{\boldsymbol{x}}_1,\ \overline{\boldsymbol{x}}_2,\ S_p$ *for* $\boldsymbol{\mu}_1,\ \boldsymbol{\mu}_2,\ \Sigma$, *respectively*

[Table 6.3.2] $\Sigma_1 = \Sigma_2$: ECM Classification Rule (Linear Classification Rule)

For observed $\boldsymbol{x}_o = (x_{o1}, x_{o2}, \cdots, x_{op})^t$

$$\widehat{L}(\boldsymbol{x}_o) > \widehat{\beta}_0 + \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \quad \Rightarrow \quad \boldsymbol{x}_o \text{ is assigned to cluster } C_1.$$

where $\widehat{L}(\boldsymbol{x}_o) = (\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2)^t S_p^{-1}\boldsymbol{x}_o,\ \ S_p = [(n_1-1)S_1+(n_2-1)S_2]/(n_1+n_2-2)$,

$$\widehat{\beta}_0 = \frac{1}{2}(\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2)^t S_p^{-1}(\overline{\boldsymbol{x}}_1+\overline{\boldsymbol{x}}_2)\ ,$$

6.3 DCA of Two Multivariate Normal Clusters

- Case 2: Unequal Covariance Matrices $\Sigma_1 \neq \Sigma_2$

$$C_1: \ -\frac{1}{2}\boldsymbol{x}_o^t(\Sigma_1^{-1}-\Sigma_2^{-1})\boldsymbol{x}_o+(\boldsymbol{\mu}_1^t\Sigma_1^{-1}-\boldsymbol{\mu}_2^t\Sigma_2^{-1})\boldsymbol{x}_o-k > \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

with $k=\frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right)+\frac{1}{2}(\boldsymbol{\mu}_1^t\Sigma_1^{-1}\boldsymbol{\mu}_1-\boldsymbol{\mu}_2^t\Sigma_2^{-1}\boldsymbol{\mu}_2)$

$\longrightarrow$ Assign new observation $\boldsymbol{x}_0$ to C_1

In practice, substitute $\bar{\boldsymbol{x}}_1$, $\bar{\boldsymbol{x}}_2$, S_1, and S_2 for $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, Σ_1 and Σ_2, respectively.

[Table 6.3.3] $\Sigma_1 \neq \Sigma_2$: ECM Classification Rule (Quadratic Classification Rule)

For observed $\boldsymbol{x}_o=(x_{o1},x_{o2},\cdots,x_{op})^t$

$$\widehat{Q}(\boldsymbol{x}_o) > \hat{k} + \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \ \Rightarrow \ \boldsymbol{x}_o \text{ is assigned to cluster } C_1.$$

where $\widehat{Q}(\boldsymbol{x}_o) = -\frac{1}{2}\boldsymbol{x}_o^t(S_1^{-1}-S_2^{-1})\boldsymbol{x}_o+(\bar{\boldsymbol{x}}_1^tS_1^{-1}-\bar{\boldsymbol{x}}_2^tS_2^{-1})\boldsymbol{x}_o,$

$$\hat{k}=\frac{1}{2}ln\left(\frac{|S_1|}{|S_2|}\right)+\frac{1}{2}(\bar{\boldsymbol{x}}_1^tS_1^{-1}\bar{\boldsymbol{x}}_1-\bar{\boldsymbol{x}}_2^tS_2^{-1}\bar{\boldsymbol{x}}_2).$$

6.3 DCA of Two Multivariate Normal Clusters

Deriving ECM classification rule (Linear Classification Rule) : $\Sigma_1 = \Sigma_2$

$$\boxed{\frac{f_1(x_o)}{f_2(x_o)} > \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)} \Leftarrow \quad f_k(x) = \frac{1}{(\sqrt{2\pi})^p |\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_k)^t \Sigma_k^{-1}(x-\mu_k)\right], \ k = 1, 2$$

$$\Leftrightarrow \frac{\frac{1}{(\sqrt{2\pi})^p |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x_o-\mu_1)^t \Sigma^{-1}(x_o-\mu_1)\right]}{\frac{1}{(\sqrt{2\pi})^p |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x_o-\mu_2)^t \Sigma^{-1}(x_o-\mu_2)\right]} > \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$\Leftrightarrow \exp\left[-\frac{1}{2}(x_o-\mu_1)^t \Sigma^{-1}(x_o-\mu_1) + \frac{1}{2}(x_o-\mu_2)^t \Sigma^{-1}(x_o-\mu_2)\right] > \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$\Leftrightarrow \boxed{(\mu_1-\mu_2)^t \Sigma^{-1} x_o - \frac{1}{2}(\mu_1-\mu_2)^t \Sigma^{-1}(\mu_1+\mu_2)] > \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]} \quad (6.3.5)$$

$\longrightarrow$ *Allocate* x_o *to* C_1

Deriving Linear Classification Rule : $\Sigma_1 = \Sigma_2$

$$LHS = -\frac{1}{2}(\boldsymbol{x}_o - \boldsymbol{\mu}_1)^t \Sigma^{-1}(\boldsymbol{x}_o - \boldsymbol{\mu}_1) + \frac{1}{2}(\boldsymbol{x}_o - \boldsymbol{\mu}_2)^t \Sigma^{-1}(\boldsymbol{x}_o - \boldsymbol{\mu}_2)$$

$$= -\frac{1}{2}\boldsymbol{x}_o^t \Sigma^{-1}\boldsymbol{x}_o + \frac{1}{2}\boldsymbol{x}_o^t \Sigma^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_1^t \Sigma^{-1}\boldsymbol{x}_o - \frac{1}{2}\boldsymbol{\mu}_1^t \Sigma^{-1}\boldsymbol{\mu}_1$$

$$+ \frac{1}{2}\boldsymbol{x}_o^t \Sigma^{-1}\boldsymbol{x}_o - \frac{1}{2}\boldsymbol{x}_o^t \Sigma^{-1}\boldsymbol{\mu}_2 - \frac{1}{2}\boldsymbol{\mu}_2^t \Sigma^{-1}\boldsymbol{x}_o + \frac{1}{2}\boldsymbol{\mu}_2^t \Sigma^{-1}\boldsymbol{\mu}_2$$

$$= \boldsymbol{\mu}_1^t \Sigma^{-1}\boldsymbol{x}_o - \boldsymbol{\mu}_2^t \Sigma^{-1}\boldsymbol{x}_o - \frac{1}{2}\left(\boldsymbol{\mu}_1^t \Sigma^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^t \Sigma^{-1}\boldsymbol{\mu}_2\right)$$

$$= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t \Sigma^{-1}\boldsymbol{x}_o - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\Sigma^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$$

6.3 DCA of Two Multivariate Normal Clusters

Deriving ECM classification rule (Quadratic Classification Rule) : $\Sigma_1 \neq \Sigma_2$

$$\boxed{\frac{f_1(\boldsymbol{x}_o)}{f_2(\boldsymbol{x}_o)} > \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)} \Leftarrow \quad f_k(\boldsymbol{x}) = \frac{1}{(\sqrt{2\pi})^p |\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)^t \Sigma_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)\right], \ k=1,\ 2$$

$$\Leftrightarrow \frac{\dfrac{1}{(\sqrt{2\pi})^p |\Sigma_1|^{1/2}} \exp\left[-\dfrac{1}{2}(\boldsymbol{x}_o-\boldsymbol{\mu}_1)^t \Sigma_1^{-1}(\boldsymbol{x}_o-\boldsymbol{\mu}_1)\right]}{\dfrac{1}{(\sqrt{2\pi})^p |\Sigma_2|^{1/2}} \exp\left[-\dfrac{1}{2}(\boldsymbol{x}_o-\boldsymbol{\mu}_2)^t \Sigma_2^{-1}(\boldsymbol{x}_o-\boldsymbol{\mu}_2)\right]} > \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)$$

$$\Leftrightarrow \frac{|\Sigma_1|^{-1/2}}{|\Sigma_2|^{-1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}_o-\boldsymbol{\mu}_1)^t \Sigma_1^{-1}(\boldsymbol{x}_o-\boldsymbol{\mu}_1) + \frac{1}{2}(\boldsymbol{x}_o-\boldsymbol{\mu}_2)^t \Sigma_2^{-1}(\boldsymbol{x}_o-\boldsymbol{\mu}_2)\right]$$

$$> \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right) \qquad k = \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2}(\boldsymbol{\mu}_1^t \Sigma_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^t \Sigma_2^{-1}\boldsymbol{\mu}_2)$$

$$\Leftrightarrow \boxed{-\frac{1}{2}\boldsymbol{x}_o^t(\Sigma_1^{-1}-\Sigma_2^{-1})\boldsymbol{x}_o + (\boldsymbol{\mu}_1^t\Sigma_1^{-1} - \boldsymbol{\mu}_2^t\Sigma_2^{-1})\boldsymbol{x}_o - k > \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]} \quad (6.3.11)$$

$\longrightarrow$ *Allocate* x_o *to* C_1

Deriving Quadratic Classification Rule : $\Sigma_1 \neq \Sigma_2$

$$LHS = -\frac{1}{2}(\boldsymbol{x}_o - \boldsymbol{\mu}_1)^t \Sigma_1^{-1}(\boldsymbol{x}_o - \boldsymbol{\mu}_1) + \frac{1}{2}(\boldsymbol{x}_o - \boldsymbol{\mu}_2)^t \Sigma_2^{-1}(\boldsymbol{x}_o - \boldsymbol{\mu}_2)$$

$$= -\frac{1}{2}\boldsymbol{x}_o^t \Sigma_1^{-1}\boldsymbol{x}_o + \frac{1}{2}\boldsymbol{x}_o^t \Sigma_1^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_1^t \Sigma_1^{-1}\boldsymbol{x}_o - \frac{1}{2}\boldsymbol{\mu}_1^t \Sigma_1^{-1}\boldsymbol{\mu}_1$$

$$+ \frac{1}{2}\boldsymbol{x}_o^t \Sigma_2^{-1}\boldsymbol{x}_o - \frac{1}{2}\boldsymbol{x}_o^t \Sigma_2^{-1}\boldsymbol{\mu}_2 - \frac{1}{2}\boldsymbol{\mu}_2^t \Sigma_2^{-1}\boldsymbol{x}_o + \frac{1}{2}\boldsymbol{\mu}_2^t \Sigma_2^{-1}\boldsymbol{\mu}_2$$

$$= -\frac{1}{2}\boldsymbol{x}_o^t(\Sigma_1^{-1} - \Sigma_2^{-1})\boldsymbol{x}_o + (\boldsymbol{\mu}_1^t \Sigma_1^{-1} - \boldsymbol{\mu}_2^t \Sigma_2^{-1})\boldsymbol{x}_o - \frac{1}{2}\left(\boldsymbol{\mu}_1^t \Sigma_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^t \Sigma_2^{-1}\boldsymbol{\mu}_2\right)$$

6.3 DCA of Two Multivariate Normal Clusters

Homogeneity Test of Covariance Matrices : $H_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_g = \Sigma$

$$\boldsymbol{x}_{ki} = (x_{ki1}, \ldots, x_{kip})', \; i = 1, \ldots, n_k \sim N_p(\boldsymbol{\mu}_k, \Sigma_k), \; k = 1, \ldots, g$$

L-R test statistics : $\Lambda = \prod_{k=1}^{g} \left(\frac{|S_k|}{|S_p|} \right)^{(n_k-1)/2}$ with $S_p = \frac{\sum_{k=1}^{g} (n_k - 1) S_k}{\sum_{k=1}^{g} (n_k - 1)}$

Box's M-test :

$$M = -2\gamma \ln \Lambda = \gamma \sum_{k=1}^{g} (n_k - 1) \log \left| S_k^{-1} S_p \right| > \chi^2_{(p(p+1)((g-1)/2)}(\alpha)$$

with $\gamma = 1 - \left[\sum_{k=1}^{g} \frac{1}{(n_k - 1)} - \frac{1}{\sum_{k=1}^{g} (n_k - 1)} \right] \left[\frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right]$

Note : library(biotools) > boxM(x, cluster)

Definition of Likelihood Ratio Test

- $\boldsymbol{\theta}$: a vector consisting of all the unknown population parameters
- The distribution of $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})^t,\ i = 1, \dots, n$ depends on a $\boldsymbol{\theta}$
- $L(\boldsymbol{\theta})$: the likelihood function with joint density of $\boldsymbol{x}_i \in R^p,\ i = 1, \dots, n.$
- $\boxed{H_0 : \boldsymbol{\theta} \in \Omega_0 \text{ vs. } H_1 : \boldsymbol{\theta} \in \Omega_1}$: Hypotheses
- L-R test statistic for testing H_0 against H_1 :

$$\Lambda = \frac{Max_{\boldsymbol{\theta} \in \Omega_0} L(\boldsymbol{\theta})}{Max_{\boldsymbol{\theta} \in \Omega_1} L(\boldsymbol{\theta})}$$

$\Rightarrow$ As $n \to \infty$, under H_0, $\boxed{-2\log\Lambda \simeq \chi^2_{(df)} \text{ with } df = \dim(\Omega_1) - \dim(\Omega_0)}$

$\Rightarrow$ $\boxed{-2\log\Lambda > \chi^2_{(df)}(\alpha) \Rightarrow \text{Reject } H_0 : \boldsymbol{\theta} \in \Omega_0 \text{ at significant level } \alpha.}$

Deriving Box's M- test based on the L-R Test

- $$\boxed{\begin{aligned} &H_o : \Sigma_1 = \cdots = \Sigma_g = \Sigma \text{ vs. } H_1 : \Sigma_1 \neq \cdots \neq \Sigma_g \\ &\Rightarrow \quad \Omega_0 = \{\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_g, \Sigma\}, \quad \Omega_1 = \{\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_g, \Sigma_1, \cdots, \Sigma_g\} \end{aligned}}$$

- $$L(\boldsymbol{\theta}) = L(\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_g ; \Sigma_1, \cdots, \Sigma_g) = \prod_{k=1}^{g} L_k(\boldsymbol{\mu}_k ; \Sigma_k)$$

- $$\begin{aligned} L_k(\boldsymbol{\mu}_k ; \Sigma_k) &= \prod_{i=1}^{n_k} f_k(\boldsymbol{x}_i) \\ &= \prod_{i=1}^{n_k} \frac{1}{(\sqrt{2\pi})^p |\Sigma_k|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right\} \\ &= \frac{1}{(\sqrt{2\pi})^{n_k p} |\Sigma_k|^{n_k/2}} exp\left\{-\frac{1}{2}\sum_{i=1}^{n_k}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right\} \\ &= \boxed{|2\pi\Sigma|^{-n_k/2} \exp\left\{-\frac{1}{2}\left[tr\left(\Sigma_k^{-1} V_k\right) + n_k(\overline{\boldsymbol{x}}_k - \boldsymbol{\mu}_k)^T \Sigma_k^{-1}(\overline{\boldsymbol{x}}_k - \boldsymbol{\mu}_k)\right]\right\}} \end{aligned}$$

with $V_k = (n_k - 1) S_k$

Deriving Box's M- test based on the L-R Test

- $$\Lambda = \frac{Max_{\boldsymbol{\theta} \in \Omega_0} L(\boldsymbol{\theta})}{Max_{\boldsymbol{\theta} \in \Omega_1} L(\boldsymbol{\theta})} = \frac{Max_{\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_g, \Sigma} L(\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_g ; \Sigma, \cdots, \Sigma)}{Max_{\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_g, \Sigma_1, \cdots, \Sigma_g} L(\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_g ; \Sigma_1, \cdots, \Sigma_g)}$$

$$= \prod_{k=1}^{g} \left(\frac{|S_k|}{|S_p|} \right)^{(n_k - 1)/2} \quad \text{with} \quad S_p = \frac{\sum_{k=1}^{g} (n_k - 1) S_k}{\sum_{k=1}^{g} (n_k - 1)}$$

$$\Rightarrow \quad \boxed{-2 \log \Lambda \simeq \chi^2_{(df)} \quad \text{with} \quad df = \dim(\Omega_1) - \dim(\Omega_0)}$$

$$= p \times g + \frac{p(p+1)}{2} g - \left(p \times g + \frac{p(p+1)}{2} \right)$$

$$= \frac{p(p+1)(g-1)}{2}$$

6.3 DCA of Two Multivariate Normal Clusters

- **[Example 6.3.1]** Two iris flowers **(setosa, versicolor)**

$$H_0 : \Sigma_1 = \Sigma_2 = \Sigma$$

Sepal length width Petal length width

Setosa				Versicolor			
X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4
5.1	3.5	1.4	0.2	7.0	3.2	4.7	1.4
4.9	3.0	1.4	0.2	6.4	3.2	4.5	1.5
4.7	3.2	1.3	0.2	6.9	3.1	4.9	1.5
4.6	3.1	1.5	0.2	5.5	2.3	4.0	1.3
5.0	3.6	1.4	0.2	6.5	2.8	4.6	1.5
5.4	3.9	1.7	0.4	5.7	2.8	4.5	1.3
4.6	3.4	1.4	0.3	6.3	3.3	4.7	1.6
5.0	3.4	1.5	0.2	4.9	2.4	3.3	1.0
4.4	2.9	1.4	0.2	6.6	2.9	4.6	1.3
4.9	3.1	1.5	0.1	5.2	2.7	3.9	1.4
5.4	3.7	1.5	0.2	5.0	2.0	3.5	1.0
4.8	3.4	1.6	0.2	5.9	3.0	4.2	1.5
4.8	3.0	1.4	0.1	6.0	2.2	4.0	1.0
4.8	3.0	1.4	0.1	6.1	2.9	4.7	1.4
5.8	4.0	1.2	0.2	5.6	2.9	3.6	1.3
5.7	4.4	1.5	0.4	6.7	3.1	4.4	1.4
5.4	3.9	1.3	0.4	5.6	3.0	4.5	1.5
5.1	3.5	1.4	0.3	5.8	2.7	4.1	1.0
5.7	3.8	1.7	0.3	6.2	2.2	4.5	1.5
5.1	3.8	1.5	0.3	5.6	2.5	3.9	1.1
5.4	3.4	1.7	0.2	5.9	3.2	4.8	1.8
5.1	3.7	1.5	0.4	6.1	2.8	4.0	1.3
4.6	3.6	1.0	0.2	6.3	2.5	4.9	1.5
5.1	3.3	1.7	0.5	6.1	2.8	4.7	1.2
4.8	3.4	1.9	0.2	6.4	2.9	4.3	1.3
5.0	3.0	1.6	0.2	6.6	3.0	4.4	1.4
5.0	3.4	1.6	0.4	6.8	2.8	4.8	1.4
5.2	3.5	1.5	0.2	6.7	3.0	5.0	1.7
5.2	3.4	1.4	0.2	6.0	2.9	4.5	1.5
4.7	3.2	1.6	0.2	5.7	2.6	3.5	1.0
4.8	3.1	1.6	0.2	5.5	2.4	3.8	1.1
5.4	3.4	1.5	0.4	5.5	2.4	3.7	1.0
5.2	4.1	1.5	0.1	5.8	2.7	3.9	1.2
5.5	4.2	1.4	0.2	6.0	2.7	5.1	1.6
4.9	3.1	1.5	0.2	5.4	3.0	4.5	1.5
5.0	3.2	1.2	0.2	6.0	3.4	4.5	1.6
5.5	3.5	1.3	0.2	6.7	3.1	4.7	1.5
4.9	3.6	1.4	0.1	6.3	2.3	4.4	1.3
4.4	3.0	1.3	0.2	5.6	3.0	4.1	1.3
5.1	3.4	1.5	0.2	5.5	2.5	4.0	1.3
5.0	3.5	1.3	0.3	5.5	2.6	4.4	1.2
4.5	2.3	1.3	0.3	6.1	3.0	4.6	1.4
4.4	3.2	1.3	0.2	5.8	2.6	4.0	1.2
5.0	3.5	1.6	0.6	5.0	2.3	3.3	1.0
5.1	3.8	1.9	0.4	5.6	2.7	4.2	1.3
4.8	3.0	1.4	0.3	5.7	3.0	4.2	1.2
5.1	3.8	1.6	0.2	5.7	2.9	4.2	1.3
4.6	3.2	1.4	0.2	6.2	2.9	4.3	1.3
5.3	3.7	1.5	0.2	5.1	2.5	3.0	1.1
5.0	3.3	1.4	0.2	5.7	2.8	4.1	1.3

[R-code 6.3.1]

$$S_1 = \begin{bmatrix} \boxed{6.0882} & 4.8616 \\ 4.8616 & \boxed{7.0408} \end{bmatrix}, \quad S_2 = \begin{bmatrix} \boxed{13.0552} & 4.1746 \\ 4.1740 & \boxed{4.8250} \end{bmatrix}$$

$$S_p = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2} = \frac{49S_1 + 49S_2}{98}$$

$$\Rightarrow \quad S_p = \begin{bmatrix} 0.1953408 & 0.0922000 \\ 0.0922000 & 0.1210796 \end{bmatrix}$$

$$\gamma = 1 - \left[\frac{1}{49} + \frac{1}{49} - \frac{1}{98}\right]\left[\frac{2\times 2^2 + 3\times 2 - 1}{6(2+1)(2-1)}\right]$$
$$= 0.9778912$$

$$M = -2\gamma \ln \Lambda = \gamma \sum_{k=1}^{2} (n_k - 1)\log\left|S_k^{-1} S_p\right| = 19.74945$$

$$\chi_3^2(0.05) = 7.814728$$

$$df = p(p+1)(g-1)/2 = 2(2+1)(2-1) = 3$$

Reject

$$M > \chi_3^2(0.05) \Leftrightarrow 19.74945 > 7.814728 \Rightarrow H_0 : \Sigma_1 = \Sigma_2 = \Sigma$$

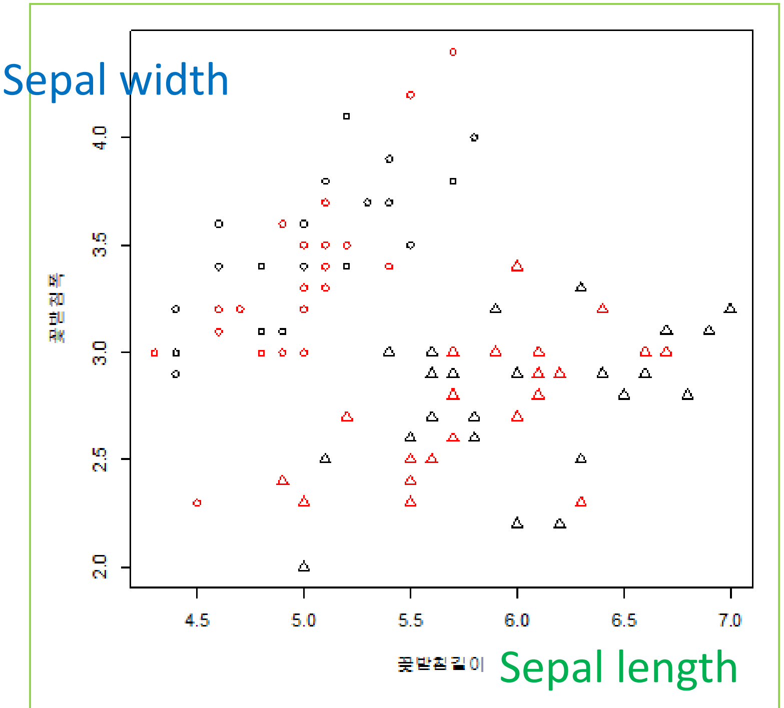
$$\boxed{p\text{-value} = 0.0001912932 < 0.05}$$

6.3 DCA of Two Multivariate Normal Clusters

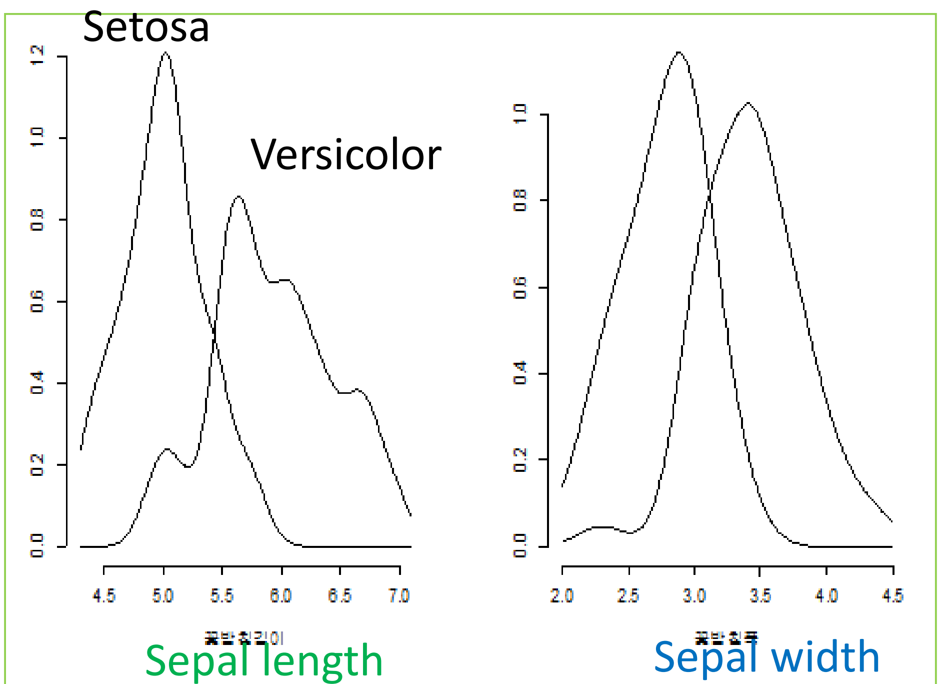
- [Example 6.3.1] Equal Covariance Matrices of Fisher's two iris flower (setosa, versicolor)

- [Figure 6.3.1] Scatter plot and Probability density function

$$\Sigma_1 \neq \Sigma_2$$



(a) scatter plot : O=setosa, Δ=versicolor



(b) Probability density function

6.3 DCA of Two Multivariate Normal Clusters

- **[Example 6.3.2] Equal Covariance Matrices of riding mower** $\boxed{H_0 : \Sigma_1 = \Sigma_2 = \Sigma}$

Discriminating 12 owners from 12 non-owners of riding mowers [Table 6.3.1]

Owners Income	Owners Lot size	Non-owners Income	Non-owners Lot size
60.0	18.4	75.0	19.6
85.5	16.8	52.8	20.8
64.8	21.6	64.8	17.2
61.5	20.8	43.2	20.4
87.0	23.6	84.0	17.6
110.1	19.2	49.2	17.6
108.0	17.6	59.4	16.0
82.8	22.4	66.0	18.4
69.0	20.0	47.4	16.4
93.0	20.8	33.0	18.8
51.0	22.0	51.0	14.0
81.0	20.0	63.0	14.8

[R-code 6.3.2]

$$S_1 = \begin{bmatrix} \boxed{3879.08} & -130.00 \\ -130.00 & \boxed{44.91} \end{bmatrix}, \quad S_2 = \begin{bmatrix} \boxed{2207.76} & -28.48 \\ -28.48 & \boxed{49.11} \end{bmatrix}$$

$$S_p = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2} = \frac{11 \times S_1 + 11 \times S_2}{22}$$

$$\Rightarrow \quad S_p = \begin{bmatrix} 276.674659 & -7.203636 \\ -7.203636 & 4.273333 \end{bmatrix}$$

$$M = -2\gamma \ln \Lambda = \gamma \sum_{k=1}^{2} (n_k - 1) \log \left| S_k^{-1} S_p \right| = 0.9934638$$

$$\chi_3^2(0.05) = 7.814728$$

$$df = p(p+1)(g-1)/2 = 2(2+1)(2-1) = 3$$

Not reject

$$M < \chi_3^2(0.05) \Leftrightarrow 0.9934638 < 7.814728 \Rightarrow H_0 : \Sigma_1 = \Sigma_2 = \Sigma$$

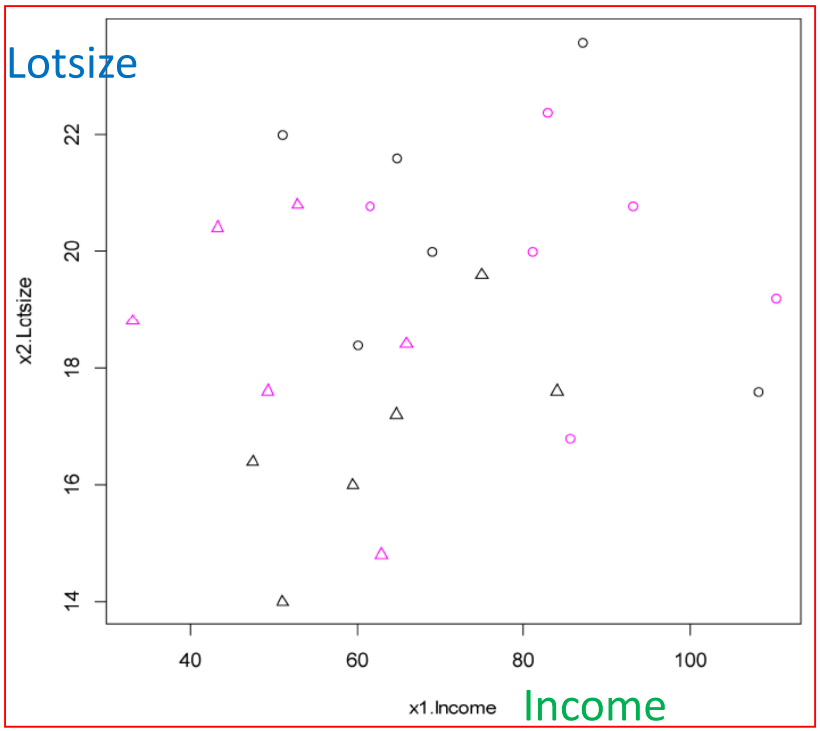
$$\boxed{p\text{-value} = 0.8028335 > 0.05}$$

6.3 DCA of Two Multivariate Normal Clusters

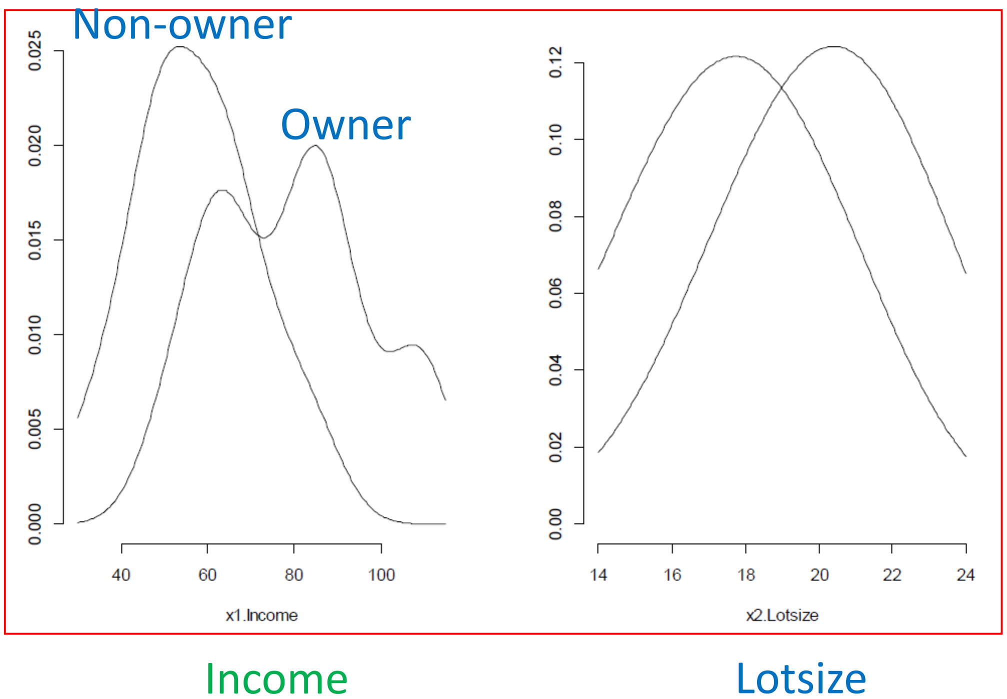
- **[Example 6.3.2] Equal Covariance Matrices of riding mower**

- **[Figure 6.3.2] Scatter plot and Probability density function**

$$\Sigma_1 = \Sigma_2$$



(a) Scatter plot: Owner = ○, $Nonowner = \Delta$

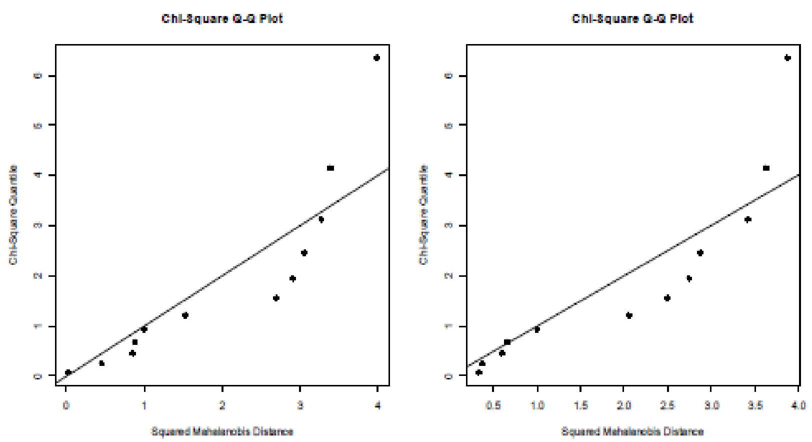


(b) Probability density function

6.3 DCA of Two Multivariate Normal Clusters

- **[Example 6.3.3] Multivariate Normality Test and LDA of riding mower data : $\Sigma_1 = \Sigma_2$**

- **[Figure 6.3.3] Normality test of riding mower data**




```
Mardia's Multivariate Normality Test
---------------------------------------
data: owner

g1p:             0.23663
chi.skew:        0.47326
p.value.skew:    0.9760514

g2p:             5.661766
z.kurtosis:      -1.012485
p.value.kurt:    0.3113062

chi.small.skew:  0.6991341
p.value.small:   0.9514357

Result: Data are multivariate normal.
---------------------------------------
```

```
Mardia's Multivariate Normality Test
---------------------------------------
data: nonowner

g1p:             0.5969362
chi.skew:        1.193872
p.value.skew:    0.8791064

g2p:             5.659681
z.kurtosis:      -1.013388
p.value.kurt:    0.3108749

chi.small.skew:  1.763675
p.value.small:   0.779121

Result: Data are multivariate normal.
---------------------------------------
```

mvn(owner, mvnTest="mardia", multivariatePlot="qq")

[R-code 6.3.3](ridingmower-Mardiatest-LDA.R)

```
# MVD test and Linear DA for two groups(owner, nonowner)
# of riding mower on two variables(x1.Income x2.Lotsize)
library(MASS)
library(MVN)

ridingmower<-read.table("ridingmower.txt", header=T)
attach(ridingmower)
ridingmower
# MVN tests based on the Skewness and Kurtosis Statistics
par(mfrow=c(1, 2))
owner=ridingmower[1:12, 2:3]
nonowner=ridingmower[13:24, 2:3]
owner
nonowner
result_owner=mardiaTest(owner, qqplot=TRUE)
result_nonowner=mardiaTest(nonowner, qqplot=TRUE)
result_owner
result_nonowner

# Linear DA
LDA=lda(pop~x1.Income+x2.Lotsize, data=ridingmower)
LDA
lcluster=predict(LDA, ridingmower)$class
lct=table(pop, lcluster)
lct
# Total percent correct
mean(pop==lcluster)
```

6.3 DCA of Two Multivariate Normal Clusters

- **[Example 6.3.3] LDA of riding mower data :** $\Sigma_1 = \Sigma_2$

- sample mean : $\bar{\boldsymbol{x}}_1 = (79.475, 20.267)^t$, $\bar{\boldsymbol{x}}_2 = (57.400, 17.633)^t$
- covariance matrix :

$$S_1 = \begin{bmatrix} 3879.08 & -130.00 \\ -130.00 & 44.91 \end{bmatrix}, \; S_2 = \begin{bmatrix} 2207.76 & -28.48 \\ -28.48 & 49.11 \end{bmatrix}$$

- pooled covariance matrix : $S_p = \begin{bmatrix} 276.674659 & -7.203636 \\ -7.203636 & 4.273333 \end{bmatrix}$

$$\boldsymbol{x}_o = (x_{o1}, x_{o2})^T \Rightarrow \widehat{L}(\boldsymbol{x}_o) = \boxed{(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^T S_p^{-1}} \boldsymbol{x}_o = \boxed{\hat{\boldsymbol{l}}^T} \boldsymbol{x}_o \quad \text{with} \quad \hat{\boldsymbol{l}} = (\hat{l_1}, \hat{l_2})^T$$

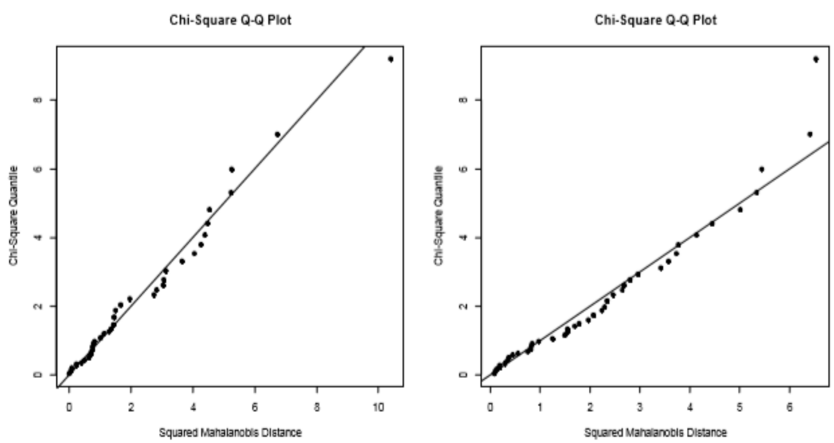
- linear discriminant function : $\widehat{L}(\boldsymbol{x}_o) = \hat{l}_1 x_{o1} + \hat{l}_p x_{o2} = -0.048 \times \text{Income} - 0.380 \times \text{Lotsize}$
- a classified table : MER = (1 + 2)/24 x 100% = 12.5%

Confusion Table(Matrix)		Classification by linear discriminant function		
		Owner	Nonowner	
Clusters	Owner	11	1	12
	Nonowner	2	10	12

6.3 DCA of Two Multivariate Normal Clusters

- **[Example 6.3.4] Multivariate Normality Test and QDA of iris data :** $\Sigma_1 \neq \Sigma_2$

- **[Figure 6.3.4] Normality test of iris data(setosa, versicolor)**




```
Mardia's Multivariate Normality Test
---------------------------------------
data: setosa

g1p:              0.09114042
chi.skew:         0.7595035
p.value.skew:     0.9437932

g2p:              8.105738
z.kurtosis:       0.09346006
p.value.kurt:     0.9255381

chi.small.skew:   0.8379339
p.value.small:    0.9332914

Result: Data are multivariate normal.
---------------------------------------
```

```
Mardia's Multivariate Normality Test
---------------------------------------
data: versicolor

g1p:              0.2112977
chi.skew:         1.760814
p.value.skew:     0.7796433

g2p:              6.974427
z.kurtosis:       -0.9064869
p.value.kurt:     0.3646782

chi.small.skew:   1.942645
p.value.small:    0.7463072

Result: Data are multivariate normal.
---------------------------------------
```

mvn(setosa, mvnTest=“mardia”, multivariatePlot=“qq”)

[R-code 6.3.4](setosaversicolor-mardiaTest-QDA.R)

```
# Quadratic DA for two groups(setosa, vesicolor) on two variables(꽃받침길이, 꽃받침폭)
library(MASS)
library(MVN)

setosa_versi<-read.table("setosaversicolor.txt", header=T)
attach(setosa_versi)

# MVN tests based on the Skewness and Kurtosis Statistics
par(mfrow=c(1, 2))
setosa=setosa_versi[1:50, 1:2]
versicolor=setosa_versi[51:100, 1:2]
setosa
versicolor
result_setosa=mardiaTest(setosa, qqplot=TRUE)
result_versicolor=mardiaTest(versicolor, qqplot=TRUE)
result_setosa
result_versicolor

# Quadratic DA with 3 groups applying
 # resubstitution prediction and equal prior probabilities.
QDA=qda(종류~꽃받침길이+꽃받침폭, data=setosa_versi, prior=c(1, 1)/2)
QDA
qcluster=predict(QDA, setosa_versi)$class
qct=table(종류, qcluster)
qct
# Total percent correct
mean(종류==qcluster)
```

6.3 DCA of Two Multivariate Normal Clusters

- **[Example 6.3.4] Multivariate normality test and QDA of iris data:** $\Sigma_1 \neq \Sigma_2$

- number of C_1(setosa) and C_2(versicolor) : $n_1 = 50,\ n_2 = 50$
- sample mean : $\bar{\boldsymbol{x}}_1 = (5.006, 3.428)^t,\ \bar{\boldsymbol{x}}_2 = (5.936, 2.770)^t$
- covariance matrix :

$$S_1 = \begin{bmatrix} 6.0882 & 4.8616 \\ 4.8616 & 7.0408 \end{bmatrix},\ S_2 = \begin{bmatrix} 13.0552 & 4.1746 \\ 4.1740 & 4.8250 \end{bmatrix}$$

$\boldsymbol{x}_o = (x_{o1}, x_{o2})^T$

- Quadratic discriminant function : $\widehat{Q}(\boldsymbol{x}_o) = -\frac{1}{2}\boldsymbol{x}_o^t(S_1^{-1} - S_2^{-1})\boldsymbol{x}_o + (\bar{\boldsymbol{x}}_1^t S_1^{-1} - \bar{\boldsymbol{x}}_2^t S_2^{-1})\boldsymbol{x}_o$
- a classified table : MER = (1 + 0)/100 x 100% = 1%

Confusion Table		Classification by quadratic discriminant function		
		setosa	versicolor	
Clusters	setosa	49	1	50
	versicolor	0	50	50

6.4 DCA with Several Clusters

Total Expected Cost of Misclassification

[Table 6.4.1] TECM classification rule of several clusters
Classification rule that minimizes the TECM is that for the observed $\boldsymbol{x}_o = (x_{o1}, x_{o2}, \cdots, x_{op})^t$ $\sum_{i=1, i \neq k}^{g} p_i f_i(\boldsymbol{x}_o) c(k\|i)$: Minimum value $\Rightarrow$ $\boldsymbol{x}_o$ is assigned to cluster $C_{k'}$ $k = 1, \ldots, g$ $\longleftrightarrow$ $p_k f_k(\boldsymbol{x}_o) c(k\|i)$: *largest*

$c(k|i) = 1$ $\longrightarrow$ Total Probability of Misclassification

[Table 6.4.2] TPM classification rule of several clusters
The classification rule that minimizes the TPM is for the observed $\boldsymbol{x}_o = (x_{o1}, x_{o2}, \cdots, x_{op})^t$ $\sum_{i=1, i \neq k}^{g} p_i f_i(\boldsymbol{x}_o)$: Minimum value $\Rightarrow$ $\boldsymbol{x}_o$ is assigned to cluster $C_{k'}$ $k = 1, \ldots, g$

$\longleftrightarrow$ $p_k f_k(\boldsymbol{x}_o)$: *largest*

Deriving TECM Classification Rule

ECM_i = Conditional ECM of $\boldsymbol{x} \in C_i$ who can be misclassified in to

C_1, or $C_2, \ldots,$ or C_{i-1}, or $C_i, \ldots,$ or C_g

$$= \sum_{k=1, k \neq i}^{g} P(k|i)c(k|i)$$

$$\Rightarrow \; TECM = p_1 ECM_i + \cdots + P_g ECM_g = \sum_{i=1}^{g} p_i ECM_i$$

with $p_i = prior\ probability\ of\ C_i$

$$\Leftrightarrow \; TECM = \sum_{i=1}^{g} p_i \left(\sum_{k=1, k \neq i}^{g} P(k|i)c(k|i) \right)$$

$$= \sum_{i=1}^{g} p_i \left(\sum_{k=1, k \neq i}^{g} \int_{C_k} f_i(\boldsymbol{x})\, c(k|i)\, d\boldsymbol{x} \right) \left(\because P(k|i) = \int_{C_k} f_i(\boldsymbol{x})\, d\boldsymbol{x} \right)$$

$$= \sum_{k=1, k \neq i}^{g} \int_{C_k} \sum_{i=1, i \neq k}^{g} p_i f_i(\boldsymbol{x})\, c(k|i)\, d\boldsymbol{x}$$

$$\Rightarrow \; Min\, TECM \;\Leftrightarrow\; \sum_{i=1, i \neq k}^{g} p_i f_i(\boldsymbol{x})\, c(k|i) : smallest$$

$$\Leftrightarrow \; p_k f_k(\boldsymbol{x})\, c(k|i) : largest$$

6.4 DCA with Several Clusters

[Table 6.4.3] Special cases of TECM classification rules

For observed $\boldsymbol{x}_o = (x_{o1}, x_{o2}, \cdots, x_{op})^t$

- $c(k|i)/c(i|k) = 1$:

  for all i $(\neq k)$, $\boxed{p_k f_k(\boldsymbol{x}_o) > p_i f_i(\boldsymbol{x}_o) \Rightarrow \boldsymbol{x}_o \text{ is assigned to cluster } C_k,}$ $k = 1, \ldots, g$

- $p_i/p_k = 1$:

  for all i $(\neq k)$, $\boxed{f_k(\boldsymbol{x}_o) c(i|k) > f_i(\boldsymbol{x}_o) c(k|i) \Rightarrow \boldsymbol{x}_o \text{ is assigned to cluster } C_k,}$ $k = 1, \ldots, g$

- $p_i/p_k = c(k|i)/c(i|k) = 1$:

  for all i $(\neq k)$, $\boxed{f_k(\boldsymbol{x}_o) > f_i(\boldsymbol{x}_o) \Rightarrow \boldsymbol{x}_o \text{ is assigned to cluster } C_k,}$ $k = 1, \ldots, g$

6.5 DCA of Several Multivariate Normal Clusters

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_p)^T \sim N_p(\boldsymbol{\mu}_i, \Sigma_i), \; i = 1, 2, \ldots, g$$

[Table 6.5.1]	TECM Classification Rule : Misclassification Costs are same

For observed $\boldsymbol{x}_o = (x_{o1}, x_{o2}, \cdots, x_{op})^t$

- $\Sigma_1 \neq \cdots \neq \Sigma_g$:

$$\widehat{Q}_k(\boldsymbol{x}_o) = \max\{\widehat{Q}_1(\boldsymbol{x}_o), \widehat{Q}_2(\boldsymbol{x}_o), \ldots, \widehat{Q}_g(\boldsymbol{x}_o)\}$$

$\Rightarrow$ $\boldsymbol{x}_o$ is assigned to cluster C_k

where $\widehat{Q}_i(\boldsymbol{x}) = -\frac{1}{2}\ln|S_i| - \frac{1}{2}(\boldsymbol{x} - \overline{\boldsymbol{x}}_i)^t S_i^{-1}(\boldsymbol{x} - \overline{\boldsymbol{x}}_i) + \ln p_i$

: quadratic discriminant score of cluster C_i

- $\Sigma_1 = \cdots = \Sigma_g$:

linear discriminant score : $\widehat{L}_i(\boldsymbol{x}) = \overline{\boldsymbol{x}}_i^t S_p^{-1}\boldsymbol{x} - \frac{1}{2}\overline{\boldsymbol{x}}_i^t S_p^{-1}\overline{\boldsymbol{x}}_i + \ln p_i, \; i = 1, 2, \ldots, g$

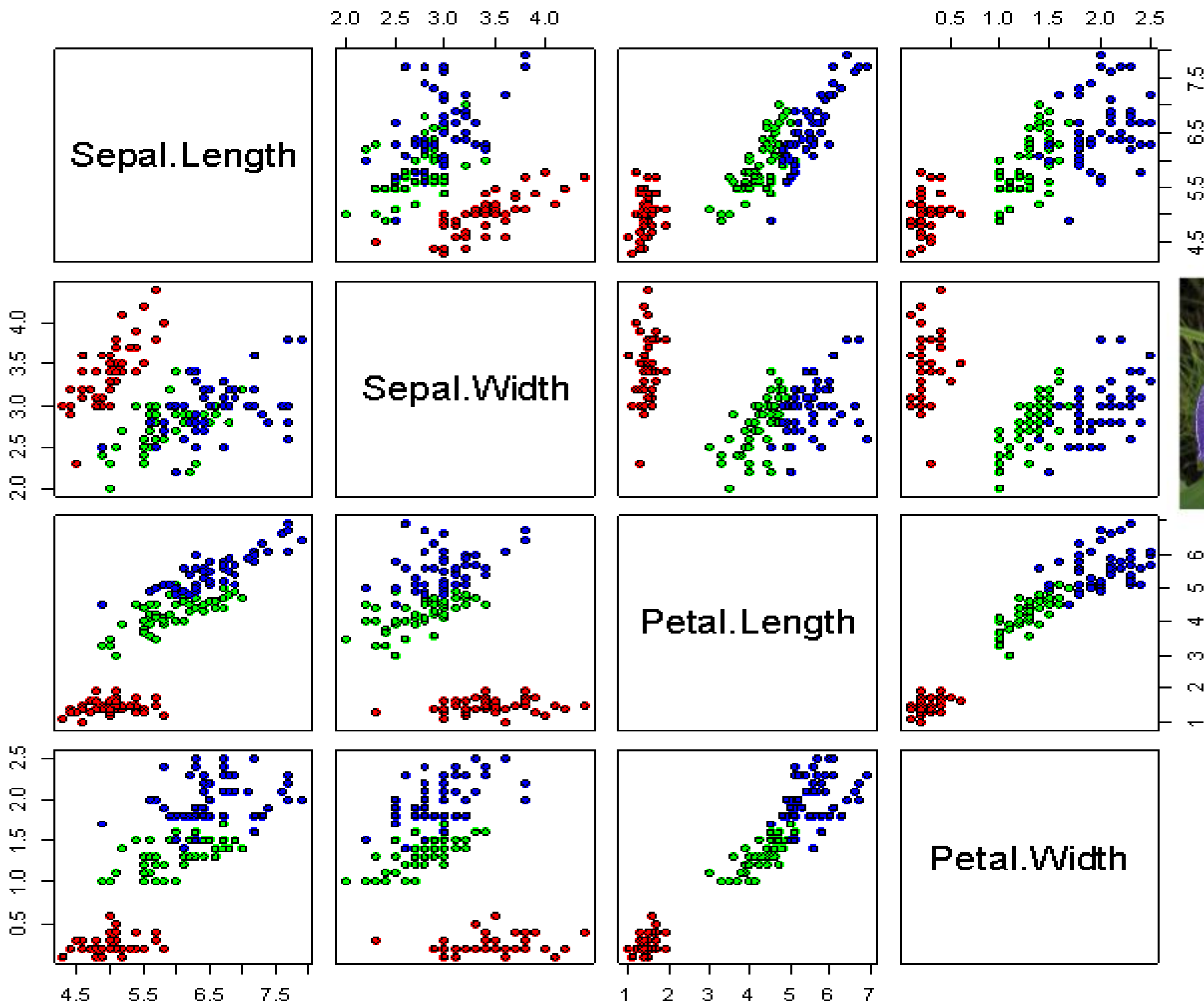
$$\widehat{L}_k(\boldsymbol{x}_o) = \max\{\widehat{L}_1(\boldsymbol{x}_o), \widehat{L}_2(\boldsymbol{x}_o), \ldots, \widehat{L}_g(\boldsymbol{x}_o)\}$$

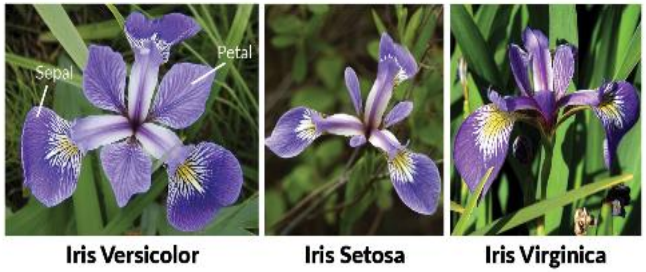
$\Rightarrow$ $\boldsymbol{x}_o$ is assigned to cluster C_k

6.5 DCA of Several Multivariate Normal Clusters

- [Example 6.5.1] DCA of Fisher's three iris species

- [Figure 6.5.1] Multiple scatter plot for three kinds of iris (setosa, versicolor, virginica)





6.5 DCA of Several Multivariate Normal Clusters

- **[Example 6.5.1] DCA of Fisher's three iris species**

[R-code 6.5.1]

STEP 1

[STEP 1] Prepare multivariate data matrix.

- three cluster of iris data : C_1(setosa), C_2(versicolor), C_3(virginica)
- number of each cluster : $n_1 = n_2 = n_3 = 50$
- mean of three clusters : $\bar{x}_1 = (5.006, 3.428, 1.462, 0.246)^t$,

$$\bar{x}_2 = (5.936, 2.770, 4.260, 1.326)^t$$

$$\bar{x}_3 = (6.588, 2.974, 5.552, 2.026)^t$$

- Covariance matrix :

covariance matrix	Sepal.Length	Sepal.Width	Petal.length	Petal.width
$S_1 =$	0.124 0.099 0.016 0.010	0.099 0.144 0.012 0.009	0.016 0.012 0.030 0.006	0.010 0.009 0.006 0.011
$S_2 =$	0.266 0.085 0.183 0.056	0.085 0.098 0.083 0.041	0.183 0.083 0.221 0.073	0.056 0.041 0.03 0.039
$S_3 =$	0.404 0.094 0.303 0.049	0.094 0.104 0.071 0.048	0.303 0.071 0.305 0.049	0.049 0.048 0.049 0.075

6.5 DCA of Several Multivariate Normal Clusters

- **[Example 6.5.1] DCA of Fisher's three iris species**

STEP 2

Test whether each cluster data follows multivariate normality

- In [Example 1.8.2], all three clusters were satisfied with multivariate normality.

STEP 3

If [STEP 2] is satisfied, test the homogeneity of each **Covariance matrix**

- $S_p = \dfrac{(n_1-1)S_1+(n_2-1)S_2+(n_3-1)S_3}{n_1+n_2+n_3-3} = \dfrac{(49)S_1+(49)S_2+(49)S_3}{150-3}$

pooled covariance matrix	Sepal.Length	Sepal.Width	Petal.length	Petal.width
$S_p =$	0.265	0.093	0.168	0.038
	0.093	0.115	0.055	0.033
	0.168	0.055	0.185	0.043
	0.038	0.033	0.043	0.042

box M-test: $M = -2\gamma \ln \Lambda = \gamma \sum_{k=1}^{3} (n_k - 1)\log\left|S_k^{-1} S_p\right| = 140.943$

$df = p(p+1)(g-1)/2 = 4(4+1)(3-1) = 20$

p-value = 0.000

Reject $H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3$

6.5 DCA of Several Multivariate Normal Clusters

- **[Example 6.5.1] DCA of Fisher's three iris species**

STEP 4

If the homogeneity of the covariance matrix is satisfied, LDA(Linear Discriminant Analysis) is performed. If not, QDA(Quadratic Discriminant Analysis) is performed.

- a classified table : MER = (0 + 2 + 1)/150 x 100% = 2%

Confusion Table		classification by QDA			
		setosa	versicolor	virginica	
cluster	setosa	50	0	0	50
	versicolor	0	48	2	50
	virginica	0	1	49	50

6.6 Evaluation of DCA

Apparent Error Rate

Criteria $\begin{cases} APER_g - RSM(resubstritution\ method) \\ EAER_g - CVM(cross-validation\ method) \end{cases}$

Expected Actual Error Rate

$$TPM = p_1 P(2|1) + p_2 P(1|2) = p_2 \int_{C_2} f_1(\boldsymbol{x})d\boldsymbol{x} + p_1 \int_{C_1} f_2(\boldsymbol{x})d\boldsymbol{x} \Rightarrow \text{MinTPM} = \text{OER st } C_1 \wedge C_2 \text{ satisfying Classification Rule}$$

- OER (optimum error rate) : $\text{OER} = p_2 \int_{C_2} f_1(\boldsymbol{x})d\boldsymbol{x} + p_1 \int_{C_1} f_2(\boldsymbol{x})d\boldsymbol{x}$
- AER(actual error rate) = $\widehat{OER}$ with $\widehat{f_1} \wedge \widehat{f_2}$ but unknown $f_1 \wedge f_2$
- APER (apparent error rate) : $APER_2 = \left(\dfrac{n_{12M} + n_{21M}}{n_1 + n_2}\right) 100\%$

[Table 6.6.1] Table of classification in DCA of two clusters

Confusion Table		Classification of two clusters by discriminant function		
		C_1	C_2	
cluster	C_1	n_{11C}	n_{12M}	n_1
	C_2	n_{21M}	n_{22C}	n_2

6.6 Evaluation of DCA

(1) Criterion for performance of RSM(resubstitution method)

$$APER_g = \left(\frac{\sum_{k=1}^{g} (n_k - n_{kkC})}{\sum_{k=1}^{g} n_k} \right) 100\% = 100\% - \boxed{CCR_g}$$

Correct Classification Rate

[Table 6.6.2] Table of classification in DCA of $g(>2)$

Confusion Table		Classification of two clusters by discriminant function				
		C_1	C_2	$\cdots$	C_g	
cluster	C_1	n_{11C}	n_{12M}	$\cdots$	n_{1gM}	n_1
	C_2	n_{21M}	n_{22C}	$\cdots$	n_{1gM}	n_2
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	C_g	n_{g1M}	n_{g2M}	$\cdots$	n_{ggC}	n_g

6.6 Evaluation of DCA

- **[Example 6.6.1]** Evaluation of misclassification rate of LDF by RSM in riding mower data

[Example 6.3.3]		Classification by linear discriminant function		
		Owner	Nonowner	
Clusters	Owner	11	1	12
	Nonowner	2	10	12

$$APER_2 = \left(\frac{1+2}{12+12}\right)100\% = 12.5\%.$$

[Example 6.3.4]		Classification by quadratic discriminant function		
		setosa	versicolor	
Clusters	setosa	49	1	50
	versicolor	0	50	50

$$APER_2 = \left(\frac{1+0}{50+50}\right)100\% = 1\%.$$

[Example 6.5.1]		classification by QDA			
		setosa	versicolor	virginica	
cluster	setosa	50	0	0	50
	versicolor	0	48	2	50
	virginica	0	1	49	50

$$APER_3 = \left(\frac{0+2+1}{50+50+50}\right)100\% = 2\%$$

Note: APER is intuitively appealing and easy to calculate.
Unfortunately, it trends to underestimate the AER

6.6 Evaluation of DCA

- Classical Cross-Validation Method C_1 C_2

Note: CV = TRUE in LDA() or QDA()

Sample C : $n = n_1 + n_2$

Spliting the total sample

Training Sample *New* C_1^*	Validation Sample *New* C_2^*
N_1 Observations	N_2 Observations

Classification function

$n_M^{(V)}$ = Number of observations of C_2^* misclassified into training sample C_1^*

$$\frac{n_M^{(V)}}{N_2} = Error\,Rate$$

Two main defects:

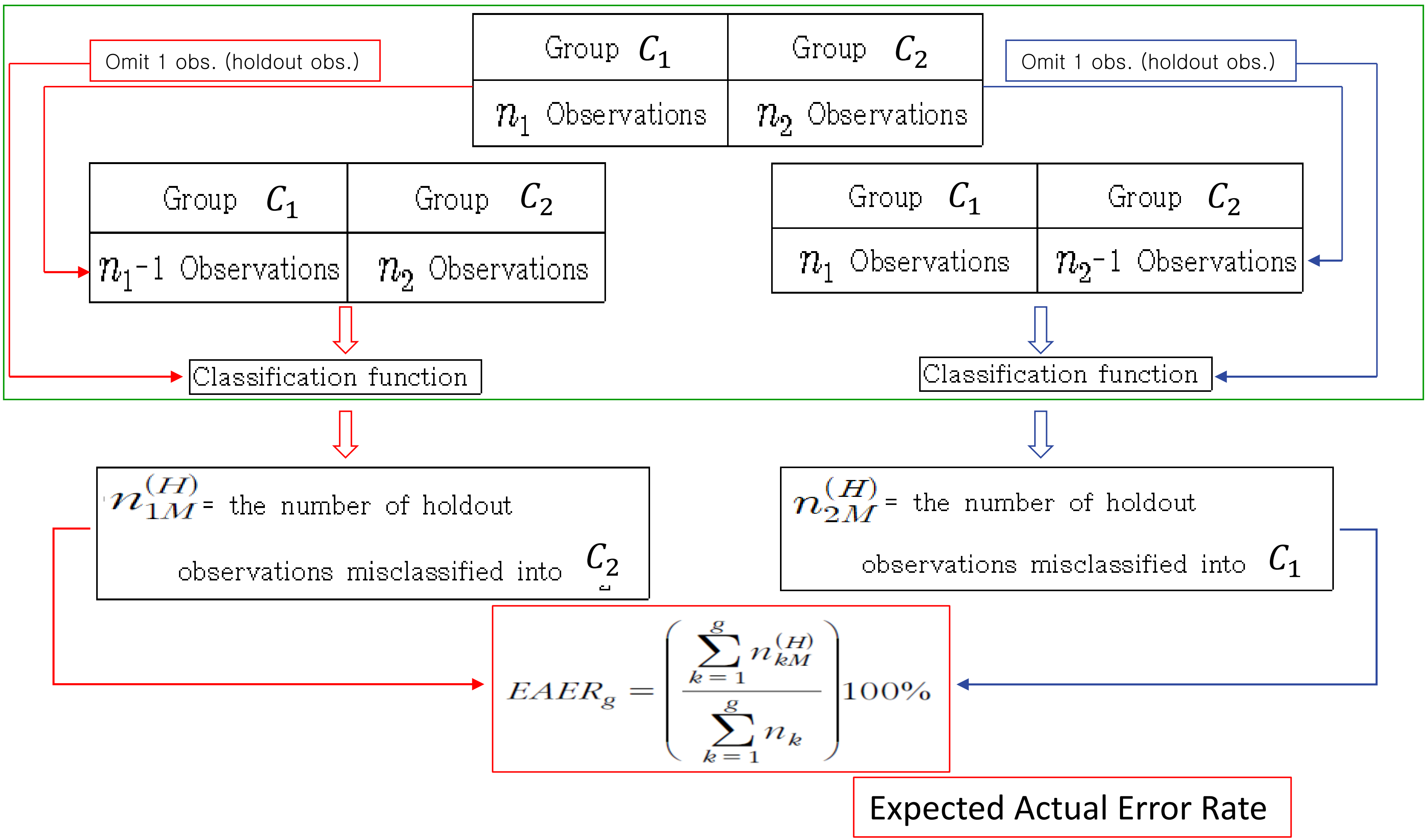
1) Large samples are required.
2) Discriminant(classification) function of training sample is not the function of interested.

=> Ultimately, almost all of the data must be used to construct the classification function

6.6 Evaluation of DCA

(2) Criterion for performance of CVM(cross-validation method)

- **[Table 6.6.3]** Lachenbruch Holdout Procedure by CVM



6.6 Evaluation of DCA

- **[Example 6.6.2]** Expected misclassification rate by CVM [R-code 6.6.1] : CV = TRUE

[Example 6.6.1]		Classification of CVM by LDA		
		Owner	Nonowner	
Clusters	Owner	10	2	12
	Nonowner	3	9	12

$$EAER_2 = \left(\frac{2+3}{12+12}\right)100\% = 20.83\%$$

$$APER_2 = \left(\frac{1+2}{12+12}\right)100\% = 12.5\%.$$

[Example 6.6.2]		Classification of CVM by QDA			
		setosa	versicolor	virginica	합계
Clusters	setosa	50	0	0	50
	versicolor	0	47	3	50
	virginica	0	1	49	50

RSM

$$EAER_3 = \left(\frac{0+3+1}{50+50+50}\right)100\% = 2.67\%$$

$$APER_3 = \left(\frac{0+2+1}{50+50+50}\right)100\% = 2\%$$

6.11 R for DCA : Practice Time

R-Code:

Linear DCA	Quadratic DCA
library(MASS) lda()	library(MASS) qda()
Homogeneity Test of Covariance Matrices : library(biotools) > boxM() Multivariate Normality Test : library(MVN) > mvn() or mardiaTest()	

6.11 R for DCA : Practice Time

R-code list of Chapter 6 Discriminant Analysis

setosaversicolor-Boxtest.R	[R-코드 6.3.1]	두 종류 붓꽃 자료의 공분산행렬 동질성 검정
ridingmower-Boxtest.R	[R-코드 6.3.2]	예초기 자료에 대한 공분산행렬의 동질성 검정
ridingmower-Mardiatest-LDA.R	[R-코드 6.3.3]	예초기 자료의 다변량정규성 검정과 LDA
setosaversicolor-mardiaTest-QDA.R	[R-코드 6.3.4]	두 종류 붓꽃 자료에 대한 다변량정규성 검정과 QDA
iris-QDA.R	[R-코드 6.5.1]	세 종류 붓꽃 자료의 QDA
wine-DA.R	[R-코드 6.5.2]	세 종류 와인 자료의 QDA
ridingmower-iris-DA-CVM.R	[R-코드 6.6.1]	예초기 자료와 붓꽃 자료의 DA에서 CVM에 의한 기대오분류율
ridingmower-FisherLDA.R	[R-코드 6.8.1]	예초기 자료에 대한 피셔의 선형판별분석
fisherexample-FisherLDA.R	[R-코드 6.8.2]	세 군집 자료의 피셔 선형판별분석
ridingmower-LogisticDA.R	[R-코드 6.9.1]	예초기 자료에 대한 로지스틱 판별분석
blood-LogisticDA-QDA.R	[R-코드 6.9.2]	적혈구 자료에 대한 이차 및 로지스틱 판별분석
iris-CART.R	[R-코드 6.10.1]	세 종류 붓꽃 자료의 분류나무
wine-DA.R	[R-코드 6.10.2]	세 종류 와인 자료의 분류나무
kyphosis-CRAT.R	[R-코드 6.10.3]	척추수술 자료의 분류나무와 가지치기

6.11 R for DCA : Practice Time

[R-code 6.3.2] Homogeneity Test of Covariance Matrices(ridingmower-Boxtest.R)

```
# Box's test for the hypothesis test of the equality of covariance matrices
# in Riding-Mowers data.
library(MASS)
ridingmower<-read.table("ridingmower.txt", header=T)
attach(ridingmower)

# Scatter Plot
plot(ridingmower[, 2:3], pch=unclass(pop), col=1:2)

# Density function of Variables
par(mfrow=c(1, 2))
ldahist(data= x1.Income, g=pop, width=30, type="density")
ldahist(data= x2.Lotsize, g=pop, width=10, type="density")

cov.Mtest=function(x, ina, a=0.05){
## x is the data set
## ina is a numeric vector indicating the groups of the data set
## a is the significance level, set to 0.05 by default
x=as.matrix(x)
p=ncol(x) ## dimension of the data set
n=nrow(x) ## total sample size
k=max(ina) ## number of groups
nu=rep(0, k) ## the sample size of each group will be stored here later
pame=rep(0, k) ## the determinant of each covariance will be stored here
## the next "for" function calculates the covariance matrix of each group
nu=as.vector(table(ina))
mat=mat1=array(dim=c(p, p, k))
for (i in 1:k) {
mat[, , i]=cov(x[ina==i, ])
pame[i]=det(mat[, , i]) ## the detemirnant of each covariance matrix
mat1[, , i]=(nu[i]-1)*cov(x[ina==i, ]) }
## the next 2 lines calculate the pooled covariance matrix
Sp=apply(mat1, 1:2, sum)
Sp=Sp/(n-k)
for (i in 1:k)
pamela=det(Sp) ## determinant of the pooled covariance matrix
test1=sum((nu-1)*log(pamela/pame))
gama1=(2*(p^2)+3*p-1)/(6*(p+1)*(k-1))
gama2=(sum(1/(nu-1))-1/(n-k))
gama=1-gama1*gama2
test=gama*test1 ## this is the M (test statistic)
df=0.5*p*(p+1)*(k-1) ## degrees of freedom of the chi-square distribution
pvalue=1-pchisq(test, df) ## p-value of the test statistic
crit=qchisq(1-a, df) ## critical value of the chi-square distribution
list(M.test=test, degrees=df, critical=crit, p.value=pvalue) }

ina=as.numeric(ridingmower[, 1])
x=ridingmower[, 2:3]
cov.Mtest(x, ina)
```

6.11 R for DCA : Practice Time

[R-code 6.5.1] QDA of Iris flowers with Homogeneity Test of Covariance Matrices(iris-QDA.R)

```
llibrary("MVN")
library(biotools)

#[Step 1] Iris flower data : setosa, versicolor, virginica
iris
attach(iris)
pairs(iris[1:4], pch=21, bg=c("red", "green", "blue")[unclass(iris$Species)])

#[Step 2] MVN tests based on the Skewness and Kurtosis Ststistics
setosa = iris[1:50, 1:4] # Iris data only for setosa and four variables
versicolor = iris[51:100, 1:4] # Iris data only for versicolor and four variables
virginica = iris[101:150, 1:4] # Iris data only for virginica and four variables
result_setosa = mardiaTest(setosa, qqplot = TRUE)
result_versicolor = mardiaTest(versicolor, qqplot = TRUE)
result_virginica = mardiaTest(virginica, qqplot = TRUE)
list(result_setosa, result_versicolor, result_virginica)

#[Step 3] Box's M-Test for Equlaity of Covariance Matrices
boxM(iris[, -5], iris[, 5])
S1=cov(setosa);S2=cov(versicolor);S3=cov(virginica)
Sp=(49*S1+49*S2+49*S3)/(150-3)
list(S1, S2, S3, Sp)

# [Step 4] Quadratic DA with 3 groups applying
# resubstitution prediction and equal prior probabilities.
QDA=qda(Species~ Sepal.Length+Sepal.Width+Petal.Length+Petal.Width,
                 data=iris, prior=c(1,1,1)/3)
qcluster=predict(QDA, iris)$class
```

```
mvn(setosa , mvnTest="mardia", multivariatePlot="qq")
```

```
table(Species, qcluster)

# Total percent correct
mean(Species==qcluster)
```

6.11 R for DCA : Practice Time

[R-code 6.3.3] ridingmower-Mardiatest-LDA.R

```
# MVD test and Linear DA for two groups(owner, nonowner)
# of riding mower on two variables(x1.Income x2.Lotsize)
library(MASS)
library(MVN)

ridingmower<-read.table("ridingmower.txt", header=T)
attach(ridingmower)
ridingmower
# MVN tests based on the Skewness and Kurtosis Statistics
par(mfrow=c(1, 2))
owner=ridingmower[1:12, 2:3]
nonowner=ridingmower[13:24, 2:3]
owner
nonowner
result_owner=mardiaTest(owner, qqplot=TRUE)
result_nonowner=mardiaTest(nonowner, qqplot=TRUE)
result_owner
result_nonowner

# Linear DA
LDA=lda(pop~x1.Income+x2.Lotsize, data=ridingmower)
LDA
lcluster=predict(LDA, ridingmower)$class
lct=table(pop, lcluster)
lct
# Total percent correct
mean(pop==lcluster)
```

[R-code 6.3.4] setosaversicolor-mardiaTest-QDA.R

```
# Quadratic DA for two groups(setosa, vesicolor) on two variables(꽃받침길이,꽃받침폭)
library(MASS)
library(MVN)
setosa_versi<-read.table("setosaversicolor.txt", header=T)
attach(setosa_versi)

# MVN tests based on the Skewness and Kurtosis Statistics
par(mfrow=c(1,2))
setosa=setosa_versi[1:50, 1:2]
versicolor=setosa_versi[51:100, 1:2]
setosa
versicolor
result_setosa = mardiaTest(setosa, qqplot = TRUE)
result_versicolor = mardiaTest(versicolor, qqplot = TRUE)
result_setosa
result_versicolor

# Quadratic DA with 3 groups applying
 # resubstitution prediction and equal prior probabilities.
QDA=qda(종류~꽃받침길이+꽃받침폭, data=setosa_versi, prior=c(1,1)/2)
QDA
qcluster=predict(QDA, setosa_versi)$class
qct=table(종류, qcluster)
qct
# Total percent correct
mean(종류==qcluster)
```

```
mvn(data, mvnTest="mardia", multivariatePlot="qq")
```

6.11 R for DCA : Practice Time

Testing Ho : MVN

```
> mvn(owner,mvnTest="mardia", multivariatePlot="qq")
$multivariateNormality
             Test          Statistic           p value Result
1 Mardia Skewness  0.699134098116459  0.951435661555086    YES
2 Mardia Kurtosis -1.01248511159407   0.311306161804624    YES
3             MVN               <NA>               <NA>    YES
```

```
> mvn(nonowner,mvnTest="mardia", multivariatePlot="qq")
$multivariateNormality
             Test          Statistic           p value Result
1 Mardia Skewness   1.76367521596657  0.779121047019934    YES
2 Mardia Kurtosis  -1.01338802018451  0.310874859294073    YES
3             MVN               <NA>               <NA>    YES
```

```
> mvn(setosa,mvnTest="mardia", multivariatePlot="qq")
$multivariateNormality
             Test          Statistic           p value Result
1 Mardia Skewness  0.759503524380438  0.943793240544741    YES
2 Mardia Kurtosis 0.0934600553610254  0.925538081956867    YES
3             MVN               <NA>               <NA>    YES
```

```
> mvn(versicolor,mvnTest="mardia", multivariatePlot="qq")
$multivariateNormality
             Test          Statistic           p value Result
1 Mardia Skewness   1.76081423820927  0.779643267719773    YES
2 Mardia Kurtosis -0.906486857998 67  0.364678217772254    YES
3             MVN               <NA>               <NA>    YES
```