EXERCISES

11.1. Consider the two data sets

$$\mathbf{X}_1 = \begin{bmatrix} 3 & 7 \\ 2 & 4 \\ 4 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{X}_2 = \begin{bmatrix} 6 & 9 \\ 5 & 7 \\ 4 & 8 \end{bmatrix}$$

for which

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$$

and

$$\mathbf{S}_{\text{pooled}} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

(a) Calculate the linear discriminant function in (11-19).

(b) Classify the observation $\mathbf{x}_0' = [2 \quad 7]$ as population π_1 or population π_2, using Rule (11-18) with equal priors and equal costs.

11.2. (a) Develop a linear classification function for the data in Example 11.1 using (11-19).

(b) Using the function in (a) and (11-20), construct the "confusion matrix" by classifying the given observations. Compare your classification results with those of Figure 11.1, where the classification regions were determined "by eye." (See Example 11.6.)

(c) Given the results in (b), calculate the apparent error rate (APER).

(d) State any assumptions you make to justify the use of the method in Parts a and b.

11.3. Prove Result 11.1.

Hint: Substituting the integral expressions for $P(2|1)$ and $P(1|2)$ given by (11-1) and (11-2), respectively, into (11-5) yields

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

Noting that $\Omega = R_1 \cup R_2$, so that the total probability

$$1 = \int_{\Omega} f_1(\mathbf{x})\,d\mathbf{x} = \int_{R_1} f_1(\mathbf{x})\,d\mathbf{x} + \int_{R_2} f_1(\mathbf{x})\,d\mathbf{x}$$

we can write

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

By the additive property of integrals (volumes),

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

Now, p_1, p_2, $c(1|2)$, and $c(2|1)$ are nonnegative. In addition, $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are nonnegative for all $\mathbf{x}$ and are the only quantities in ECM that depend on $\mathbf{x}$. Thus, ECM is minimized if R_1 includes those values $\mathbf{x}$ for which the integrand

$$[c(1|2)p_2 f_2(\mathbf{x}) - c(2|1)p_1 f_1(\mathbf{x})] \leq 0$$

and excludes those $\mathbf{x}$ for which this quantity is positive.

11.4. A researcher wants to determine a procedure for discriminating between two multivariate populations. The researcher has enough data available to estimate the density functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ associated with populations π_1 and π_2, respectively. Let $c(2|1) = 50$ (this is the cost of assigning items as π_2, given that π_1 is true) and $c(1|2) = 100$.

In addition, it is known that about 20% of all possible items (for which the measurements $\mathbf{x}$ can be recorded) belong to π_2.

(a) Give the minimum ECM rule (in general form) for assigning a new item to one of the two populations.

(b) Measurements recorded on a new item yield the density values $f_1(\mathbf{x}) = .3$ and $f_2(\mathbf{x}) = .5$. Given the preceding information, assign this item to population π_1 or population π_2.

11.5. Show that

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

[see Equation (11-13).]

11.6. Consider the linear function $Y = \mathbf{a}'\mathbf{X}$. Let $E(\mathbf{X}) = \boldsymbol{\mu}_1$ and $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$ if $\mathbf{X}$ belongs to population π_1. Let $E(\mathbf{X}) = \boldsymbol{\mu}_2$ and $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$ if $\mathbf{X}$ belongs to population π_2. Let $m = \tfrac{1}{2}(\mu_{1Y} + \mu_{2Y}) = \tfrac{1}{2}(\mathbf{a}'\boldsymbol{\mu}_1 + \mathbf{a}'\boldsymbol{\mu}_2)$. Given that $\mathbf{a}' = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}$, show each of the following.

(a) $E(\mathbf{a}'\mathbf{X}|\pi_1) - m = \mathbf{a}'\boldsymbol{\mu}_1 - m > 0$

(b) $E(\mathbf{a}'\mathbf{X}|\pi_2) - m = \mathbf{a}'\boldsymbol{\mu}_2 - m < 0$

Hint: Recall that $\boldsymbol{\Sigma}$ is of full rank and is positive definite, so $\boldsymbol{\Sigma}^{-1}$ exists and is positive definite.

11.7. Let $f_1(x) = (1 - |x|)$ for $|x| \le 1$ and $f_2(x) = (1 - |x - .5|)$ for $-.5 \le x \le 1.5$.

(a) Sketch the two densities.

(b) Identify the classification regions when $p_1 = p_2$ and $c(1|2) = c(2|1)$.

(c) Identify the classification regions when $p_1 = .2$ and $c(1|2) = c(2|1)$.

11.8. Refer to Exercise 11.7. Let $f_1(x)$ be the same as in that exercise, but take $f_2(x) = \tfrac{1}{4}(2 - |x - .5|)$ for $-1.5 \le x \le 2.5$.

(a) Sketch the two densities.

(b) Determine the classification regions when $p_1 = p_2$ and $c(1|2) = c(2|1)$.

11.9. For $g = 2$ groups, show that the ratio in (11-59) is proportional to the ratio

$$\frac{\begin{pmatrix}\text{squared distance}\\ \text{between means of } Y\end{pmatrix}}{(\text{variance of } Y)} = \frac{(\mu_{1Y} - \mu_{2Y})^2}{\sigma_Y^2} = \frac{(\mathbf{a}'\boldsymbol{\mu}_1 - \mathbf{a}'\boldsymbol{\mu}_2)^2}{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}$$
$$= \frac{\mathbf{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\mathbf{a}}{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}} = \frac{(\mathbf{a}'\boldsymbol{\delta})^2}{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}$$

where $\boldsymbol{\delta} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ is the difference in mean vectors. This ratio is the population counterpart of (11-23). Show that the ratio is maximized by the linear combination

$$\mathbf{a} = c\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta} = c\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

for any $c \ne 0$.

Hint: Note that $(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})' = \frac{1}{4}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'$ for $i = 1,2$, where $\bar{\boldsymbol{\mu}} = \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$.

11.10. Suppose that $n_1 = 11$ and $n_2 = 12$ observations are made on two random variables X_1 and X_2, where X_1 and X_2 are assumed to have a bivariate normal distribution with a common covariance matrix $\boldsymbol{\Sigma}$, but possibly different mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ for the two samples. The sample mean vectors and pooled covariance matrix are

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}; \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\mathbf{S}_{\text{pooled}} = \begin{bmatrix} 7.3 & -1.1 \\ -1.1 & 4.8 \end{bmatrix}$$

(a) Test for the difference in population mean vectors using Hotelling's two-sample T^2-statistic. Let $\alpha = .10$.

(b) Construct Fisher's (sample) linear discriminant function. [See (11-19) and (11-25).]

(c) Assign the observation $\mathbf{x}_0' = [0 \quad 1]$ to either population π_1 or π_2. Assume equal costs and equal prior probabilities.

11.11. Suppose a univariate random variable X has a normal distribution with variance 4. If X is from population π_1, its mean is 10; if it is from population π_2, its mean is 14. Assume equal prior probabilities for the events $A1 = X$ is from population π_1 and $A2 = X$ is from population π_2, and assume that the misclassification costs $c(2|1)$ and $c(1|2)$ are equal (for instance, \$10). We decide that we shall allocate (classify) X to population π_1 if $X \le c$, for some c to be determined, and to population π_2 if $X > c$. Let $B1$ be the event X is classified into population π_1 and $B2$ be the event X is classified into population π_2. Make a table showing the following: $P(B1|A2)$, $P(B2|A1)$, $P(A1 \text{ and } B2)$, $P(A2 \text{ and } B1)$; $P(\text{misclassification})$, and expected cost for various values of c. For what choice of c is expected cost minimized? The table should take the following form:

c	$P(B1\|A2)$	$P(B2\|A1)$	$P(A1 \text{ and } B2)$	$P(A2 \text{ and } B1)$	$P(\text{error})$	Expected cost
10						
⋮						
14						

What is the value of the minimum expected cost?

11.12. Repeat Exercise 11.11 if the prior probabilities of $A1$ and $A2$ are equal, but $c(2|1) = \$5$ and $c(1|2) = \$15$.

11.13. Repeat Exercise 11.11 if the prior probabilities of $A1$ and $A2$ are $P(A1) = .25$ and $P(A2) = .75$ and the misclassification costs are as in Exercise 11.12.

11.14. Consider the discriminant functions derived in Example 11.3. Normalize $\hat{\mathbf{a}}$ using (11-21) and (11-22). Compute the two midpoints $\hat{m}_1^*$ and $\hat{m}_2^*$ corresponding to the two choices of normalized vectors, say, $\hat{\mathbf{a}}_1^*$ and $\hat{\mathbf{a}}_2^*$. Classify $\mathbf{x}_0' = [-.210, -.044]$ with the function $\hat{y}_0^* = \hat{\mathbf{a}}^{*\prime} x_0$ for the two cases. Are the results consistent with the classification obtained for the case of equal prior probabilities in Example 11.3? Should they be?

11.15. Derive the expressions in (11-27) from (11-6) when $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are multivariate normal densities with means $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ and covariances $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$, respectively.

11.16. Suppose $\mathbf{x}$ comes from one of two populations:

π_1: Normal with mean $\boldsymbol{\mu}_1$ and covariance matrix $\boldsymbol{\Sigma}_1$

π_2: Normal with mean $\boldsymbol{\mu}_2$ and covariance matrix $\boldsymbol{\Sigma}_2$

If the respective density functions are denoted by $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$, find the expression for the quadratic discriminator

$$Q = \ln\left[\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}\right]$$

If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, for instance, verify that Q becomes

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

11.17. Suppose populations π_1 and π_2 are as follows:

	Population	
	π_1	π_2
Distribution	Normal	Normal
Mean $\boldsymbol{\mu}$	$[10, 15]'$	$[10, 25]'$
Covariance $\boldsymbol{\Sigma}$	$\begin{bmatrix} 18 & 12 \\ 12 & 32 \end{bmatrix}$	$\begin{bmatrix} 20 & -7 \\ -7 & 5 \end{bmatrix}$

Assume equal prior probabilities and misclassifications costs of $c(2|1) = \$10$ and $c(1|2) = \$73.89$. Find the posterior probabilities of populations π_1 and π_2, $P(\pi_1|\mathbf{x})$ and $P(\pi_2|\mathbf{x})$, the value of the quadratic discriminator Q in Exercise 11.16, and the classification for each value of $\mathbf{x}$ in the following table:

$\mathbf{x}$	$P(\pi_1\|\mathbf{x})$	$P(\pi_2\|\mathbf{x})$	Q	Classification
$[10, 15]'$				
$[12, 17]'$				
$\vdots$				
$[30, 35]'$				

(*Note:* Use an increment of 2 in each coordinate—11 points in all.)

Show each of the following on a graph of the x_1, x_2 plane.

(a) The mean of each population

(b) The ellipse of minimal area with probability .95 of containing $\mathbf{x}$ for each population

(c) The region R_1 (for population π_1) and the region $\Omega - R_1 = R_2$ (for population π_2)

(d) The 11 points classified in the table

11.18. If $\mathbf{B}$ is defined as $c(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'$ for some constant c, verify that $\mathbf{e} = c\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ is in fact an (unscaled) eigenvector of $\boldsymbol{\Sigma}^{-1}\mathbf{B}$, where $\boldsymbol{\Sigma}$ is a covariance matrix.

11.19. (a) Using the original data sets $\mathbf{X}_1$ and $\mathbf{X}_2$ given in Example 11.7, calculate $\bar{\mathbf{x}}_i$, $\mathbf{S}_i$, $i = 1, 2$, and $\mathbf{S}_{\text{pooled}}$, verifying the results provided for these quantities in the example.

(b) Using the calculations in Part a, compute Fisher's linear discriminant function, and use it to classify the sample observations according to Rule (11-25). Verify that the confusion matrix given in Example 11.7 is correct.

(c) Classify the sample observations on the basis of smallest squared distance $D_i^2(\mathbf{x})$ of the observations from the group means $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$. [See (11-54).] Compare the results with those in Part b. Comment.

11.20. The matrix identity (see Bartlett [3])

$$\mathbf{S}_{H,\text{pooled}}^{-1} = \frac{n-3}{n-2}\left(\mathbf{S}_{\text{pooled}}^{-1} + \frac{c_k}{1 - c_k(\mathbf{x}_H - \bar{\mathbf{x}}_k)'\mathbf{S}_{\text{pooled}}^{-1}(\mathbf{x}_H - \bar{\mathbf{x}}_k)} \cdot \mathbf{S}_{\text{pooled}}^{-1}(\mathbf{x}_H - \bar{\mathbf{x}}_k)(\mathbf{x}_H - \bar{\mathbf{x}}_k)'\mathbf{S}_{\text{pooled}}^{-1}\right)$$

where

$$c_k = \frac{n_k}{(n_k - 1)(n - 2)}$$

allows the calculation of $\mathbf{S}_{H,\text{pooled}}^{-1}$ from $\mathbf{S}_{\text{pooled}}^{-1}$. Verify this identity using the data from Example 11.7. Specifically, set $n = n_1 + n_2$, $k = 1$, and $\mathbf{x}_H' = [2, 12]$. Calculate $\mathbf{S}_{H,\text{pooled}}^{-1}$ using the full data $\mathbf{S}_{\text{pooled}}^{-1}$ and $\bar{\mathbf{x}}_1$, and compare the result with $\mathbf{S}_{H,\text{pooled}}^{-1}$ in Example 11.7.

11.21. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s > 0$ denote the $s \leq \min(g-1, p)$ nonzero eigenvalues of $\boldsymbol{\Sigma}^{-1}\mathbf{B}_\mu$ and $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_s$ the corresponding eigenvectors (scaled so that $\mathbf{e}'\boldsymbol{\Sigma}\mathbf{e} = 1$). Show that the vector of coefficients $\mathbf{a}$ that maximizes the ratio

$$\frac{\mathbf{a}'\mathbf{B}_\mu\mathbf{a}}{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}} = \frac{\mathbf{a}'\left[\sum_{i=1}^{g}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\right]\mathbf{a}}{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}$$

is given by $\mathbf{a}_1 = \mathbf{e}_1$. The linear combination $\mathbf{a}_1'\mathbf{X}$ is called the *first discriminant*. Show that the value $\mathbf{a}_2 = \mathbf{e}_2$ maximizes the ratio subject to $\text{Cov}(\mathbf{a}_1'\mathbf{X}, \mathbf{a}_2'\mathbf{X}) = 0$. The linear combination $\mathbf{a}_2'\mathbf{X}$ is called the *second discriminant*. Continuing, $\mathbf{a}_k = \mathbf{e}_k$ maximizes the ratio subject to $0 = \text{Cov}(\mathbf{a}_k'\mathbf{X}, \mathbf{a}_i'\mathbf{X})$, $i < k$, and $\mathbf{a}_k'\mathbf{X}$ is called the kth *discriminant*. Also, $\text{Var}(\mathbf{a}_i'\mathbf{X}) = 1$, $i = 1, \ldots, s$. [See (11-62) for the sample equivalent.]

Hint: We first convert the maximization problem to one already solved. By the spectral decomposition in (2-20), $\boldsymbol{\Sigma} = \mathbf{P}'\boldsymbol{\Lambda}\mathbf{P}$ where $\boldsymbol{\Lambda}$ is a diagonal matrix with positive elements λ_i. Let $\boldsymbol{\Lambda}^{1/2}$ denote the diagonal matrix with elements $\sqrt{\lambda_i}$. By (2-22), the symmetric square-root matrix $\boldsymbol{\Sigma}^{1/2} = \mathbf{P}'\boldsymbol{\Lambda}^{1/2}\mathbf{P}$ and its inverse $\boldsymbol{\Sigma}^{-1/2} = \mathbf{P}'\boldsymbol{\Lambda}^{-1/2}\mathbf{P}$ satisfy $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1/2} = \mathbf{I} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{1/2}$ and $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$. Next, set

$$\mathbf{u} = \boldsymbol{\Sigma}^{1/2}\mathbf{a}$$

so $\mathbf{u}'\mathbf{u} = \mathbf{a}'\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{a} = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$ and $\mathbf{u}'\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}\mathbf{u} = \mathbf{a}'\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{1/2}\mathbf{a} = \mathbf{a}'\mathbf{B}_\mu\mathbf{a}$. Consequently, the problem reduces to maximizing

$$\frac{\mathbf{u}'\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}\mathbf{u}}{\mathbf{u}'\mathbf{u}}$$

over $\mathbf{u}$. From (2-51), the maximum of this ratio is λ_1, the largest eigenvalue of $\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}$. This maximum occurs when $\mathbf{u} = \mathbf{e}_1$, the normalized eigenvector associated with λ_1. Because $\mathbf{e}_1 = \mathbf{u} = \boldsymbol{\Sigma}^{1/2}\mathbf{a}_1$, or $\mathbf{a}_1 = \boldsymbol{\Sigma}^{-1/2}\mathbf{e}_1$, $\text{Var}(\mathbf{a}_1'\mathbf{X}) = \mathbf{a}_1'\boldsymbol{\Sigma}\mathbf{a}_1 = \mathbf{e}_1'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2}\mathbf{e}_1 = \mathbf{e}_1'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{e}_1 = \mathbf{e}_1'\mathbf{e}_1 = 1$. By (2-52), $\mathbf{u} \perp \mathbf{e}_1$ maximizes the preceding ratio when $\mathbf{u} = \mathbf{e}_2$, the normalized eigenvector corresponding to λ_2. For this choice, $\mathbf{a}_2 = \boldsymbol{\Sigma}^{-1/2}\mathbf{e}_2$, and $\text{Cov}(\mathbf{a}_2'\mathbf{X}, \mathbf{a}_1'\mathbf{X}) = \mathbf{a}_2'\boldsymbol{\Sigma}\mathbf{a}_1 = \mathbf{e}_2'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1/2}\mathbf{e}_1 = \mathbf{e}_2'\mathbf{e}_1 = 0$, since $\mathbf{e}_2 \perp \mathbf{e}_1$. Similarly, $\text{Var}(\mathbf{a}_2'\mathbf{X}) = \mathbf{a}_2'\boldsymbol{\Sigma}\mathbf{a}_2 = \mathbf{e}_2'\mathbf{e}_2 = 1$. Continue in this fashion for the remaining discriminants. Note that if λ and $\mathbf{e}$ are an eigenvalue–eigenvector pair of $\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}$, then

$$\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}\mathbf{e} = \lambda\mathbf{e}$$

and multiplication on the left by $\boldsymbol{\Sigma}^{-1/2}$ gives

$$\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}\mathbf{e} = \lambda\boldsymbol{\Sigma}^{-1/2}\mathbf{e} \quad \text{or} \quad \boldsymbol{\Sigma}^{-1}\mathbf{B}_\mu(\boldsymbol{\Sigma}^{-1/2}\mathbf{e}) = \lambda(\boldsymbol{\Sigma}^{-1/2}\mathbf{e})$$

Thus, $\boldsymbol{\Sigma}^{-1}\mathbf{B}_\mu$ has the same eigenvalues as $\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}$, but the corresponding eigenvector is proportional to $\boldsymbol{\Sigma}^{-1/2}\mathbf{e} = \mathbf{a}$, as asserted.

11.22. Show that $\Delta_S^2 = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \lambda_1 + \lambda_2 + \cdots + \lambda_s$, where $\lambda_1, \lambda_2, \ldots, \lambda_s$ are the nonzero eigenvalues of $\boldsymbol{\Sigma}^{-1}\mathbf{B}_\mu$ (or $\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}$) and Δ_S^2 is given by (11-68). Also, show that $\lambda_1 + \lambda_2 + \cdots + \lambda_r$ is the resulting separation when only the first r discriminants, $Y_1, Y_2, \ldots, Y_r$ are used.

Hint: Let $\mathbf{P}$ be the orthogonal matrix whose ith row $\mathbf{e}_i'$ is the eigenvector of $\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}$ corresponding to the ith largest eigenvalue, $i = 1, 2, \ldots, p$. Consider

$$\underset{(p\times 1)}{\mathbf{Y}} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_s \\ \vdots \\ Y_p \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1'\boldsymbol{\Sigma}^{-1/2}\mathbf{X} \\ \vdots \\ \mathbf{e}_s'\boldsymbol{\Sigma}^{-1/2}\mathbf{X} \\ \vdots \\ \mathbf{e}_p'\boldsymbol{\Sigma}^{-1/2}\mathbf{X} \end{bmatrix} = \mathbf{P}\boldsymbol{\Sigma}^{-1/2}\mathbf{X}$$

Now, $\boldsymbol{\mu}_{iY} = E(\mathbf{Y} \mid \pi_i) = \mathbf{P}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}_i$, and $\bar{\boldsymbol{\mu}}_Y = \mathbf{P}\boldsymbol{\Sigma}^{-1/2}\bar{\boldsymbol{\mu}}$, so

$$\begin{aligned}(\boldsymbol{\mu}_{iY} - \bar{\boldsymbol{\mu}}_Y)'(\boldsymbol{\mu}_{iY} - \bar{\boldsymbol{\mu}}_Y) &= (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1/2}\mathbf{P}'\mathbf{P}\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}}) \\ &= (\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})\end{aligned}$$

Therefore, $\Delta_S^2 = \sum_{i=1}^{g} (\boldsymbol{\mu}_{iY} - \bar{\boldsymbol{\mu}}_Y)'(\boldsymbol{\mu}_{iY} - \bar{\boldsymbol{\mu}}_Y)$. Using Y_1, we have

$$\begin{aligned}\sum_{i=1}^{g} (\mu_{iY_1} - \bar{\mu}_{Y_1})^2 &= \sum_{i=1}^{g} \mathbf{e}_1'\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1/2}\mathbf{e}_1 \\ &= \mathbf{e}_1'\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}\mathbf{e}_1 = \lambda_1\end{aligned}$$

because $\mathbf{e}_1$ has eigenvalue λ_1. Similarly, Y_2 produces

$$\sum_{i=1}^{g} (\mu_{iY_2} - \bar{\mu}_{Y_2})^2 = \mathbf{e}_2'\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}\mathbf{e}_2 = \lambda_2$$

and Y_p produces

$$\sum_{i=1}^{g} (\mu_{iY_p} - \bar{\mu}_{Y_p})^2 = \mathbf{e}_p'\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_\mu\boldsymbol{\Sigma}^{-1/2}\mathbf{e}_p = \lambda_p$$

Thus,

$$\Delta_S^2 = \sum_{i=1}^{g} (\boldsymbol{\mu}_{iY} - \bar{\boldsymbol{\mu}}_Y)'(\boldsymbol{\mu}_{iY} - \bar{\boldsymbol{\mu}}_Y)$$
$$= \sum_{i=1}^{g} (\mu_{iY_1} - \bar{\mu}_{Y_1})^2 + \sum_{i=1}^{g} (\mu_{iY_2} - \bar{\mu}_{Y_2})^2 + \cdots + \sum_{i=1}^{g} (\mu_{iY_p} - \bar{\mu}_{Y_p})^2$$
$$= \lambda_1 + \lambda_2 + \cdots + \lambda_p = \lambda_1 + \lambda_2 + \cdots + \lambda_s$$

since $\lambda_{s+1} = \cdots = \lambda_p = 0$. If only the first r discriminants are used, their contribution to Δ_S^2 is $\lambda_1 + \lambda_2 + \cdots + \lambda_r$.

The following exercises require the use of a computer.

11.23. Consider the data given in Exercise 1.14.

(a) Check the marginal distributions of the x_i's in both the multiple-sclerosis (MS) group and non-multiple-sclerosis (NMS) group for normality by graphing the corresponding observations as normal probability plots. Suggest appropriate data transformations if the normality assumption is suspect.

(b) Assume that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$. Construct Fisher's linear discriminant function. Do all the variables in the discriminant function appear to be important? Discuss your answer. Develop a classification rule assuming equal prior probabilities and equal costs of misclassification.

(c) Using the results in (b), calculate the apparent error rate. If computing resources allow, calculate an estimate of the expected actual error rate using Lachenbruch's holdout procedure. Compare the two error rates.

11.24. Annual financial data are collected for bankrupt firms approximately 2 years prior to their bankruptcy and for financially sound firms at about the same time. The data on four variables, X_1 = CF/TD = (cash flow)/(total debt), X_2 = NI/TA = (net income)/(total assets), X_3 = CA/CL = (current assets)/(current liabilities), and X_4 = CA/NS = (current assets)/(net sales), are given in Table 11.4.

(a) Using a different symbol for each group, plot the data for the pairs of observations (x_1, x_2), (x_1, x_3) and (x_1, x_4). Does it appear as if the data are approximately bivariate normal for any of these pairs of variables?

(b) Using the $n_1 = 21$ pairs of observations (x_1, x_2) for bankrupt firms and the $n_2 = 25$ pairs of observations (x_1, x_2) for nonbankrupt firms, calculate the sample mean vectors $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ and the sample covariance matrices $\mathbf{S}_1$ and $\mathbf{S}_2$.

(c) Using the results in (b) and assuming that both random samples are from bivariate normal populations, construct the classification rule (11-29) with $p_1 = p_2$ and $c(1|2) = c(2|1)$.

(d) Evaluate the performance of the classification rule developed in (c) by computing the apparent error rate (APER) from (11-34) and the estimated expected actual error rate $\hat{E}$ (AER) from (11-36).

(e) Repeat Parts c and d, assuming that $p_1 = .05$, $p_2 = .95$, and $c(1|2) = c(2|1)$. Is this choice of prior probabilities reasonable? Explain.

(f) Using the results in (b), form the pooled covariance matrix $\mathbf{S}_{\text{pooled}}$, and construct Fisher's sample linear discriminant function in (11-19). Use this function to classify the sample observations and evaluate the APER. Is Fisher's linear discriminant function a sensible choice for a classifier in this case? Explain.

(g) Repeat Parts b–e using the observation pairs (x_1, x_3) and (x_1, x_4). Do some variables appear to be better classifiers than others? Explain.

(h) Repeat Parts b–e using observations on all four variables (X_1, X_2, X_3, X_4).

Table 11.4 Bankruptcy Data

Row	$x_1 = \frac{CF}{TD}$	$x_2 = \frac{NI}{TA}$	$x_3 = \frac{CA}{CL}$	$x_4 = \frac{CA}{NS}$	Population $\pi_i, i = 1, 2$
1	−.45	−.41	1.09	.45	0
2	−.56	−.31	1.51	.16	0
3	.06	.02	1.01	.40	0
4	−.07	−.09	1.45	.26	0
5	−.10	−.09	1.56	.67	0
6	−.14	−.07	.71	.28	0
7	.04	.01	1.50	.71	0
8	−.06	−.06	1.37	.40	0
9	.07	−.01	1.37	.34	0
10	−.13	−.14	1.42	.44	0
11	−.23	−.30	.33	.18	0
12	.07	.02	1.31	.25	0
13	.01	.00	2.15	.70	0
14	−.28	−.23	1.19	.66	0
15	.15	.05	1.88	.27	0
16	.37	.11	1.99	.38	0
17	−.08	−.08	1.51	.42	0
18	.05	.03	1.68	.95	0
19	.01	−.00	1.26	.60	0
20	.12	.11	1.14	.17	0
21	−.28	−.27	1.27	.51	0
1	.51	.10	2.49	.54	1
2	.08	.02	2.01	.53	1
3	.38	.11	3.27	.35	1
4	.19	.05	2.25	.33	1
5	.32	.07	4.24	.63	1
6	.31	.05	4.45	.69	1
7	.12	.05	2.52	.69	1
8	−.02	.02	2.05	.35	1
9	.22	.08	2.35	.40	1
10	.17	.07	1.80	.52	1
11	.15	.05	2.17	.55	1
12	−.10	−.01	2.50	.58	1
13	.14	−.03	.46	.26	1
14	.14	.07	2.61	.52	1
15	.15	.06	2.23	.56	1
16	.16	.05	2.31	.20	1
17	.29	.06	1.84	.38	1
18	.54	.11	2.33	.48	1
19	−.33	−.09	3.01	.47	1
20	.48	.09	1.24	.18	1
21	.56	.11	4.29	.45	1
22	.20	.08	1.99	.30	1
23	.47	.14	2.92	.45	1
24	.17	.04	2.45	.14	1
25	.58	.04	5.06	.13	1

Legend: $\pi_1 = 0$: bankrupt firms; $\pi_2 = 1$: nonbankrupt firms.
Source: 1968, 1969, 1970, 1971, 1972 Moody's Industrial Manuals.

11.25. The annual financial data listed in Table 11.4 have been analyzed by Johnson [19] with a view toward detecting influential observations in a discriminant analysis. Consider variables X_1 = CF/TD and X_3 = CA/CL.

(a) Using the data on variables X_1 and X_3, construct Fisher's linear discriminant function. Use this function to classify the sample observations and evaluate the APER. [See (11-25) and (11-34).] Plot the data and the discriminant line in the (x_1, x_3) coordinate system.

(b) Johnson [19] has argued that the multivariate observations in rows 16 for bankrupt firms and 13 for sound firms are influential. Using the X_1, X_3 data, calculate Fisher's linear discriminant function with *only* data point 16 for bankrupt firms deleted. Repeat this procedure with *only* data point 13 for sound firms deleted. Plot the respective discriminant lines on the scatter in part a, and calculate the APERs, ignoring the deleted point in each case. Does deleting either of these multivariate observations make a difference? (Note that neither of the potentially influential data points is particularly "distant" from the center of its respective scatter.)

11.26. Using the data in Table 11.4, define a binary response variable Z that assumes the value 0 if a firm is bankrupt and 1 if a firm is not bankrupt. Let X = CA/CL, and consider the straight-line regression of Z on X.

(a) Although a binary response variable does not meet the standard regression assumptions, consider using least squares to determine the fitted straight line for the X, Z data. Plot the fitted values for bankrupt firms as a dot diagram on the interval [0, 1]. Repeat this procedure for nonbankrupt firms and overlay the two dot diagrams. A reasonable discrimination rule is to predict that a firm will go bankrupt if its fitted value is closer to 0 than to 1. That is, the fitted value is less than .5. Similarly, a firm is predicted to be sound if its fitted value is greater than .5. Use this decision rule to classify the sample firms. Calculate the APER.

(b) Repeat the analysis in Part a using all four variables, $X_1, \ldots, X_4$. Is there any change in the APER? Do data points 16 for bankrupt firms and 13 for nonbankrupt firms stand out as influential?

(c) Perform a logistic regression using all four variables.

11.27. The data in Table 11.5 contain observations on X_2 = sepal width and X_4 = petal width for samples from three species of iris. There are $n_1 = n_2 = n_3 = 50$ observations in each sample.

(a) Plot the data in the (x_2, x_4) variable space. Do the observations for the three groups appear to be bivariate normal?

Table 11.5 Data on Irises

π_1: *Iris setosa*				π_2: *Iris versicolor*				π_3: *Iris virginica*			
Sepal length x_1	Sepal width x_2	Petal length x_3	Petal width x_4	Sepal length x_1	Sepal width x_2	Petal length x_3	Petal width x_4	Sepal length x_1	Sepal width x_2	Petal length x_3	Petal width x_4
5.1	3.5	1.4	0.2	7.0	3.2	4.7	1.4	6.3	3.3	6.0	2.5
4.9	3.0	1.4	0.2	6.4	3.2	4.5	1.5	5.8	2.7	5.1	1.9
4.7	3.2	1.3	0.2	6.9	3.1	4.9	1.5	7.1	3.0	5.9	2.1
4.6	3.1	1.5	0.2	5.5	2.3	4.0	1.3	6.3	2.9	5.6	1.8
5.0	3.6	1.4	0.2	6.5	2.8	4.6	1.5	6.5	3.0	5.8	2.2
5.4	3.9	1.7	0.4	5.7	2.8	4.5	1.3	7.6	3.0	6.6	2.1

(continues on next page)

Table 11.5 *(continued)*

π_1: *Iris setosa*				π_2: *Iris versicolor*				π_3: *Iris virginica*			
Sepal length x_1	Sepal width x_2	Petal length x_3	Petal width x_4	Sepal length x_1	Sepal width x_2	Petal length x_3	Petal width x_4	Sepal length x_1	Sepal width x_2	Petal length x_3	Petal width x_4
4.6	3.4	1.4	0.3	6.3	3.3	4.7	1.6	4.9	2.5	4.5	1.7
5.0	3.4	1.5	0.2	4.9	2.4	3.3	1.0	7.3	2.9	6.3	1.8
4.4	2.9	1.4	0.2	6.6	2.9	4.6	1.3	6.7	2.5	5.8	1.8
4.9	3.1	1.5	0.1	5.2	2.7	3.9	1.4	7.2	3.6	6.1	2.5
5.4	3.7	1.5	0.2	5.0	2.0	3.5	1.0	6.5	3.2	5.1	2.0
4.8	3.4	1.6	0.2	5.9	3.0	4.2	1.5	6.4	2.7	5.3	1.9
4.8	3.0	1.4	0.1	6.0	2.2	4.0	1.0	6.8	3.0	5.5	2.1
4.3	3.0	1.1	0.1	6.1	2.9	4.7	1.4	5.7	2.5	5.0	2.0
5.8	4.0	1.2	0.2	5.6	2.9	3.6	1.3	5.8	2.8	5.1	2.4
5.7	4.4	1.5	0.4	6.7	3.1	4.4	1.4	6.4	3.2	5.3	2.3
5.4	3.9	1.3	0.4	5.6	3.0	4.5	1.5	6.5	3.0	5.5	1.8
5.1	3.5	1.4	0.3	5.8	2.7	4.1	1.0	7.7	3.8	6.7	2.2
5.7	3.8	1.7	0.3	6.2	2.2	4.5	1.5	7.7	2.6	6.9	2.3
5.1	3.8	1.5	0.3	5.6	2.5	3.9	1.1	6.0	2.2	5.0	1.5
5.4	3.4	1.7	0.2	5.9	3.2	4.8	1.8	6.9	3.2	5.7	2.3
5.1	3.7	1.5	0.4	6.1	2.8	4.0	1.3	5.6	2.8	4.9	2.0
4.6	3.6	1.0	0.2	6.3	2.5	4.9	1.5	7.7	2.8	6.7	2.0
5.1	3.3	1.7	0.5	6.1	2.8	4.7	1.2	6.3	2.7	4.9	1.8
4.8	3.4	1.9	0.2	6.4	2.9	4.3	1.3	6.7	3.3	5.7	2.1
5.0	3.0	1.6	0.2	6.6	3.0	4.4	1.4	7.2	3.2	6.0	1.8
5.0	3.4	1.6	0.4	6.8	2.8	4.8	1.4	6.2	2.8	4.8	1.8
5.2	3.5	1.5	0.2	6.7	3.0	5.0	1.7	6.1	3.0	4.9	1.8
5.2	3.4	1.4	0.2	6.0	2.9	4.5	1.5	6.4	2.8	5.6	2.1
4.7	3.2	1.6	0.2	5.7	2.6	3.5	1.0	7.2	3.0	5.8	1.6
4.8	3.1	1.6	0.2	5.5	2.4	3.8	1.1	7.4	2.8	6.1	1.9
5.4	3.4	1.5	0.4	5.5	2.4	3.7	1.0	7.9	3.8	6.4	2.0
5.2	4.1	1.5	0.1	5.8	2.7	3.9	1.2	6.4	2.8	5.6	2.2
5.5	4.2	1.4	0.2	6.0	2.7	5.1	1.6	6.3	2.8	5.1	1.5
4.9	3.1	1.5	0.2	5.4	3.0	4.5	1.5	6.1	2.6	5.6	1.4
5.0	3.2	1.2	0.2	6.0	3.4	4.5	1.6	7.7	3.0	6.1	2.3
5.5	3.5	1.3	0.2	6.7	3.1	4.7	1.5	6.3	3.4	5.6	2.4
4.9	3.6	1.4	0.1	6.3	2.3	4.4	1.3	6.4	3.1	5.5	1.8
4.4	3.0	1.3	0.2	5.6	3.0	4.1	1.3	6.0	3.0	4.8	1.8
5.1	3.4	1.5	0.2	5.5	2.5	4.0	1.3	6.9	3.1	5.4	2.1
5.0	3.5	1.3	0.3	5.5	2.6	4.4	1.2	6.7	3.1	5.6	2.4
4.5	2.3	1.3	0.3	6.1	3.0	4.6	1.4	6.9	3.1	5.1	2.3
4.4	3.2	1.3	0.2	5.8	2.6	4.0	1.2	5.8	2.7	5.1	1.9
5.0	3.5	1.6	0.6	5.0	2.3	3.3	1.0	6.8	3.2	5.9	2.3
5.1	3.8	1.9	0.4	5.6	2.7	4.2	1.3	6.7	3.3	5.7	2.5
4.8	3.0	1.4	0.3	5.7	3.0	4.2	1.2	6.7	3.0	5.2	2.3
5.1	3.8	1.6	0.2	5.7	2.9	4.2	1.3	6.3	2.5	5.0	1.9
4.6	3.2	1.4	0.2	6.2	2.9	4.3	1.3	6.5	3.0	5.2	2.0
5.3	3.7	1.5	0.2	5.1	2.5	3.0	1.1	6.2	3.4	5.4	2.3
5.0	3.3	1.4	0.2	5.7	2.8	4.1	1.3	5.9	3.0	5.1	1.8

Source: Anderson [1].

(b) Assume that the samples are from bivariate normal populations with a common covariance matrix. Test the hypothesis $H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$ versus H_1: at least one $\boldsymbol{\mu}_i$ is different from the others at the $\alpha = .05$ significance level. Is the assumption of a common covariance matrix reasonable in this case? Explain.

(c) Assuming that the populations are bivariate normal, construct the quadratic discriminate scores $\hat{d}_i^Q(\mathbf{x})$ given by (11-47) with $p_1 = p_2 = p_3 = \frac{1}{3}$. Using Rule (11-48), classify the new observation $\mathbf{x}_0' = [3.5 \quad 1.75]$ into population π_1, π_2, or π_3.

(d) Assume that the covariance matrices $\boldsymbol{\Sigma}_i$ are the same for all three bivariate normal populations. Construct the linear discriminate score $\hat{d}_i(\mathbf{x})$ given by (11-51), and use it to assign $\mathbf{x}_0' = [3.5 \quad 1.75]$ to one of the populations π_i, $i = 1, 2, 3$ according to (11-52). Take $p_1 = p_2 = p_3 = \frac{1}{3}$. Compare the results in Parts c and d. Which approach do you prefer? Explain.

(e) Assuming equal covariance matrices and bivariate normal populations, and supposing that $p_1 = p_2 = p_3 = \frac{1}{3}$, allocate $\mathbf{x}_0' = [3.5 \quad 1.75]$ to π_1, π_2, or π_3 using Rule (11-56). Compare the result with that in Part d. Delineate the classification regions $\hat{R}_1$, $\hat{R}_2$, and $\hat{R}_3$ on your graph from Part a determined by the linear functions $\hat{d}_{ki}(\mathbf{x}_0)$ in (11-56).

(f) Using the linear discriminant scores from Part d, classify the sample observations. Calculate the APER and $\hat{E}(\text{AER})$. (To calculate the latter, you should use Lachenbruch's holdout procedure. [See (11-57).])

11.28. Darroch and Mosimann [6] have argued that the three species of iris indicated in Table 11.5 can be discriminated on the basis of "shape" or scale-free information alone. Let $Y_1 = X_1/X_2$ be sepal shape and $Y_2 = X_3/X_4$ be petal shape.

(a) Plot the data in the $(\log Y_1, \log Y_2)$ variable space. Do the observations for the three groups appear to be bivariate normal?

(b) Assuming equal covariance matrices and bivariate normal populations, and supposing that $p_1 = p_2 = p_3 = \frac{1}{3}$, construct the linear discriminant scores $\hat{d}_i(\mathbf{x})$ given by (11-51) using both variables $\log Y_1, \log Y_2$ and each variable individually. Calculate the APERs.

(c) Using the linear discriminant functions from Part b, calculate the holdout estimates of the expected AERs, and fill in the following summary table:

Variable(s)	Misclassification rate
$\log Y_1$	
$\log Y_2$	
$\log Y_1, \log Y_2$	

Compare the preceding misclassification rates with those in the summary tables in Example 11.12. Does it appear as if information on shape alone is an effective discriminator for these species of iris?

(d) Compare the corresponding error rates in Parts b and c. Given the scatter plot in Part a, would you expect these rates to differ much? Explain.

11.29. The GPA and GMAT data alluded to in Example 11.11 are listed in Table 11.6.

(a) Using these data, calculate $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \bar{\mathbf{x}}_3, \bar{\mathbf{x}}$, and $\mathbf{S}_{\text{pooled}}$ and thus verify the results for these quantities given in Example 11.11.

Table 11.6 Admission Data for Graduate School of Business

π_1: Admit			π_2: Do not admit			π_3: Borderline		
Applicant no.	GPA (x_1)	GMAT (x_2)	Applicant no.	GPA (x_1)	GMAT (x_2)	Applicant no.	GPA (x_1)	GMAT (x_2)
1	2.96	596	32	2.54	446	60	2.86	494
2	3.14	473	33	2.43	425	61	2.85	496
3	3.22	482	34	2.20	474	62	3.14	419
4	3.29	527	35	2.36	531	63	3.28	371
5	3.69	505	36	2.57	542	64	2.89	447
6	3.46	693	37	2.35	406	65	3.15	313
7	3.03	626	38	2.51	412	66	3.50	402
8	3.19	663	39	2.51	458	67	2.89	485
9	3.63	447	40	2.36	399	68	2.80	444
10	3.59	588	41	2.36	482	69	3.13	416
11	3.30	563	42	2.66	420	70	3.01	471
12	3.40	553	43	2.68	414	71	2.79	490
13	3.50	572	44	2.48	533	72	2.89	431
14	3.78	591	45	2.46	509	73	2.91	446
15	3.44	692	46	2.63	504	74	2.75	546
16	3.48	528	47	2.44	336	75	2.73	467
17	3.47	552	48	2.13	408	76	3.12	463
18	3.35	520	49	2.41	469	77	3.08	440
19	3.39	543	50	2.55	538	78	3.03	419
20	3.28	523	51	2.31	505	79	3.00	509
21	3.21	530	52	2.41	489	80	3.03	438
22	3.58	564	53	2.19	411	81	3.05	399
23	3.33	565	54	2.35	321	82	2.85	483
24	3.40	431	55	2.60	394	83	3.01	453
25	3.38	605	56	2.55	528	84	3.03	414
26	3.26	664	57	2.72	399	85	3.04	446
27	3.60	609	58	2.85	381			
28	3.37	559	59	2.90	384			
29	3.80	521						
30	3.76	646						
31	3.24	467						

(b) Calculate $\mathbf{W}^{-1}$ and $\mathbf{B}$ and the eigenvalues and eigenvectors of $\mathbf{W}^{-1}\mathbf{B}$. Use the linear discriminants derived from these eigenvectors to classify the new observation $\mathbf{x}_0' = [3.21 \quad 497]$ into one of the populations π_1: admit; π_2: not admit; and π_3: borderline. Does the classification agree with that in Example 11.11? Should it? Explain.

11.30. Gerrild and Lantz [13] chemically analyzed crude-oil samples from three zones of sandstone:

π_1: Wilhelm

π_2: Sub-Mulinia

π_3: Upper

The values of the trace elements

$$
\begin{aligned}
X_1 &= \text{vanadium (in percent ash)} \\
X_2 &= \text{iron (in percent ash)} \\
X_3 &= \text{beryllium (in percent ash)}
\end{aligned}
$$

and two measures of hydrocarbons,

$$X_4 = \text{saturated hydrocarbons (in percent area)}$$

$$X_5 = \text{aromatic hydrocarbons (in percent area)}$$

are presented for 56 cases in Table 11.7. The last two measurements are determined from areas under a gas–liquid chromatography curve.

(a) Obtain the estimated minimum TPM rule, assuming normality. Comment on the adequacy of the assumption of normality.

(b) Determine the estimate of $E(\text{AER})$ using Lachenbruch's holdout procedure. Also give the confusion matrix.

(c) Consider various transformations of the data to normality (see Example 11.14), and repeat Parts a and b.

Table 11.7 Crude-Oil Data

	x_1	x_2	x_3	x_4	x_5
π_1	3.9	51.0	0.20	7.06	12.19
	2.7	49.0	0.07	7.14	12.23
	2.8	36.0	0.30	7.00	11.30
	3.1	45.0	0.08	7.20	13.01
	3.5	46.0	0.10	7.81	12.63
	3.9	43.0	0.07	6.25	10.42
	2.7	35.0	0.00	5.11	9.00
π_2	5.0	47.0	0.07	7.06	6.10
	3.4	32.0	0.20	5.82	4.69
	1.2	12.0	0.00	5.54	3.15
	8.4	17.0	0.07	6.31	4.55
	4.2	36.0	0.50	9.25	4.95
	4.2	35.0	0.50	5.69	2.22
	3.9	41.0	0.10	5.63	2.94
	3.9	36.0	0.07	6.19	2.27
	7.3	32.0	0.30	8.02	12.92
	4.4	46.0	0.07	7.54	5.76
	3.0	30.0	0.00	5.12	10.77
π_3	6.3	13.0	0.50	4.24	8.27
	1.7	5.6	1.00	5.69	4.64
	7.3	24.0	0.00	4.34	2.99
	7.8	18.0	0.50	3.92	6.09
	7.8	25.0	0.70	5.39	6.20
	7.8	26.0	1.00	5.02	2.50
	9.5	17.0	0.05	3.52	5.71
	7.7	14.0	0.30	4.65	8.63
	11.0	20.0	0.50	4.27	8.40
	8.0	14.0	0.30	4.32	7.87
	8.4	18.0	0.20	4.38	7.98

(continues on next page)

Table 11.7 *(continued)*

x_1	x_2	x_3	x_4	x_5
10.0	18.0	0.10	3.06	7.67
7.3	15.0	0.05	3.76	6.84
9.5	22.0	0.30	3.98	5.02
8.4	15.0	0.20	5.02	10.12
8.4	17.0	0.20	4.42	8.25
9.5	25.0	0.50	4.44	5.95
7.2	22.0	1.00	4.70	3.49
4.0	12.0	0.50	5.71	6.32
6.7	52.0	0.50	4.80	3.20
9.0	27.0	0.30	3.69	3.30
7.8	29.0	1.50	6.72	5.75
4.5	41.0	0.50	3.33	2.27
6.2	34.0	0.70	7.56	6.93
5.6	20.0	0.50	5.07	6.70
9.0	17.0	0.20	4.39	8.33
8.4	20.0	0.10	3.74	3.77
9.5	19.0	0.50	3.72	7.37
9.0	20.0	0.50	5.97	11.17
6.2	16.0	0.05	4.23	4.18
7.3	20.0	0.50	4.39	3.50
3.6	15.0	0.70	7.00	4.82
6.2	34.0	0.07	4.84	2.37
7.3	22.0	0.00	4.13	2.70
4.1	29.0	0.70	5.78	7.76
5.4	29.0	0.20	4.64	2.65
5.0	34.0	0.70	4.21	6.50
6.2	27.0	0.30	3.97	2.97

11.31. Refer to the data on salmon in Table 11.2.

(a) Plot the bivariate data for the two groups of salmon. Are the sizes and orientation of the scatters roughly the same? Do bivariate normal distributions with a common covariance matrix appear to be viable population models for the Alaskan and Canadian salmon?

(b) Using a linear discriminant function for two normal populations with equal priors and equal costs [see (11-19)], construct dot diagrams of the discriminant scores for the two groups. Does it appear as if the growth ring diameters separate for the two groups reasonably well? Explain.

(c) Repeat the analysis in Example 11.8 for the male and female salmon separately. Is it easier to discriminate Alaskan male salmon from Canadian male salmon than it is to discriminate the females in the two groups? Is gender (male or female) likely to be a useful discriminatory variable?

11.32. Data on hemophilia A carriers, similar to those used in Example 11.3, are listed in Table 11.8 on page 664. (See [15].) Using these data,

(a) Investigate the assumption of bivariate normality for the two groups.

Table 11.8 Hemophilia Data

Noncarriers (π_1)			Obligatory carriers (π_2)		
Group	$\log_{10}$ (AHF activity)	$\log_{10}$ (AHF antigen)	Group	$\log_{10}$ (AHF activity)	$\log_{10}$ (AHF antigen)
1	−.0056	−.1657	2	−.3478	.1151
1	−.1698	−.1585	2	−.3618	−.2008
1	−.3469	−.1879	2	−.4986	−.0860
1	−.0894	.0064	2	−.5015	−.2984
1	−.1679	.0713	2	−.1326	.0097
1	−.0836	.0106	2	−.6911	−.3390
1	−.1979	−.0005	2	−.3608	.1237
1	−.0762	.0392	2	−.4535	−.1682
1	−.1913	−.2123	2	−.3479	−.1721
1	−.1092	−.1190	2	−.3539	.0722
1	−.5268	−.4773	2	−.4719	−.1079
1	−.0842	.0248	2	−.3610	−.0399
1	−.0225	−.0580	2	−.3226	.1670
1	.0084	.0782	2	−.4319	−.0687
1	−.1827	−.1138	2	−.2734	−.0020
1	.1237	.2140	2	−.5573	.0548
1	−.4702	−.3099	2	−.3755	−.1865
1	−.1519	−.0686	2	−.4950	−.0153
1	.0006	−.1153	2	−.5107	−.2483
1	−.2015	−.0498	2	−.1652	.2132
1	−.1932	−.2293	2	−.2447	−.0407
1	.1507	.0933	2	−.4232	−.0998
1	−.1259	−.0669	2	−.2375	.2876
1	−.1551	−.1232	2	−.2205	.0046
1	−.1952	−.1007	2	−.2154	−.0219
1	.0291	.0442	2	−.3447	.0097
1	−.2228	−.1710	2	−.2540	−.0573
1	−.0997	−.0733	2	−.3778	−.2682
1	−.1972	−.0607	2	−.4046	−.1162
1	−.0867	−.0560	2	−.0639	.1569
			2	−.3351	−.1368
			2	−.0149	.1539
			2	−.0312	.1400
			2	−.1740	−.0776
			2	−.1416	.1642
			2	−.1508	.1137
			2	−.0964	.0531
			2	−.2642	.0867
			2	−.0234	.0804
			2	−.3352	.0875
			2	−.1878	.2510
			2	−.1744	.1892
			2	−.4055	−.2418
			2	−.2444	.1614
			2	−.4784	.0282

Source: See [15].

(b) Obtain the sample linear discriminant function, assuming equal prior probabilities, and estimate the error rate using the holdout procedure.

(c) Classify the following 10 new cases using the discriminant function in Part b.

(d) Repeat Parts a–c, assuming that the prior probability of obligatory carriers (group 2) is $\frac{1}{4}$ and that of noncarriers (group 1) is $\frac{3}{4}$.

New Cases Requiring Classification

Case	$\log_{10}$(AHF activity)	$\log_{10}$(AHF antigen)
1	−.112	−.279
2	−.059	−.068
3	.064	.012
4	−.043	−.052
5	−.050	−.098
6	−.094	−.113
7	−.123	−.143
8	−.011	−.037
9	−.210	−.090
10	−.126	−.019

11.33. Consider the data on bulls in Table 1.10.

(a) Using the variables YrHgt, FtFrBody, PrctFFB, Frame, BkFat, SaleHt, and SaleWt, calculate Fisher's linear discriminants, and classify the bulls as Angus, Hereford, or Simental. Calculate an estimate of $E(\text{AER})$ using the holdout procedure. Classify a bull with characteristics YrHgt = 50, FtFrBody = 1000, PrctFFB = 73, Frame = 7, BkFat = .17, SaleHt = 54, and SaleWt = 1525 as one of the three breeds. Plot the discriminant scores for the bulls in the two-dimensional discriminant space using different plotting symbols to identify the three groups.

(b) Is there a subset of the original seven variables that is almost as good for discriminating among the three breeds? Explore this possibility by computing the estimated $E(\text{AER})$ for various subsets.

11.34. Table 11.9 on pages 666–667 contains data on breakfast cereals produced by three different American manufacturers: General Mills (G), Kellogg (K), and Quaker (Q). Assuming multivariate normal data with a common covariance matrix, equal costs, and equal priors, classify the cereal brands according to manufacturer. Compute the estimated $E(\text{AER})$ using the holdout procedure. Interpret the coefficients of the discriminant functions. Does it appear as if some manufacturers are associated with more "nutritional" cereals (high protein, low fat, high fiber, low sugar, and so forth) than others? Plot the cereals in the two-dimensional discriminant space, using different plotting symbols to identify the three manufacturers.

11.35. Table 11.10 on page 668 contains measurements on the gender, age, tail length (mm), and snout to vent length (mm) for Concho Water Snakes.

Define the variables

$$X_1 = \text{Gender}$$
$$X_2 = \text{Age}$$
$$X_3 = \text{TailLength}$$
$$X_4 = \text{SntoVnLength}$$

Table 11.9 Data on Brands of Cereal

Brand	Manufacturer	Calories	Protein	Fat	Sodium	Fiber	Carbohydrates	Sugar	Potassium	Group
1 Apple_Cinnamon_Cheerios	G	110	2	2	180	1.5	10.5	10	70	1
2 Cheerios	G	110	6	2	290	2.0	17.0	1	105	1
3 Cocoa_Puffs	G	110	1	1	180	0.0	12.0	13	55	1
4 Count_Chocula	G	110	1	1	180	0.0	12.0	13	65	1
5 Golden_Grahams	G	110	1	1	280	0.0	15.0	9	45	1
6 Honey_Nut_Cheerios	G	110	3	1	250	1.5	11.5	10	90	1
7 Kix	G	110	2	1	260	0.0	21.0	3	40	1
8 Lucky_Charms	G	110	2	1	180	0.0	12.0	12	55	1
9 Multi_Grain_Cheerios	G	100	2	1	220	2.0	15.0	6	90	1
10 Oatmeal_Raisin_Crisp	G	130	3	2	170	1.5	13.5	10	120	1
11 Raisin_Nut_Bran	G	100	3	2	140	2.5	10.5	8	140	1
12 Total_Corn_Flakes	G	110	2	1	200	0.0	21.0	3	35	1
13 Total_Raisin_Bran	G	140	3	1	190	4.0	15.0	14	230	1
14 Total_Whole_Grain	G	100	3	1	200	3.0	16.0	3	110	1
15 Trix	G	110	1	1	140	0.0	13.0	12	25	1
16 Wheaties	G	100	3	1	200	3.0	17.0	3	110	1
17 Wheaties_Honey_Gold	G	110	2	1	200	1.0	16.0	8	60	1
18 All_Bran	K	70	4	1	260	9.0	7.0	5	320	2
19 Apple_Jacks	K	110	2	0	125	1.0	11.0	14	30	2
20 Corn_Flakes	K	100	2	0	290	1.0	21.0	2	35	2
21 Corn_Pops	K	110	1	0	90	1.0	13.0	12	20	2

continued

22 Cracklin'_Oat_Bran	K	110	3	3	140	4.0	10.0	7	160	2
23 Crispix	K	110	2	0	220	1.0	21.0	3	30	2
24 Froot_Loops	K	110	2	1	125	1.0	11.0	13	30	2
25 Frosted_Flakes	K	110	1	0	200	1.0	14.0	11	25	2
26 Frosted_Mini_Wheats	K	100	3	0	0	3.0	14.0	7	100	2
27 Fruitful_Bran	K	120	3	0	240	5.0	14.0	12	190	2
28 Just_Right_Crunchy_Nuggets	K	110	2	1	170	1.0	17.0	6	60	2
29 Mueslix_Crispy_Blend	K	160	3	2	150	3.0	17.0	13	160	2
30 Nut&Honey_Crunch	K	120	2	1	190	0.0	15.0	9	40	2
31 Nutri-grain_Almond-Raisin	K	140	3	2	220	3.0	21.0	7	130	2
32 Nutri-grain_Wheat	K	90	3	0	170	3.0	18.0	2	90	2
33 Product_19	K	100	3	0	320	1.0	20.0	3	45	2
34 Raisin Bran	K	120	3	1	210	5.0	14.0	12	240	2
35 Rice_Krispies	K	110	2	0	290	0.0	22.0	3	35	2
36 Smacks	K	110	2	1	70	1.0	9.0	15	40	2
37 Special_K	K	110	6	0	230	1.0	16.0	3	55	2
38 Cap'n'Crunch	Q	120	1	2	220	0.0	12.0	12	35	3
39 Honey_Graham_Ohs	Q	120	1	2	220	1.0	12.0	11	45	3
40 Life	Q	100	4	2	150	2.0	12.0	6	95	3
41 Puffed_Rice	Q	50	1	0	0	0.0	13.0	0	15	3
42 Puffed_Wheat	Q	50	2	0	0	1.0	10.0	0	50	3
43 Quaker_Oatmeal	Q	100	5	2	0	2.7	1.0	1	110	3

Source: Data courtesy of Chad Dacus.

Table 11.10 Concho Water Snake Data

	Gender	Age	TailLength	Snto VnLength		Gender	Age	TailLength	Snto VnLength
1	Female	2	127	441	1	Male	2	126	457
2	Female	2	171	455	2	Male	2	128	466
3	Female	2	171	462	3	Male	2	151	466
4	Female	2	164	446	4	Male	2	115	361
5	Female	2	165	463	5	Male	2	138	473
6	Female	2	127	393	6	Male	2	145	477
7	Female	2	162	451	7	Male	3	145	507
8	Female	2	133	376	8	Male	3	145	493
9	Female	2	173	475	9	Male	3	158	558
10	Female	2	145	398	10	Male	3	152	495
11	Female	2	154	435	11	Male	3	159	521
12	Female	3	165	491	12	Male	3	138	487
13	Female	3	178	485	13	Male	3	166	565
14	Female	3	169	477	14	Male	3	168	585
15	Female	3	186	530	15	Male	3	160	550
16	Female	3	170	478	16	Male	4	181	652
17	Female	3	182	511	17	Male	4	185	587
18	Female	3	172	475	18	Male	4	172	606
19	Female	3	182	487	19	Male	4	180	591
20	Female	3	172	454	20	Male	4	205	683
21	Female	3	183	502	21	Male	4	175	625
22	Female	3	170	483	22	Male	4	182	612
23	Female	3	171	477	23	Male	4	185	618
24	Female	3	181	493	24	Male	4	181	613
25	Female	3	167	490	25	Male	4	167	600
26	Female	3	175	493	26	Male	4	167	602
27	Female	3	139	477	27	Male	4	160	596
28	Female	3	183	501	28	Male	4	165	611
29	Female	4	198	537	29	Male	4	173	603
30	Female	4	190	566					
31	Female	4	192	569					
32	Female	4	211	574					
33	Female	4	206	570					
34	Female	4	206	573					
35	Female	4	165	531					
36	Female	4	189	528					
37	Female	4	195	536					

Source: Data courtesy of Raymond J. Carroll.

(a) Plot the data as a scatter plot with tail length (x_3) as the horizontal axis and snout to vent length (x_4) as the vertical axis. Use different plotting symbols for female and male snakes, and different symbols for different ages. Does it appear as if tail length and snout to vent length might usefully discriminate the genders of snakes? The different ages of snakes?

(b) Assuming multivariate normal data with a common covariance matrix, equal priors, and equal costs, classify the Concho Water Snakes according to gender. Compute the estimated $E(\text{AER})$ using the holdout procedure.

(c) Repeat part (b) using age as the groups rather than gender.

(d) Repeat part (b) using only snout to vent length to classify the snakes according to age. Compare the results with those in part (c). Can effective classification be achieved with only a single variable in this case? Explain.

11.36. Refer to Example 11.17. Using logistic regression, refit the salmon data in Table 11.2 with only the covariates freshwater growth and marine growth. Check for the significance of the model and the significance of each individual covariate. Set $\alpha = .05$. Use the fitted function to classify each of the observations in Table 11.2 as Alaskan salmon or Canadian salmon using rule (11-77). Compute the apparent error rate, APER, and compare this error rate with the error rate from the linear classification function discussed in Example 11.8.

References

1. Anderson, E. "The Irises of the Gaspé Peninsula." *Bulletin of the American Iris Society*, **59** (1939), 2–5.
2. Anderson, T. W. *An Introduction to Multivariate Statistical Analysis* (3rd ed.). New York: John Wiley, 2003.
3. Bartlett, M. S. "An Inverse Matrix Adjustment Arising in Discriminant Analysis." *Annals of Mathematical Statistics*, **22** (1951), 107–111.
4. Bouma, B. N., et al. "Evaluation of the Detection Rate of Hemophilia Carriers." *Statistical Methods for Clinical Decision Making*, **7**, no. 2 (1975), 339–350.
5. Breiman, L., J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees*. Belmont, CA: Wadsworth, Inc., 1984.
6. Darroch, J. N., and J. E. Mosimann. "Canonical and Principal Components of Shape." *Biometrika*, **72**, no. 1 (1985), 241–252.
7. Efron, B. "The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis." *Journal of the American Statistical Association*, **81** (1975), 321–327.
8. Eisenbeis, R. A. "Pitfalls in the Application of Discriminant Analysis in Business, Finance and Economics." *Journal of Finance*, **32**, no. 3 (1977), 875–900.
9. Fisher, R. A. "The Use of Multiple Measurements in Taxonomic Problems." *Annals of Eugenics*, **7** (1936), 179–188.
10. Fisher, R. A. "The Statistical Utilization of Multiple Measurements." *Annals of Eugenics*, **8** (1938), 376–386.
11. Ganesalingam, S. "Classification and Mixture Approaches to Clustering via Maximum Likelihood." *Applied Statistics*, **38**, no. 3 (1989), 455–466.
12. Geisser, S. "Discrimination, Allocatory and Separatory, Linear Aspects." In *Classification and Clustering*, edited by J. Van Ryzin, pp. 301–330. New York: Academic Press, 1977.
13. Gerrild, P. M., and R. J. Lantz. "Chemical Analysis of 75 Crude Oil Samples from Pliocene Sand Units, Elk Hills Oil Field, California." *U.S. Geological Survey Open-File Report*, 1969.
14. Gnanadesikan, R. *Methods for Statistical Data Analysis of Multivariate Observations* (2nd ed.). New York: Wiley-Interscience, 1997.
15. Habbema, J. D. F., J. Hermans, and K. Van Den Broek. "A Stepwise Discriminant Analysis Program Using Density Estimation." In *Compstat 1974, Proc. Computational Statistics*, pp. 101–110. Vienna: Physica, 1974.

16. Hills, M. "Allocation Rules and Their Error Rates." *Journal of the Royal Statistical Society (B)*, **28** (1966), 1–31.
17. Hosmer, D. W. and S. Lemeshow. *Applied Logistic Regression* (2nd ed.). New York: Wiley-Interscience, 2000.
18. Hudlet, R., and R. A. Johnson. "Linear Discrimination and Some Further Results on Best Lower Dimensional Representations." In *Classification and Clustering*, edited by J. Van Ryzin, pp. 371–394. New York: Academic Press, 1977.
19. Johnson, W. "The Detection of Influential Observations for Allocation, Separation, and the Determination of Probabilities in a Bayesian Framework." *Journal of Business and Economic Statistics*, **5**, no. 3 (1987), 369–381.
20. Kendall, M. G. *Multivariate Analysis*. New York: Hafner Press, 1975.
21. Kim, H. and Loh, W. Y., "Classification Trees with Unbiased Multiway Splits," *Journal of the American Statistical Association*, **96**, (2001), 589–604.
22. Krzanowski, W. J. "The Performance of Fisher's Linear Discriminant Function under Non-Optimal Conditions." *Technometrics*, **19**, no. 2 (1977), 191–200.
23. Lachenbruch, P. A. *Discriminant Analysis*. New York: Hafner Press, 1975.
24. Lachenbruch, P. A., and M. R. Mickey. "Estimation of Error Rates in Discriminant Analysis." *Technometrics*, **10**, no. 1 (1968), 1–11.
25. Loh, W. Y. and Shih, Y. S., "Split Selection Methods for Classification Trees," *Statistica Sinica*, **7**, (1997), 815–840.
26. McCullagh, P., and J. A. Nelder. *Generalized Linear Models* (2nd ed.). London: Chapman and Hall, 1989.
27. Mucciardi, A. N., and E. E. Gose. "A Comparison of Seven Techniques for Choosing Subsets of Pattern Recognition Properties." *IEEE Trans. Computers*, **C20** (1971), 1023–1031.
28. Murray, G. D. "A Cautionary Note on Selection of Variables in Discriminant Analysis." *Applied Statistics*, **26**, no. 3 (1977), 246–250.
29. Rencher, A. C. "Interpretation of Canonical Discriminant Functions, Canonical Variates and Principal Components." *The American Statistician*, **46** (1992), 217–225.
30. Stern, H. S. "Neural Networks in Applied Statistics." *Technometrics*, **38**, (1996), 205–214.
31. Wald, A. "On a Statistical Problem Arising in the Classification of an Individual into One of Two Groups." *Annals of Mathematical Statistics*, **15** (1944), 145–162.
32. Welch, B. L. "Note on Discriminant Functions." *Biometrika*, **31** (1939), 218–220.