Chapter 11

11.1 **(a)** The linear discriminant function given in (11-19) is

$$\hat{y} = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}_{\text{pooled}}^{-1} \boldsymbol{x} = \hat{\boldsymbol{a}}' \boldsymbol{x}$$

where

$$\boldsymbol{S}_{\text{pooled}}^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$$

so the the linear discriminant function is

$$\left(\begin{bmatrix} 3 \\ 6 \end{bmatrix} - \begin{bmatrix} 5 \\ 8 \end{bmatrix} \right)' \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \boldsymbol{x} = [-2 \quad 0] \boldsymbol{x} = -2x_1$$

(b)

$$\hat{m} = \frac{1}{2}(\hat{y}_1 + \hat{y}_2) = \frac{1}{2}(\hat{\boldsymbol{a}}'\bar{\boldsymbol{x}}_1 + \hat{\boldsymbol{a}}'\bar{\boldsymbol{x}}_2) = -8$$

Assign $\boldsymbol{x}_0'$ to π_1 if

$$\hat{y}_0 = [2 \quad 7]\boldsymbol{x}_0 \geq \hat{m} = -8$$

and assign $\boldsymbol{x}_0$ to π_2 otherwise.

Since $[-2 \quad 0]\boldsymbol{x}_0 = -4$ is greater than $\hat{m} = -8$, assign $\boldsymbol{x}_0'$ to population π_1.

11.2 (a) $\pi_1 \equiv$ Riding-mower owners; $\pi_2 \equiv$ Nonowners

Here are some summary statistics for the data in Example 11.1:

$$\bar{x}_1 = \begin{bmatrix} 109.475 \\ 20.267 \end{bmatrix}, \qquad \bar{x}_2 = \begin{bmatrix} 87.400 \\ 17.633 \end{bmatrix}$$

$$S_1 = \begin{bmatrix} 352.644 & -11.818 \\ -11.818 & 4.082 \end{bmatrix}, \qquad S_2 = \begin{bmatrix} 200.705 & -2.589 \\ -2.589 & 4.464 \end{bmatrix}$$

$$S_{\text{pooled}} = \begin{bmatrix} 276.675 & -7.204 \\ -7.204 & 4.273 \end{bmatrix}, \qquad S_{\text{pooled}}^{-1} = \begin{bmatrix} .00378 & .00637 \\ .00637 & .24475 \end{bmatrix}$$

The linear classification function for the data in Example 11.1 using (11-19) is

$$\left(\begin{bmatrix} 109.475 \\ 20.267 \end{bmatrix} - \begin{bmatrix} 87.400 \\ 17.633 \end{bmatrix}\right)' \begin{bmatrix} .00378 & .00637 \\ .00637 & .24475 \end{bmatrix} x = \begin{bmatrix} .100 & .785 \end{bmatrix} x$$

where

$$\hat{m} = \frac{1}{2}(\bar{y}_1 + \bar{y}_2) = \frac{1}{2}(\hat{a}'\bar{x}_1 + \hat{a}'\bar{x}_2) = 24.719$$

(b) Assign an observation x to π_1 if

$$0.100x_1 + 0.785x_2 \geq 24.72$$

Otherwise, assign x to π_2

Here are the observations and their classifications:

Owners			Nonowners		
Observation	$a'x_0$	Classification	Observation	$a'x_0$	Classification
1	23.444	nonowner	1	25.886	owner
2	24.738	owner	2	24.608	nonowner
3	26.436	owner	3	22.982	nonowner
4	25.478	owner	4	23.334	nonowner
5	30.226	owner	5	25.216	owner
6	29.082	owner	6	21.736	nonowner
7	27.616	owner	7	21.500	nonowner
8	28.864	owner	8	24.044	nonowner
9	25.600	owner	9	20.614	nonowner
10	28.628	owner	10	21.058	nonowner
11	25.370	owner	11	19.090	nonowner
12	26.800	owner	12	20.918	nonowner

From this, we can construct the confusion matrix:

		Predicted Membership		
		π_1	π_2	Total
Actual membership	π_1	11	1	12
	π_2	2	10	12

(c) The apparent error rate is $\frac{1+2}{12+12} = 0.125$

(d) The assumptions are that the observations from π_1 and π_2 are from multivariate normal distributions with equal covariance matrices, $\Sigma_1 = \Sigma_2 = \Sigma$.

11.3 We need to show that the regions R_1 and R_2 that minimize the ECM are defined by the values x for which the following inequalities hold:

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

$$R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)$$

Substituting the expressions for $P(2|1)$ and $P(1|2)$ into (11-5) gives

$$\text{ECM} = c(2|1)p_1 \int_{R_2} f_1(x)dx + c(1|2)p_2 \int_{R_1} f_2(x)dx$$

And since $\Omega = R_1 \cup R_2$,

$$1 = \int_{R_1} f_1(x)dx + \int_{R_2} f_1(x)dx$$

and thus,

$$\text{ECM} = c(2|1)p_1 \left[1 - \int_{R_1} f_1(x)dx\right] + c(1|2)p_2 \int_{R_1} f_2(x)dx$$

Since both of the integrals above are over the same region, we have

$$\text{ECM} = \int_{R_1} [c(1|2)p_2 f_2(x)dx - c(2|1)p_1 f_1(x)]dx + c(2|1)p_1$$

The minimum is obtained when R_1 is chosen to be the region where the term in brackets is less than or equal to 0. So choose R_1 so that

$$c(2|1)p_1 f_1(x) \geq c(1|2)p_2 f_2(x) \quad \text{or}$$

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

11.4 **(a)** The minimum ECM rule is given by assigning an observation $\boldsymbol{x}$ to π_1 if

$$\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} \geq \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right) = \left(\frac{100}{50}\right)\left(\frac{.2}{.8}\right) = .5$$

and assigning $\boldsymbol{x}$ to π_2 if

$$\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} < \left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right) = \left(\frac{100}{50}\right)\left(\frac{.2}{.8}\right) = .5$$

(b) Since $f_1(\boldsymbol{x}) = .3$ and $f_2(\boldsymbol{x}) = .5$,

$$\frac{f_1(\boldsymbol{x})}{f_2(\boldsymbol{x})} = .6 \geq .5$$

and assign $\boldsymbol{x}$ to π_1.

11.5

$$-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2) =$$

$$-\frac{1}{2}[\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}-2\boldsymbol{\mu}_1'\boldsymbol{\Sigma}^{-1}\mathbf{x}+\boldsymbol{\mu}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1-\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}+2\boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}\mathbf{x}-\boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2]$$

$$= -\frac{1}{2}[-2(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\mathbf{x}+\boldsymbol{\mu}_1'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1-\boldsymbol{\mu}_2'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2]$$

$$= (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2) .$$

11.6 **a)** $E(\underset{\sim}{a}'\underset{\sim}{X}|\pi_1) - m = \underset{\sim}{a}'\underset{\sim}{\mu}_1 - m = \underset{\sim}{a}'\underset{\sim}{\mu}_1 - \frac{1}{2}\underset{\sim}{a}'(\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2)$

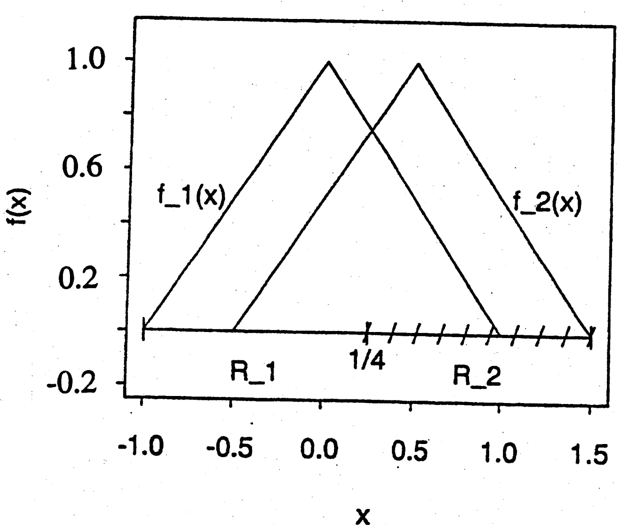
$$= \frac{1}{2}\underset{\sim}{a}'(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2) = \frac{1}{2}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)'\Sigma^{-1}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2) > 0 \text{ since}$$

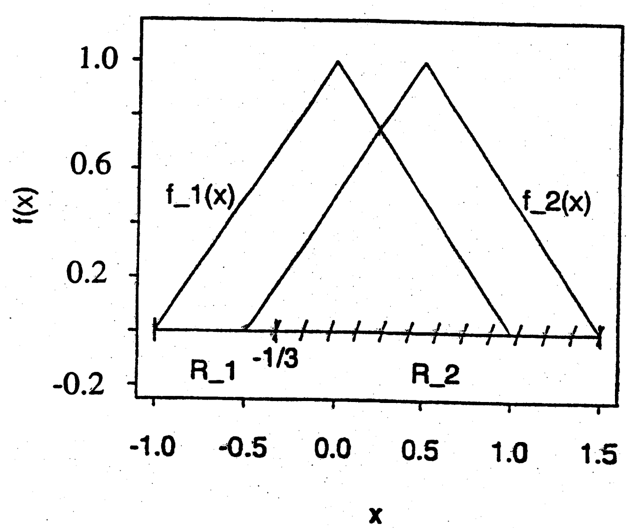
Σ^{-1} is positive definite.

b) $E(\underset{\sim}{a}'\underset{\sim}{X}|\pi_2) - m = \underset{\sim}{a}'\underset{\sim}{\mu}_2 - m = \frac{1}{2}\underset{\sim}{a}'(\underset{\sim}{\mu}_2 - \underset{\sim}{\mu}_1)$

$$= -\frac{1}{2}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)'\Sigma^{-1}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2) < 0.$$

11.7 **(a)** Here are the densities:





(b) When $p_1 = p_2$ and $c(1|2) = c(2|1)$, the classification regions are

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq 1 \qquad R_2 : \frac{f_1(x)}{f_2(x)} < 1$$

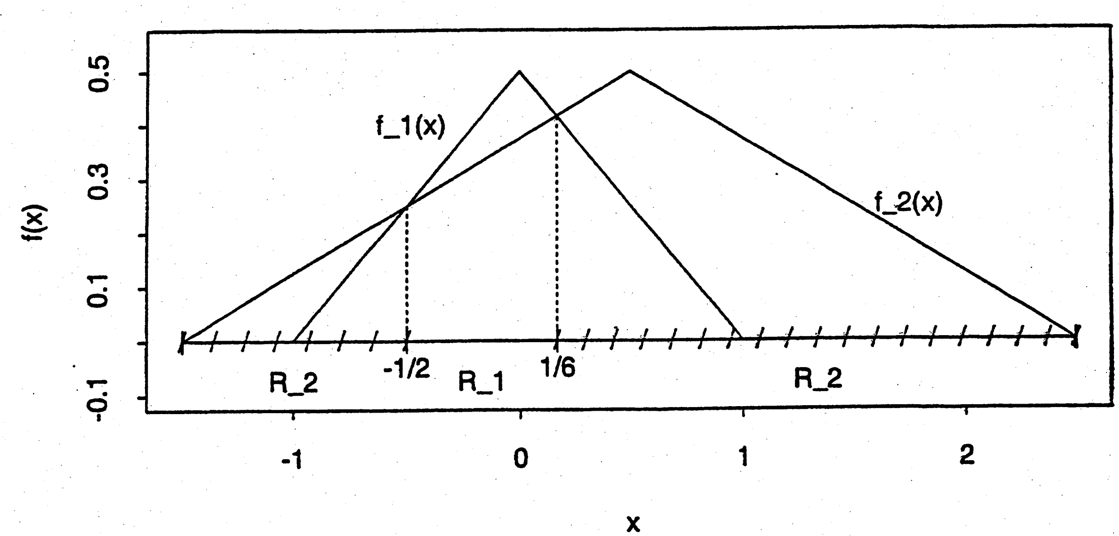
These regions are given by $R_1 : -1 \leq x \leq .25$ and $R_2 : .25 < x \leq 1.5$.

(c) When $p_1 = .2$, $p_2 = .8$, and $c(1|2) = c(2|1)$, the classification regions are

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \frac{p_2}{p_1} = .4 \qquad R_2 : \frac{f_1(x)}{f_2(x)} < .4$$

These regions are given by $R_1 : -1 \leq x \leq -1/3$ and $R_2 : -1/3 < x \leq 1.5$.

11.8 **(a)** Here are the densities:



(b) When $p_1 = p_2$ and $c(1|2) = c(2|1)$, the classification regions are

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq 1 \qquad R_2 : \frac{f_1(x)}{f_2(x)} < 1$$

These regions are given by

$$R_1 : -1/2 \le x < 1/6 \quad \text{and} \quad R_2 = -1.5 \le x < -1/2, \quad 1/6 \le x \le 2.5$$

11.9

$$\frac{\mathbf{a}'\mathbf{B}_\mu\mathbf{a}}{\mathbf{a}'\Sigma\mathbf{a}} = \frac{\underset{\sim}{a}'[(\underset{\sim}{\mu}_1-\bar{\underset{\sim}{\mu}})(\underset{\sim}{\mu}_1-\bar{\underset{\sim}{\mu}})' + (\underset{\sim}{\mu}_2-\bar{\underset{\sim}{\mu}})(\underset{\sim}{\mu}_2-\bar{\underset{\sim}{\mu}})']\underset{\sim}{a}}{\underset{\sim}{a}'\Sigma\underset{\sim}{a}}$$

where $\bar{\underset{\sim}{\mu}} = \frac{1}{2}(\underset{\sim}{\mu}_1 + \underset{\sim}{\mu}_2)$. Thus $\underset{\sim}{\mu}_1 - \bar{\underset{\sim}{\mu}} = \frac{1}{2}(\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2)$ and $\underset{\sim}{\mu}_2 - \bar{\underset{\sim}{\mu}} = \frac{1}{2}(\underset{\sim}{\mu}_2 - \underset{\sim}{\mu}_1)$ so

$$\frac{\mathbf{a}'\mathbf{B}_\mu\mathbf{a}}{\mathbf{a}'\Sigma\mathbf{a}} = \frac{\frac{1}{2}\,\underset{\sim}{a}'(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)'\underset{\sim}{a}}{\underset{\sim}{a}'\Sigma\underset{\sim}{a}} \quad .$$

11.10 **(a)** Hotelling's two-sample T^2-statistic is

$$
\begin{aligned}
T^2 &= (\bar{x}_1 - \bar{x}_2)' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right) S_{\text{pooled}}\right]^{-1} (\bar{x}_1 - \bar{x}_2) \\
&= [-3 \quad -2] \left[\left(\frac{1}{11} + \frac{1}{12}\right) \begin{bmatrix} 7.3 & -1.1 \\ -1.1 & 4.8 \end{bmatrix}\right]^{-1} \begin{bmatrix} -3 \\ -2 \end{bmatrix} = 14.52
\end{aligned}
$$

Under $H_0 : \underset{\sim}{\mu}_1 = \underset{\sim}{\mu}_2$,

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

Since $T^2 = 14.52 \geq \frac{(11+12-2)2}{11+12-2-1} F_{2,20}(.1) = 5.44$, we reject the null hypothesis $H_0 : \mu_1 = \mu_2$ at the $\alpha = 0.1$ level of significance.

(b) Fisher's linear discriminant function is

$$
\hat{y}_0 = \hat{a}' x_0 = -.49 x_1 - .53 x_2
$$

(c) Here, $\hat{m} = -.25$. Assign x_0' to π_1 if $-.49x_1 - .53x_2 + .25 \geq 0$. Otherwise assign x_0' to π_2.

For $x_0' = [0 \quad 1], \hat{y}_0 = -.53(1) = -.53$ and $\hat{y}_0 - \hat{m} = -.28 < 0$. Thus, assign x_0 to π_2.

11.11 Assuming equal prior probabilities $p_1 = p_2 = \frac{1}{2}$, and equal misclassification costs $c(2|1) = c(1|2) = \$10$:

c	$P(B1\|A2)$	$P(B2\|A1)$	$P(A2$ and $B1)$	$P(A1$ and $B2)$	$P(\text{error})$	Expected cost
9	.006	.691	.346	.003	.349	3.49
10	.023	.500	.250	.011	.261	2.61
11	.067	.309	.154	.033	.188	1.88
12	.159	.159	.079	.079	.159	1.59
13	.309	.067	.033	.154	.188	1.88
14	.500	.023	.011	.250	.261	2.61

Using (11- 5) , the expected cost is minimized for $c = 12$ and the minimum expected cost is \$1.59.

11.12 Assuming equal prior probabilities $p_1 = p_2 = \frac{1}{2}$, and misclassification costs $c(2|1) = \$5$ and $c(1|2) = \$10$,

expected cost $= \$5P(A1$ and $B2) + \$15P(A2$ and $B1)$.

c	$P(B1\|A2)$	$P(B2\|A1)$	$P(A2$ and $B1)$	$P(A1$ and $B2)$	$P(\text{error})$	Expected cost
9	0.006	0.691	0.346	0.003	0.349	1.78
10	0.023	0.500	0.250	0.011	0.261	1.42
11	0.067	0.309	0.154	0.033	0.188	1.27
12	0.159	0.159	0.079	0.079	0.159	1.59
13	0.309	0.067	0.033	0.154	0.188	2.48
14	0.500	0.023	0.011	0.250	0.261	3.81

Using (11- 5) , the expected cost is minimized for $c = 10.90$ and the minimum expected cost is \$1.27.

11.13 Assuming prior probabilities $P(A1) = 0.25$ and $P(A2) = 0.75$, and misclassification costs $c(2|1) = \$5$ and $c(1|2) = \$10$,

expected cost $= \$5P(B2|A1)(.25) + \$15P(B1|A2)(.75)$.

c	$P(B1\|A2)$	$P(B2\|A1)$	$P(A2 \text{ and } B1)$	$P(A1 \text{ and } B2)$	$P(\text{error})$	Expected cost
9	0.006	0.691	0.173	0.005	0.178	0.93
10	0.023	0.500	0.125	0.017	0.142	0.88
11	0.067	0.309	0.077	0.050	0.127	1.14
12	0.159	0.159	0.040	0.119	0.159	1.98
13	0.309	0.067	0.017	0.231	0.248	3.56
14	0.500	0.023	0.006	0.375	0.381	5.65

Using (11-5) , the expected cost is minimized for $c = 9.80$ and the minimum expected cost is \$0.88.

11.14 Using (11-21),

$$\hat{a}_1^* = \frac{\hat{a}}{\sqrt{\hat{a}'\hat{a}}} = \begin{bmatrix} .79 \\ -.61 \end{bmatrix} \quad \text{and} \quad \hat{m}_1^* = -0.10$$

Since $\hat{a}_1^* x_0 = -0.14 < \hat{m}_1^* = -0.1$, classify x_0 as π_2.

Using (11-22),

$$\hat{a}_2^* = \frac{\hat{a}}{\hat{a}_1} = \begin{bmatrix} 1.00 \\ -.77 \end{bmatrix} \quad \text{and} \quad \hat{m}_2^* = -0.12$$

Since $\hat{a}_2^* x_0 = -0.18 < \hat{m}_2^* = -0.12$, classify x_0 as π_2.

These results are consistent with the classification obtained for the case of equal prior probabilities in Example 11.3. These two classification results should be identical to those of Example 11.3.

11.15 $\dfrac{f_1(\underset{\sim}{x})}{f_2(\underset{\sim}{x})} \geq \left[\dfrac{c(1|2)}{c(2|1)}\,\dfrac{p_2}{p_1}\right]$ defines the same region as

$\ln f_1(\underset{\sim}{x}) - \ln f_2(\underset{\sim}{x}) \geq \ln\left[\dfrac{c(1|2)}{c(2|1)}\,\dfrac{p_2}{p_1}\right]$. For a multivariate normal distribution

$$\ln f_i(\underset{\sim}{x}) = -\frac{1}{2}\ln|\Sigma_i| - \frac{p}{2}\ln 2\pi - \frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu}_i)'\Sigma_i^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}_i),\ i = 1,2$$

so

$$\begin{aligned}
\ln f_1(\underset{\sim}{x}) - \ln f_2(\underset{\sim}{x}) &= -\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu}_1)'\Sigma_1^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}_1) \\
&\quad + \frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu}_2)'\Sigma_2^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}_2) - \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) \\
&= -\frac{1}{2}[\underset{\sim}{x}'\Sigma_1^{-1}\underset{\sim}{x} - 2\underset{\sim}{\mu}_1'\Sigma_1^{-1}\underset{\sim}{x} + \underset{\sim}{\mu}_1'\Sigma_1^{-1}\underset{\sim}{\mu}_1 \\
&\quad - \underset{\sim}{x}'\Sigma_2^{-1}\underset{\sim}{x} + 2\underset{\sim}{\mu}_2'\Sigma_2^{-1}\underset{\sim}{x} - \underset{\sim}{\mu}_2'\Sigma_2^{-1}\underset{\sim}{\mu}_2] - \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) \\
&= -\frac{1}{2}\underset{\sim}{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\underset{\sim}{x} + (\underset{\sim}{\mu}_1'\Sigma_1^{-1} - \underset{\sim}{\mu}_2'\Sigma_2^{-1})\underset{\sim}{x} - k
\end{aligned}$$

where $k = \dfrac{1}{2}\ln\left(\dfrac{|\Sigma_1|}{|\Sigma_2|}\right) + \dfrac{1}{2}(\underset{\sim}{\mu}_1'\Sigma_1^{-1}\underset{\sim}{\mu}_1 - \underset{\sim}{\mu}_2'\Sigma_2^{-1}\underset{\sim}{\mu}_2)$.

11.16

$$Q = \ln\left[\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}\right] = -\frac{1}{2}\ln|\Sigma_1| - \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)'\Sigma_1^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)$$

$$+\frac{1}{2}\ln|\Sigma_2| + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)'\Sigma_1^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)$$

$$= -\frac{1}{2}\mathbf{x}'(\Sigma_1^{-1}-\Sigma_2^{-1})\mathbf{x} + \mathbf{x}'\Sigma_1^{-1}\boldsymbol{\mu}_1 - \mathbf{x}'\Sigma_2^{-1}\boldsymbol{\mu}_2 - k$$

where $k = \frac{1}{2}\left[\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \boldsymbol{\mu}_1'\Sigma_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2'\Sigma_2^{-1}\boldsymbol{\mu}_2\right]$.

When $\Sigma_1 = \Sigma_2 = \Sigma$,

$$Q = \mathbf{x}'\Sigma^{-1}\boldsymbol{\mu}_1 - \mathbf{x}'\Sigma^{-1}\boldsymbol{\mu}_2 - \frac{1}{2}(\boldsymbol{\mu}_1'\Sigma^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2'\Sigma^{-1}\boldsymbol{\mu}_2)$$

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

11.17 Assuming equal prior probabilities and misclassification costs $c(2|1) = \$10$ and $c(1|2) = \$73.89$. In the table below,

$$\begin{aligned} Q &= -\frac{1}{2}\boldsymbol{x}_0'(\Sigma_1^{-1} - \Sigma_2^{-1})\boldsymbol{x}_0 + (\boldsymbol{\mu}_1'\Sigma_1^{-1} - \boldsymbol{\mu}_2'\Sigma_2^{-1})\boldsymbol{x}_0 \\ &\quad -\frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) - \frac{1}{2}(\boldsymbol{\mu}_1'\Sigma_1^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2'\Sigma_2^{-1}\boldsymbol{\mu}_2) \end{aligned}$$

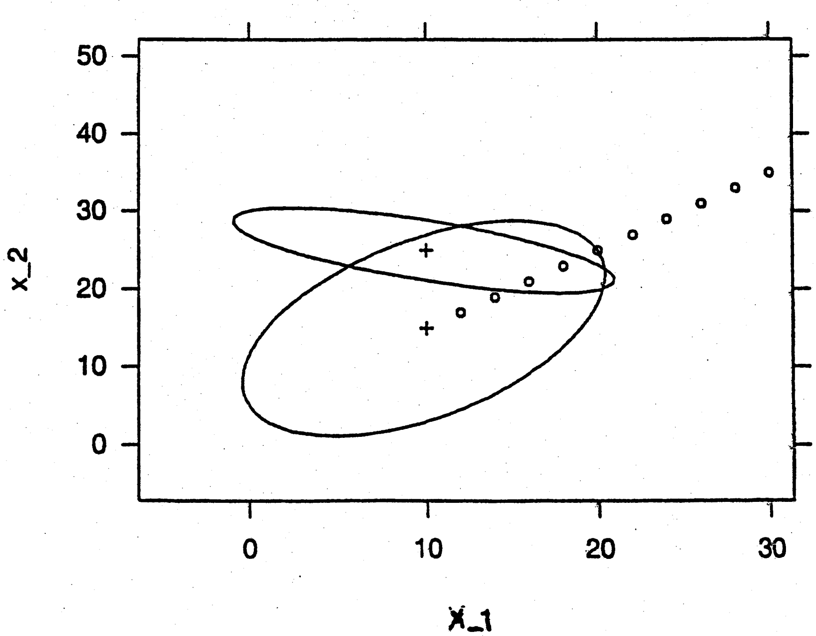
x	$P(\pi_1\|x)$	$P(\pi_2\|x)$	Q	Classification
$[10,\ 15]'$	1.00000	0	18.54	π_1
$[12,\ 17]'$	0.99991	0.00009	9.36	π_1
$[14,\ 19]'$	0.95254	0.04745	3.00	π_1
$[16,\ 21]'$	0.36731	0.63269	-0.54	π_2
$[18,\ 23]'$	0.21947	0.78053	-1.27	π_2
$[20,\ 25]'$	0.69517	0.30483	0.87	π_2
$[22,\ 27]'$	0.99678	0.00322	5.74	π_1
$[24,\ 29]'$	1.00000	0.00000	13.46	π_1
$[26,\ 31]'$	1.00000	0.00000	24.01	π_1
$[28,\ 33]'$	1.00000	0.00000	37.38	π_1
$[30,\ 35]'$	1.00000	0.00000	53.56	π_1

The quadratic discriminator was used to classify the observations in the above table. An observation x is classified as π_1 if

$$Q \geq \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] = \ln\left(\frac{73.89}{10}\right) = 2.0$$

Otherwise, classify x as π_2.

For (a), (b), (c) and (d), see the following plot.



11.18 The vector $\underset{\sim}{e}$ is an (unscaled) eigenvector of $\Sigma^{-1}B$ since

$$\Sigma^{-1}B\underset{\sim}{e} = \Sigma^{-1}c(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)'c\Sigma^{-1}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)$$

$$= c^2\Sigma^{-1}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)'\Sigma^{-1}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)$$

$$= \lambda\,\Sigma^{-1}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2) = \lambda\,\underset{\sim}{e}$$

where $\lambda = c^2\,(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)'\Sigma^{-1}(\underset{\sim}{\mu}_1-\underset{\sim}{\mu}_2)$.

11.19 (a) The calculated values agree with those in Example 11.7.

(b) Fisher's linear discriminant function is

$$\hat{y}_0 = \hat{\boldsymbol{a}}'\boldsymbol{x}_0 = -\frac{1}{3}x_1 + \frac{2}{3}x_2$$

where

$$\overline{y}_1 = \frac{17}{3};\ \overline{y}_2 = \frac{10}{3};\ \hat{m} = \frac{27}{6} = 4.5$$

Assign $\boldsymbol{x}_0'$ to π_1 if $-\frac{1}{3}x_1 + \frac{2}{3}x_2 - 4.5 \geq 0$

Otherwise assign $\boldsymbol{x}_0'$ to π_2.

π_1

Observation	$\hat{\boldsymbol{a}}'\boldsymbol{x}_0 - \hat{m}$	Classification
1	2.83	π_1
2	0.83	π_1
3	-0.17	π_2

π_2

Observation	$\hat{\boldsymbol{a}}'\boldsymbol{x}_0 - \hat{m}$	Classification
1	-1.50	π_2
2	0.50	π_1
3	-2.50	π_2

The results from this table verify the confusion matrix given in Example 11.7.

(c) This is the table of squared distances $\hat{D}_i^2(\boldsymbol{x})$ for the observations, where

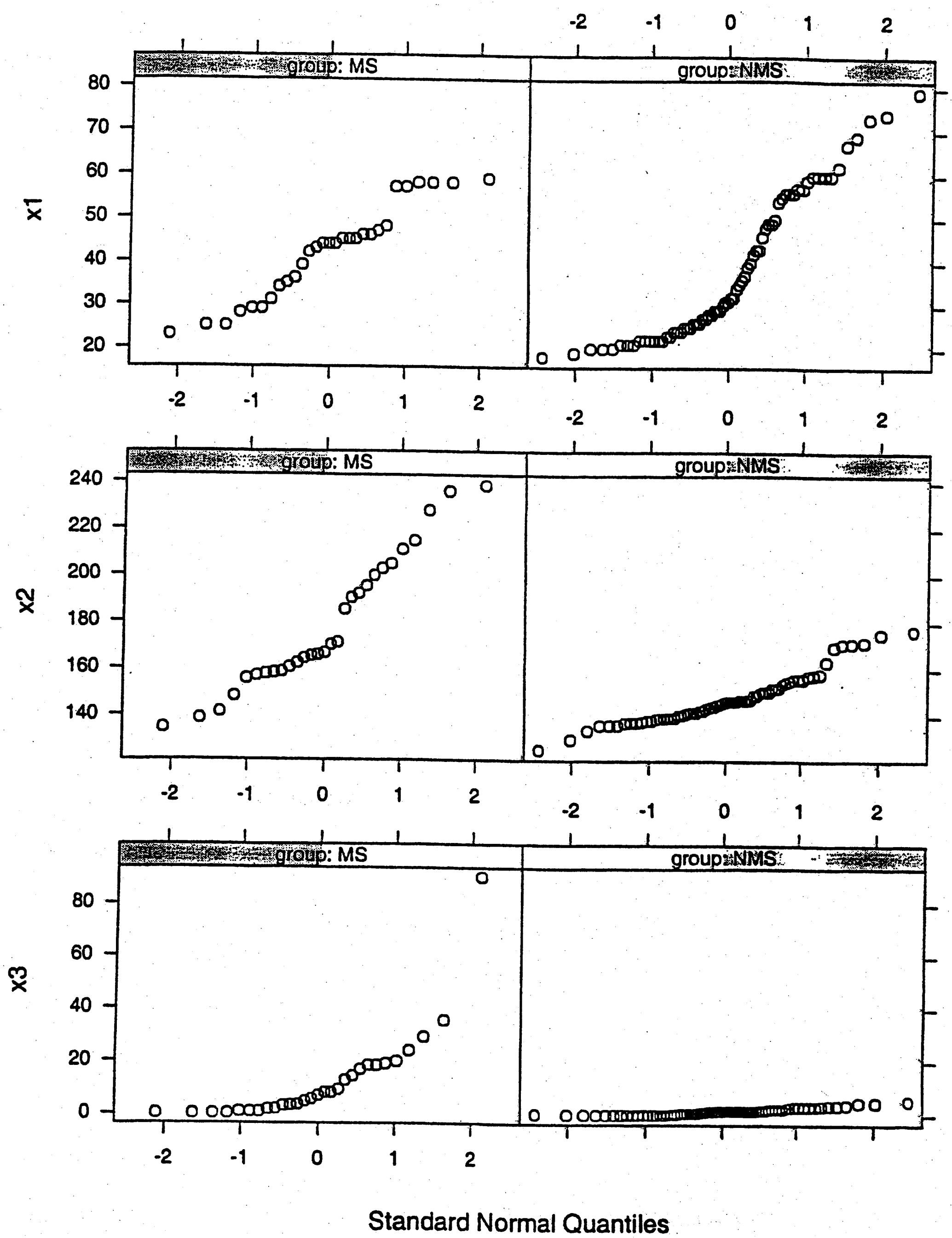
$$D_i^2(\boldsymbol{x}) = (\boldsymbol{x} - \overline{\boldsymbol{x}}_i)' \boldsymbol{S}_{\text{pooled}}^{-1} (\boldsymbol{x} - \overline{\boldsymbol{x}}_i)$$

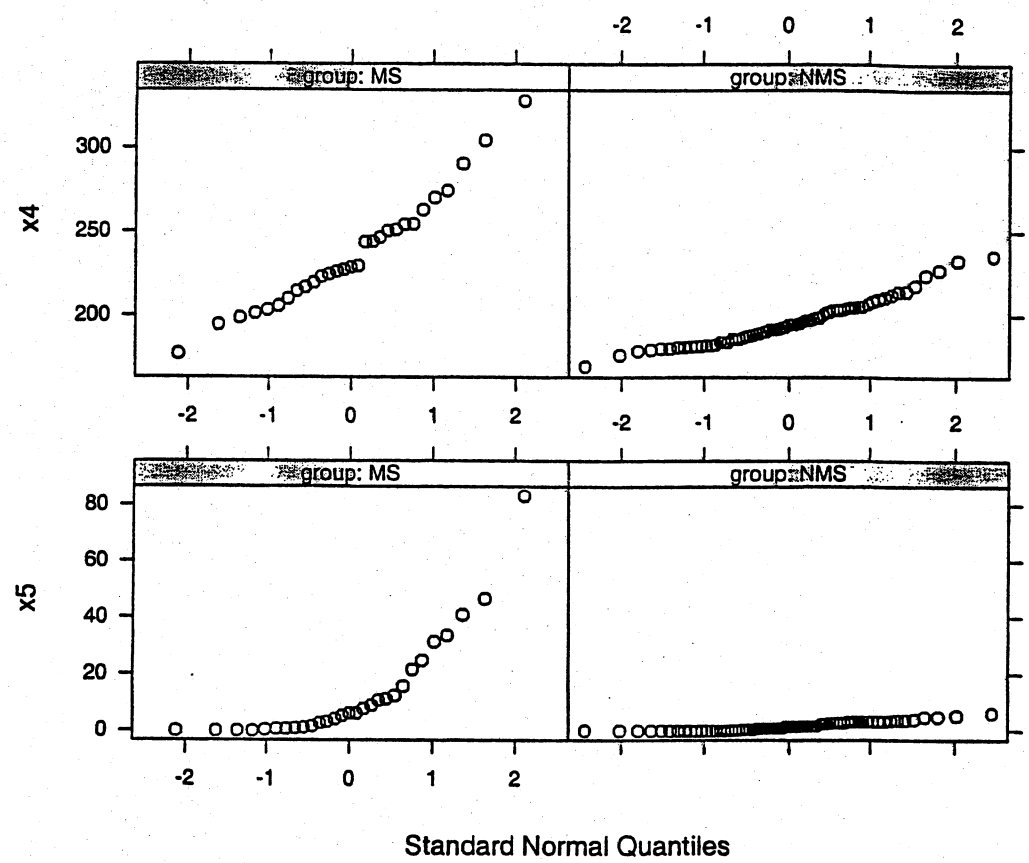
π_1				π_2			
Obs.	$\hat{D}_1^2(\boldsymbol{x})$	$\hat{D}_2^2(\boldsymbol{x})$	Classification	Obs.	$\hat{D}_1^2(\boldsymbol{x})$	$\hat{D}_2^2(\boldsymbol{x})$	Classification
1	$\frac{4}{3}$	$\frac{21}{3}$	π_1	1	$\frac{13}{3}$	$\frac{4}{3}$	π_2
2	$\frac{4}{3}$	$\frac{9}{3}$	π_1	2	$\frac{1}{3}$	$\frac{4}{3}$	π_1
3	$\frac{4}{3}$	$\frac{3}{3}$	π_2	3	$\frac{19}{3}$	$\frac{4}{3}$	π_2

The classification results are identical to those obtained in **(b)**

11.20 The result obtained from this matrix identity is identical to the result of Example 11.7.

11.23 (a) Here are the normal probability plots for each of the variables x_1, x_2, x_3, x_4, x_5


group: MS
group: NMS
x1
x2
x3
Standard Normal Quantiles



Variables x_1, x_3, and x_5 appear to be nonnormal. The transformations $\ln(x_1), \ln(x_3+1)$, and $\ln(x_5+1)$ appear to slightly improve normality.

(b) Using the original data, the linear discriminant function is:

$$\hat{y} = \hat{\boldsymbol{a}}'\boldsymbol{x} = 0.023x_1 - 0.034x_2 + 0.21x_3 - 0.08x_4 - 0.25x_5$$

where

$$\hat{m} = -23.23$$

Thus, we allocate x_0 to π_1 (NMS group) if

$$\hat{a}x_0 - \hat{m} = 0.023x_1 - 0.034x_2 + 0.21x_3 - 0.08x_4 - 0.25x_5 + 23.23 \geq 0$$

Otherwise, allocate x_0 to π_2 (MS group).

(c) Confusion matrix:

		Predicted Membership π_1	Predicted Membership π_2	Total
Actual membership	π_1	66	3	69
	π_2	7	22	29

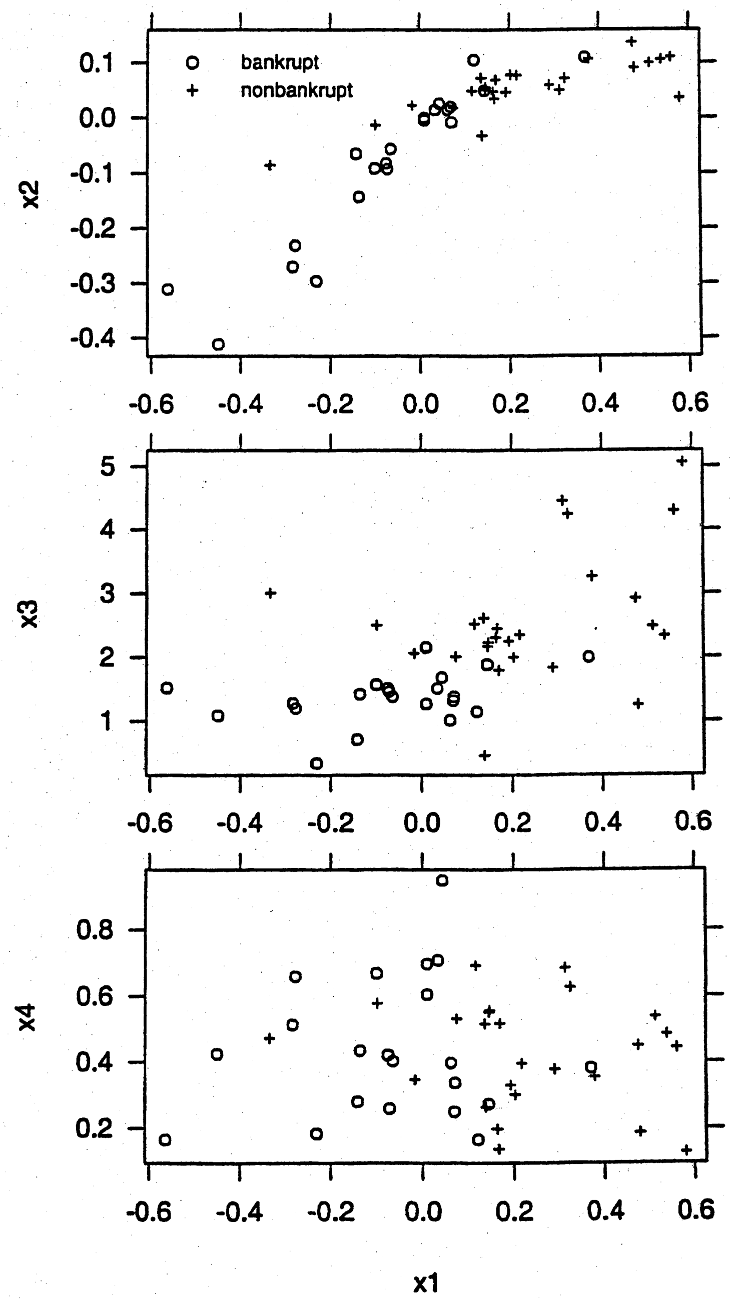
APER$= \frac{3+7}{69+29} = .102$

This is the holdout confusion matrix:

		Predicted Membership π_1	Predicted Membership π_2	Total
Actual membership	π_1	64	5	69
	π_2	8	21	29

$\hat{E}(\text{AER}) = \frac{5+8}{69+29} = .133$

11.24 (a) Here are the scatterplots for the pairs of observations (x_1, x_2), (x_1, x_3), and (x_1, x_4):



The data in the above plot appear to form fairly elliptical shapes, so bivariate normality does not seem like an unreasonable assumption.

(b) $\pi_1 \equiv$ bankrupt firms, $\pi_2 \equiv$ nonbankrupt firms. For (x_1, x_2):

$$\overline{\boldsymbol{x}}_1 = \begin{bmatrix} -0.0688 \\ -0.0819 \end{bmatrix}, \quad \boldsymbol{S}_1 = \begin{bmatrix} 0.04424 & 0.02847 \\ 0.02847 & 0.02092 \end{bmatrix}$$

$$\overline{\boldsymbol{x}}_2 = \begin{bmatrix} 0.2354 \\ 0.0551 \end{bmatrix}, \quad \boldsymbol{S}_2 = \begin{bmatrix} 0.04735 & 0.00837 \\ 0.00837 & 0.00231 \end{bmatrix}$$

(c), (d), (e) See the tables of part (g)

(f)

$$\boldsymbol{S}_{\text{pooled}} = \begin{bmatrix} 0.04594 & 0.01751 \\ 0.01751 & 0.01077 \end{bmatrix}$$

Fisher's linear discriminant function is

$$\hat{y} = \hat{\boldsymbol{a}}'\boldsymbol{x} = -4.67x_1 - 5.12x_2$$

where

$$\hat{m} = -.32$$

Thus, we allocate $\boldsymbol{x}_0$ to π_1 (Bankrupt group) if

$$\hat{\boldsymbol{a}}\boldsymbol{x}_0 - \hat{m} = -4.67x_1 - 5.12x_2 + .32 \geq 0$$

Otherwise, allocate $\boldsymbol{x}_0$ to π_2 (Nonbankrupt group).

APER$= \frac{9}{46} = .196$.

Since S_1 and S_2 look quite different, Fisher's linear discriminant function may not be appropriate. However the performance of this linear discriminant function is as good as that of the quadratic discriminant function, based on the APER criterion.

(g) For (x_1, x_3),

$$\bar{x}_1 = \begin{bmatrix} -0.0688 \\ 1.3675 \end{bmatrix}, \quad S_1 = \begin{bmatrix} 0.04424 & 0.03428 \\ 0.03428 & 0.16455 \end{bmatrix}$$
$$\bar{x}_2 = \begin{bmatrix} 0.2354 \\ 2.5939 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 0.04735 & 0.07543 \\ 0.07543 & 1.04596 \end{bmatrix}$$

For (x_1, x_4),

$$\bar{x}_1 = \begin{bmatrix} -0.0688 \\ 0.4368 \end{bmatrix}, \quad S_1 = \begin{bmatrix} 0.04424 & 0.00431 \\ 0.00431 & 0.04441 \end{bmatrix}$$
$$\bar{x}_2 = \begin{bmatrix} 0.2354 \\ 0.4264 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 0.04735 & -0.00662 \\ -0.00662 & 0.02618 \end{bmatrix}$$

For the various classification rules and error rates for these variable pairs, see the following tables.

This is the table of quadratic functions for the variable pairs (x_1, x_2), (x_1, x_3), and (x_1, x_5), both with $p_1 = 0.5$ and $p_1 = 0.05$. The classification rule for any of these functions is to classify a new observation into π_1 (bankrupt firms) if the quadratic function is ≥ 0, and to classify the new observation into

π_2 (nonbankrupt firms) otherwise. Notice in the table below that only the constant term changes when the prior probabilities change.

Variables	Prior	Quadratic function	
(x_1, x_2)	$p_1 = 0.5$	$-61.77x_1^2 + 35.84x_1x_2 + 407.20x_2^2 + 5.64x_1 - 30.60x_2$	$-$ 0.17
	$p_1 = 0.05$		$-$ 3.11
(x_1, x_3)	$p_1 = 0.5$	$-1.55x_1^2 + 3.89x_1x_3 - 3.08x_3^2 - 10.69x_1 + 7.90x_3$	$-$ 3.14
	$p_1 = 0.05$		$-$ 6.08
(x_1, x_4)	$p_1 = 0.5$	$-0.46x_1^2 + 7.75x_1x_4 + 8.43x_4^2 - 10.05x_1 - 8.11x_4$	$+$ 2.23
	$p_1 = 0.05$		$-$ 0.71

Here is a table of the APER and $\hat{E}$(AER) for the various variable pairs and prior probabilities.

	APER		$\hat{E}$(APR)	
Variables	$p_1 = 0.5$	$p_1 = 0.05$	$p_1 = 0.5$	$p_1 = 0.05$
(x_1, x_2)	0.20	0.26	0.22	0.26
(x_1, x_3)	0.11	0.37	0.13	0.39
(x_1, x_4)	0.17	0.39	0.22	0.46

For equal priors, it appears that the (x_1, x_3) variable pair is the best classifier, as it has the lowest APER. For unequal priors, $p_1 = 0.05$ and $p_2 = 0.95$, the variable pair (x_1, x_2) has the lowest APER.

(h) When using all four variables (X_1, X_2, X_3, X_4),

$$\overline{\boldsymbol{x}}_1 = \begin{bmatrix} -0.0688 \\ -0.0819 \\ 1.3675 \\ 0.4368 \end{bmatrix}, \quad \boldsymbol{S}_1 = \begin{bmatrix} 0.04424 & 0.02847 & 0.03428 & 0.00431 \\ 0.02847 & 0.02092 & 0.02580 & 0.00362 \\ 0.03428 & 0.02580 & 0.16455 & 0.03300 \\ 0.00431 & 0.00362 & 0.03300 & 0.04441 \end{bmatrix}$$

$$\overline{\boldsymbol{x}}_2 = \begin{bmatrix} 0.2354 \\ 0.0551 \\ 2.5939 \\ 0.4264 \end{bmatrix}, \quad \boldsymbol{S}_2 = \begin{bmatrix} 0.04735 & 0.00837 & 0.07543 & -0.00662 \\ 0.00837 & 0.00231 & 0.00873 & 0.00031 \\ 0.07543 & 0.00873 & 1.04596 & 0.03177 \\ -0.00662 & 0.00031 & 0.03177 & 0.02618 \end{bmatrix}$$

Assign a new observation $\boldsymbol{x}_0$ to π_1 if its quadratic function given below is less than 0:

Prior	Quadratic function
$p_1 = 0.5$	$\boldsymbol{x}_0' \begin{bmatrix} -49.232 & -20.657 & -2.623 & 14.050 \\ -20.657 & 526.336 & 11.412 & -52.493 \\ -2.623 & 11.412 & -3.748 & 1.4337 \\ 14.050 & -52.493 & 1.434 & 11.974 \end{bmatrix} \boldsymbol{x}_0 + \begin{bmatrix} 4.91 \\ -28.42 \\ 8.65 \\ -11.80 \end{bmatrix}' \boldsymbol{x}_0 - 2.69$
$p_1 = 0.05$	$-\ 5.64$

For $p_1 = 0.5$: APER $= \frac{3}{46} = .07$, $\quad \hat{E}(\text{AER}) = \frac{5}{46} = .11$

For $p_1 = 0.05$: APER $= \frac{9}{46} = .20$, $\quad \hat{E}(\text{AER}) = \frac{11}{46} = .24$

11.25 **(a)** Fisher's linear discriminant function is

$$\hat{y}_0 = a' x_0 - \hat{m} = -4.80x_1 - 1.48x_3 + 3.33$$

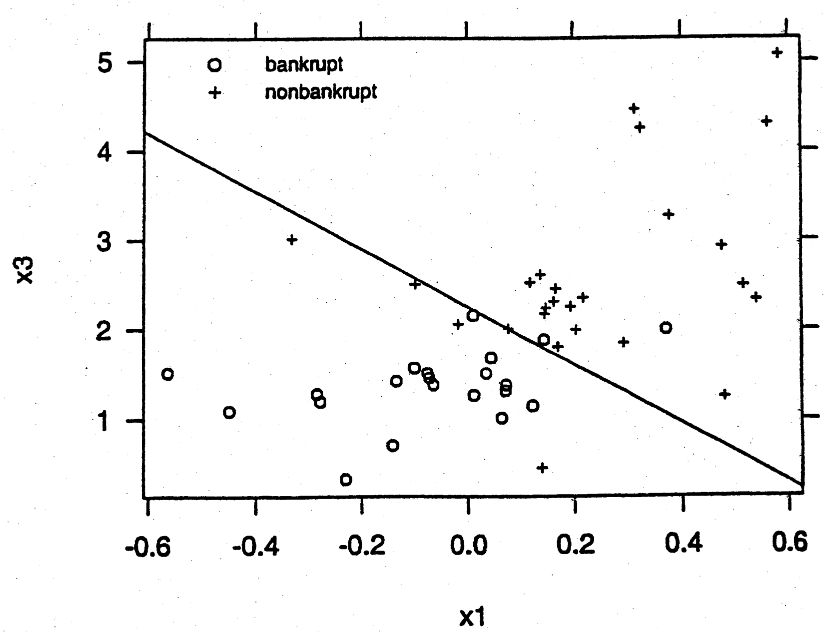
Classify x_0 to π_1 (bankrupt firms) if

$$a' x_0 - \hat{m} \geq 0$$

Otherwise classify x_0 to π_2 (nonbankrupt firms).

The APER is $\frac{2+4}{46} = .13$.

This is the scatterplot of the data in the (x_1, x_3) coordinate system, along with the discriminant line.



(b) With data point 16 for the bankrupt firms deleted, Fisher's linear discriminant function is given by

$$\hat{y}_0 = \boldsymbol{a}'\boldsymbol{x}_0 - \hat{m} = -5.93x_1 - 1.46x_3 + 3.31$$

Classify $\boldsymbol{x}_0$ to π_1 (bankrupt firms) if

$$\boldsymbol{a}'\boldsymbol{x}_0 - \hat{m} \geq 0$$

Otherwise classify $\boldsymbol{x}_0$ to π_2 (nonbankrupt firms).

The APER is $\frac{1+4}{45} = .11$.

With data point 13 for the nonbankrupt firms deleted, Fisher's linear discriminant function is given by

$$\hat{y}_0 = \boldsymbol{a}'\boldsymbol{x}_0 - \hat{m} = -4.35x_1 - 1.97x_3 + 4.36$$

Classify $\boldsymbol{x}_0$ to π_1 (bankrupt firms) if

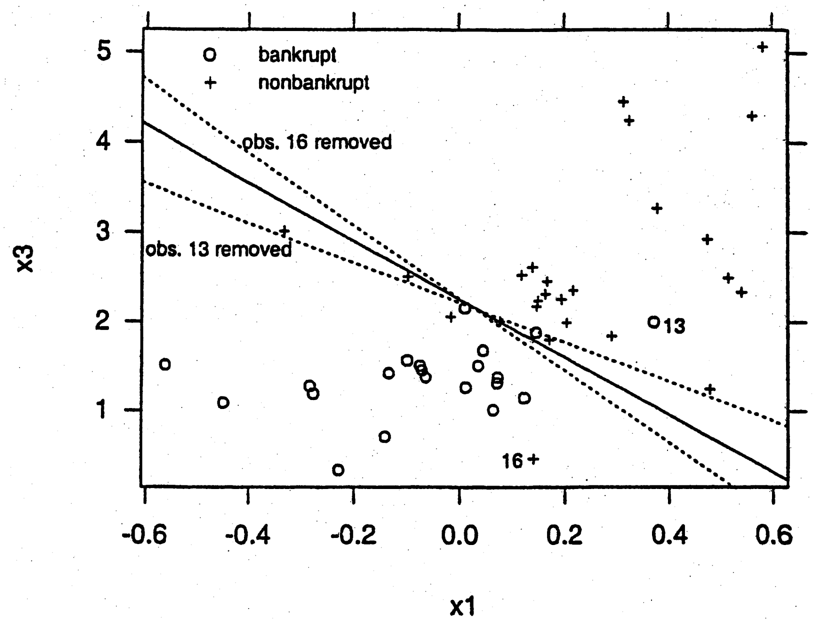
$$\boldsymbol{a}'\boldsymbol{x}_0 - \hat{m} \geq 0$$

Otherwise classify $\boldsymbol{x}_0$ to π_2 (nonbankrupt firms).

The APER is $\frac{1+3}{45} = .089$.

This is the scatterplot of the observations in the (x_1, x_3), coordinate system with the discriminant lines for the three linear discriminant functions given above. Also labelled are observation 16 for bankrupt firms and observation

13 for nonbankrupt firms.



It appears that deleting these observations has changed the line significantly.

11.26 (a) The least squares regression results for the X, Z data are:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	-0.081412	0.13488497	-0.604	0.5492
X3	1	0.307221	0.05956685	5.158	0.0001

Here are the dot diagrams of the fitted values for the bankrupt firms and for the nonbankrupt firms:

```
                   .  :
     .     .   ::::::.  .  ..
+---------+---------+---------+---------+---------+-------Bankrupt
                          .  ..  .
    .          .  .   :  :  ::.:.   ..   .           :  .      .
+---------+---------+---------+---------+---------+-------Nonbankrupt
0.00     0.30      0.60      0.90      1.20      1.50
```

This table summarizes the classification results using the fitted values:

OBS	GROUP	FITTED	CLASSIFICATION
13	bankrupt	0.57896	misclassify
16	bankrupt	0.53122	misclassify
31	nonbankr	0.47076	misclassify
34	nonbankr	0.06025	misclassify
38	nonbankr	0.48329	misclassify
41	nonbankr	0.30089	misclassify

The confusion matrix is:

		Predicted Membership π_1	Predicted Membership π_2	Total
Actual membership	π_1	19	2	21
	π_2	4	21	25

Thus, the APER is $\frac{2+4}{46} = .13$.

(b) The least squares regression results using all four variables X_1, X_2, X_3, X_4 are:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	0.208915	0.18615284	1.122	0.2683
X1	1	0.156317	0.46653100	0.335	0.7393
X2	1	1.149093	0.90606395	1.268	0.2119
X3	1	0.225972	0.07030479	3.214	0.0026
X4	1	-0.305175	0.32336357	-0.944	0.3508

Here are the dot diagrams of the fitted values for the bankrupt firms and for the nonbankrupt firms:

```
                        .:
         .  .   .. .   ... :: ... .  .   .
 -+---------+---------+---------+---------+---------+-----Bankrupt
                                .
                                :  .  .
                  .        .::: :: :     ..  :   .   .
 -+---------+---------+---------+---------+---------+-----Nonbankrupt
-0.35      0.00      0.35      0.70      1.05      1.40
```

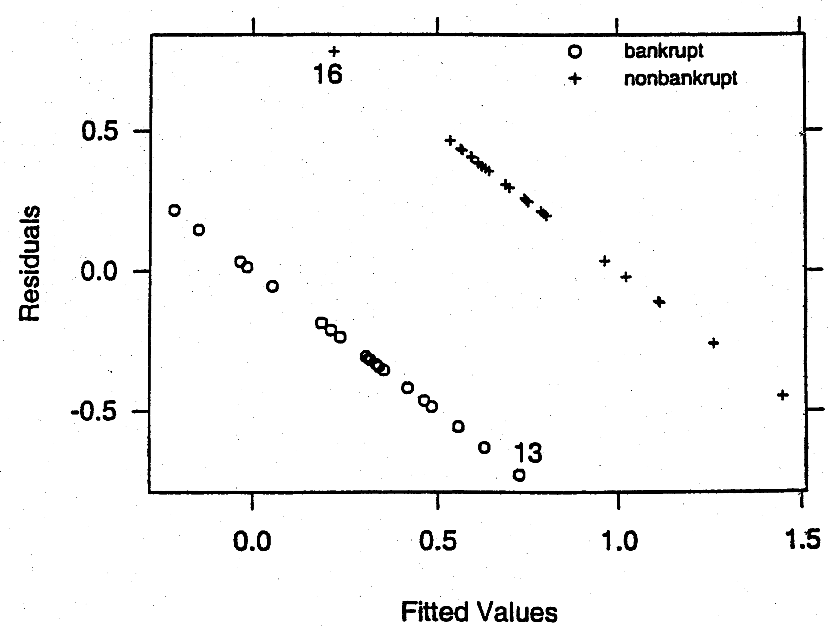
This table summarizes the classification results using the fitted values:

OBS	GROUP	FITTED	CLASSIFICATION
15	bankrupt	0.62997	misclassify
16	bankrupt	0.72676	misclassify
20	bankrupt	0.55719	misclassify
34	nonbankr	0.21845	misclassify

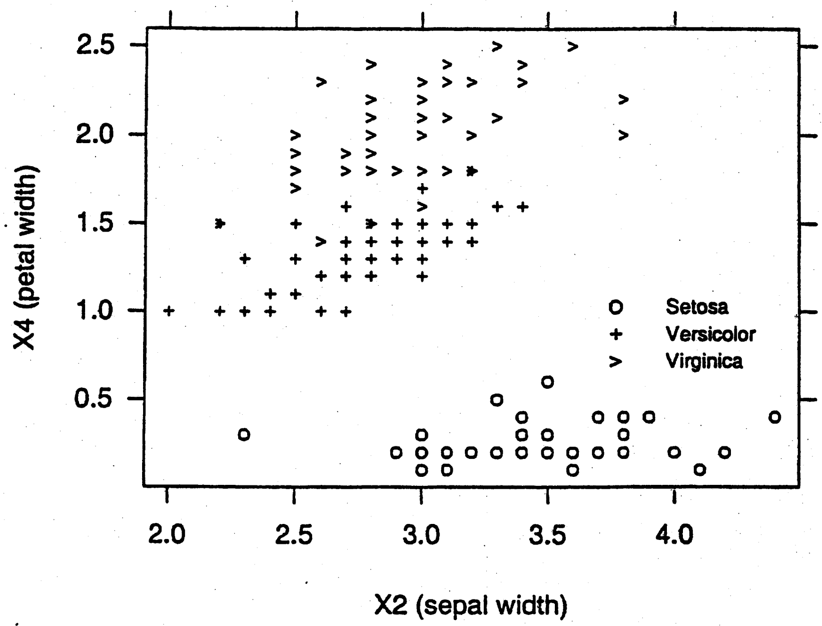
The confusion matrix is:

		Predicted Membership π_1	π_2	Total
Actual membership	π_1	18	3	21
	π_2	1	24	25

Thus, the APER is $\frac{3+1}{46} = .087$. Here is a scatterplot of the residuals against the fitted values, with points 16 of the bankrupt firms and 13 of the non-bankrupt firms labelled. It appears that point 16 of the bankrupt firms is an outlier.



11.27 (a) Plot of the data in the (x_2, x_4) variable space:



The points from all three groups appear to form an elliptical shape. However, it appears that the points of π_1 *(Iris setosa)* form an ellipse with a different orientation than those of π_2 *(Iris versicolor)* and π_3 *(Iris virginica)*. This indicates that the observations from π_1 may have a different covariance matrix from the observations from π_2 and π_3.

(b) Here are the results of a test of the null hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$ versus H_1 : at least one of the $\boldsymbol{\mu}_i$'s is different from the others at the $\alpha = 0.05$ level of significance:

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.02343863	199.145	8	288	0.0001

Thus, the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ is rejected at the $\alpha = 0.05$ level of significance. As discussed earlier, the plots give us reason to doubt the assumption of equal covariance matrices for the three groups.

(c) $\pi_1 \equiv$ *Iris setosa*; $\pi_2 \equiv$ *Iris versicolor* $\pi_3 \equiv$ *Iris virginica*

The quadratic discriminant scores $\hat{d}_i^Q(\boldsymbol{x})$ given by (11-47) with $p_1 = p_2 = p_3 = \frac{1}{3}$ are:

population	$\hat{d}_i^Q(\boldsymbol{x}) = -\frac{1}{2}\ln\|\boldsymbol{S}_i\| - \frac{1}{2}(\boldsymbol{x}-\overline{\boldsymbol{x}}_i)'\boldsymbol{S}_i^{-1}(\boldsymbol{x}-\overline{\boldsymbol{x}}_i)$
π_1	$-3.68x_2^2 + 6.16x_2x_4 - 47.60x_4^2 + 23.71x_2 + 2.30x_4 - 37.67$
π_2	$-9.09x_2^2 + 19.57x_2x_4 - 22.87x_4^2 + 24.94x_2 + 7.63x_4 - 36.53$
π_3	$-6.76x_2^2 + 8.54x_2x_4 - 9.32x_4^2 + 22.92x_2 + 12.38x_4 - 44.04$

To classify the observation $\boldsymbol{x}_0' = [3.5 \quad 1.75]$, compute $\hat{d}_i^Q(\boldsymbol{x}_0)$ for $i = 1, 2, , 3$, and classify $\boldsymbol{x}_0$ to the population for which $\hat{d}_i^Q(\boldsymbol{x}_0)$ is the largest.

$$\hat{d}_1^Q(\boldsymbol{x}_0) = -103.77$$

$$\hat{d}_2^Q(\boldsymbol{x}_0) = 0.043$$

$$\hat{d}_3^Q(\boldsymbol{x}_0) = -1.23$$

So classify $\boldsymbol{x}_0$ to π_2 (*Iris versicolor*).

(d) The linear discriminant scores $\hat{d}_i(\boldsymbol{x})$ are:

population	$\hat{d}_i(\boldsymbol{x}) = \overline{\boldsymbol{x}}_i'\boldsymbol{S}_{\text{pooled}}\boldsymbol{x} - \frac{1}{2}\overline{\boldsymbol{x}}_i'\boldsymbol{S}_{\text{pooled}}\overline{\boldsymbol{x}}_i$	$\hat{d}_i(\boldsymbol{x}_0)$
π_1	$36.02x_2 - 22.26x_4 - 59.00$	28.12
π_2	$19.31x_2 + 16.58x_4 - 37.73$	58.86
π_3	$15.49x_2 + 36.28x_4 - 59.78$	57.92

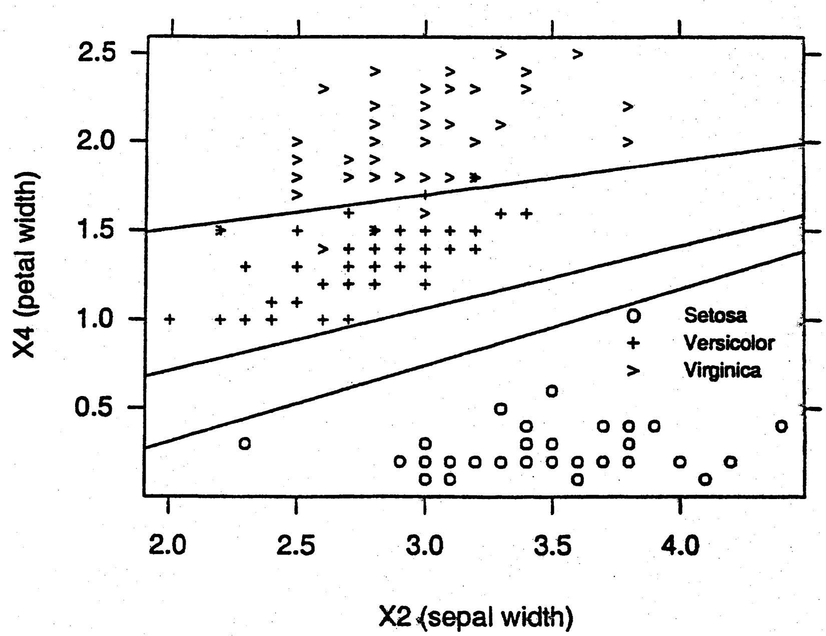
Since $\hat{d}_i(\boldsymbol{x}_0)$ is the largest for $i = 2$, we classify the new observation $\boldsymbol{x}_0' = [3.5 \quad 1.75]$ to π_1 according to (11-52). The results are the same for (c) and (d).

(e) To use rule (11-56), construct $\hat{d}_{ki}(\boldsymbol{x}) = \hat{d}_k(\boldsymbol{x}) - \hat{d}_i(\boldsymbol{x})$ for all $i \neq k$. Then classify $\boldsymbol{x}$ to π_k if $\hat{d}_{ki}(\boldsymbol{x}) \geq 0$ for all $i = 1, 2, 3$. Here is a table of $\hat{d}_{ki}(\boldsymbol{x}_0)$ for $i, k = 1, 2, 3$:

			i	
		1	2	3
	1	0	-30.74	-29.80
j	2	30.74	0	0.94
	1	29.80	-0.94	0

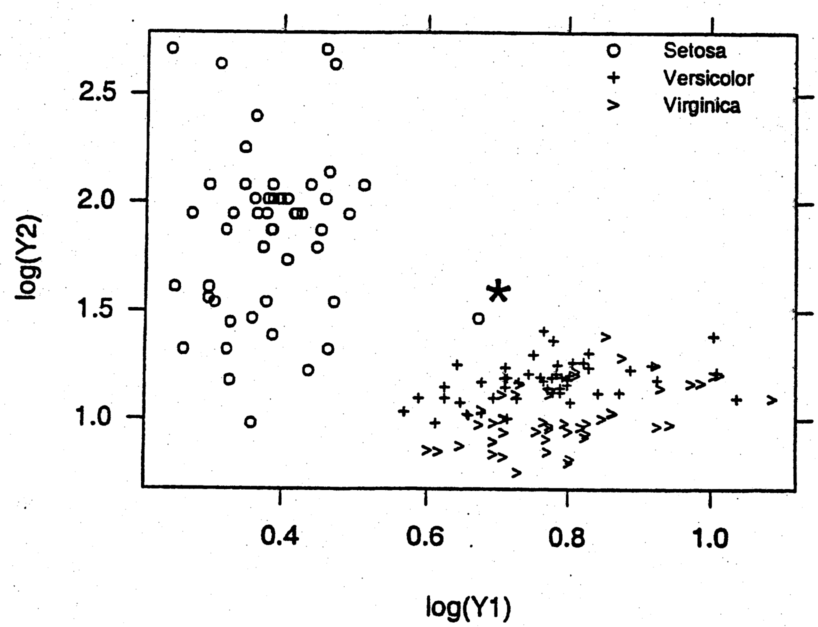
Since $\hat{d}_{ki}(\boldsymbol{x}_0) \geq 0$ for all $i \neq 2$, we allocate $\boldsymbol{x}_0$ to π_2, using (11-52)

Here is the scatterplot of the data in the (x_2, x_4) variable space, with the classification regions $\hat{R}_1, \hat{R}_2$, and $\hat{R}_3$ delineated.



(f) The APER $= \frac{1+4}{150} = .033$. $\hat{E}(\text{AER}) = \frac{4+2}{150} = .04$

11.28 (a) This is the plot of the data in the $(\log Y_1, \log Y_2)$ variable space:



The points of all three groups appear to follow roughly an ellipse-like pattern. However, the orientation of the ellipse appears to be different for the observations from π_1 (*Iris setòsa*), from the observations from π_2 and π_3. In π_1, there also appears to be an outlier, labelled with a "*".

(b), (c) Assuming equal covariance matrices and ivariate normal populations, these are the linear discriminant scores $\hat{d}_i(\boldsymbol{x})$ for $i = 1, 2, 3$.

For both variables $\log Y_1$, and $\log Y_2$:

population	$\hat{d}_i(\boldsymbol{x}) = \overline{\boldsymbol{x}}_i' \boldsymbol{S}_{\text{pooled}} \boldsymbol{x} - \frac{1}{2} \overline{\boldsymbol{x}}_i' \boldsymbol{S}_{\text{pooled}} \overline{\boldsymbol{x}}_i$
π_1	$26.81 \log Y_1 + 28.90 \log Y_2 - 31.97$
π_2	$75.10 \log Y_1 + 13.82 \log Y_2 - 36.83$
π_3	$79.94 \log Y_1 + 10.80 \log Y_2 - 37.30$

For variable $\log Y_1$ only:

population	$\hat{d}_i(\boldsymbol{x}) = \overline{\boldsymbol{x}}_i' \boldsymbol{S}_{\text{pooled}} \boldsymbol{x} - \frac{1}{2} \overline{\boldsymbol{x}}_i' \boldsymbol{S}_{\text{pooled}} \overline{\boldsymbol{x}}_i$
π_1	$40.90 \log Y_1 - 7.82$
π_2	$81.84 \log Y_1 - 31.30$
π_3	$85.20 \log Y_1 - 33.93$

For variable $\log Y_2$ only:

population	$\hat{d}_i(\boldsymbol{x}) = \overline{\boldsymbol{x}}_i' \boldsymbol{S}_{\text{pooled}} \boldsymbol{x} - \frac{1}{2} \overline{\boldsymbol{x}}_i' \boldsymbol{S}_{\text{pooled}} \overline{\boldsymbol{x}}_i$
π_1	$30.93 \log Y_2 - 28.73$
π_2	$19.52 \log Y_2 - 11.44$
π_3	$16.87 \log Y_2 + 8.54$

Variables	APER	E(AER)
$\log Y_1$, $\log Y_2$	$\frac{26}{150} = .17$	$\frac{27}{150} = .18$
$\log Y_1$	$\frac{49}{150} = .33$	$\frac{49}{150} = .33$
$\log Y_2$,	$\frac{34}{150} = .23$	$\frac{34}{150} = .23$

The preceeding misclassification rates are not nearly as good as those in Example 11.12. Using "shape" is effective in discriminating π_1 (*iris versicolor*) from π_2 and π_3. It is not as good at discriminating π_2 from π_3, because of the overlap of π_1 and π_2 in both shape variables. Therefore, shape is not an effective discriminator of all three species of iris.

(d) Given the bivariate normal-like scatter and the relatively large samples, we do not expect the error rates in parts (b) and (c) to differ much.

11.29 (a) The calculated values of $\bar{x}_1, \bar{x}_1, \bar{x}_3, \bar{x}$, and S_{pooled} agree with the results for these quantities given in Example 11.11

(b)

$$W^{-1} = \begin{bmatrix} 0.348899 & 0.000193 \\ 0.000193 & .000003 \end{bmatrix}, \quad B = \begin{bmatrix} 12.50 & 1518.74 \\ 1518.74 & 258471.12 \end{bmatrix}$$

The eigenvalues and scaled eigenvectors of $W^{-1}B$ are

$$\hat{\lambda}_1 = 5.646, \quad \hat{a}_1' = \begin{bmatrix} 5.009 \\ 0.009 \end{bmatrix}$$

$$\hat{\lambda}_2 = 0.191, \quad \hat{a}_2' = \begin{bmatrix} 0.207 \\ -0.014 \end{bmatrix}$$

To classify $x_0' = [3.21 \quad 497]$, use (11-67) and compute

$$\Sigma_{j=1}^{2}[\hat{a}_j'(x - \bar{x}_i)]^2 \qquad i = 1, 2, 3$$

Allocate x_0' to π_k if

$$\Sigma_{j=1}^{2}[\hat{a}_j'(x - \bar{x}_k)]^2 \leq \Sigma_{j=1}^{2}[\hat{a}_j'(x - \bar{x}_i)]^2 \qquad \text{for all} i \neq k$$

For x_0,

k	$\Sigma_{j=1}^{2}[\hat{a}_j'(x - \bar{x}_k)]^2$
1	2.63
2	16.99
3	2.43

Thus, classify x_0 to π_3 This result agrees with the classification given in Example 11.11. Any time there are three populations with only two discriminants, classification results using Fisher's Discriminants will be identical to those using the sample distance method of Example 11.11.

11.30 **(a)** Assuming normality and equal covariance matrices for the three populations π_1, π_2, and π_3, the minimum TPM rule is given by:

Allocate $\boldsymbol{x}$ to π_k if the linear discriminant score $\hat{d}_k(\boldsymbol{x})$ = the largest of $\hat{d}_1(\boldsymbol{x}), \hat{d}_2(\boldsymbol{x}), \hat{d}_3(\boldsymbol{x})$

where $\hat{d}_i(\boldsymbol{x})$ is given in the following table for $i = 1, 2, 3$.

population	$\hat{d}_i(\boldsymbol{x}) = \overline{\boldsymbol{x}}_i' \boldsymbol{S}_{\text{pooled}} \boldsymbol{x} - \frac{1}{2}\overline{\boldsymbol{x}}_i' \boldsymbol{S}_{\text{pooled}} \overline{\boldsymbol{x}}_i$
π_1	$0.70x_1 + 0.58x_2 - 13.52x_3 + 6.93x_4 + 1.44x_5 - 44.78$
π_2	$1.85x_1 + 0.32x_2 - 12.78x_3 + 8.33x_4 - 0.14x_5 - 35.20$
π_3	$2.64x_1 + 0.20x_2 - 2.16x_3 + 5.39x_4 - 0.08x_5 - 23.61$

(b) Confusion matrix is:

		Predicted Membership			
		π_1	π_2	π_3	Total
Actual membership	π_1	7	0	0	7
	π_2	1	10	0	11
	π_3	0	3	35	38

And the APER $\frac{0+1+3}{56} = .071$

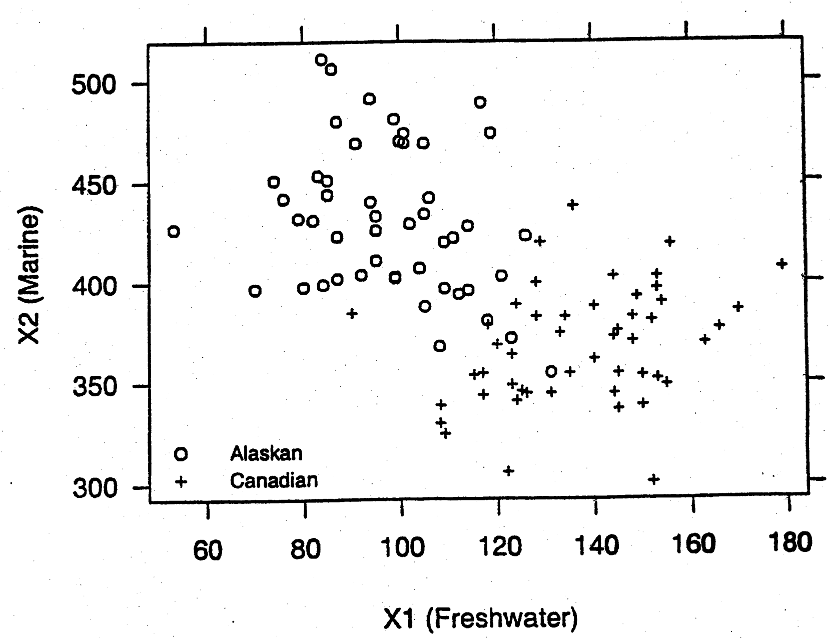
The holdout confusion matrix is:

		Predicted Membership			
		π_1	π_2	π_3	Total
Actual membership	π_1	7	0	0	7
	π_2	2	7	2	11
	π_3	0	3	35	38

$\hat{E}(\text{AER}) = \frac{2+2+3}{56} = .125$

(c) One choice of transformations, $x_1, \log x_2, \sqrt{x_3}, \log x_4, \sqrt{x_5}$ appears to improve the normality of the data but the classification rule from these data has slightly higher error rates than the rule derived from the original data. The error rates (APER, $\hat{E}$(AER)) for the linear discriminants in Example 11.14 are also slightly higher than those for the original data.

11.31 (a) The data look fairly normal.



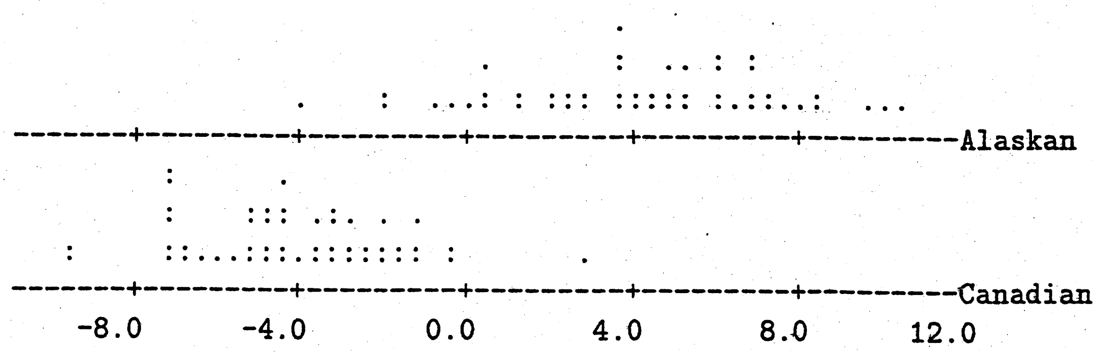
Although the covariances have different signs for the two groups, the correlations are small. Thus the assumption of bivariate normal distributions with equal covariance matrices does not seem unreasonable.

(b) The linear discriminant function is

$$\hat{a}'x - \hat{m} = -0.13x_1 + 0.052x_2 - 5.54$$

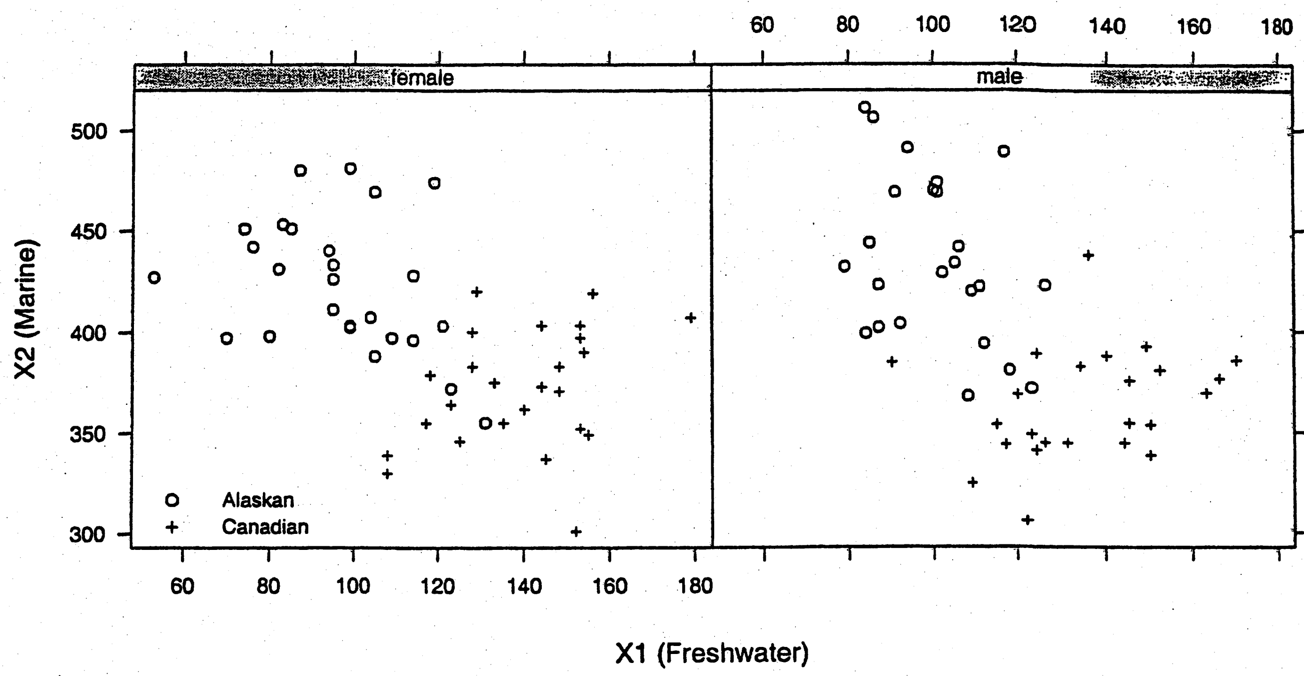
Classify an observation x_0 to π_1 (Alaskan salmon) if $\hat{a}'x_0 - \hat{m} \geq 0$ and classify x_0 to π_2 (Canadian salmon) otherwise.

Dot diagrams of the discriminant scores:



It does appear that growth ring diameters separate the two groups reasonably well, as APER$= \frac{6+1}{100} = .07$ and E(AER)$= \frac{6+1}{100} = .07$

(c) Here are the bivariate plots of the data for male and female salmon separately.



For the male salmon, these are some summary statistics

$$\overline{x}_1 = \begin{bmatrix} 100.3333 \\ 436.1667 \end{bmatrix}, \quad S_1 = \begin{bmatrix} 181.97101 & -197.71015 \\ -197.71015 & 1702.31884 \end{bmatrix}$$

$$\overline{x}_2 = \begin{bmatrix} 135.2083 \\ 364.0417 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 370.17210 & 141.64312 \\ 141.64312 & 760.65036 \end{bmatrix}$$

The linear discriminant function for the male salmon only is

$$\hat{a}'x - \hat{m} = -0.12x_1 + 0.056x_2 - 8.12$$

Classify an observation x_0 to π_1 (Alaskan salmon) if $\hat{a}'x_0 - \hat{m} \geq 0$ and classify x_0 to π_2 (Canadian salmon) otherwise.

Using this classification rule, APER= $\frac{3+1}{48} = .08$ and E(AER)= $\frac{3+2}{48} = .10$.

For the female salmon, these are some summary statistics

$$\overline{\boldsymbol{x}}_1 = \begin{bmatrix} 96.5769 \\ 423.6539 \end{bmatrix}, \quad \boldsymbol{S}_1 = \begin{bmatrix} 336.33385 & -210.23231 \\ -210.23231 & 1097.91539 \end{bmatrix}$$

$$\overline{\boldsymbol{x}}_2 = \begin{bmatrix} 139.5385 \\ 369.0000 \end{bmatrix}, \quad \boldsymbol{S}_2 = \begin{bmatrix} 289.21846 & 120.64000 \\ 120.64000 & 1038.72000 \end{bmatrix}$$

The linear discriminant function for the female salmon only is

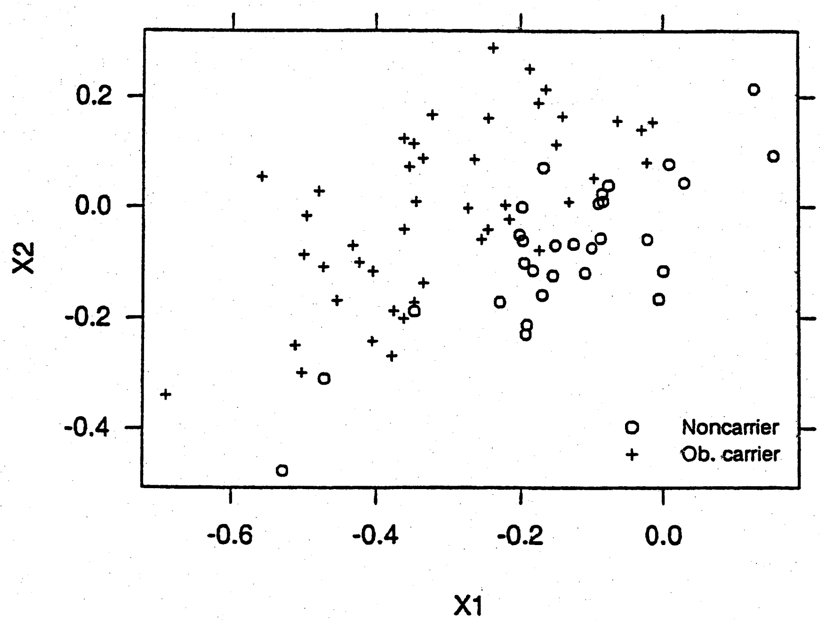
$$\hat{\boldsymbol{a}}'\boldsymbol{x} - \hat{m} = -0.13x_1 + 0.05x_2 - 2.66$$

Classify an observation $\boldsymbol{x}_0$ to π_1 (Alaskan salmon) if $\hat{\boldsymbol{a}}'\boldsymbol{x}_0 - \hat{m} \geq 0$ and classify $\boldsymbol{x}_0$ to π_2 (Canadian salmon) otherwise.

Using this classification rule, APER= $\frac{3+0}{52} = .06$ and E(AER)= $\frac{3+0}{52} = .06$.

It is unlikely that gender is a useful discriminatory variable, as splitting the data into female and male salmon did not improve the classification results greatly.

11.32 **(a)** Here is the bivariate plot of the data for the two groups:



Because the points for both groups form fairly elliptical shapes, the bivariate normal assumption appears to be a reasonable one. Normal score plots for each group confirm this.

(b) Assuming equal prior probabilities, the sample linear discriminant function is

$$\hat{\boldsymbol{a}}'\boldsymbol{x} - \hat{m} = 19.32x_1 - 17.12x_2 + 3.56$$

Classify an observation $\boldsymbol{x}_0$ to π_1 (Noncarriers) if $\hat{\boldsymbol{a}}'\boldsymbol{x}_0 - \hat{m} \geq 0$ and classify $\boldsymbol{x}_0$ to π_2 (Obligatory carriers) otherwise.

The holdout confusion matrix is

		Predicted Membership		
		π_1	π_2	Total
Actual membership	π_1	26	4	30
	π_2	8	37	45

$\hat{E}(\text{AER}) = \frac{4+8}{75} = .16$

(c) The classification results for the 10 new cases using the discriminant function in part (b):

Case	x_1	x_2	$\hat{a}'x - \hat{m}$	Classification
1	-0.112	-0.279	6.17	π_1
2	-0.059	-0.068	3.58	π_1
3	0.064	0.012	4.59	π_1
4	-0.043	-0.052	3.62	π_1
5	-0.050	-0.098	4.27	π_1
6	-0.094	-0.113	3.68	π_1
7	-0.123	-0.143	3.63	π_1
8	-0.011	-0.037	3.98	π_1
9	-0.210	-0.090	1.04	π_1
10	-0.126	-0.019	1.45	π_1

(d) Assuming that the prior probability of obligatory carriers is $\frac{1}{4}$ and that of noncarriers is $\frac{3}{4}$, the sample linear discriminant function is

$$\hat{a}'x - \hat{m} = 19.32x_1 - 17.12x_2 + 4.66$$

Classify an observation x_0 to π_1 (Noncarriers) if $\hat{a}'x_0 - \hat{m} \geq 0$ and classify x_0 to π_2 (Obligatory carriers) otherwise.

The holdout confusion matrix is

		Predicted Membership π_1	π_2	Total
Actual membership	π_1	30	0	30
	π_2	18	27	45

$\hat{E}(\text{AER}) = \frac{18+0}{75} = 0.24$

The classification results for the 10 new cases using the discriminant function in part (b):

Case	x_1	x_2	$\hat{a}'x - \hat{m}$	Classification
1	-0.112	-0.279	7.27	π_1
2	-0.059	-0.068	4.68	π_1
3	0.064	0.012	5.69	π_1
4	-0.043	-0.052	4.72	π_1
5	-0.050	-0.098	5.37	π_1
6	-0.094	-0.113	4.78	π_1
7	-0.123	-0.143	4.73	π_1
8	-0.011	-0.037	5.08	π_1
9	-0.210	-0.090	2.14	π_1
10	-0.126	-0.019	2.55	π_1

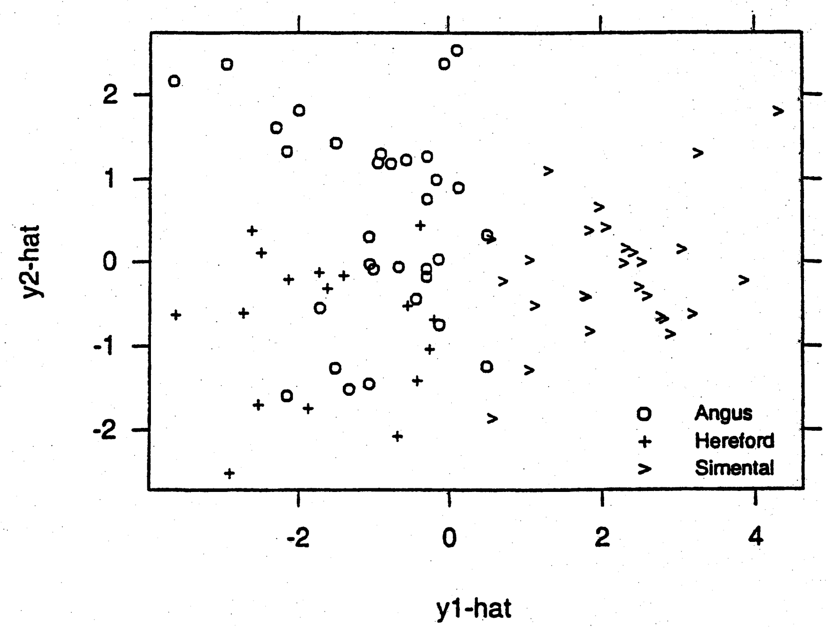
11.33 Let $x_3 \equiv$ YrHgt, $x_4 \equiv$ FtFrBody, $x_6 \equiv$ Frame, $x_7 \equiv$ BkFat, $x_8 \equiv$ SaleHt, and $x_9 \equiv$ SaleWt.

(a) For $\pi_1 \equiv$ Angus, $\pi_2 \equiv$ Hereford, and $\pi_3 \equiv$ Simental, here are Fisher's linear discriminants

$$\begin{aligned}
\hat{d}_1 &= -3737 + 126.88x_3 - 0.48x_4 + 19.08x_5 - 205.22x_6 \\
&\quad +275.84x_7 + 28.15x_8 - 0.03x_9 \\
\hat{d}_2 &= -3686 + 127.70x_3 - 0.47x_4 + 18.65x_5 - 206.18x_6 \\
&\quad +265.33x_7 + 26.80x_8 - 0.03x_9 \\
\hat{d}_1 &= -3881 + 128.08x_3 - 0.48x_4 + 19.39x_5 - 206.36x_6 \\
&\quad +245.50x_7 + 29.47x_8 - 0.03x_9
\end{aligned}$$

When $x_0' = [50, 1000, 73, 7, .17, 54, 1525]$ we obtain $\hat{d}_1 = 3596.31$, $\hat{d}_2 = 3593.32$, and $\hat{d}_3 = 3594.13$, so assign the new observation to π_2, Hereford.

This is the plot of the discriminant scores in the two-dimensional discriminant space:



(b) Here is the APER and $\hat{E}$(AER) for different subsets of the variables:

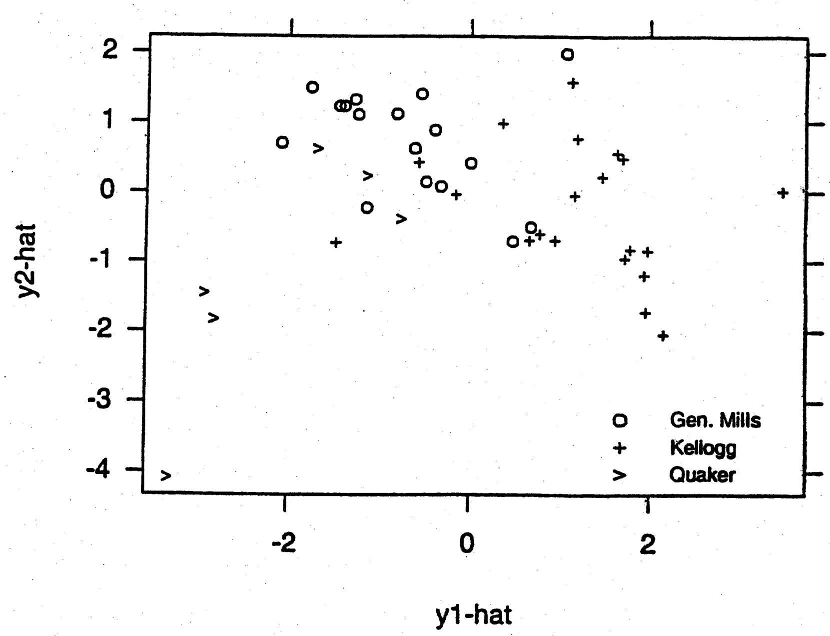
Subset	APER	$\hat{E}$(AER)
$x_3, x_4, x_5, x_6, x_7, x_8, x_9$	.13	.25
x_4, x_5, x_7, x_8	.14	.20
x_5, x_7, x_8	.21	.24
x_4, x_5	.43	.46
x_4, x_7	.36	.39
x_4, x_8	.32	.36
x_7, x_8	.22	.22
x_5, x_7	.25	.29
x_5, x_8	.28	.32

11.34 For $\pi_1 \equiv$ General Mills, $\pi_2 \equiv$ Kellogg, and $\pi_3 \equiv$ Quaker and assuming multivariate normal data with a common covariance matrix, equal costs, and equal priors, these are Fisher's linear discriminant functions:

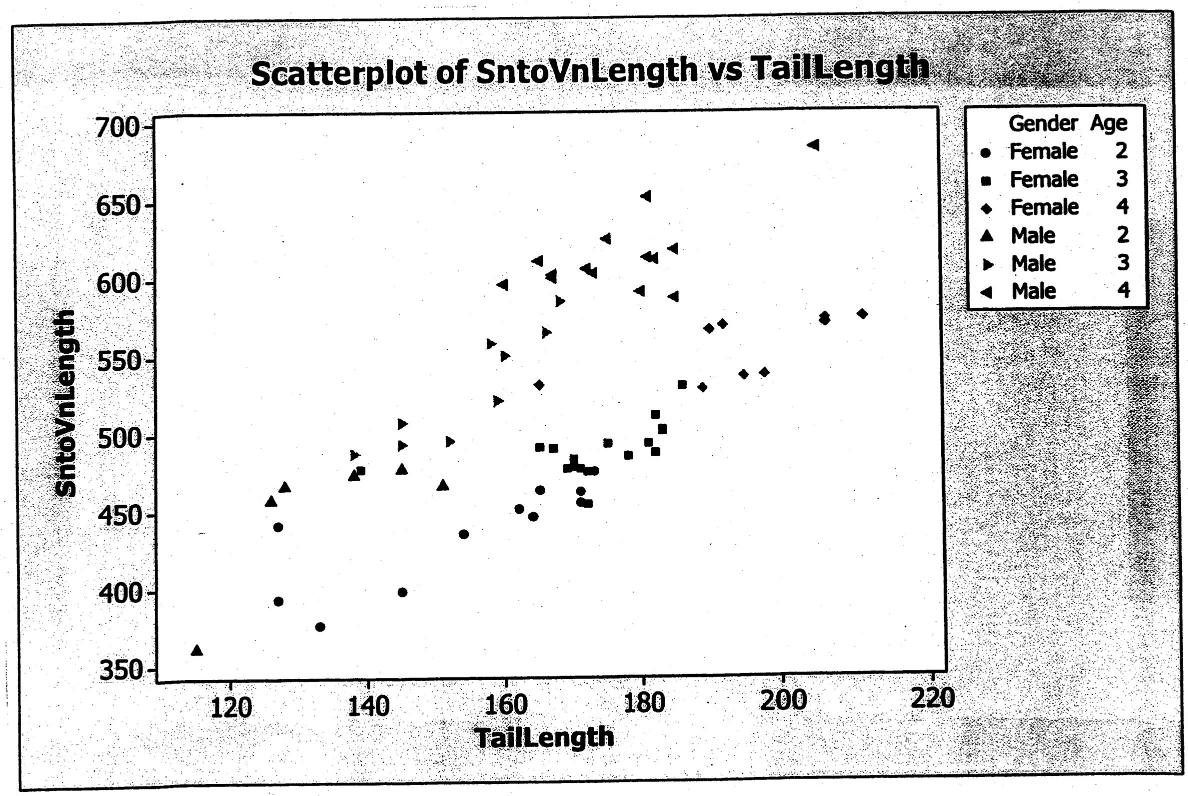
$$
\begin{aligned}
\hat{d}_1 &= .23x_3 + 3.79x_4 - 1.69x_5 - .01x_6 5.53x_7 \\
&\quad 1.90x_8 + 1.36x_9 - 0.12x_{10} - 33.14 \\
\hat{d}_2 &= .32x_3 + 4.15x_4 - 3.62x_5 - .02x_6 9.20x_7 \\
&\quad 2.07x_8 + 1.50x_9 - 0.20x_{10} - 43.07 \\
\hat{d}_3 &= .29x_3 + 2.64x_4 - 1.20x_5 - .02x_6 5.43x_7 \\
&\quad 1.22x_8 + .65x_9 - 0.13x_{10}
\end{aligned}
$$

The Kellogg cereals appear to have high protein, fiber, and carbohydrates, and low fat. However, they also have high sugar. The Quaker cereals appear to have low sugar, but also have low protein and carbohydrates.

Here is a plot of the cereal data in two-dimension discriminant space:



11.35 (a) Scatter plot of tail length and snout to vent length follows. It appears as if these variables will effectively discriminate gender but will be less successful in discriminating the age of the snakes.



(b) Linear Discriminant Function for Groups

	Female	Male
Constant	-36.429	-41.501
SntoVnLength	0.039	0.163
TailLength	0.310	-0.046

Summary of Classification with Cross-validation

	True Group	
Put into Group	Female	Male
Female	34	2
Male	3	27
Total N	37	29
N correct	34	27
Proportion	0.919	0.931

N = 66 N Correct = 61 Proportion Correct = 0.924

$E(\text{AER}) = 1 - .924 = .076 \rightarrow 7.6\%$

(c) Linear Discriminant Function for Groups

	2	3	4
Constant	-112.44	-145.76	-193.14
SntoVnLength	0.33	0.38	0.45
TailLength	0.53	0.60	0.65

Summary of Classification with Cross-validation

	True Group		
Put into Group	2	3	4
2	13	2	0
3	4	21	2
4	0	3	21
Total N	17	26	23
N correct	13	21	21
Proportion	0.765	0.808	0.913

N = 66 N Correct = 55 Proportion Correct = 0.833

$E(\text{AER}) = 1 - .833 = .167 \rightarrow 16.7\%$

(d) Linear Discriminant Function for Groups

	2	3	4
Constant	-79.11	-102.76	-141.94
SntoVnLength	0.36	0.41	0.48

Summary of Classification with Cross-validation

	True Group		
Put into Group	2	3	4
2	14	1	0
3	3	21	4
4	0	4	19
Total N	17	26	23
N correct	14	21	19
Proportion	0.824	0.808	0.826

N = 66 N Correct = 54 Proportion Correct = 0.818

$E(\text{AER}) = 1 - .818 = .182 \rightarrow 18.2\%$

Using only snout to vent length to discriminate the ages of the snakes is about as effective as using both tail length and snout to vent length. Although in both cases, there is a reasonably high proportion of misclassifications.

11.36 Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Constant	3.92484	6.31500	0.62	0.534			
Freshwater	0.126051	0.0358536	3.52	0.000	1.13	1.06	1.22
Marine	-0.0485441	0.0145240	-3.34	0.001	0.95	0.93	0.98

```
Log-Likelihood = -19.394
Test that all slopes are zero: G = 99.841, DF = 2, P-Value = 0.000
```

The regression is significant (p-value = 0.000) and retaining the constant term the fitted function is

$$\ln\left(\frac{\hat{p}(z)}{1-\hat{p}(z)}\right) = 3.925 + .126(\text{freshwater growth}) - .049(\text{marine growth})$$

Consequently:

Assign z to population 2 (Canadian) if $\ln\left(\frac{\hat{p}(z)}{1-\hat{p}(z)}\right) \geq 0$; otherwise assign z to population 1 (Alaskan).

The confusion matrix follows.

		Predicted 1	Predicted 2	Total
Actual	1	46	4	50
	2	3	47	50

$\text{APER} = \frac{7}{100} = .07 \rightarrow 7\%$ This is the same APER produced by the linear classification function in Example 11.8.