# 8 How Do Bone Shapes Indicate Arthritis?

## 8.1 Introduction

In this chapter we return to the analysis of the bone shape data discussed in Chapter 4. The *intercondylar notch*, the inside of the inverted U-shape shown in Figure 1.5, is considered important by medical specialists. The anterior cruciate ligament runs through the intercondylar notch, and damage to this ligament is known to be a risk factor for osteoarthritis of the knee. Although other studies have examined large-scale features of the intercondylar notch, there has not been very much examination of its detailed shape, nor of its direct relationship to the incidence of osteoarthritis.

In Chapter 4 we studied the shape of the bone outline by considering a number of landmarks and interpolating between them. In this chapter we look much more closely at the shape of the intercondylar notch, by taking a more subtle approach to the detailed representation of the shapes.

We consider a set of 96 notch outlines, on each of which we have some concomitant information, such as the age of the individual and whether there is evidence of arthritic bone change. Our concentration on the notch alone allows us to include a number of partly damaged bones that could not be considered in Chapter 4; as long as any damage does not affect the notch it is no longer a problem. In the sample we consider there are 21 femora from arthritic individuals and 75 from individuals showing no signs of arthritic bone change. We use the data to demonstrate three aspects of functional data analysis.

Y	$X_{\rm L}$	$X_{\mathrm{M}}$	Y	$X_{\rm L}$	$X_{\rm M}$	Y	$X_{\rm L}$	$X_{\rm M}$	Y	$X_{\rm L}$	$X_{\rm M}$
80	59	61	92	49	83	104	45	87	116	41	85
81	54	61	93	48	84	105	45	87	117	40	85
82	52	63	94	48	85	106	45	87	118	40	85
83	52	64	95	48	86	107	45	86	119	39	86
84	52	66	96	47	87	108	44	86	120	37	87
85	52	68	97	47	87	109	44	86	121	36	87
86	52	73	98	47	87	110	43	86	122	35	88
87	51	77	99	46	87	111	43	86	123	35	89
88	51	78	100	46	87	112	44	86	124	33	90
89	51	79	101	46	87	113	43	86	125	30	91
90	50	80	102	46	87	114	42	86			
91	49	82	103	46	87	115	41	86			

Table 8.1. The coordinates of the lateral and medial edges of the intercondylar notch for one particular femur. The values Y give the pixel rows, numbered from top to bottom of the image. The values  $X_{\rm L}$  and  $X_{\rm M}$  give the lateral and medial positions within row Y of the edges of the intercondylar notch.

- 1. How do we handle curves and shapes without making use of landmarks?
- 2. What does principal components analysis tell us about the variability of these data?
- 3. What are the issues involved in developing a functional analogue of discriminant analysis?

Part of the object of the study is to understand the way in which arthritic and nonarthritic bones differ. We have information on this aspect in that some bones display eburnation, which is a consequence of arthritis. In this chapter we take eburnation to be synonymous with arthritis, but it could well be that some of the noneburnated bones are from individuals with arthritis that is mild or in its early stages. This means that any conclusions we reach about the differences between arthritic and nonarthritic bones are conservative.

# 8.2 Analyzing shapes without landmarks

The bone shapes are stored as  $128 \times 128$  pixel images, obtained by processing pictures such as that in Figure 1.5. The pixels are numbered from the top to the bottom of the picture in the vertical direction, and from the lateral to the medial (the outside to the inside) in the horizontal direction. To record the shape of the notch, we move row by row up the pixel image, starting with the first row of pixels that touches the notch, and for each



Figure 8.1. The raw data for the intercondylar notch for a particular individual. The lateral side is on the left and the medial on the right.

row find the pixel positions at either side of the notch. A specific example is given in Table 8.1 and plotted in Figure 8.1.

It can be seen from the figure that Y cannot be written as a simple function of X. For instance, as we move down the notch on the medial side, we first move out to pixel 87, then back to 85, and finally out to 91 again. Furthermore, a large part of this edge is vertical or nearly so. Merely considering Y as a function of X will not work; instead we will have to find a better way of parameterizing the shape of the notch.

A fruitful approach is parameterization by arc length. We define functions x(t) and y(t) such that as t increases from 0 to 1 the point  $\{x(t), y(t)\}$  moves at a constant speed along the curve. We then regard the two-dimensional function  $\mathbf{z}(t) = \{x(t), y(t)\}$  as being our functional datum. Landmarks are not required; instead the distance along the curve is used to yield the points whose coordinates are used for the subsequent analysis. Distance measured along a curve is called *arc length*.

To apply this approach to the data given in Table 8.1, first we connect the dots to obtain a continuous outline, as shown in Figure 8.2. In this figure, there is some rapid variation in the part of the curve on the lateral side, partly due to the pixelation of the image. We are not interested in variation on this scale. However, we wish to calculate distances along the curve in order to define the functions x(t) and y(t), and small scale variations will increase such arc lengths in a spurious way. Therefore we perform some



Figure 8.2. Joining the centers of the boundary pixels: the first step in producing a curve parameterized by arc length.



Figure 8.3. Smoothing by joining the midpoints of the line segments in Figure 8.2: the next step in producing a curve parameterized by arc length.



Figure 8.4. Fifty points equally spaced along the curve shown in Figure 8.3 with the curve rescaled to start and finish at standard positions. Interpolating these positions linearly is the final step in producing a curve parameterized by arc length.

very light smoothing, joining the midpoints between the dots, instead of the dots themselves. The effect, shown in Figure 8.3, is to reduce the local variability noticeably without changing the structure in any substantial way.

Only the shape of the notch is of interest, so we rescale the curve equally in both coordinate directions, and also shift it, to make it run from (0,0) to (1,0) in the X-Y plane. By calculating the distance along all the small line segments that make up Figure 8.3 we find 50 points  $(x_k, y_k)$  at equal arc length along the curve, as plotted in Figure 8.4. This process has yielded a fine grid of 50 points evenly spaced along the notch outline, capturing all the essential features of the shape of the intercondylar notch. Let  $t_1 =$  $0, t_2 = 1/49, t_3 = 2/49, \ldots, t_{50} = 1$ . To complete the specification of the shape as a curve parameterized by arc length, define the functions x(t) and y(t) by setting  $x(t_k) = x_k$  and  $y(t_k) = y_k$  for each k, and interpolating linearly between these points.

This process is applied to each of the N = 96 outlines in the sample. For each j = 1, 2, ..., N we obtain a pair of functions  $\{X_j(t), Y_j(t)\}$ , written as the vector function  $\mathbf{Z}_j(t)$ . Each  $X_j$  and  $Y_j$  is held in discretized form, so the actual data are held in an  $N \times 50 \times 2$  array, recording the coordinates of the 50 points picked out along each curve. The (j, k, 1) element of this array is the X-coordinate of the kth point on the *j*th curve, and the (j, k, 2) element the corresponding Y-coordinate. The choice of the number 50 is somewhat arbitrary, and our analyses are not particularly sensitive to this choice; because of the original pixelation of the data there is no point in trying to recover information on any smaller scale.

## 8.3 Investigating shape variation

#### 8.3.1 Looking at means alone

We can define the notion of a mean shape, by finding the functions

$$\bar{X}(t) = N^{-1} \sum_{i} X_i(t)$$
 and  $\bar{Y}(t) = N^{-1} \sum_{i} Y_i(t)$ ,

and letting the mean shape be the curve traced out by the two-dimensional function  $\overline{\mathbf{Z}}(t) = \{\overline{X}(t), \overline{Y}(t)\}$ . In practice, we average over the first dimension of the data array to yield a 50 × 2 matrix giving the coordinates of 50 points along the mean curve; joining these points gives the mean curve  $\overline{\mathbf{Z}}(t)$  plotted in Figure 8.5. The halfway point along this curve, for instance, is the average of all the halfway points on the individual curves.<sup>1</sup>

The means of the eburnated and noneburnated groups are plotted in Figure 8.6. It might appear that the distinguishing feature of the arthritic bones is that they have a shallower notch, because this is the way that the mean shapes differ. However, we show that a more careful statistical analysis does not yield the same conclusion, and that the mode of variability that best distinguishes the two groups is quite different.

#### 8.3.2 Principal components analysis

Before considering further the subdivision into arthritic and nonarthritic bones, we investigate the ways in which the data set as a whole varies. Regarding the two-dimensional functions  $\mathbf{Z}_i(t)$  as our functional data, functional PCA yields an expansion in terms of two-dimensional functions  $\boldsymbol{\xi}_j(t) = \{\xi_j^X(t), \xi_j^Y(t)\}$ . There are coefficients  $z_{ij}$  such that the observations can be expanded as

$$\mathbf{Z}_{i}(t) = \sum_{j \ge 1} z_{ij} \boldsymbol{\xi}_{j}(t).$$
(8.1)

<sup>&</sup>lt;sup>1</sup>There is an interesting wrinkle here that is not relevant to our particular application: the points along the mean curve need not actually be themselves equally spaced, and in some cases it may be a good idea to go back and reparameterize the individual curves by reference to the way that the mean curve turns out. In our case this is not a problem.



Figure 8.5. The mean notch shape curve.



Figure 8.6. Solid: the mean curve for arthritic bones; dashed: the mean curve for nonarthritic bones.



Figure 8.7. The first four principal components of variability of the notch shapes. The solid curves are the outlines corresponding to adding a multiple of the relevant weight function to the mean, and the dashed curves those obtained by subtracting the same multiple. The percentages of variability explained by these components are, respectively, 72.5, 13.9, 5.9, and 3.9%.

The actual PCA is performed by carrying out a standard PCA of the 100vectors giving the coordinates of the points along the curves. It turns out that no smoothing is necessary.

To understand the principal component weight functions  $\boldsymbol{\xi}_j(t)$ , we can, as usual, plot  $\overline{\mathbf{Z}}(t) \pm c\boldsymbol{\xi}_j(t)$  for some suitable multiple c. In this case the perturbed functions  $\overline{\mathbf{Z}}(t) \pm c\boldsymbol{\xi}_j(t)$  are two-dimensional functions, and we plot their path in X-Y space as t varies. In Figure 8.7 the effects of the first four principal components of variability are displayed. These components together explain 96% of the variability in the data, with no other component explaining more than about 1% of the variability.

The displayed components all have simple interpretations. The first component corresponds to the depth of the notch, and the second to the shift of the notch relative to the bottom points of the condyles. The third component gives information about the width of the notch, and the fourth shows how convex the medial part of the notch tends to be.

The depth of the notch accounts for a great deal of the variability in the sample, and so in plotting Figure 8.7 the size of the perturbation shown in

each part of the plot is not chosen by reference to the amount of variability in the original sample. Instead, the same multiple c of the principal component curve is used in each case, the multiple being chosen to make the mode of variability clear without grossly exaggerating it. Further details are given in the Web page associated with this chapter.

Thus far, the use of functional PCA for functions parameterized by arc length has no particular relation to the concomitant information that some of the bones are arthritic and some are not. In order to explore this aspect, we consider a different functional data analysis method, an extension of discriminant analysis. Apart from the way in which it identifies particular modes of variability within the population, the principal components analysis provides a convenient basis for the expression of the shapes in the sample and of other notch shapes.

## 8.4 The shape of arthritic bones

#### 8.4.1 Linear discriminant analysis

Suppose that  $\delta_i$  is a sequence of numbers such that  $\delta_i = 1$  if the *i*th bone is arthritic and -1 if it is not. In the present context, the object of functional discriminant analysis is to find a vector function  $\boldsymbol{\alpha}(t) = (\alpha_X(t), \alpha_Y(t))$  such that we can predict  $\delta$  for any given bone (drawn either from the sample or from a new set of data) by calculating the discriminant values

$$\hat{\delta}_{i} = \int_{0}^{1} \{X_{i}(t)\alpha_{X}(t)dt + Y_{i}(t)\alpha_{Y}(t)\}dt,$$
(8.2)

and checking whether it lies above or below some critical value C.

The function  $\alpha(t)$  characterizes the mode of variability that best discriminates between the two populations. Moving away from the mean in the direction of  $\alpha(t)$  is the way of increasing the integral in (8.2) as fast as possible. But how is  $\alpha$  to be found?

Suppose the data were vectors  $Z_i$  rather than functions  $\mathbf{Z}_i(t)$ . The corresponding problem would be to find a vector a and a constant C such that we could predict the population from which a vector Z was drawn by calculating whether a'Z > C. The classical method called *linear discriminant analysis* finds the vector a that minimizes the ratio of the within-group sum of squares to the between-group sum of squares. Let  $\overline{Z}^{(1)}$  and  $\overline{Z}^{(2)}$  be the means of the two populations and let  $\hat{S}$  be the pooled estimate of the variance matrix. Then the linear discriminant method yields

$$a = \hat{S}^{-1}(\bar{Z}^{(2)} - \bar{Z}^{(1)})$$

and

$$C = \frac{1}{2}a'(\bar{Z}^{(2)} + \bar{Z}^{(1)}).$$



Figure 8.8. The mode of variability corresponding to a linear discriminant analysis carried out directly on the matrix of coordinates defining the notch shapes. The arrows show how the 50 defining points on the mean curve are perturbed in the direction defined by the discriminant vector. The way to increase the discriminant score most quickly is to move away from the mean shape in the direction of the arrows.

In the functional case, we have observations on 100 variables, the Xand Y-coordinates of the points around the notch, for each of the N individuals in the sample. Naively, we could apply the linear discriminant method to these high-dimensional vectors. The resulting 100-vector a can be translated back into a  $50 \times 2$  matrix of weights corresponding to a mode of variability in the space of possible notch shapes.

Unfortunately this approach does not give a meaningful result. See Figure 8.8 for the mode of variability that it yields. This mode of variability clearly cannot be associated with any genuine feature of the problem in hand. Furthermore, this discriminant has the property that it classifies every bone in the sample perfectly; every arthritic bone has a'Z > C and every nonarthritic bone has a'Z < C. However superficially attractive such performance may be, it is scarcely credible as a result of the study.

This phenomenon—gross overfitting combined with an apparently meaningless discriminant function—is an intrinsic feature of the naive approach, and has nothing to do with the arthritis data in particular. It has a mathematical explanation touched upon in Chapter 12 of Ramsay and Silverman (1997) and discussed in more detail in references given there. In the present context a more informal explanation is given in Section 8.6.2 below.

#### 8.4.2 Regularizing the discriminant analysis

We have to apply some regularization in order to give meaningful answers. A simple method is to expand the data in terms of some suitable basis, and only to consider a finite number of terms in this basis, both in the expansion of the data themselves and in the specification of the discriminant weight function  $(\alpha_X(t), \alpha_Y(t))$ .

In the present case, the principal components analysis gives a lowdimensional representation of the data that preserves as much as possible of the sample variability. For this reason we use as our basis expansion the harmonics provided by the functional PCA of the data themselves. Fix some fairly small integer J and consider only the first J terms in the principal components expansion (8.1) of each of the functions. For concreteness we choose J = 6. For each bone, we then have six principal component scores on which to base our linear discriminant, and we apply standard discriminant analysis to the  $N \times 6$  matrix  $(z_{ij}, i = 1, \ldots, N; j = 1, \ldots, 6)$ . This yields a vector a of length 6, giving a linear discriminant in terms of the principal component scores,

$$\hat{\delta}_i = \sum_{j=1}^6 a_j z_{ij}.$$
(8.3)

We can express the discriminant value in terms of the notch curves themselves. By standard properties of principal component expansions,

$$z_{ij} = \int_0^1 \{X_i(t)\xi_j^X(t) + Y_i(t)\xi_j^Y(t)\}dt$$

for each *i* and *j*. Substituting into (8.3), the linear discriminant value  $\delta_i$  satisfies

$$\hat{\delta}_{i} = \sum_{j=1}^{6} a_{j} \int_{0}^{1} \{X_{i}(t)\xi_{j}^{X}(t) + Y_{i}(t)\xi_{j}^{Y}(t)\}dt$$
$$= \int_{0}^{1} \{\alpha_{X}(t)X_{i}(t) + \alpha_{Y}(t)Y_{i}(t)\}dt,$$

where

$$\begin{bmatrix} \alpha_X(t) \\ \alpha_Y(t) \end{bmatrix} = \sum_{j=1}^6 a_j \begin{bmatrix} \xi_j^X(t) \\ \xi_j^Y(t) \end{bmatrix}.$$
(8.4)

Comparing with equation (8.2), we can consider the two-dimensional function  $\boldsymbol{\alpha}(t) = \{\alpha_X(t), \alpha_Y(t)\}$  as defining the functional linear discriminant between the two groups of bones.



Figure 8.9. The mode of variability corresponding to a functional linear discriminant  $\alpha(t)$  based on the first six principal components of the notch shape data. The solid curve is the mean shape, and the arrows show the direction in which the discriminant score increases most rapidly.

The mode of variability corresponding to the resulting  $\alpha(t)$  is displayed in Figure 8.9. Bones with a higher discriminant score will have an intercondyle notch twisted to the left in the way that the figure is plotted. Because the mean is somewhat twisted to the right, this will tend to make the notch more symmetrical and to have a right edge that is less concave. The arthritic bones will tend to be in this category, and the average difference between the two groups of bones is approximately that corresponding to the lengths of the arrows in Figure 8.9.

The number J may be thought of as a regularization parameter, which determines how far we regularize the problem in order to produce our estimate. If we set J very small, equal to 1, for example, then the discrimination can only be based on a single principal component and important information may be lost. On the other hand, if J is chosen too large, then we will get the kind of spurious results discussed in Section 8.4.1 above. As in many smoothing and regularization contexts it is often sufficient to experiment with different values of the regularization parameter and choose between them by inspection, and in this case such inspection will immediately rule out values of J greater than about 12. However, it is also helpful to have criteria to help make this choice, and one of these is a cross-validation method described further in Section 8.6.3. This method confirms our choice J = 6.



Figure 8.10. Box plots of discriminant scores. The two plots on the left give linear discriminant scores based on the first six principal components. Those on the right give scores based on the difference between the group means. The scores are scaled so that the arthritic bones have mean 1 and the nonarthritic mean -1. Note that the boxes in the first two plots do not overlap at all, whereas there is considerable overlap between the boxes in the last two plots. In every case the box covers the middle 50% of the relevant sample.

#### 8.4.3 Why not just look at the group means?

The mode of variability that best discriminates between the arthritic and nonarthritic bones picks out features that are not at all apparent in the simple comparison of the means in Figure 8.6. Is this a contradiction?

The two curves in Figure 8.6 differ almost entirely along the lines of the first principal component of variability of the population as a whole, shown in Figure 8.7 to correspond to the depth of the notch. There is considerable population variation in this component, and hence in the notch depth, and this general variation is reflected in the differences in the mean notch depths for the two subpopulations. If we project all the data on the direction of the difference between the mean curves, the *t*-statistic for the difference between the two subpopulations is about 3.1.

On the other hand, if we consider the linear discriminant scores based on the first six principal components, the *t*-statistic for the difference between the two groups is 4.8. The regularized linear discriminant is much better at separating the two groups than is the direction of variability defined by the group means. Figure 8.10 gives a graphic presentation of this: the two scores are each rescaled so that the mean of the arthritic bones is +1 and the mean of the controls is -1. The box plots show that the "six principal component linear discriminant" approach separates the subpopulations far better than the "mean difference projection direction."

## 8.5 What have we seen?

The right way to express shapes in functional form may not always be obvious. If our object is a curve in two dimensions then parameterization by arc length can be a convenient way of representing the functional observations as vector-valued functions  $\{x(t), y(t)\}$  of a scalar parameter t. Standard methods such as functional PCA can then be used to analyze the data. Without such a parameterization even the notion of a mean curve has no obvious definition.

Linear discriminant analysis can be extended to the functional context, but regularization is necessary to give meaningful results. Intuitively, if an entire function is used to predict a single quantity, such as the class to which the function belongs, then a totally spurious feature of the function may give perfect prediction for the particular data set observed. One possible regularization approach is to concentrate on the first few principal components, or some other finite-dimensional representation of the data. Whatever method of regularization is used, the regularization parameter can be chosen by inspection or by an approach like cross-validation.

Functional discriminant analysis can distinguish groups better than consideration of the group mean curves alone. The group means may differ in ways that reflect modes of variability in the population generally, rather than those that specifically separate the groups within the population. The means of the two subpopulations might suggest that it is the depth of the notch that is associated with the symptoms of arthritis. However, the functional discriminant analysis indicates that the best discriminating characteristic is the differing amout of "twist" in the notch shape. This aspect of the shape could affect the way that the anterior cruciate ligament lies in the intercondylar notch, with a possible link to arthritis as discussed in Section 8.1. Within the present study, we cannot disentangle the influence of bone shape on arthritis from the possibility that arthritis causes a change in bone shape. However, our results give clues and pointers for future work in the fields of rheumatology and biomechanics.

# 8.6 Notes and further issues

### 8.6.1 Bibliography

The notch shape study discussed is a reworking of Shepstone, Rogers, Kirwan, and Silverman (2001), which deals with the same data and the same clinical issues, but uses a somewhat different approach to the parameterization of the notch shapes and to the subsequent analysis. That paper contains full details of the medical background, including key references to work in the rheumatological, biomechanical, and veterinary literature. Functional discriminant analysis is a particular example of the use of functions as predictors, as discussed broadly by Ramsay and Silverman (1997, Chapter 10). They treat in detail the general necessity for regularization in such problems, and consider various approaches to regularization, including roughness penalty methods. An early paper in the FDA literature dealing with these issues is Leurgans, Moyeed, and Silverman (1993), who demonstrate and investigate the need for regularization in another functional context, canonical correlation analysis. Hastie, Buja and Tibshirani (1995) set out the general idea of functional discriminant analysis making use of a roughness penalty approach to regularization. They apply their methods to a problem in speech recognition and to the classification of digits in handwritten postal addresses. Both functional canonical correlation analysis and functional discriminant analysis are treated in detail in Ramsay and Silverman (1997, Chapter 12).

#### 8.6.2 Why is regularization necessary?

We can give an intuitive argument for the necessity of regularization for the bone shape discriminant problem. The discretized coordinates of the data provide N points in 100-dimensional space. Four of the coordinates are fixed, because the notches are all scaled to start at (0,1) and end at (1,1), so the points are essentially in 96-dimensional space. We set the elements of a corresponding to these four fixed coordinates to zero. Now consider any division of the points into two groups, red and blue, say, and suppose that we want to find a vector a such that  $a'Z_i = 1$  if  $Z_i$  is a red point, and  $a'Z_i = -1$  if  $Z_i$  is a blue point. These are N equations in the 96 unknowns in a, and so, because N = 96, there is a solution that gives perfect discrimination between the populations. If we had used a finer discretization of the notches then there would have been N equations in even more unknowns, and hence an infinite set of such solutions. To put it less precisely, there is so much freedom in the choice of the vector a that it is not surprising that some completely uninteresting direction happens to give a discriminant function that works excellently on the given data but is in fact spurious—of course it will not have any value for classifying any new data collected.

This intuitive argument points to the qualitative difference between the regularization of functional discriminant analysis and roughness penalty smoothing as applied to PCA (as discussed in Chapter 2). For discriminant analysis, regularization is a mathematical necessity, however well behaved the original data—indeed, for mathematical reasons we do not go into here, the smoother the data the more acute the need for regularization. On the other hand, for functional PCA, smoothing is only important when we have data of high intrinsic variability, as we did in Chapter 2; an unsmoothed analysis will often suffice.

#### 8.6.3 Cross-validation in classification problems

The best approach to assessing the quality of a discriminant is to go out and collect completely new data and to see how well the discriminant rule based on the original data works on these new data. Unfortunately, in many contexts there are no new data available, and so we have to make use of the data we have. The simplest assessment of the discriminant is the *resubstitution* approach: feed the original data back through the discriminant, and see how well classified they are. This approach will usually be optimistic. The *leave-one-out cross-validation* method attempts to avoid the use of the same data both to train and to test the discriminant as follows: classify each data point using a discriminant constructed from all the data except that particular point. This requires a separate discriminant function for each data point in the sample and so may be computationally intensive, although there are some computational shortcuts that can be used. The approach is reminiscent of the cross-validation method when estimating the mean in the way described in Section 2.6.

Table 8.2. The cross-validation counts of false positives and false negatives for various values of the number J of principal components used in the discriminant algorithm. To get misclassification rates, divide the first row by 75 and the second row by 21.

J	1	2	3	4	5	6	7	8	9	10	11	12
False pos	26	27	22	22	23	19	23	21	21	22	22	22
False neg	10	8	8	8	8	7	7	7	7	7	8	9

Because the cross-validation approach gives a classification for each point individually, we can count both the number of false positives (nonarthritic bones that are classified as arthritic) and the number of false negatives (arthritic bones that fail to be so classified). The results are tabulated for various values of J in Table 8.2. In some circumstances we might need to combine false positive and false negative rates into a single score, but the choice J = 6 is the unique value that minimizes both scores, and so would be the minimum whatever linear combination of the two scores we were to choose.

A final comparison relevant to the discussion of Section 8.4.3 can be obtained by calculating the leave-one-out cross-validation scores for the approach of projecting on the difference between the two group means. This yields false positive and negative rates of 25 and 9, respectively, noticeably worse than the values of 19 and 7 yielded by the discriminant based on the first six principal components.