Chapter 4
Discriminating Groups of Organisms

Richard E. Strauss

Idea and Aims

A common problem in morphometric studies is to determine whether, and in what ways, two or more previously established groups of organisms differ. Discrimination of predefined groups is a very different problem than trying to characterize the patterns of morphological variation among individuals, and so the kinds of morphometric tools used for these two kinds of questions differ. In this paper I review the basic procedures used for discriminating groups of organisms based on morphological characteristics – measures of size and shape. A critical reading of morphometric discrimination studies of various kinds of organisms in recent years suggests that a review of procedures is warranted, particularly with regard to the kinds of assumptions being made. I will discuss the main concepts and methods used in problems of discrimination, first using conventional morphometric characters (measured distances between putatively homologous landmarks), and then using landmarks directly with geometric morphometric approaches.

Introduction

Suppose that we have several sets of organisms representing two or more known groups. Individuals from the groups must be recognizable on the basis of extrinsic criteria. For example, if the groups represent females and males of some species of fish, then we might identify individuals using pigmentation patterns or other kinds of sexual sex characteristics or, lacking those, by examination of the gonads. The key idea is that we must be able to unambiguously assign individuals to previously recognized groups. We still might wish to know a number of things about them. Can we discriminate the groups based on morphometric traits? If so, how well? How different are the groups? Are the groups "significantly" different in morphology? How

R.E. Strauss (✉)
Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409-3131, USA
e-mail: Rich.Strauss@ttu.edu

do we assess such significance in the presence of correlations among the morphometric characters? Which characters are the most important in discriminating the groups? Can group membership be predicted for "unknown" individuals? If so, how reliable are the predictions?

These questions can be answered (or at least approached) using three related kinds of methods: discriminant analysis (also called discriminant function analysis or canonical variate analysis), Mahalanobis distance, and multivariate analysis of variance. Discriminant analysis (DA) is used to estimate the linear combinations of characters that best discriminate the groups. Mahalanobis distance (D^2) estimates the distances between a pair of groups within the multivariate character space, in the presence of correlations among variables. And multivariate analysis of variance (MANOVA) determines whether the samples differ non-randomly (that is, significantly). It's interesting that the three kinds of methods were developed independently by three mathematicians: Fisher (DA) in England, Hotelling (MANOVA) in the United States, and Mahalanobis (D^2) in India. Due to differences in notation, underlying similarities between the methods were not noticed for some 20 years, but they now have a common algebraic formulation.

Conventional Morphometrics

Kinds of Data

Traditionally, before the onset of geometric morphometrics, morphometric studies were done using distances measured directly on specimens, often with calipers or microscopes, often in combination with meristic counts, angles, and other kinds of quantitative characters. Bookstein (Bookstein 1978; Bookstein et al. 1985; Strauss and Bookstein 1982) was the first to systematically stress the distinction between distances and other kinds of data, and the need to measure distances between comparable anatomical landmarks rather than arbitrarily on the form.

In the last decade or so, the use of digitizing equipment to record the positions of landmarks has become commonplace, and distances on specimens are usually calculated as Euclidean distances between landmarks. But directly measured distances continue to be used, sometimes mixed with other kinds of data.

For the following discussions I will assume that the variables (characters) consist entirely of distances measured between landmarks. Such distances are usually logarithmically transformed prior to analysis to improve their statistical properties and to characterize allometric relationships (Bookstein et al. 1985; Bryant 1986; Jungers et al. 1995; Keene 1995; Strauss 1993). However, use of log-transformations remains a somewhat controversial topic, and I won't pursue it here.

Principal Component Analysis

It's not uncommon for researchers to use principal component analysis (PCA) to attempt to discriminate groups of individuals. However, PCA is inherently a

single-group procedure and is not guaranteed to find group differences even if they exist. PCA is used to redistribute the total variance among a set of data points onto a set of mutually orthogonal axes (i.e., at right angles to one another) that merely redescribe the patterns of variation among the data. The new axes are the principal components, which are statistically independent of one another and so can be examined one at a time. The data points can be projected onto the axes (at right angles) to provide numerical scores of individuals on the components (Fig. 4.1). The principal components are calculated such that the variance of scores of individuals on the first axis (PC1) is as great as possible, so that PC1 can be said to account for the maximum variance in the data. Because the second component is by definition at right angles to the first, the scores of individuals on PC2 are uncorrelated with
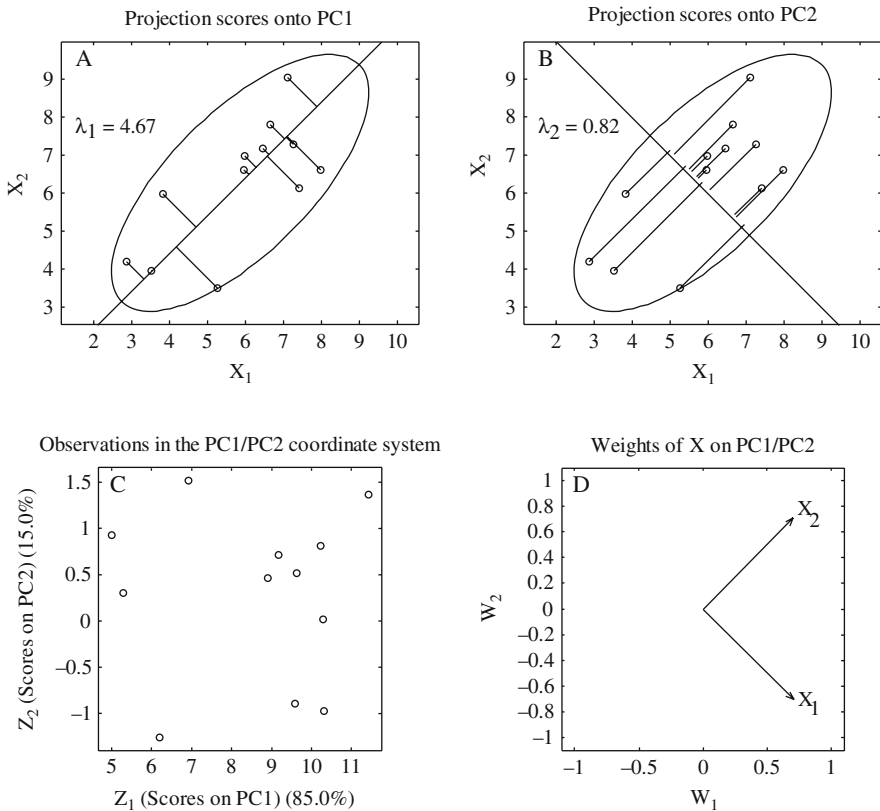

Fig. 4.1 Example of a principal component analysis for a scatter of data points for two variables. (**a**) The data points as projected onto PC1 to give scores on PC1. The ellipse is a 95% confidence interval on the data. The value λ_1 is the first eigenvalue, the variance of the scores on PC1. (**b**) The same data points as projected onto PC2. The value λ_2 is the second eigenvalue, the variance of the scores on PC2. (**c**) The data points plotted as scores in the space of components PC1 and PC2. (**d**) Projection of the axes for the two variables as unit vectors onto the space of components PC1 and PC2. These vectors indicate the maximum direction of variation in the corresponding variables in the PC1/PC2 space of **Panel C**

those on PC1. PC2 is the axis, orthogonal to PC1, on which the variance of scores is as great as possible. PC3 is the axis, mutually orthogonal to both PC1 and PC2, on which the variance of scores is as great as possible. And so on. PCA is usually used as a dimension-reduction procedure, because a scatterplot of points on the first two or three components may characterize most of the variation among the data point.

This procedure is a simple description of an eigenanalysis: the principal components are eigenvectors, and the variance of projection scores onto each component is the corresponding eigenvalue (Fig. 4.1). In practice, all components are calculated as a set rather than sequentially. The procedure can be viewed geometrically as a translation and solid rotation of the coordinate system. The origin of the coordinate system is moved (translated) to the center of the cloud of points, and then the coordinate axes are rotated as a set, at right angles to one another, so as to maximize the variance components. The data points maintain their original configuration, while the coordinate system moves around them. Thus the number of principal component axes is equal to the number of variables. The principal components are specified by sets of coefficients (weights, one per variable), the weights being computed so as to compensate for redundancy of information due to intercorrelations between variables. A principal component score for an individual is essentially a weighted average of the variables. The coefficients can be rescaled as vector correlations (Fig. 4.1), which are often more informative. The coefficients allow interpretation of the contributions of individual variables to variation in projection scores on the principal components. See Jolicoeur and Mosimann (1960) and Smith (1973) for early and very intuitive descriptions of the use of PCA in morphometric analyses.

The procedure inherently assumes that the data represent a single homogeneous sample from a population, although such structure isn't necessary to calculate the principal-component solution. (However, the assumption that the data were sampled from a multivariate-normally distributed population is necessary for classical tests of the significance of eigenvalues or eigenvectors.) Even if multiple groups are present in the data, the procedure does not take group structure into consideration. PCA maximizes variance on the components, regardless of its source. If the among-group variation is greater than the within-group variation, the PCA scatterplots might depict group differences. However, PCA is not guaranteed to discriminate groups. If group differences fail to show up on a scatterplot, it does not follow that group differences don't exist in the data.

Multiple-group modifications of PCA such as common principal components (CPC) have been developed (Flury 1988; Thorpe 1988), but these are generally not for purposes of discrimination. Rather, such methods assume that the same principal components exist in multiple groups (possibly with different eigenvalues) and allow estimation of the common components. Multiple-group methods are useful, for example, for adjusting morphometric data for variation in body size or other sources of extraneous variation prior to discrimination (Burnaby 1966; Humphries et al. 1981; Klingenberg et al. 1996).

Discriminant Analysis

In contrast to principal components analysis, discriminant analysis is explicitly a multiple-group procedure, and assumes that the groups are known (correctly) before analysis on the basis of extrinsic criteria and that all individuals are members of one (and only one) of the known groups. The terminology of discriminant analysis can be somewhat confusing. Fisher (1936) originally developed the "linear discriminant" for two groups. This was later generalized to the case of three or more groups independently by Bartlett, Hotelling, Mahalanobis, Rao and others to solve several related problems that are relevant to morphometric studies: the discrimination groups of similar organisms, the description of the morphological differences among groups, the measurement of overall difference between groups, and the allocation of "unknown" individuals to known groups. The allocation of unknown individuals is generally called classification, though this term is often used in a different way by systematic biologists, which by itself can cause confusion. The discrimination problem for three or more groups came to be known as "canonical variate analysis" ("canonical" in the sense of providing rules for classification), although this phrase has also been used synonymously with a related statistical procedure usually known as canonical correlation analysis. The tendency in recent years is to use "discriminant analysis" or "discriminant function analysis" for discrimination of any number of groups, although the term "canonical variate analysis" is still widely used.

Discriminant analysis (DA or DFA) optimizes discrimination between groups by one or more axes, the discriminant functions (DFs). These are mathematical functions in the sense that the projection scores of data points on the axes are linear combinations of the variables, as in PCA. Like PCA, DA is a form of eigenanalysis, except that in this case the axes are eigenvectors of the among-group covariance matrix rather than the total covariance matrix. For k groups, DA finds the k–1 discriminant axes that maximally separate the k groups (one axis for two groups, two for three groups, etc.). Like PCs, DFs have corresponding eigenvalues that specify the amount of among-group variance (rather than total variance) accounted for by the scores on each DF. Also like PCs, discriminant axes are linear combinations of the variables and are specified by sets of coefficients, or weights, that allow interpretation of contributions of individual variables. See Albrecht (1980, 1992) and Campbell and Atchley (1981) for geometric interpretations of discriminant analysis.

The first discriminant axis has a convenient interpretation in terms of analysis of variance of the projection scores (Fig. 4.2). Rather than being the axis that maximizes the total variance among scores, as in PCA (Fig. 4.1), the discriminant axis is positioned so as to maximize the total variance among groups relative to that within groups, which is the quantity measured by the ANOVA F-statistic. The projection scores on DF1 give an F-statistic value greater than that of any other possible axis. The same is true for three or more groups (Fig. 4.3). The DF1 axis is positioned so as to maximize the dispersion of scores of groups along it. The dispersion giving the
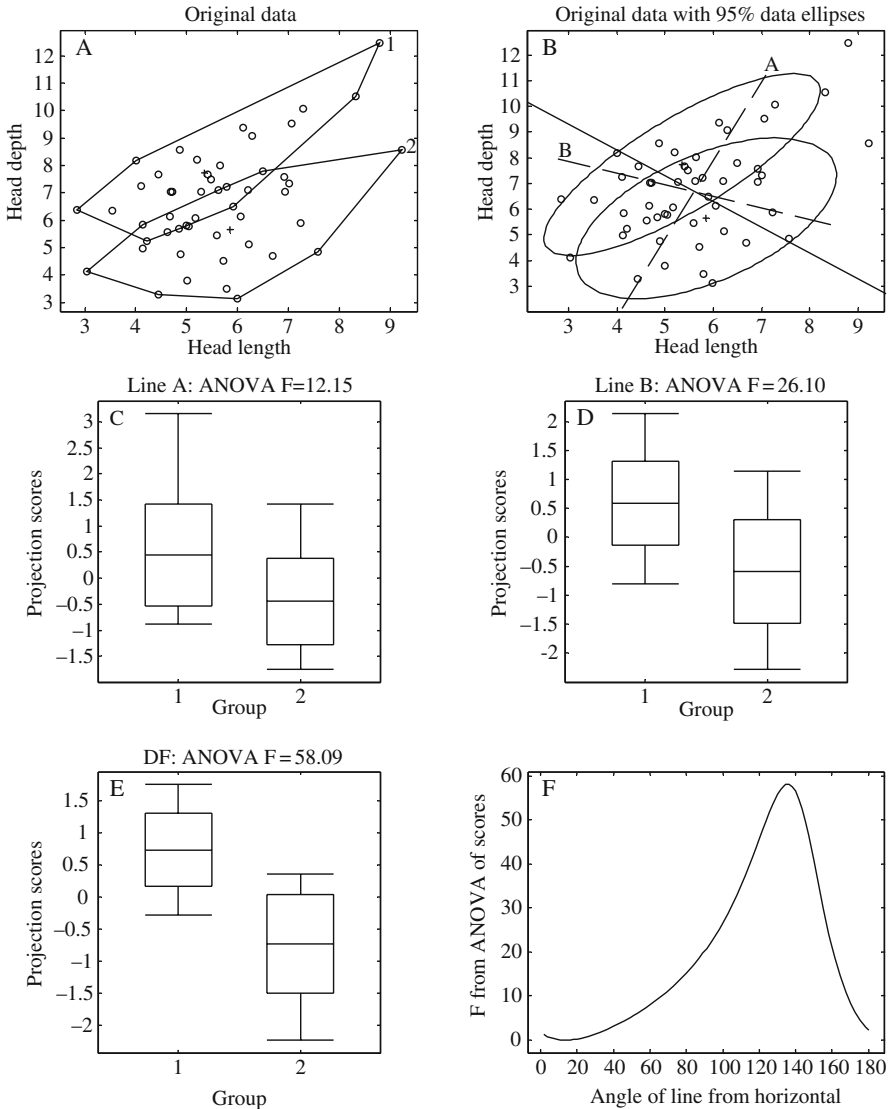
Fig. 4.2 Example of a discriminant analysis for samples of two species of *Poecilia*, in terms of two variables: head length and head width, both in mm. (**a**) Original data, with convex hulls indicating dispersion of data points for the two groups. (**b**) Data and 95% confidence intervals for the two groups. *Dotted lines* A and B are arbitrarily chosen axes; the *solid line* is the discriminant axis for the two groups. (**c**) Box plots for the two groups of projection scores onto *dotted line* A, and corresponding ANOVA F-statistic. (**d**) Box plots of projection scores onto *dotted line* B, and corresponding ANOVA F-statistic. (**e**) Box plots of projection scores onto the discriminant axis, and corresponding ANOVA F-statistic. (**f**) F-statistic from ANOVAs of projection scores onto all possible axes, as a function of angle (in degrees) from the horizontal (head-length axis) of **Panel B**. The discriminant axis is that having the maximum ANOVA F-statistic value
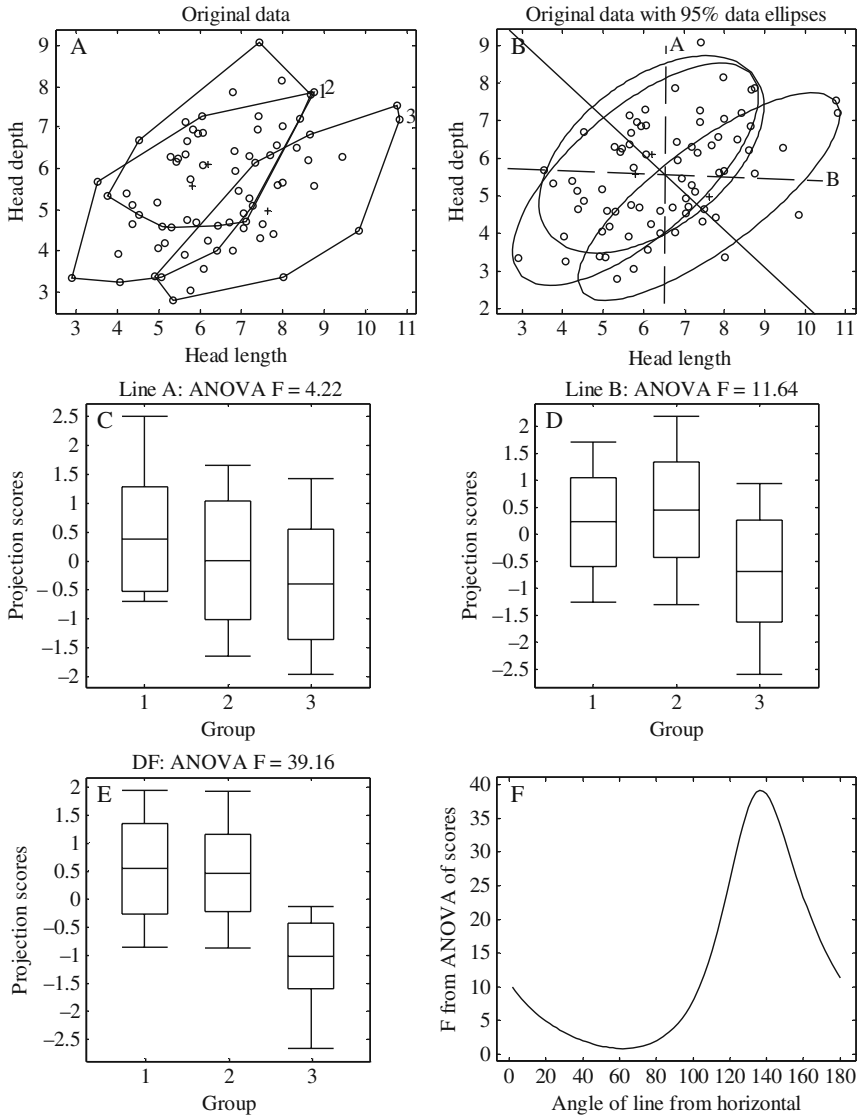
Fig. 4.3 Example of a discriminant analysis for samples of three species of *Poecilia*, in terms of two variables: head length and head width, both in mm. (**a**) Original data, with convex hulls indicating dispersion of data points for the three groups. (**b**) Data and 95% confidence intervals for the two groups. *Dotted lines* A and B are arbitrarily chosen axes; the *solid line* is the discriminant axis for the two groups. (**c**) Box plots for the three groups of projection scores onto *dotted line* A, and corresponding ANOVA F-statistic. (**d**) Box plots of projection scores onto *dotted line* B, and corresponding ANOVA F-statistic. (**e**) Box plots of projection scores onto the discriminant axis, and corresponding ANOVA F-statistic. (**f**) F-statistic from ANOVAs of projection scores onto all possible axes, as a function of angle (in degrees) from the horizontal (head-length axis) of **Panel B**. The discriminant axis is that having the maximum ANOVA F-statistic value

maximum F-statistic might distinguish one group from the others (as in Fig. 4.3e), or might separate all groups by a small amount; the particular pattern depends on the structure of the data.

As with PCA, a unique set of discriminant axes can be calculated for any set of data if the sample sizes are sufficiently large. However, inferences about the populations from which the data were sampled are reasonable only if the populations are assumed to be multivariate-normally distributed with equal covariance matrices (the multivariate extensions of the normality and homoscedasticity assumptions of ANOVA). In particular, discrimination of samples will be optimal with respect to their populations only if this distributional assumption is true. Because a topological cross-section through a multivariate normal distribution is an ellipse, confidence ellipses on the sample data are often depicted on scatterplots to visually assess this underlying assumption (Owen and Chmielewski 1985; Figs. 4.2b and 4.3b). If the assumption about population distributions is true, then the sample ellipses will be approximately of the same size and shape because they will differ only randomly (i.e., they will be homogeneous). Bootstrap and other randomization methods can give reliable confidence intervals on estimates of discriminant functions and related statistics even if the distributional assumption is violated (Dalgleish 1994; Ringrose 1996; Von Zuben et al. 1998; Weihs 1995).

The minimum sample sizes required for a discriminant analysis can sometimes be limiting, particularly if there are many variables relative to the number of specimens, as is often the case in morphometric studies. In the same way that an analysis of variance of a single variable is based on the among-group variance relative to the pooled (averaged) within-group variance, in a DA the eigenvectors and eigenvalues are derived from the among-group covariance matrix relative to the pooled within-group covariance matrix, which is the averaged covariance matrix across all groups. Using the pooled matrix is reasonable if the separate matrices differ only randomly, as assumed. But if the separate matrices are quite different, they can average out to a "circular" rather than elliptical distribution, for which the net correlations are approximately zero. In this case the DA results would not differ much from those of a PCA.

The minimum sample size requirement for a DA relates to the fact that the pooled within-group matrix must be inverted (because it's "in the denominator", so to speak), and inversion can't be done unless the degrees of freedom of the within-group matrix be greater than the number of variables. The within-group degrees of freedom is typically N-p-1, where N is the total sample size and p is the number of variables. However, this is the minimum requirement for a solution to be found. The number of specimens should be much larger than the number of variables for a stable solution – one that wouldn't change very much if a new set of samples from the same populations were taken. A typical rule of thumb is that the number of specimens should be at least five or so times the number of variables. However, the minimally reasonable sample size depends on how distinctive the groups are (because subtle differences require more statistical power to detect). In addition, it requires larger sample sizes to determine the nature of the differences among groups than just to demonstrate that the difference is significant.
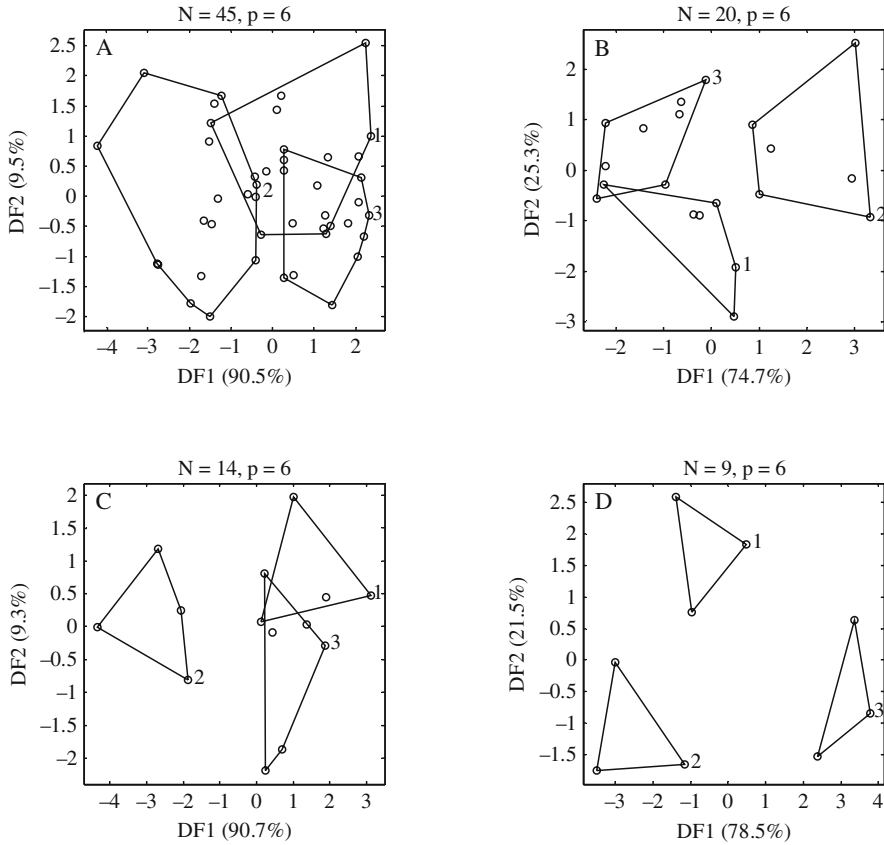
Fig. 4.4 Example of the effect of sample size on the apparent discrimination among three groups. (**a**) Scatterplot of scores on the two discriminant axes for 45 specimens and 6 variables. (**b**) Scatterplot of scores for a random 20 of 45 specimens. (**c**) Scatterplot of scores for a random 14 of 45 specimens. (**d**) Scatterplot of scores for a random 9 of 45 specimens

Because of this matrix-inversion problem, the degree of discrimination among groups can become artificially inflated for small sample sizes (relative to the number of variables) (Fig. 4.4). Scatterplots on discriminant axes can suggest that groups are highly distinctive even though the group means might actually differ by little more than random variation. Because of this, discriminant scatterplots must be interpreted with caution, and never without supporting statistics (described below).

Another factor that enters into the minimum-sample-size issue is variation in the number of specimens per group. When the covariance matrices for the separate groups are pooled, the result is a weighted average covariance matrix, weighted by sample size per group. This makes sense because the precision of any statistical estimate increases as sample size increases, and so a covariance matrix for a large sample is a more reliable estimate of the "real" covariance matrix. Because

variances and covariances can be estimated for as few as three specimens, very small groups can in principle be included in a discriminant analysis. In practice, however, it is often beneficial to omit groups having sample sizes of less than five or so.

Some recently developed methods for performing discriminant analysis with relatively small sample sizes (e.g. Anderson and Robinson 2003; Howland and Park 2004; Ye et al. 2004) seem promising, but none have yet been applied to morphometric data.

Size-Free Discriminant Analysis

In systematics it has long been considered desirable to be able to discriminate among groups of organisms (populations, species, etc.) on the basis of "size-free" or size-invariant shape measures (dos Reis et al. 1990; Humphries et al. 1981). This is particularly important when the organisms display indeterminant growth, in which case discrimination among taxa might represent merely a sampling artifact if different samples comprise different proportions of age classes. Discrimination among samples in which variation in size cannot be easily controlled may lead to spurious results, since the size-frequency distribution of different taxa will be a function of the ontogenetic development of individuals present in different samples. In this case one way of correcting the problem would be to statistically "correct" or adjust for the effect of size present within samples of each group. However, a number of different definitions of "size-free" shape have been applied. The terms shape and size have been used in various and sometimes conflicting ways (Bookstein 1989a).

In size adjustment the effects of size variation are to be partitioned or removed from the data, usually by some form of regression, and residuals are subsequently used as size-independent shape variables (Jolicoeur et al. 1984; Jungers et al. 1995). In distance-based morphometrics, the most common methods for size adjustment have involved bivariate regression (Albrecht et al. 1993; Schulte-Hostedde et al. 2005; Thorpe 1983) multiple-group principal components (Pimentel 1979; Thorpe and Leamy 1983), sheared principal components (Bookstein et al. 1985; Humphries et al. 1981; Rohlf and Bookstein 1987), and Burnaby's procedure (Burnaby 1966; Gower 1976; Rohlf and Bookstein 1987). Although many different methods have been proposed, there has been little agreement on which method should be used. This issue is important because different size-adjustment methods often yield slightly different results.

In the case of size-adjustment for multiple taxa, the issue arises as to whether and how group structure (e.g., presence of multiple species) should be taken into consideration (Klingenberg and Froese 1991) – whether the correction should be separately by group or should be based on the pooled within-group regression. The latter implicitly assumes that all within-group covariance matrices are identical, although this assumption can be relaxed with use of common principal components (Airoldi and Flury 1988; Bartoletti et al. 1999; Klingenberg et al. 1996).

Mahalanobis Distances

Whereas discriminant analysis scores can provide a visualization of group separation, Mahalanobis distances (D^2) measure the distances between group centroids on a scale that is adjusted to the (pooled) within-group variance in the direction of the group difference. (D, the square root of D^2, measures the distance between group centroids adjusted by the standard deviation rather than the variance.) In Fig. 4.5, for example, the Euclidean (straight-line) distance from centroid A to centroid B is that same as that from A to C. However, the Mahalanobis distances are quite different because the distance from A to B is measured "with the grain" while that from A to C is measured "across the grain". In terms of variation, the relative distance from A to C is much greater than that from A to B.

This is often said to be analogous to using an F-statistic to measure the difference between two group means, although that is not quite correct – an F-statistic increases as sample size increases, whereas a Mahalanobis distance approaches its "true" value with increasing sample size. The Mahalanobis distance is essentially a distance in a geometric space in which the variables are uncorrelated and equally scaled. It also possesses all of the characteristics that a measure must have to be a metric: the distance between two identical points must be zero, the distance between two non-identical points must be greater than zero, the distance from A to B must be the same as that from B to A (symmetry), and the pairwise distances among three points must satisfy the triangle inequality. For morphometric data, such a measure of group separation is more informative than the simple Euclidean distance between groups.
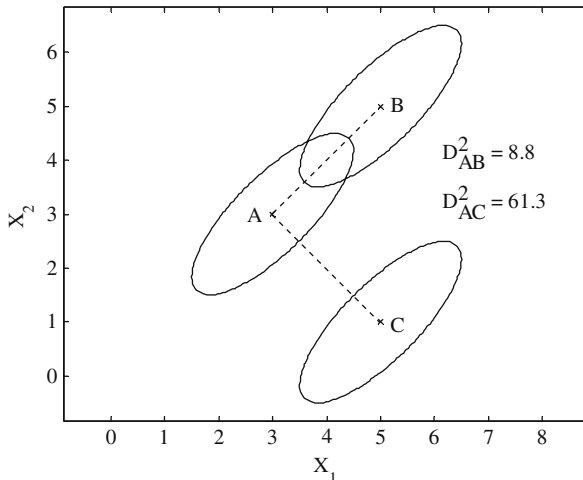

Fig. 4.5 Mahalanobis distances between centroids of groups. Variation within groups is indicated by 95% confidence ellipses for the data. Euclidean distances between centroids of A and B and of A and C are both 2.83. Corresponding Mahalanobis distances are indicated on plot

Mahalanobis distances can also be measured between a point and a group centroid or between two points. In both cases the distance is relative to the covariance matrix of the group.

Confidence intervals for Mahalanobis distances can be estimated by comparison to a theoretical F distribution if the distribution of the group(s) is assumed to be multivariate normal (Reiser 2001). More robust confidence intervals for real biological data can be estimated by bootstrapping the data within-group (Edgington 1995; Manly 1997; Wilcox 2005).

MANOVA

Analysis of variance (ANOVA) is the univariate case of the more general multivariate analysis of variance (MANOVA). Instead of a "univariate F" statistic measuring the heterogeneity among a set of means with respect to the pooled within-group variance, the resulting "multivariate F" measures the heterogeneity among a set of multivariate centroids with respect to the pooled within-group covariance matrix. The covariance matrix accounts for the observed correlations among variables. As with ANOVA, the samples can be cross-classified with respect to two or more factors, or can be structured with respect to other kinds of sampling designs (Gower and Krzanowski 1999).

In practice the actual test statistic calculated is Wilks' lambda, which is related to the computations involved in discriminant functions and Mahalanobis distances. It is a direct measure of the proportion of total variance in the variables that is not accounted for by the grouping of specimens. If Wilks' lambda is small, then a large proportion of the total variance is accounted for by the grouping, which in turns suggests that the groups have different mean values for one or more of the variables. Because the sampling distribution of Wilks' lambda is rather difficult to evaluate, lambda is usually transformed approximately to an F statistic. There are a number of alternative statistics that are similar in purpose to Wilks' lambda but that have somewhat different statistical properties, such as Pillai's trace and Roy's greatest root. These are often reported by statistical software, but in general are not widely used (Everitt and Dunn 2001).

Under the null hypothesis that all groups have been sampled randomly from the same population, and therefore differ only randomly in all of their statistical properties, the F statistic can be used to estimate a "P-value", the probability of sampling the observed amount of heterogeneity among centroids if the null hypothesis is true. The P-value is accurate only if the population from which the groups have been sample is multivariate-normal in distribution. If the null hypothesis is true, then the covariance matrices for all groups will differ only randomly (i.e., they will be homogeneous), and thus can be pooled for the test. If the within-group covariance matrices differ significantly, then the pooled covariance matrix may be biased, as will the P-value. As with statistical tests in general, violated assumptions will often (but not necessarily) lead to P-values that are too small, and thus will lead to the rejection of the null hypothesis too often.

Since claiming significant differences when they don't exist is counterproductive in science, the dependence of MANOVA on such stringent assumptions is a problem. This can be circumvented to some degree by using randomization procedures (e.g., random permutation) to estimate the null sampling distribution of the test statistic rather than theoretical distributions (such as the F distribution) (Anderson 2001). Such "non-parametric" tests, although not assumption-free, tend to be much more robust to statistical assumptions than are conventional statistical hypothesis tests.

It is often assumed that a series of separate ANOVAs, one per variable, is equivalent to a MANOVA. However, this is not the case, for several reasons (Willig and Owen 1987). First, if the variables are correlated, then the separate ANOVAs are not statistically independent. For example, if the ANOVA for one variable is statistically significant, then the ANOVAs for variables correlated with it will also tend to be significant. Thus the results from the ANOVAs will be redundant to an unknown extent and difficult to integrate. Second, the overall ("family-wise") Type I error rate become artificially high as the number of statistical tests increases, so that the probability of obtaining a significant results due to chance increases (the "multiple-comparisons" problem; Hochberg and Tamhane 1987).

If the overall MANOVA is statistically significant, then separate ANOVAs can be done to assess which of the variables has contributed to the group differences. But the multiple-comparisons issues remain, and subsequent statistical testing must be done carefully.

Classification

A procedure closely related to discriminant functions and Mahalanobis distances is that of classifying "unknown" specimens to known, predefined groups. (Note that this use of "classification" is related to, but different from, the common use of the term in systematics.) A strong assumption of any classification procedure is that the individual being classified is actually a member of one of the groups included in the analysis. If this assumption is ignored or wrong, then any estimated probabilities of group membership may be misleading (Albrecht 1992).

There are two basic approaches to classifying unknowns with morphometric data. The first, and most conventional, is based in principle on means: calculate the Mahalanobis distance from the unknown to the centroid of each group, and assign it to the closest group (Hand 1981). Because Mahalanobis distances are based on pooled covariance matrices, correct assignments depend on the assumptions of homogeneous covariance matrices and, to a lesser degree, of multivariate normality. This approach can be viewed as subdividing the data space into mutually exclusive "decision spaces", one for each predefined group, and classifying each unknown according to the decision space in which it lies. Each Mahalanobis distance has an associated chi-square probability, which can be used to estimate probabilities of group membership (or their complements, probabilities of misclassification; Williams 1982). More robust estimates of classification probabilities can be

approximated by bootstrapping the "known" specimens within-group (Davison and Hinkley 1996; Fu et al. 2005; Higgins and Strauss 2004).

The second approach is to view the data space in terms of mixtures of multivariate-normal distributions, one for each predefined group. Such methods tend to be much more sensitive to deviations from the assumptions of multivariate normality and homogeneous covariance matrices, but can better accommodate differences in sample size among groups (White and Ruttenberg 2007).

Cross Validation

Cross-validation is a widely used resampling technique that is often used for the assessment of statistical models (Stone 1974). Like other randomization methods such as the bootstrap and jackknife, it is almost distribution-free in the sense that it evaluates the performance of a statistical procedure given the actual structure of the data. It is necessary because whenever predictions from a statistical model are evaluated with the same data used to estimate the model, the fit is "too good"; this is known as over-fitting. When new data are used, the model almost always performs worse than expected. In the case of discriminant analysis and related methods, overfitting comes into play both in the assessment of group differences (discriminant-score plots and MANOVA) and in estimates of probabilities of group membership.

The basic idea behind cross-validation is simply to use a portion of the data (the "training" or "calibration" set) to fit the model and estimate parameters, and use the remaining data (the "test" set) to evaluate the performance of the model. For classification problems, for example, the group identities of all specimens are known in advance, and so they can be used to check whether the predicted identities are correct. This is typically done in a "leave-one-out" manner: one specimen is set aside and all N-1 others are used to estimate Mahalanobis distances. The omitted specimen is then treated as an unknown and its group membership is predicted. The procedure is repeated for all specimens, sequentially leaving each one out of the analysis and estimating distances from the others, then predicting the group membership of the omitted specimen. The overall proportions of correct predictions are unbiased estimates of the probabilities of correct classification, given the actual structure of the data. Cross-validation methods are particularly appropriate for small samples (Fu et al. 2005).

Related Methods

The most commonly used alternative to discriminant analysis is logistic regression, which usually involves fewer violations of assumptions, is robust, handles discrete and categorical data as well as continuous variables, and has coefficients that are somewhat easier to interpret (Hosmer and Lemeshow 2000). However, discriminant analysis is preferable when its assumptions are reasonably met because it has consistently greater statistical power (Press and Wilson 1978).

Quadratic discriminant analysis (QDA) is closely related to linear discriminant analysis (LDA), except that there is no assumption that the covariance matrices of the groups are homogeneous (Meshbane and Morris 1995). When the covariance matrices are homogeneous, LDA is systematically better than QDA both at group separation and classification. When the covariance matrices vary significantly, QDA is usually better, but not always, especially for small samples (Flury et al. 1994; Marks and Dunn 1974). This is apparently due to the greater robustness of LDA to violation of assumptions. In any case, there have been few morphometric applications of quadratic discriminant analysis.

There are several different versions of nonlinear discriminant analysis, which finds nonlinear functions that best discriminate among known groups. Most nonlinear methods work by finding some linear transformation of the character space that produces optimal linear discriminant functions. Generalized discriminant analysis (Baudat and Anouar 2000) has become the most widely used method.

And finally, neural networks have been used successfully in both linear and nonlinear discrimination and classification problems (Baylac et al. 2003; Dobigny et al. 2002; Higgins and Strauss 2004; Kiang 2003; Raudys 2001; Ripley 1994).

Geometric Morphometrics

Whereas conventional morphometric studies utilize distances as variables, geometric morphometrics (Bookstein 1991; Dryden and Mardia 1998; Rohlf 1993) is based directly on the digitized x,y,(z)-coordinate positions of landmarks, points representing the spatial positions of putatively homologous structures in two or three dimensions. Bookstein (1991) has characterized the types of landmarks, their configurations, and limitations, and Adams (1999) has extended their utility.

Once landmark coordinates have been obtained for a set of forms, they must be standardized to be directly comparable. This is typically done using a generalized Procrustes analysis in two or three dimensions, in which the sum of squared distances between homologous landmarks of each form and a reference configuration is iteratively minimized by translations and rigid rotations of the landmark configurations (Goodall 1995; Gower 1975; Penin and Baylac 1995; Rohlf and Slice 1990).

Isometric size differences are eliminated by dividing the coordinates of each form by its centroid size, defined as the square root of the sum of the squared distances between the geometric center of the form and its landmarks (Bookstein 1991). The residual variation in landmark positions among forms (deviations from the reference form) are referred to as "Procrustes residuals" in the x and y (and possibly z) coordinate directions. The square root of the sum of the squared distances between corresponding landmarks of two aligned configurations is an approximation of Procrustes distance, which plays a central role in the theory of shape analysis (Small 1996). It is also the measure that binds together the collection of methods for the analysis of shape variation that comprises the "morphometric synthesis" (Bookstein 1996).

To characterize and visualize differences between pairs of reference forms, the aligned landmark coordinates are often fitted to an interpolation function such as a thin-plate spline (Bookstein 1989b; Rohlf and Slice 1990), which can be decomposed into global (affine) and local (nonaffine) components. The nonaffine component can be further decomposed into partial or relative warps, geometrically orthogonal (and thus independent) components that correspond to shape deformations at different scales.

However, for the purpose of discrimination among groups of forms, the Procrustes residuals can be used directly as variables for discriminant analysis, MANOVA, and classification, as described above. In this case the number of variables for two-dimensional forms is twice the number of landmarks, one set for the x coordinates and one set for the y coordinates. For three-dimensional forms the number of variables would be three times the number of landmarks.

The large number of variables relative to the number of specimens therefore presents even more of a problem in geometric morphometrics than it tends to do in conventional morphometrics. The usual procedure is to use the Procrustes residuals in a principal component analysis, and then use the projection scores on the first few components as derived variables (e.g., Depecker et al. 2006). Since these derived variables are uncorrelated across all observations, the covariance matrices have zeros in the off-diagonal positions, and Mahalanobis distances are equivalent to Euclidean distances.

Conclusion

The multivariate methods reviewed here remain a powerful set of tools for morphometric studies, and their importance in the field cannot be overemphasized. Although the widespread availability of computer software has permitted their use by biologists of varying levels of statistical background and sophistication, it remains true that it is the responsibility of individual researchers to understand the properties and underlying assumptions of the methods they use.

References

Adams DC (1999) Methods for shape analysis of landmark data from articulated structures. Evolutionary Ecology Research 1: 959–970

Airoldi JP, Flury BK (1988) An application of common principal component analysis to cranial morphometry of *Microtus californicus* and *Microtus ochrogaster* (Mammalia, Rodentia). Journal of Zoology (London) 216: 21–36

Albrecht GH (1980) Multivariate analysis and the study of form, with special reference to canonical variate analysis. American Zoologist 20: 679–693

Albrecht GH (1992) Assessing the affinities of fossils using canonical variates and generalized distances. Human Evolution 7: 49–69

Albrecht GH, Gelvin BR, Hartman SE (1993) Ratios as a size adjustment in morphometrics. American Journal of Physical Anthropology 91: 441–468

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. Austral Ecology 26: 32–46

Anderson MJ, Robinson J (2003) Generalized discriminant analysis based on distances. Australian and New Zealand Journal of Statistics 45: 301–318

Bartoletti S, Flury BD, Nel DG (1999) Allometric extension. Biometrics 55: 1210–1214

Baudat G, Anouar F (2000) Generalized discriminant analysis using a kernel approach. Neural Computation 12: 2385–2404

Baylac M, Villemant C, Simbolotti G (2003) Combining geometric morphometrics with pattern recognition for the investigation of species complexes. Biological Journal of the Linean Society 80: 89–98

Bookstein FL (1978) The Measurement of biological shape and shape change. Lecture Notes in Biomathematics 24. Springer-Verlag, New York

Bookstein FL (1989a) "Size and shape": a comment on semantics. Systematic Zoology 38: 173–180

Bookstein FL (1989b) Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11: 567–585

Bookstein FL (1991) Morphometric tools for landmark data: geometry and biology, Cambridge University Press, Cambridge

Bookstein FL (1996) Combining the tools of geometric morphometrics. In Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE (Eds.) Advances in morphometrics, Plenum Press, New York, 131–151 pp

Bookstein FL, Chernoff B, Elder RL, Humphries JM, Smith GR, Strauss RE (1985) Morphometrics in evolutionary biology: the geometry of size and shape change. Academy of Natural Sciences, Philadelphia

Bryant EH (1986) On use of logarithms to accommodate scale. Systematic Zoology 35: 552–559

Burnaby TP (1966) Growth-invariant discriminant functions and generalized distances. Biometrics 22: 96–110

Campbell NA, Atchley WR (1981) The geometry of canonical variates analysis. Systematic Zoology 30: 268–280

Dalgleish LI (1994) Discriminant analysis: statistical inference using the jackknife and bootstrap procedures. Psychological Bulletin 116: 498–508

Davison AC, Hinkley DV (1996) Bootstrap methods and their application, Cambridge University Press, Cambridge

Depecker M, Berge C, Penin X, Renous S (2006) Geometric morphometrics of the shoulder girdle in extant turtles (Chelonii). Journal of Anatomy 208: 35–45

Dobigny G, Baylac M, Denys C (2002) Geometric morphometrics, neural networks and diagnosis of sibling *Taterillus* species (Rodentia, Gerbillinae). Biological Journal of the Linnean Society 77: 319–327

dos Reis SF, Pessoa LM, Strauss RE (1990) Application of size-free canonical discriminant analysis to studies of geographic differentiation. Revista Brasileira de Genética 13: 509–520

Dryden IL, Mardia KV (1998) Statistical shape analysis. John Wiley, New York

Edgington ES (1995) Randomization tests. Marcel Dekker, New York

Everitt BS, Dunn G (2001) Applied multivariate data analysis, Second Edition. Wiley & Sons, New York

Fisher RA (1936) The use of multiple measurements in taxonomic problems. Annals of Eugenics 7: 179–188

Flury B, Schmid MJ, Narayanan A (1994) Error rates in quadratic discrimination with constraints on the covariance matrices. Journal of Classification 11: 101–120

Flury BK (1988) Common principal components and related multivariate models, Wiley, New York

Fu WJ, Carroll RJ, Wang S (2005) Estimating misclassification error with small samples via bootstrap cross-validation. Bioinformatics 21: 1979–1986

Goodall CR (1995) Procrustes methods in the statistical analysis of shape revisited. In Mardia KV, Gill CA (Eds.) Current issues in statistical shape analysis, University of Leeds Press, Leeds, 18–33 pp

Gower JC (1975) Generalized procrustes analysis. Psychometrika 40: 33–51

Gower JC (1976) Growth-free canonical variates and generalized inverses. Bulletin of the Geological Institutions of the University of Uppsala 7: 1–10

Gower JC, Krzanowski WJ (1999) Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance. Royal Statistical Society: Series C (Applied Statistics) 48: 505–519

Hand DJ (1981) Discrimination and classification. John Wiley, New York

Higgins CL, Strauss RE (2004) Discrimination and classification of foraging paths produced by search-tactic models. Behavioral Ecology 15: 248–254

Hochberg Y, Tamhane AC (1987) Multiple comparison procedures, Wiley & Sons, New York

Hosmer DW, Lemeshow S (2000) Applied logistic regression analysis, Second Edition, John Wiley & Sons, New York

Howland P, Park H (2004) Generalizing discriminant analysis using the generalized singular value decomposition. IEEE Transactions on Pattern Analysis and Machine Intelligence 26: 995–1006

Humphries JM, Bookstein FL, Chernoff B, Smith GR, Elder RL, Poss S (1981) Multivariate discrimination by shape in relation to size. Systematic Zoology 30: 291–308

Jolicoeur P, Mosimann JE (1960) Size and shape variation in the painted turtle: a principal component analysis. Growth 24: 399–354

Jolicoeur P, Pirlot P, Baron G, Stephan H (1984) Brain structure and correlation patterns in Insectivora, Chiroptera, and primates. Systematic Zoology 33: 14–29

Jungers WL, Falsetti AB, Wall CE (1995) Shape, relative size, and size-adjustments in morphometrics. Yearbook of Physical Anthropology 38: 137–161

Keene ON (1995) The log transformation is special. Statistics in Medicine 14: 811–819

Kiang MY (2003) A comparative assessment of classification methods. Decision Support Systems 35: 441–454

Klingenberg CP, Froese R (1991) A multivariate comparison of allometric growth patterns. Systematic Zoology 40: 410–419

Klingenberg CP, Neuenschwander BE, Flury BD (1996) Ontogeny and individual variation: analysis of patterned covariance matrices with common principal components. Systematic Biology 45: 135–150

Manly BFJ (1997) Randomization, bootstrap and Monte Carlo methods in biology. Chapman & Hall, London

Marks S, Dunn OJ (1974) Discriminant functions when covariance matrices are unequal. Journal of the American Statistical Association 69: 555–559

Meshbane A, Morris JD (1995) A method for selecting between linear and quadratic classification models in discriminant analysis. Journal of Experimental Education 63: 263–273

Owen JG, Chmielewski MA (1985) On canonical variates analysis and the construction of confidence ellipses in systematic studies. Systematic Zoology 34: 366–374

Penin X, Baylac M (1995) Analysis of skull shape changes in apes, using 3D Procrustes superimposition. In Mardia KV, Gill CA (Eds.) Current issues in statistical shape analysis, Leeds University Press, Leeds, England, 208–210 pp

Pimentel RA (1979) Morphometrics: the multivariate analysis of biological data. Kendall-Hunt, Dubuque

Press SJ, Wilson S (1978) Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association 73: 699–705

Raudys S (2001) Statistical and neural classifiers: an integrated approach. Springer, New York

Reiser B (2001) Confidence intervals for the Mahalanobis distance. Communications in Statistics – Simulation and. Computation 30: 37–45

Ringrose TJ (1996) Alternative confidence regions for canonical variate analysis. Biometrika 83: 575–587

Ripley BD (1994) Neural networks and related methods for classification. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 56: 409–456

Rohlf FJ (1993) Relative warp analysis and an example of its application to mosquito wings. In Marcus LF, Bello E, Garcia-Valdecasas A (Eds.) Contributions to morphometrics, Museo Nacional de Ciencias Naturales, Madrid, Spain, 131–159 pp

Rohlf FJ, Bookstein FL (1987) A comment on shearing as a method for "size correction". Systematic Zoology 36: 356–367

Rohlf FJ, Slice D (1990) Extensions of the Procrustes method for the optimal superposition of landmarks. Systematic Zoology 39: 40–59

Schulte-Hostedde AI, Zinner B, Millar JS, Hickling GJ (2005) Restitution of mass-size residuals: validating body condition indices. Ecology 86: 155–163

Small CG (1996) The statistical theory of shape. Springer, New York

Smith GR (1973) Analysis of several hybrid cyprinid fishes from western North America. Copeia 1973: 395–410

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 36: 111–147

Strauss RE (1993) The study of allometry since Huxley. In Huxley JS (Ed.) Problems of relative growth, Johns Hopkins University Press, Baltimore, 47–75 pp

Strauss RE, Bookstein FL (1982) The truss: body form reconstruction in morphometrics. Systematic Zoology 31: 113–135

Thorpe RS (1983) A review of the numerical methods for recognising and analysing racial differentiation. In Felsenstein J (Ed.) Numerical taxonomy, Springer-Verlag, Berlin, 404–423 pp

Thorpe RS (1988) Multiple group principal components analysis and population differentiation. Journal of Zoology (London) 216: 37–40

Thorpe RS, Leamy L (1983) Morphometric studies in inbred and hybrid housemice (*Mus* sp.): multivariate analysis of size and shape. Journal of Zoology (London) 199: 421–432

Von Zuben FJ, Duarte LC, Stangenhaus G, Pessoa LM, dos Reis SF (1998) Bootstrap confidence regions for canonical variates: application to studies of evolutionary differentiation. Biometrical Journal 40: 327–339

Weihs C (1995) Canonical discriminant analysis: comparison of resampling methods and convex-hull approximation. In Krzanowski WJ (Ed.) Recent advances in descriptive multivariate analysis, Clarendon Press, Oxford, UK, 34–50 pp

White JW, Ruttenberg BI (2007) Discriminant function analysis in marine ecology: some oversights and their solutions. Marine Ecology Progress Series 329: 301–305

Wilcox RR (2005) Introduction to robust estimation and hypothesis testing, Academic Press, New York

Williams BK (1982) A simple demonstration of the relationship between classification and canonical variates analysis. American Statistician 36: 363–365

Willig MR, Owen RD (1987) Univariate analyses of morphometric variation do not emulate the results of multivariate analyses. Systematic Zoology 36: 398–400

Ye J, Janardan R, Park CH, Park H (2004) An optimization criterion for generalized discriminant analysis on undersampled problems. IEEE Transactions on Pattern Analysis and Machine Intelligence 26: 982–994