

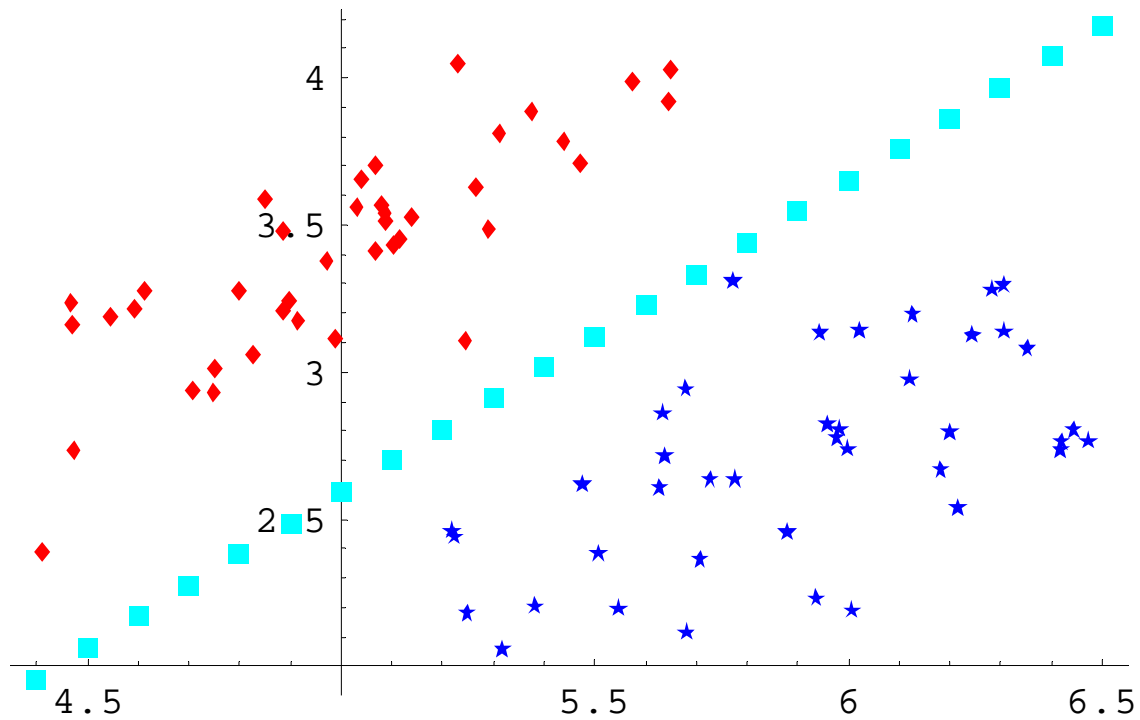
# Discriminant Analysis

In discriminant analysis, we try to find functions of the data that optimally discriminate between two or more groups. Discriminant analysis is, in a sense, MANOVA in reverse. In MANOVA, we ask whether two or more groups differ on two or more variables, and try to predict scores on the dependent variables from a knowledge of group membership. In discriminant analysis, we try to predict group membership from the data.

## ***A Caveat***

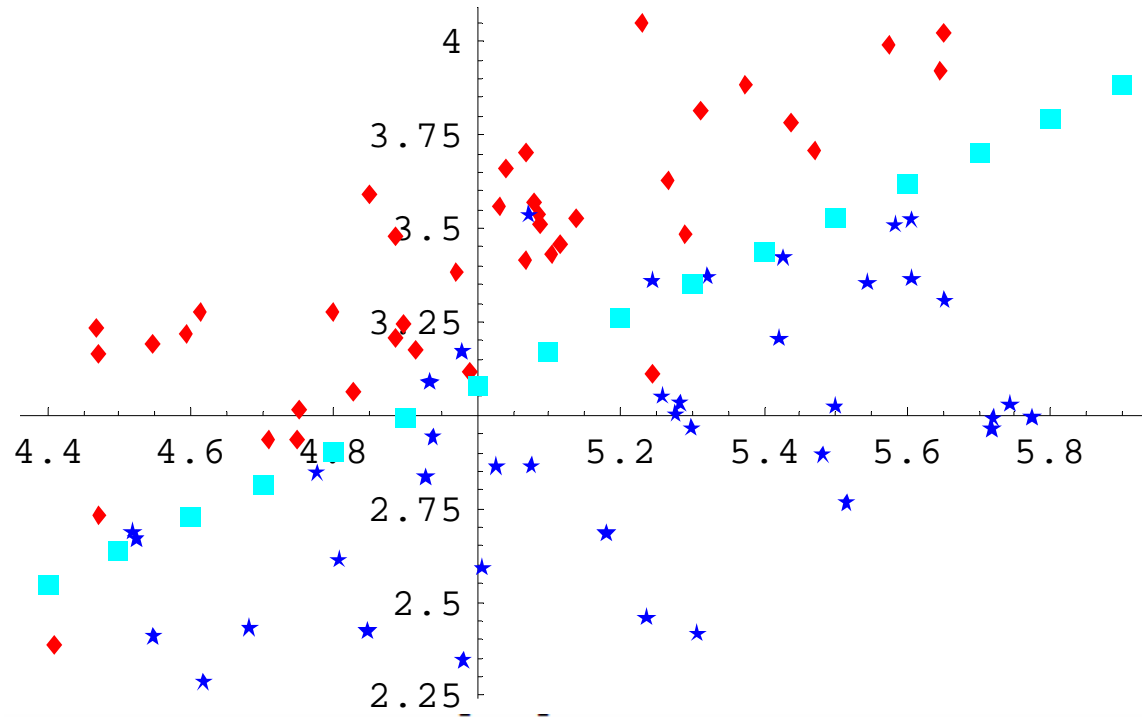
There are a number of different ways of arriving at formulae that produce essentially the same result in discriminant analysis. Consequently, different computer programs or books may give different formulae that yield different numerical values for some quantities. This can be very confusing.

# Linear Discriminant Function – Two Group Case



$$\mu_1 = \{5.006, 3.428\}; \mu_2 = \{5.936, 2.770\};$$

$$\Sigma_1 = \begin{pmatrix} 0.1218 & 0.0972 \\ 0.0972 & 0.1408 \end{pmatrix};$$



$$\mu_1 = \{5.006, 3.428\}; \mu_2 = \{5.236, 2.997\};$$

$$\Sigma_1 = \begin{pmatrix} 0.1218 & 0.0972 \\ 0.0972 & 0.1408 \end{pmatrix};$$

The *linear discriminant function* was proposed by Fisher (1936). Suppose we have  $N_1$  independent observations from population 1 and  $N_2$  independent observations from population 2, and we have recorded  $p$  measurements. The sample mean vectors are  $\bar{\mathbf{x}}_1$  and  $\bar{\mathbf{x}}_2$ , and the grand mean is

$$\bar{\mathbf{x}} = \frac{N_1 \bar{\mathbf{x}}_1 + N_2 \bar{\mathbf{x}}_2}{N_1 + N_2} \quad (1)$$

Following Morrison (1983), suppose we indicate group membership with the dummy variable

$$y_i = \begin{cases} \frac{N_2}{N_1 + N_2}, (\text{group 1}) \\ \frac{-N_1}{N_1 + N_2}, (\text{group 2}) \end{cases} \quad (2)$$

One may easily show (Morrison, 1983, p. 258) that the vector of estimated regression coefficients for predicting the  $y$  scores from the  $\mathbf{x}$  variates is

$$\hat{\boldsymbol{\beta}} = c\mathbf{A}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \quad (3)$$

where

$$c = \frac{N_1 N_2 / (N_1 + N_2)}{1 + [N_1 N_2 / (N_1 + N_2)](\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{A}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)} \quad (4)$$

The predicted  $y$  scores are

$$\hat{y}_i = \hat{\boldsymbol{\beta}}'(\mathbf{x}_i - \bar{\mathbf{x}}) \quad (5)$$

We can use the regression formula (5) to classify scores, i.e., attempt to categorize them into groups. A score  $\hat{y}_i$  classified as being in the group whose predicted score mean is closest to it. Since the group means are

$$\hat{\beta}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}) \text{ and } \hat{\beta}'(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}) \quad (6)$$

the midpoint, or cutpoint, is

$$\frac{\hat{\beta}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}) + \hat{\beta}'(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}})}{2} = \hat{\beta}'\left(\frac{\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2}{2} - \bar{\mathbf{x}}\right) \quad (7)$$

Recall that group 1 is associated with positive scores and group 2 negative scores. Consequently, if a predicted score  $\hat{y}_i$  is *above* the cutpoint in Equation (7), it is classified in group 1, otherwise in group 2. That is, a score is classified in group 1 if

$$\hat{\beta}'(\mathbf{x}_i - \bar{\mathbf{x}}) > \hat{\beta}'\left(\frac{\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2}{2} - \bar{\mathbf{x}}\right) \quad (8)$$

or

$$\hat{\boldsymbol{\beta}}' \mathbf{x}_i > \hat{\boldsymbol{\beta}}' \left( \frac{\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2}{2} \right) \quad (9)$$

Notice that the regression coefficients can all be multiplied by a common constant  $c$  without affecting the inequality. Moreover, the pooled estimate  $\mathbf{S}$  of the common covariance matrix can be calculated as

$$\mathbf{S} = \frac{1}{N_1 + N_2 - 2} \mathbf{A} \quad (10)$$

so  $\hat{\boldsymbol{\beta}}$  in Equation (9) can be replaced by  $\mathbf{A}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$  or  $\mathbf{a} = \mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ , since either substitution involves eliminating a multiplicative constant. With that substitution, we get

$$w = \mathbf{a}'\mathbf{x} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1}\mathbf{x} \quad (11)$$

which is known as the *linear discriminant function*. The cutoff point is halfway between the averages of  $w$ , or at

$$\frac{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} = \mathbf{a}' \frac{\mathbf{S}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \quad (12)$$

So effectively, the classification rule becomes  
assign to population 1 if

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \mathbf{x} - \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) > 0 \quad (13)$$

and assign to population 2 otherwise.

Of course, we *could* generate a different discriminant function for each group and use a different decision rule: assign a subject to the group whose function value is higher. Equation (13) can be broken down into two formulae,

$$f_1 = \bar{\mathbf{x}}_1' \mathbf{S}^{-1} \mathbf{x} - \frac{1}{2} \bar{\mathbf{x}}_1' \mathbf{S}^{-1} \bar{\mathbf{x}}_1 = \mathbf{b}_1' \mathbf{x} + a_1 \quad (14)$$

and

$$f_2 = \bar{\mathbf{x}}_2' \mathbf{S}^{-1} \mathbf{x} - \frac{1}{2} \bar{\mathbf{x}}_2' \mathbf{S}^{-1} \bar{\mathbf{x}}_2 = \mathbf{b}_2' \mathbf{x} + a_2 \quad (15)$$

with, for example,

$$\mathbf{b}'_1 = \bar{\mathbf{x}}'_1 \mathbf{S}^{-1} \quad (16)$$

and

$$a_1 = -\frac{1}{2} \bar{\mathbf{x}}'_1 \mathbf{S}^{-1} \bar{\mathbf{x}}_1 = -\frac{1}{2} \mathbf{b}'_1 \bar{\mathbf{x}}_1 \quad (17)$$

Equations (14)–(17) yield the “Fisher discriminant function” weights and constant printed by SPSS, *except for one additional element*. If the groups have a different prior likelihood of occurrence, the

above function values will lead to a substantial amount of classification error. This can be corrected by incorporating the probabilities  $p_j$  of being in group  $j$  by using the following formula

$$a_j^* = a_j + \ln(p_j) \quad (18)$$

This constant is used along with

$$\mathbf{b}'_j = \bar{\mathbf{x}}_j' \mathbf{S}^{-1} \quad (19)$$

to generate the scores for group  $j$ .

The individual is classified into the group whose score is highest. In practice, prior probabilities are often not known, in which case the estimates

$$\hat{p}_j = \frac{N_j}{N_\bullet} \quad (20)$$

are often employed as a default.

*Example.* Morrison (1990, page 143) gives data for 49 subjects, 12 diagnosed with “senile factor present” and 37 diagnosed with “no senile factor.” The data are available online in the file *morrisonEx43.sav*. The Wechsler Adult Intelligence scale was administered to all subjects by independent investigators, and scores for 4 subtests (*Information, Similarities, Arithmetic, Picture Completion*) recorded. So the data set consists of 49 observations on 5 variables.

This data set is analyzed several times in Morrison's text. In this case, we will examine a standard 2-group linear discriminant analysis the way Morrison reports it, and the way SPSS reports it.

Morrison computes the linear discriminant function using Equation (11), and, for each subject, compares the computed function to the cutoff value in Equation (12).

In this case,  $\mathbf{S}^{-1}$  is given by

```
In[12]:= Inverse[S] // MatrixForm
```

```
Out[12]//MatrixForm=
```

$$\begin{pmatrix} 0.259832 & -0.136477 & -0.0586981 & -0.0658857 \\ -0.136477 & 0.186734 & -0.0383237 & 0.0162299 \\ -0.0586981 & -0.0383237 & 0.150966 & -0.0173444 \\ -0.0658857 & 0.0162299 & -0.0173444 & 0.211142 \end{pmatrix}$$

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' = [3.81757 \quad 4.23423 \quad 2.98649 \quad 3.22297]$$

So the discriminant function is

$$y = .0264x_1 + .2075x_2 + .0086x_3 + .4459x_4 \quad (21)$$

and the cutoff point is 4.750.

SPSS reports coefficients (“Unstandardized Canonical Coefficients”) that are proportional to those in Equation (11). These are divided by the standard deviation of the predicted scores, i.e.,

$$\mathbf{a}^* = \left( \mathbf{a}'\mathbf{S}^{-1}\mathbf{a} \right)^{-1/2} \mathbf{a} \quad (22)$$

Note that variables  $x_1$  and  $x_3$  do not appear to have much influence in discriminating between the senile and non-senile group.

Incidentally, one important outcome of the analysis is the *classification matrix*, which shows the result of applying the discriminant function classification rule.

Using all 4 variables, we get the following:

**Classification Results<sup>a</sup>**

			Predicted Group Membership		Total
			0	1	
Original	Count	SENILE			
		0	29	8	37
		1	4	8	12
	%	0	78.4	21.6	100.0
		1	33.3	66.7	100.0

a. 75.5% of original grouped cases correctly classified.

In this case, the misclassification rates are rather high. Moreover, these classification rates are probably unduly optimistic. We can improve things.

But first, let's perform the analysis using the more general approach employed by SPSS.

SPSS can report a linear discriminant function for each group, as in Equations (14)–(15).

### Classification Function Coefficients

	SENILE	
	0	1
INFO	.760	.734
SIMILAR	-.239	-.447
ARITH	.491	.483
PICCOMPL	.811	.366
(Constant)	-10.382	-5.632

Fisher's linear discriminant functions

To perform classification, you compute the two functions, and assign an individual to the group with the higher score.

Now, if we drop the two non-contributing variables and redo the analysis, we get

**Classification Results<sup>a</sup>**

			Predicted Group Membership		Total
			0	1	
Original	Count	SENILE			
		0	29	8	37
		1	4	8	12
	%	0	78.4	21.6	100.0
		1	33.3	66.7	100.0

a. 75.5% of original grouped cases correctly classified.

Exactly the same as before.

However, we have not yet employed the correction for prior probabilities. If we do that, we get

**Classification Results<sup>a</sup>**

			Predicted Group Membership		Total
			0	1	
Original	Count	SENILE			
		0	37	0	37
		1	6	6	12
	%	0	100.0	.0	100.0
		1	50.0	50.0	100.0

a. 87.8% of original grouped cases correctly classified.