

## 12 Two-Sample Discriminant Analysis

### 12.1 Introduction

In the simplest type of discrimination problem, we wish to allocate an individual or sampling unit to one of two populations; the allocation to be based on a set of measurements associated with the individual. For example, a doctor may take a series of measurements on a patient, on the basis of which a diagnosis will be made. This is a decision problem, and, as with most decisions, we may wish to take account of prior probabilities and the costs associated with classification errors.

### 12.2 Problem Formulation and Decision Rule

Suppose that we have two populations, population 1 and population 2. Let the distributional behaviour of observations from these two populations be characterized by their probability density functions,  $f_1(\mathbf{x})$  and  $f_2(\mathbf{x})$ , respectively, and let  $\pi_i$  be the (prior) probability that a randomly selected individual, from the ‘population at large’, belongs to population  $i$ ,  $i = 1, 2$ .

Thus, for the population at large, we have the joint random vector  $(C, \mathbf{X})^T$ , where the random variable  $C$  denotes the class label of an observation and  $\mathbf{X}$  is the associated random vector of features. Then  $\{\pi_1, \pi_2\}$  is the marginal p.m.f for  $C$ , and the  $f_i(\mathbf{x})$  are the densities for the *conditional* random vectors  $\mathbf{X}|C = i$ . Clearly the joint p.d.f. is given by

$$f(i, \mathbf{x}) = \pi_i f_i(\mathbf{x}).$$

The classification task is to estimate the true value of  $C$ , having observed  $\mathbf{X} = \mathbf{x}$ .

Suppose  $c$  is a classification procedure or classifier mapping the domain of  $\mathbf{X}$  into the set  $\{1, 2\}$ , and let  $C(i|j)$  be the cost of misallocating an individual from population  $j$  to population  $i$ . Clearly  $C(i|i) = 0$ .

If we define  $p(i|\mathbf{x})$  to be the posterior probability of population  $i$  given  $\mathbf{X} = \mathbf{x}$ , then

$$\begin{aligned} p(i|\mathbf{x}) &\equiv \mathbb{P}(C = i | \mathbf{X} = \mathbf{x}) \\ &= \frac{\pi_i f_i(\mathbf{x})}{\pi_1 f_1(\mathbf{x}) + \pi_2 f_2(\mathbf{x})}. \end{aligned} \tag{1}$$

**Theorem 12.1** The classification rule that minimizes total expected misclassification cost is the rule that sets  $c(\mathbf{x}) = i$  where  $i$  minimizes

$$\sum_{j=1}^2 C(i|j) p(j|\mathbf{x}) \quad \text{with respect to } i = 1, 2.$$

That is, the rule sets  $c(\mathbf{x}) = 1$  if  $C(1|2)p(2|\mathbf{x}) \leq C(2|1)p(1|\mathbf{x})$ , and  $c(\mathbf{x}) = 2$  otherwise.

This is known as Bayes rule, and the corresponding total expected misclassification cost is known as the Bayes risk.

**Proof:** See, for example, Ripley (1996) where this is proved for the general case of  $K$  classes.

Using Theorem 12.1, and the expression for  $p(i|\mathbf{x})$  given in equation (1), Bayes rule becomes:

Allocate the individual with observation vector  $\mathbf{x}$  to population 1 if

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{\pi_2 C(1|2)}{\pi_1 C(2|1)}. \quad (2)$$

Otherwise allocate to population 2.

### Remarks 12.2

(i) If  $C(1|2) = C(2|1)$ , then (2) becomes

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{\pi_2}{\pi_1},$$

or  $\pi_1 f_1(\mathbf{x}) \geq \pi_2 f_2(\mathbf{x})$ , i.e. allocate to the population with the greatest *posterior* probability.

(ii) If  $C(1|2) = C(2|1)$  and  $\pi_1 = \pi_2$  then (2) becomes

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq 1$$

or  $f_1(\mathbf{x}) \geq f_2(\mathbf{x})$ , i.e. allocate to the population with the greatest *likelihood*.

(iii) When  $\pi_2 \neq \pi_1$  and  $C(1|2) \neq C(2|1)$  then the decision takes into account:

- the ratio of the misallocation costs;
- the ratio of the prior probabilities.

### Example 12.3

A healthcare centre is trying to screen patients in a certain area for a particular disease.

It is widely believed that about 2% of the population have the disease.

Not detecting the disease is considered to be ten times more serious than initially classifying a healthy individual diseased.

Formulate an appropriate allocation rule, based on  $\mathbf{x} \in \mathbb{R}^p$  taken on an individual, and the above information.

*Solution:* Let  $f_H(\cdot)$  be the p.d.f. associated with the ‘healthy’ population, and  $f_D(\cdot)$  be that associated with the ‘diseased’.

Then the decision rule is:

$$\text{If } f_H(\mathbf{x})f_D(\mathbf{x}) \geq \frac{0.02}{0.98} \times \frac{10}{1} \approx 0.2,$$

then the individual with readings given by  $\mathbf{x}$  is classified as healthy. The individual is classified as unhealthy if

$$\frac{f_H(\mathbf{x})}{f_D(\mathbf{x})} < 0.2 \text{mbxi.e.} \frac{f_D(\mathbf{x})}{f_H(\mathbf{x})} > 5.$$

Thus, we do not recall the individual for further tests unless the *likelihood ratio* in favour of allocation to the diseased population is more than 5 to 1.

### 12.3 Form of the Decision Rule under Multivariate Normality

Let us now suppose that observation vectors from the  $i$ -th population have the  $N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$  distribution,  $i=1, 2$ . (So the two populations share the same covariance matrix). Then

$$f_i(\mathbf{x}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right\} \quad (4)$$

Hence

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) + \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_2) \right\} \quad (5)$$

i.e.

$$\ln \left( \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \right) = -\frac{1}{2} [(\mathbf{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) - (\mathbf{x} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_2)] . \quad (6)$$

Note that (6) is equivalent to

$$\ln \left( \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \right) = -\frac{1}{2} [D^2(\mathbf{x}, \boldsymbol{\mu}_1) - D^2(\mathbf{x}, \boldsymbol{\mu}_2)] ,$$

so that if prior probabilities are equal and misallocation costs are also equal, the observation  $\mathbf{x}$  is allocated to population 1 if  $D^2(\mathbf{x}, \boldsymbol{\mu}_1) \leq D^2(\mathbf{x}, \boldsymbol{\mu}_2)$  and to population 2 otherwise.

The terms in the square brackets of (6) can be expanded and shown to be equal to

$$\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} - 2\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x} + 2\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 - \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2$$

which upon simplification gives

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

i.e.

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

Plugging this expression back into (6) yields:

$$\ln \left( \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \right) = \mathbf{x}^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).$$

Setting  $\mathbf{L} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$  in the foregoing expression gives

$$\begin{aligned}\ln\left(\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}\right) &= \mathbf{x}^T \mathbf{L} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)^T \mathbf{L} \\ &= \mathbf{L}^T \mathbf{x} - \frac{1}{2} \mathbf{L}^T (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2).\end{aligned}$$

Now (2) holds if, and only if,

$$\ln\left(\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}\right) \geq \ln\left(\frac{\pi_2 C(1|2)}{\pi_1 C(2|1)}\right)$$

i.e.

$$\mathbf{L}^T \mathbf{x} - \frac{1}{2} \mathbf{L}^T (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \geq k, \quad (7)$$

where

$$k = \ln\left(\frac{\pi_2 C(1|2)}{\pi_1 C(2|1)}\right).$$

#### Remarks 12.4

(i) The allocation rule is:

$$\begin{cases} \mathbf{L}^T \mathbf{x} - \frac{1}{2} \mathbf{L}^T (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \geq k & \text{allocate to population 1} \\ \text{otherwise} & \text{allocate to population 2.} \end{cases}$$

(ii)  $\mathbf{L}^T \mathbf{x}$  is FISHER'S LINEAR DISCRIMINANT FUNCTION. The unknown parameters in the L.H.S. above can be replaced by appropriate estimators from the sample, yielding the SAMPLE LINEAR DISCRIMINANT FUNCTION and the corresponding classification criterion. (See later.)

(iii)

$$k = 0 \Leftrightarrow \pi_2 C(1|2) = \pi_1 C(2|1).$$

## 12.4 Misclassification Probabilities

We can quantify the probability of *misclassification* in certain cases.

Without loss of generality, suppose  $\mathbf{L}^T \boldsymbol{\mu}_1 > \mathbf{L}^T \boldsymbol{\mu}_2$ , and define

$$\alpha = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \mathbf{L}^T (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad (= D^2(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)).$$

**Note:** That is,  $\alpha$  is a measure of the distance between the two population means - the squared Mahalanobis distance.

Also, define the scalar random variable

$$U_1 = \mathbf{L}^T \mathbf{X} - \frac{1}{2} \mathbf{L}^T (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \quad \text{where } \mathbf{X} \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}).$$

Then

$$\begin{aligned} E[U_1] &= \mathbf{L}^T \boldsymbol{\mu}_1 - \frac{1}{2} \mathbf{L}^T (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \\ &= \frac{1}{2} \mathbf{L}^T (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \frac{\alpha}{2} \end{aligned}$$

and

$$\begin{aligned} \text{var}(U_1) &= \text{var} \left( \mathbf{L}^T \mathbf{X} - \frac{1}{2} \mathbf{L}^T (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \right) \\ &= \text{var}(\mathbf{L}^T \mathbf{X}) \\ &= \mathbf{L}^T \boldsymbol{\Sigma} \mathbf{L} \\ &= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\ &= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \alpha \end{aligned}$$

Since  $\mathbf{X}$  has a MVN distribution, then the linear compound  $\mathbf{L}^T \mathbf{X}$  is Normally distributed, as is  $U_1$ . Hence  $U_1 \sim N(\frac{\alpha}{2}, \alpha)$ .

Similarly, define

$$U_2 = \mathbf{L}^T \mathbf{X} - \frac{1}{2} \mathbf{L}^T (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \quad \text{where } \mathbf{X} \sim N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}).$$

Then

$$E[U_2] = \mathbf{L}^T \boldsymbol{\mu}_2 - \frac{1}{2} \mathbf{L}^T (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) = -\frac{\alpha}{2}$$

and

$$\text{var}(U_2) = \text{var}(\mathbf{L}^T \mathbf{X}) = \mathbf{L}^T \boldsymbol{\Sigma} \mathbf{L} = \alpha.$$

Thus  $U_2 \sim N(-\frac{\alpha}{2}, \alpha)$ .

Let  $p(i|j)$  be the probability that an individual from population  $j$  is misallocated to population  $i$ . Then

$$p(1|2) = P(U_2 > k) \quad \text{and} \quad p(2|1) = P(U_1 < k).$$

This allows us to produce the following table.

		True Class	
		1	2
Predicted Class	1	$p(1 1)$	$p(1 2)$
	2	$p(2 1)$	$p(2 2)$
Associated prior probabilities		$\pi_1$	$\pi_2$

and if we vary  $k$  we can see how the two misclassification probabilities vary with it.

Overall, the total probability of misclassification is

$$\pi_1 p(2|1) + \pi_2 p(1|2)$$

i.e.

$$\pi_1 P(U_1 < k) + \pi_2 P(U_2 > k).$$

When  $k = 0$ , this becomes

$$\begin{aligned} \pi_1 P(U_1 < 0) + \pi_2 P(U_2 > 0) &= \pi_1 P(U_1 < 0) + \pi_2 P(U_1 < 0) \\ &= (\pi_1 + \pi_2) P(U_1 < 0) = P(U_1 < 0) \\ &= P\left(\frac{U_1 - \frac{\alpha}{2}}{\sqrt{\alpha}} < \frac{-\alpha/2}{\sqrt{\alpha}}\right) = \Phi\left(\frac{-\sqrt{\alpha}}{2}\right) \end{aligned}$$

### Remarks 12.5

(i) Where the population parameters are unknown, the probability of misclassification may be estimated from the data, so that a natural estimate of  $\alpha$  is

$$D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^T \mathbf{S}_U^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2),$$

the squared Mahalanobis distance between the samples. In this case the estimated total probability of misclassification is  $\Phi(-\frac{1}{2}D)$ . In reality, this approach tends to be optimistic, underestimating the ‘true’ misclassification probabilities, when the sample size is small.

(ii) As before, notice the relationship between  $D^2$  and the two-sample Hotelling  $T^2$  statistic.

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} D^2 \sim T_p^2(n_1 + n_2 - 2)$$

## 12.5 Discriminant Analysis using R

We shall assume multivariate normality. In practice, however, we may know nothing about the key parameters of the two distributions, in which case, these have to be estimated from the available data. Thus a point estimate for  $\boldsymbol{\mu}_i$  would be  $\bar{\mathbf{x}}_i$ ; and  $\boldsymbol{\Sigma}$  could be estimated by  $\mathbf{S}_U$ , the unbiased pooled ‘within-groups’ sample covariance matrix. Invoking such estimates leads to an allocation rule of the form:

$$\begin{cases} \hat{\mathbf{L}}^T \mathbf{x} - \frac{1}{2} \hat{\mathbf{L}}^T (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \geq k & \text{allocate to population 1} \\ \text{otherwise} & \text{allocate to population 2} \end{cases}$$

where  $\hat{\mathbf{L}} = \mathbf{S}_U^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ .

Consider the following example.

### Example 12.6 (Sheep)

Five diagnostic tests were carried out on two groups of sheep: those suffering from Scrapie, and the other from a more serious disease (possibly ovine CJD). The measurements arising from the tests for each sheep are summarized in the tables overleaf.

*Sheep with Scrapie:*

	T1	T2	T3	T4	T5
1	11	18	15	18	15
2	33	27	31	21	17
3	20	28	27	23	19
4	18	26	18	18	9
5	22	23	22	16	10

*Sheep with more serious disease:*

	T1	T2	T3	T4	T5
1	18	17	20	18	18
2	31	24	31	26	20
3	14	16	17	20	17
4	25	24	31	26	18
5	36	28	24	26	29

Can we construct a rule that would allow us to distinguish between the two diseases on the basis of the results of the tests (T1-T5)?

It is thought that about 50% of the sheep that show symptoms of either disease are actually suffering from scrapie. Also, the ratio of misallocation costs is gauged to be close to 1. Hence, taking  $k = 0$ , our discrimination procedure, calculated on the basis of the available data, can be found as follows:

```
> s.t1 <- c(11, 33, 20, 18, 22)
> s.t2 <- c(18, 27, 28, 26, 23)
> s.t3 <- c(15, 31, 27, 18, 22)
> s.t4 <- c(18, 21, 23, 18, 16)
> s.t5 <- c(15, 17, 19, 9, 10)
> u.t1 <- c(18, 31, 14, 25, 36)
> u.t2 <- c(17, 24, 16, 24, 28)
> u.t3 <- c(20, 31, 17, 31, 24)
> u.t4 <- c(18, 26, 20, 26, 26)
> u.t5 <- c(18, 20, 17, 18, 29)
> s.dat <- data.frame(s.t1, s.t2, s.t3, s.t4, s.t5)
> u.dat <- data.frame(u.t1, u.t2, u.t3, u.t4, u.t5)

> s.dat
  s.t1 s.t2 s.t3 s.t4 s.t5
1   11   18   15   18   15
2   33   27   31   21   17
3   20   28   27   23   19
4   18   26   18   18    9
5   22   23   22   16   10

> u.dat
  u.t1 u.t2 u.t3 u.t4 u.t5
1   18   17   20   18   18
2   31   24   31   26   20
3   14   16   17   20   17
4   25   24   31   26   18
5   36   28   24   26   29
```

```

> no.s <- 5
> no.u <- 5
> s.pooled <- ((no.s - 1) * var(s.dat) + (no.u - 1) * var(u.dat))/(no.s + no.u - 2)
> dimnames(s.pooled) <- NULL

> s.pooled
      [,1] [,2] [,3] [,4] [,5]
[1,] 72.700 33.025 41.65 18.675 22.300
[2,] 33.025 21.250 21.30 12.725 11.925
[3,] 41.650 21.300 41.30 16.350  9.850
[4,] 18.675 12.725 16.35 11.450 10.200
[5,] 22.300 11.925  9.85 10.200 21.650

> sm.s <- apply(s.dat, 2, mean)
> sm.u <- apply(u.dat, 2, mean)
> names(sm.s) <- NULL
> names(sm.u) <- NULL

> sm.s
[1] 20.8 24.4 22.6 19.2 14.0

> sm.u
[1] 24.8 21.8 24.6 23.2 20.4

> L <- solve(s.pooled) %*% (sm.s - sm.u)
> L
      [,1]
[1,] -0.7491324
[2,]  2.0307983
[3,]  0.5350933
[4,] -2.3422912
[5,]  0.2175097

```

Thus  $\hat{\mathbf{L}} = [-0.7491324, \dots, 0.2175097]^T$ , so that the L.D.F is given by

$$\hat{\mathbf{L}}^T \mathbf{x} = -0.749 T1 + 2.031 T2 + 0.535 T3 - 2.342 T4 + 0.218 T5,$$

and the allocation rule is

$$\begin{cases} \hat{\mathbf{L}}^T \mathbf{x} - \frac{1}{2} \hat{\mathbf{L}}^T (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \geq 0 & \text{allocate to population with scrapie} \\ \text{otherwise} & \text{allocate to population with more serious disease} \end{cases}$$

Let us estimate the misclassification probability in this example, noting that with these small samples this is likely to be an underestimate.

```

> alpha <- t(sm.s - sm.u) %*% solve(s.pooled) %*% (sm.s - sm.u)
> pmisclass <- pnorm(-0.5 * sqrt(alpha))
> pmisclass
      [,1]
[1,] 0.02568942

```



Thus we estimate that we will misdiagnose in about 257 cases in 10,000.

*Question: Would our rule correctly allocate the sheep (in the sample presented in the example) to the correct populations?*

```
> s.mat <- as.matrix.data.frame(s.dat)
> u.mat <- as.matrix.data.frame(u.dat)
> s.ldr <- t(L) %*% t(s.mat) - 0.5 * t(L) %*% (sm.s + sm.u) *rep(1, 5)
> u.ldr <- t(L) %*% t(u.mat) - 0.5 * t(L) %*% (sm.s + sm.u) *rep(1, 5)
> s.ldr
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] 0.8976995 4.663609 10.04319 12.20038 10.15392
> u.ldr
      [,1]      [,2]      [,3]      [,4]      [,5]
[1,] -3.04903 -10.98945 -8.590671 -6.929673 -8.399982
```

The components of `s.ldr` are all positive, and those of `u.ldr` all negative. Thus, the rule has correctly classified all of the sheep.

*Question: A sheep showing signs of scrapie is investigated by a veterinary surgeon who finds that observed values of tests T1 to T5 are 19, 13, 11, 14 and 13, respectively. Using an appropriate discrimination procedure, determine whether this single sheep should be confirmed to have scrapie, or considered to have the more serious disease.*

```
> obs <- c(19, 13, 11, 14, 13)
> drop(t(L) %*% obs - 0.5 * t(L) %*% (sm.s + sm.u))
[1] -8.455579
```

According to our rule, the negative value indicates that the single sheep, under consideration, should be classified as having the more serious disease.