16. MULTIVARIATE ANALYSIS OF VARIANCE AND DISCRIMINANT ANALYSIS

In this chapter, we will examine the relationship between two or more response variables and one or more categorical predictor variables. We are primarily interested in two research questions. First, are there differences between groups based on all the response variables taken together and second, can we successfully classify observations, particularly new observations, into the correct group.

16.1. Multivariate analysis of variance (MANOVA)

There are many situations where we record more than one response variable from each sampling or experimental unit and where these units are allocated to or occur in treatment groups. Ecologists often record the abundances of many species from each sampling or experimental unit and physiologists commonly measure more than one variable (e.g. blood pressure, heart rate etc.) on experimental animals. For example, Peckarsky *et al*. (1993) examined the sub-lethal responses of mayfly larvae in streams to three different predator treatments (no predator and normal food, no predator and reduced food, one predatory mayfly (*Megarcys*) and normal food). There were five response variables recorded for each mayfly: body mass, egg mass, % eggs, total mass, maturation time. Botanists and zoologists also often measure many morphological variables when describing organisms from different locations or to compare organisms that may or may not be taxonomically different.

If each response variable is of inherent biological interest, our research question might be whether there are group or treatment effects on each variable separately. Then the appropriate strategy is to analyse each variable using a separate univariate ANOVA to test for differences between groups. Some statisticians have argued that there is an inherent disadvantage to this approach. Because the response variables are measured from the same experimental or sampling units and may be highly correlated, the multiple ANOVA tests are not independent of each other and this can make interpretation difficult. Also, the number of univariate tests can get large if we have many variables so the familywise Type I error rate may be very high for the collection of tests (Harris 1993; see also Chapter 3). A common recommendation is to adjust the significance level of each ANOVA test by using a Bonferroni-type correction so the familywise Type I error rate stays at or below 0.05 (or whatever *a priori* significance level you choose). Unfortunately, with many response variables, this can result in unacceptably low power for each univariate test.

With multiple response variables, we might be more interested in whether there are group differences on all the response variables considered simultaneously. This is the aim of multivariate analysis of variance (MANOVA), the analogue of univariate ANOVA when we have multiple response variable for each experimental or sampling unit. Basically our hypothesis is now about group effects on a combination of the response variables and instead of comparing group means on a single variable, we now compare group centroids for two or more variables. In the Peckarsky *et al*. (1993) example, we would test whether there is an effect of predator treatment on a combination of body mass, egg mass, % eggs, total mass, and maturation time of individual mayflies.

We will illustrate MANOVA with two examples from the biological literature.

Trace metals in marine sediments

Haynes *et al*. (1995) carried out a pilot study to test for differences between sites in trace metal concentrations in marine sediments off the Victorian coast in southern Australia. They had three sites: Delray Beach, site of a proposed wastewater outfall, and two possible control sites, Seaspray and Woodside. At each site, they had four randomly chosen stations and at each station, two randomly chosen cores of sediment. They recorded the concentrations of copper, chromium, cadmium, lead, iron, nickel, manganese and mercury. We will test for

the effects of site on a subset of these response variables taken together. Although this is strictly a nested design, site would be tested against the random station effect so we will average the replicate cores for each station and use a single factor MANOVA for comparing sites. The analysis of these data is presented in Box 16-1.

Plant functional groups and leaf characters

In Chapters 9 and 15, we described the study of Reich *et al*. (1999) who examined the generality of leaf traits from different species across a range of ecosystems and geographic regions. We will analyze a subset of their data (to avoid missing cells), with two locations (Colorado and Wisconsin) and two functional groups (forbs and shrubs) in a crossed design. There were between three and eleven species in each cell and five response variables were measured: specific leaf area ($\log_{10}$ transformed), leaf nitrogen concentration, mass-based net photosynthetic capacity, area-based net photosynthetic capacity and leaf diffusive conductance at photosynthetic capacity. We will test for the effects of location and functional group, and their interaction, on these five response variables taken together. The analysis of these data is presented in Box 16-2.

16.1.1. Single factor MANOVA

Linear combination

The simplest design where a MANOVA is appropriate is when we have *n* replicate experimental or sampling units ("objects" from Chapter 15) allocated to two or more levels of a factor (groups) and we record *p* (where *p* is greater than two) response variables from each unit. The MANOVA is based on a linear combination (*z*) of the *p* response variables as defined in Chapter 15 (see equations 15.1 and 15.2). In the example from Haynes *et al*. (1995), there were *n* equals four replicate stations in each of three groups (sites) with *p* equals four response variables (trace metals). The MANOVA uses the linear combination (*z*) of response variables, out of the infinite number of possible linear combinations, which maximizes the ratio of between-group and within-group variances of *z*. This linear combination is also called the discriminant function for the difference between groups and is used in discriminant function analysis (see Section 16.2):

$$z_{ik} = \text{constant} + c_1 y_{i1} + c_2 y_{i2} + \ldots c_j y_{ij} + \ldots + c_p y_{ip} \qquad (16.1)$$

For example, from Haynes et al. (1995):

$$z_{ik} = \text{constant} + c_1(\log_{10}\text{Cu})_i + c_2(\log_{10}\text{Pb})_i + c_3(\log_{10}\text{Ni})_i + c_4(\log_{10}\text{Mn})_i \qquad (16.2)$$

From Reich et al. (1999):

$$z_{ik} = \text{constant} + c_1(\log_{10}\text{ specific leaf area})_i + c_2(\text{leaf N})_i + c_3(\text{mass-based photosynthetic capacity})_i + c_4(\text{area-based photosynthetic capacity})_i + c_5(\text{leaf diffusive capacity})_i \qquad (16.3)$$

In equations 16.1, 16.2 and 16.3, z_{ik} are the values for object *i* for linear combination *k*, the combination that maximises the ratio of between-group and within-group variances of z_{ik}. From Haynes et al. (1995), this is the value for station *i* from solving equation 16.2 for linear combination *k*. The coefficients (c_i) are the weights measuring the relative contribution of each variable to the linear combination. As described in Box 15.1, these coefficients will be scaled in some form and will be represented in matrix descriptions of MANOVA as elements of a matrix of eigenvectors (Box 15.1). Note that if the variables are standardized to zero mean and unit variance, the constant equals zero.

The determination of the linear combination that maximizes the ratio of between- and within-group variances is best done using simple matrix algebra, some of which we have already described in Chapter 15. The steps for a single factor MANOVA are:

1. The Between-groups, Within-groups and Total SS used in an ANOVA are replaced by sums-of-squares-and-cross-products matrices (SSCP or **S**; see Chapter 15), one matrix for between groups (the hypothesis or effect matrix, **H**), one for within groups (the error or residual matrix, **E**) and one for total (the total matrix, **T**). The values in the main

diagonal of these matrices are the univariate sums-of-squares for each variable, either between group means (**H**) or pooled across replicates within groups (**E**). The other elements are the sums-of-cross products between any two of the variables. For example, the cross product for the between-groups matrix for two variables is the sum of (i) the product of the differences between each group mean and the overall mean for one variable and (ii) the differences between each value and the mean for the other variable – see Table 16-1.

2. We multiply **H** by the inverse of **E** (i.e. $\mathbf{HE}^{-1}$). Matrix inversion is the multivariate analogue of division so what we are really doing here is "dividing" **H** by **E**, the between-groups SSCP matrix "divided by" the within-groups SSCP matrix.

3. We then decompose the resulting matrix product (Box 15.1) to calculate characteristic roots or eigenvalues of each linear combination (eigenvector). The eigenvalues measure how much of the total between-group variance in the variables (the sum of the between-group variances of each of the variables) is explained by each linear combination or eigenvector. The eigenvectors contain the coefficients for each linear combination.

4. The linear combination producing the largest eigenvalue is the linear combination that maximizes the ratio of between-group and within-group variance (i.e. maximizes the explained variance between groups) and the eigenvector is a vector of coefficients or weights for that linear combination.

Null hypothesis

The H_0 for a single factor MANOVA is that the population effect of the groups or treatments is zero with respect to all linear combinations of the response variables. This is equivalent to no difference between population centroids (multivariate means). This H_0 can be tested by using statistics based on one of the measures of variance of a matrix, such as the determinant or the trace (Chapter 15; see also Harris 1985, Johnson & Field 1993, Stevens 1992, Tabachnick & Fidell 1996):

- Wilk's lambda (l), which is the ratio of the determinants of the within-groups SSCP and the total SSCP: $\frac{|\mathbf{E}|}{|\mathbf{T}|}$. Remember that the determinant of a matrix is a measure of generalized variance for that matrix (Chapter 15), so Wilk's l is a measure of how much of the total variance is due to the residual, with smaller values indicating larger group differences.

- Hotelling-Lawley trace, which is the ratio of the determinants of the between-groups SSCP and the within-groups SSCP: $\frac{|\mathbf{H}|}{|\mathbf{E}|}$. This is also the sum of the eigenvalues (trace) of the matrix product $\mathbf{HE}^{-1}$. Larger values indicate greater differences between group centroids.

- Pillai trace, which is the sum of the eigenvalues (trace) of $\mathbf{HT}^{-1}$, i.e. the variance between groups.

- Roy's largest root, which is the largest eigenvalue of $\mathbf{HE}^{-1}$, i.e. the eigenvalue of the linear combination that explains most of the variance and covariance between groups. This statistic is less commonly provided by statistical software.

The sampling distributions of these statistics are not well understood and they are usually converted to approximate *F*-ratio statistics (Tabachnick & Fidell 1996). Wilk's, Hotelling's and Pillai's statistics produce identical *F* tests when there are only two groups and become Hotelling's T^2 statistic – see example based on plant functional group data in Box 16-2. This is the multivariate extension of the *t* test for comparing two groups (Harris 1985, Tabachnick & Fidell 1996). They will generally produce similar results with more than two groups,

although Pillai's trace seems to be the most robust of the tests (Johnson & Field 1993), especially when the assumption of similar variance-covariance matrices might be violated (Section 16.1.4). In our two worked examples (Box 16-1 and Box 16-2), the conclusions from Wilk's, Hotelling's and Pillai's statistics were the same. Most statistical software will provide **H**, **E**, maximum l and all the multivariate test statistics with their approximate F tests.

16.1.2. Specific comparisons

Most statistical software allow contrasts among the factor levels in MANOVA, analogous to planned contrasts in the univariate ANOVA (Chapter 8). Unplanned multiple comparisons are a more difficult problem, although use of Bonferroni-adjusted (see Chapter 3) pairwise MANOVAs is one conservative solution. Harris (1993) and Johnson & Field (1993) have reviewed other approaches for comparing specific groups after a MANOVA.

16.1.3. Relative importance of each response variable

If the null hypothesis of no difference between group centroids is rejected, we usually are interested in which of the response variables contributes most to the group differences. There are several methods of assessing the relative contribution of each response variable to the difference between groups in a MANOVA.

Univariate ANOVAs

We can examine the univariate ANOVAs on each response variable separately. Indeed, univariate hypotheses about group differences for each response variable will often be relevant. These univariate results do not necessarily indicate the relative contribution of each variable to the MANOVA result because they ignore correlations between variables. Correlations between variables can have marked effects on the power of MANOVA tests (Cole *et al*. 1994). Some authors (e.g. Harris 1985, 1993) also emphasize the problem of increasing familywise Type I error rates when doing multiple univariate ANOVAs, a problem inherent in any multiple testing situation (see Chapter 3).

Step-down analysis

Step-down analysis is an analogue of forward selection stepwise multiple regression (Chapter 6) but taking into account the group structure (Tabachnick & Fidell 1996). This procedure relies on ordering the response variables based on theoretical expectations of their importance or using univariate analyses to choose the variable that shows the greatest difference between groups. First, the response variable with the highest priority is decided; for example, this might be the variable with the largest F-ratio from univariate ANOVAs on all the response variables. Each response variable is then tested sequentially, in the order determined *a priori*, in an ANCOVA model (Chapter 12), with groups as the categorical predictor and the higher priority response variables as covariates. We are interested in how much each additional variable adds to variance explained by the variables already included.

Automated step-down analysis is available in some statistical software; otherwise, it must done with a series of ANCOVAs. Step-down analysis suffers from the problems we described in Chapter 6 for stepwise multiple regression, although we are not trying to find the "best" model in this situation, just assess the relative importance of each of the response variables. Step-down analysis also results in numerous unplanned significance tests so you need to be aware of the high familywise Type I error rate. Huberty (1994) describes similar approaches, such as deleting variables one at a time and running a MANOVA on each set of the p-1 remaining variables. The variables can be ordered based on the size of the change in MANOVA test statistic for each set.

Coefficients of linear combination

A more subjective approach is based on examining the discriminant function, i.e. the linear combination of the response variables that maximizes the ratio of between- to within-group variance. There is a coefficient for each variable in the discriminant function, plus one for

the grand mean (i.e. intercept or constant). If the different variables are measured on comparable scales (or we have values of a single variable recorded repeatedly through time in a repeated measures design), then the relative size of these coefficients (also termed "weights") provides a comparable measure of the contribution of each variable to the variance explained by the discriminant function and thus the difference between groups. If the variables are measured on very different scales, then we need to standardize them so that the coefficients can be compared. The simplest method is to standardize the discriminant function by the within-group variances, although whether this produces coefficients that are directly comparable is debatable (see Harris 1985 and Huberty 1994 for differing opinions). Standard errors can be estimated for each discriminant function coefficient (Flury & Riedwyl 1988), although they are rarely provided by statistical software.

Loadings

Loadings are the correlations between each variable and the discriminant function (see Chapters 15 and 17). These simply represent the correlations between the value of a variable and the score for the discriminant function with the units as replicates. The loadings of each variable on each discriminant function can be found by multiplying the within-group correlation matrix between variables (pooled across groups) by the matrix of standardized discriminant function coefficients (Tabachnick & Fidell 1996). Correlations automatically standardize the variables and examining loadings is popular because correlation coefficients are familiar and easily interpretable. However, these loadings are directly proportional to the univariate *F*-ratio statistics for each variable tested between groups so they ignore any relationships between the variables (Harris 1985). Note that one of the effects of highly correlated response variables can be a contradictory pattern when coefficients are compared with loadings.

Comments

Most statistical software will provide all of these coefficients and the loadings, either in MANOVA output or as part of a discriminant function analysis. The terminology used in the output does, however, vary considerably between programs and Tabachnick & Fidell (1996) provide a detailed comparison of the major software. Our experience is that unless there are many variables with some high correlations, the different approaches will produce a similar pattern. In our worked example of trace metals in sediments (Haynes et al. 1995; Box 16-1), the variables were log transformed but not standardized. The univariate *F*-ratios, loadings and function coefficients showed the same pattern, with the order of importance being log Mn, followed by log Ni, log Cu and log Pb. The step-down analysis showed that none of the variables contributed significantly to site differences besides log Mn.

16.1.4. Assumptions of MANOVA

It is important to check normality, homogeneity of variance, and outliers for each response variable using univariate exploratory data analysis procedures (boxplots, residual plots, pplots etc.; see Chapter 4). Given that the multivariate tests (especially Pillai's trace) are relatively robust to deviations from multivariate normality, particularly if each response variable has approximate univariate normality and sample sizes are equal, two multivariate assumptions are of major concern (Johnson & Field 1993, Tabachnick & Fidell 1996).

First, MANOVA tests are sensitive to multivariate outliers, which are cases with an unusual pattern of values for all the response variables considered simultaneously. Mahalanobis distance, the distance of each observation from the centroid or multivariate mean, can be used to detect multivariate outliers and is provided by most statistical software (Chapter 15).

Second, homogeneity of variances and covariances (i.e. equality of the variance-covariance matrices for each group) is an important assumption - this is the multivariate extension of univariate homogeneity of within-group variances. If this assumption is not met, then the pooled within-group matrix (**E**) will be misleading. Box's *M* test can test the H_0 of equal variance-covariance matrices but it is very sensitive to deviations from multivariate normality and is not recommended. There is no easy check for this assumption (but see

discussion in Johnson & Field 1993), although it is more likely to be met when univariate homogeneity holds for each response variable. Like univariate ANOVA tests, MANOVA tests are more reliable when sample sizes are equal. Reducing the dimensionality (reducing the number of variables) of the analysis improves the robustness of all the MANOVA tests statistics (Johnson & Field 1993).

Johnson & Field (1993) provided strong evidence from simulation studies that Pillai's trace statistic is the most robust to deviations from the assumption of homogeneity of the variance-covariance matrices across groups. Suitable transformations of individual variables should also be used where appropriate and including quadratic terms in the discriminant function can also help (Section 16.2.3).

Collinearity between variables is also a problem in the same way as for multiple regression (Chapter 6). The discriminant function coefficients, just like multiple regression coefficients, will be sensitive to which variables are included or excluded when variables are highly correlated. Most statistical software provides collinearity diagnostics, such as tolerance or variance inflation factors, and examinations of pairwise correlations between variables will be informative. Not including highly correlated (redundant) variables will help lessen the impact of collinearity and, since it reduces the dimensionality of the data matrix, will also make the MANOVA more robust to heterogeneous variance-covariance matrices (see above).

16.1.5. Robust MANOVA

Approaches to MANOVA that are robust to the underlying assumptions of multivariate normality and homogeneity of the variance-covariance matrices have been based on randomization procedures (Johnson & Field 1993). Edgington (1995) and Manly (1997) describe numerous possible test statistics for randomization MANOVA tests. These include a test based on Wilk's lambda, one using the sum of the logs of the univariate t or F statistics for each variable and one that compares the sum of squared Euclidean distances between objects and their sample centroids between groups and within groups. Manly (1997) pointed out that with large data sets (many variables and/or observations), only a subsample of all possible randomizations of observations on all variables to groups will be possible.

Another type of test is to determine distances or dissimilarities between all pairs of objects and compare the between groups and within groups dissimilarities. These tests will be described in Chapter 18 when we consider multivariate analyses based on dissimilarities in detail.

16.1.6. More complex designs

MANOVAs can also be used to test null hypotheses about combinations of variables in more complex designs. The matrix calculations described in Section 16.1.1 are done for each effect and error term that would have been used if univariate ANOVA models were fitted (Harris 1985). Separate linear combinations of variables are thus constructed for each main effect and interaction (Box 16-2) and the contribution of each variable needs to be assessed separately for each effect and its appropriate discriminant function. An ecological example is from Juenger & Bergelson (2000), who studied interactions between herbivory and pollination on various aspects of reproduction in the perennial wildflower *Ipomopsis aggregata* ssp. candida (the scarlet gilia) in Colorado, USA. Their experimental design had two factors: artificial grazing or clipping (two levels: control vs. experimentally clipped) and male function (two levels: control vs. emasculation, i.e. anther removal). There were 20 replicate plants in each combination (cell) of the two factor crossed design and a number of response variables were measured for each plant: total number of flowers, fruits and undamaged seeds and total seed mass. They used Wilk's lambda to test the two main effects and the interaction effect on the combination of the four response variables (Table 16-2).

A more complex factorial MANOVA was used by Pennings & Calloway (1996), who studied the effects of a parasitic plant (*Cuscuta salina*) on a saltmarsh community. They set up an experiment with three factors: *Cuscuta* infection, zone within saltmarsh and size of

patch. They recorded the biomass of three non-parasitic plant species and analyzed these three response variables with a three factor (infection, marsh zone, patch size) crossed MANOVA and tested the hypotheses for each main effect and interaction using Pillai's trace statistic (Table 16-3).

Note that one of the commonest applications of MANOVA in biology is in the analysis of repeated measures designs (Chapters 10 & 11), where the differences between pairs of repeated measurements are analyzed as multiple response variables using MANOVA statistics.

16.2. Discriminant function analysis

Discriminant function analysis (DFA) is a "classification" technique, introduced by Fisher (1936) and recently reviewed by Huberty (1994). DFA is used when we have observations from pre-determined groups with two or more response variables recorded for each observation. DFA generates a linear combination of variables that maximizes the probability of correctly assigning observations to their pre-determined groups and can also be used to classify new observations into one of the groups. We might also wish to classify new stations to a site based on these variables and have some measure of the likelihood of success of our classification. Examples of DFA are common in the biological literature. For example, Skelly (1995) used DFA to test how well three variables (survivorship, size and larval period) could be used to classify individuals of two species of frogs (chorus frogs and spring peepers). Petit & Petit (1996) used DFA to separate four habitats based on ten variables (canopy cover, canopy height, density of various stem sizes) measured around nest boxes occupied by warblers along the Tennessee River.

We will illustrate DFA with the same two data sets we used for MANOVA.

Trace metals in marine sediments

We will analyze the data from Haynes et al. (1995), previously used in Box 16-1 for a MANOVA, with a discriminant function analysis. Our aim is to classify stations to each of the three sites (Delray Beach, Seaspray, Woodside) based on trace metal concentrations in marine sediments off the Victorian coast in southern Australia. The DFA of these data are in Box 16-3.

Plant functional groups and leaf characters

We will also analyze the data from Reich et al. (1999), used in Box 16-2 for a MANOVA, with a discriminant function analysis to classify species into one of four location and plant functional group combinations (Colorado-forb, Colorado-shrub, Wisconsin-forb, Wisconsin-shrub) based on five response variables: specific leaf area ($\log_{10}$ transformed), leaf nitrogen concentration, mass-based net photosynthetic capacity, area-based net photosynthetic capacity and leaf diffusive conductance at photosynthetic capacity. The DFA of these data are in Box 16-4.

16.2.1. Description and hypothesis testing

Discriminant function analysis (DFA) is mathematically identical to a single factor MANOVA, although the former emphasizes classification and prediction rather than tests of hypotheses about group differences. However, the first step in any DFA is to derive discriminant functions (also called canonical discriminant functions) that are linear combinations of the original variables. The first discriminant function is the linear combination of variables that maximizes the ratio of between-groups to within-groups variance (i.e. maximises the differences between groups) and is the linear combination used for the MANOVA test of no differences between group centroids derived in Section 16.1.1. The second discriminant function is independent of (uncorrelated with) the first and best separates groups using the variation remaining (the residual variation) after the first discriminant function has been determined, and so on for the third, fourth etc. discriminant functions.

The number of discriminant functions that can be extracted depends on the number of groups and the number of variables - it is the lesser of the degrees of freedom for groups (number of groups minus one) and the number of variables (Tabachnick & Fidell 1996). In the example from Haynes *et al*. (1995), with three groups (sites) and four variables, there can be only two discriminant functions. Even in situations when there are more functions, the first one or two usually have the most discriminating power. Most statistical software also provides eigenvalues (how much of the between-group variance is explained by each function) and the proportion of total variance explained.

Determining which variables contribute most to discriminant functions, and therefore to group separation, is done in the same way as for MANOVA (Section 16.1.3). The relative sizes of the standardized coefficients for each discriminant function indicate which variables are more important to each discriminant function. Also useful are loadings, which measure the correlation between each variable and each discriminant function, although they ignore any correlation between variables. With a large number of variables, stepwise discriminant function analysis can be used, similar to stepwise multiple regression. The stepwise approach enters and removes variables in a model-building process to try and produce a discriminant function with only the "important" variables. Our criticisms of stepwise procedures (see Chapter 6) are just as applicable here and we do not recommend stepwise discriminant analysis.

The test of the H_0 of no difference between group centroids (MANOVA) is usually the first step in a discriminant function analysis because if it is not significant, the discriminant functions will not be very useful for separating groups and therefore classifying observations. Successive discriminant functions can be tested for significance (the second one is tested after the first has been extracted) using the MANOVA tests described in Section 16.1.1.

We can calculate discriminant function scores (z_i) for each observation on each function by simply solving each discriminant function as in equation 16.1. These scores can be used in a linear discriminant function (LDF) plot (Huberty 1994), with the first discriminant function scores on one axis and the second discriminant function scores on the other axis. Either individual observations or centroids can be plotted. These plots indicate subjectively how similar or different groups are in terms of the discriminant functions. For example, there was a clear separation between sites from Haynes et al. (1995) when the first two discriminant functions were plotted (Figure 16-1), although most of the difference was for function one.

LDFs can also be presented as biplots where the loadings (correlations) of each variable on each function are plotted as vectors, scaled so that the vectors are commensurate with the scale of functions scores. The direction of the vectors indicates an increase in the values of the variable towards those objects in that direction on the plot, and the length of the vector indicates the rate of increase. Biplots will be explored in more detail in Chapter 17.

16.2.2. Classification and prediction

The second purpose of a DFA is to classify each observation into one of the groups and assess the success of the classification. A classification equation is derived for each group and is a linear combination of variables like a discriminant function, including a constant (equation 16.1):

For example, for Delray using the variables from Haynes et al. (1995):

$$C_{Delray} = \text{constant} + c_1(\log_{10} \text{Cu}) + c_2(\log_{10} \text{Pb}) + c_3(\log_{10} \text{Ni}) + c_4(\log_{10} \text{Mn}) \quad (16.4)$$

There are four steps in determining and using the classification function for any group:

- The coefficients (c) of the classification equation are termed classification coefficients (Tabachnick & Fidell 1996) and are found by multiplying the within-group covariance matrix (pooled across all groups) by the matrix of means for each variable for that group.

- The constant for a group is determined by multiplying the matrix of classification coefficients for that group (i.e. the coefficients for each variable) again by the matrix of means for each variable for that group.
- A classification score for each observation for each group is then calculated by using the actual values for each variable to solve the classification equation for that group.
- Each observation is formally classified into the group for which it has the highest score. This may or may not be the actual group from which the observation came.

Tabachnick & Fidell (1996) have provided a fully worked example of the calculations and Huberty (1994) has a more detailed theoretical background.

Discriminant analysis routines in most statistical software provide classification matrices, that indicate to which group each observation was classified and whether that classification was correct. The success of classifications of observations will be greater if the groups were clearly distinguishable on the first discriminant function. For example, the stations from Haynes *et al*. (1995) were clearly separable into groups (highly significant MANOVA) and the classification success was also high (Box 16-3). In contrast, there was no significant separation of groups in the Reich *et al*. (1999) data and the classification success was lower (e.g. only six out eleven species correctly classified as being from forbs from Wisconsin – Box 16-4)

One difficulty with the classification methodology we just described is that the classification functions are calculated using all observations and these functions are then used to classify the same observations, i.e. we classify each observation with an equation that already used that observation. One way of avoiding the resulting inherent bias is to use a jackknife procedure (Chapter 2). The classification of each observation is based on group classification functions that are determined when the observation is omitted and only the remaining observations are used to calculate coefficients and constants. In our examples, the jackknife classifications were less successful, but probably more robust, than the usual classifications using all observations (Box 16-3 andBox 16-4). The biggest difference between the usual and jackknifed classifications will often be for groups with the smallest sample size, again illustrated for the classification of the Reich *et al*. (1999) data where our classification success for Colorado shrubs went from 75% to 0% when the jackknife approach was used (Box 16-4).

Most uses of DFA we have found in the biological literature have focused on description and hypothesis testing, rather than classification. For example, Petit & Petit (1996) derived three discriminant functions to separate four habitats based on ten variables (canopy cover, canopy height, density of various stem sizes) measured around nest boxes occupied by warblers along the Tennessee River. They found that the first function explained 96.7% of the variance and canopy cover was the variable most highly correlated (loading = 0.84) with this first discriminant function. Skelly (1995) used a number of discriminant function analyses in his study of tadpole behaviour and performance. In one, he tested how well three variables (survivorship, size, larval period) classified individuals into one of two species of frogs (chorus frogs and spring peepers). He presented a single discriminant function, which significantly separated the two species (MANOVA). Larval period had the highest coefficient (1.015) and loading (0.76) for this function, i.e. larval period separated the species more than size and survivorship (Table 16-4).

16.1.2. Assumptions of discriminant function analysis

DFA has the same assumptions as MANOVA (Section 16.1.4). The most important of these assumptions is homogeneity of the within-group variance-covariance matrices, especially for the classification part of discriminant analysis because this is quite sensitive to heterogeneous variance-covariance matrices between groups. This assumption is very difficult to test formally and Tabachnick & Fidell (1996) suggested plotting the scores for each observation for the first two discriminant functions (e.g. Figure 16-1) and checking if

the spread of points is similar among the groups. Transformations of variables will often help.

If there is clear heterogeneity across the within-group variance-covariance matrices, you can try fitting quadratic functions instead of the usual linear ones. Quadratic functions include coefficients for squares of the variables and do not assume equal within-group covariances; statistical software usually offers quadratic functions as an option. Quadratic terms are usually highly correlated with the linear term for the same variable. This can result in collinearity problems (Section 16.1.4) and centered variables may need to be used (Chapter 6).

16.1.3. More complex designs

Because DFA is identical to a MANOVA, DFA can be extended to more complex designs, such as factorial designs, as described in Section 16.1.6. However, when focusing on classification, we usually treat each combination of factor levels (cell) as a separate group and use methods developed for single factor designs.

16.2. MANOVA vs Discriminant function analysis

MANOVA and DFA are mathematically identical (Tabachnik & Fidell 1996), although the terminology used in the two procedures often differs. In MANOVA, we test whether population centroids, based on a number of response variables, are different between groups. In DFA, we use the response variables to try and predict group membership and also to classify new observations to one or other of the groups with some measure of success of that classification. The linear combination of variables that maximizes the ratio of between-group to within-group variation in MANOVA is the first discriminant function. Discriminant function analysis goes further than MANOVA, however, by calculating additional discriminant functions and using the functions to classify observations to groups.

16.3. General issues and hints for analysis

General issues

- MANOVA can be used to analyze any design where there is more than one response variable and one or more categorical predictor variables and the question of interest concerns the response variables considered simultaneously.
- MANOVA is also used when analyzing partly nested models for "repeated measures" designs where the differences between levels of the within-subjects factor are treated as multiple response variables.
- Although checking the assumptions is more difficult for multivariate analyses compared with univariate analyses, the former are also more sensitive to departures from the assumptions.
- MANOVA and DFA are functionally equivalent, the former emphasizing between-group differences on a single discriminant function, the latter using more than one discriminant function and focusing on classification.

Hints for analysis

- Homogeneity of between-group variances and covariances is important. Keep sample sizes similar and at least ensure homogeneity of variances for each variable separately. Check for outliers with Mahalanobis distance, tested against a c^2 distribution with p df and a strict significance level (0.001).
- Pillai's trace is the most robust of the test statistics for MANOVA and is recommended.
- The contribution of each variable to a discriminant function is best measured by the standardized coefficients.

- Loadings for each variable on each discriminant function ignore correlations between variables and will have the same pattern between groups as the univariate *F* tests for each variable
- Jackknifed classifications of each observation to each group are probably more reliable than standard classifications because the former do not include the observation being classified when calculating the classification score.

Box 16-1 Worked example of MANOVA: heavy metals in marine sediments

Haynes *et al*. (1995) carried out a pilot study to test for differences between sites in trace metal concentrations in marine sediments off the Victorian coast in southern Australia. They had three sites: Delray Beach, site of a proposed wastewater outfall, and two possible control sites, Seaspray and Woodside. At each site, they recorded the concentrations of copper, chromium, cadmium, lead, iron, nickel, manganese and mercury (means of two sediment cores) at four randomly chosen stations. We used only the 1991 data in our analyses. There were strong correlations among some of the metals (e.g. Cu & Cr, Fe & Ni) so only four variables (Cu, Ni, Pb, Mn) were included in the analysis. There was strong indication of skewness for the four variables, so all were $\log_{10}$ transformed. There were a few cases with significant (P<0.001) Mahalanobis distances (D_{ij}^2 > 16.3) but these were not extreme and remained in the analysis. All variables except Cu (Levene's test, P = 0.023) had similar variances between groups.

The multivariate test statistics all result in rejection of the H_0 that there is no difference in site group centroids:

	Statistic	df	*F*	*P*
Wilk's *l*	0.058	8, 12	4.728	0.008
Pillai trace	1.272	8, 14	3.058	0.033
Hotelling-Lawley trace	10.549	8, 10	6.593	0.004

Pairwise contrasts among the sites, with a sequential Bonferroni (Holm's method) adjustment of *P* values, indicated that only the difference between Delray Beach and Woodside were significant.

Contrast	Pillai trace	df	*F*	*P*	Adj *P*
Delray vs Seaspray	0.713	4, 6	3.719	0.074	0.078
Delray vs Woodside	0.909	4, 6	14.924	0.003	0.009
Seaspray vs Woodside	0.772	4, 6	5.092	0.039	0.078

The univariate *F* tests indicate significant differences between sites for all four metals:

Source	df	MS	*F*	*P*
Log Cu				
Site	2	0.098	5.208	0.031
Residual	9	0.019		
Log Pb				
Site	2	0.136	4.834	0.038
Residual	9	0.028		
Log Ni				
Site	2	0.083	8.655	0.008
Residual	9	0.009		
Log Mn				
Site	2	0.244	23.608	<0.001
Residual	9	0.010		

The raw and standardized coefficients for the discriminant function obviously differ but because the variables were log transformed, the difference in scales between the variables is not great the and the basic pattern is the same. The standardized coefficients suggest that lead contributes least to the difference between sites and manganese the most. The loadings simply reflect the univariate *F*-ratio statistics from above, and the pattern is the same as for the coefficients. Mn and Ni are most important, and Pb least important, at separating the sites.

Variable	Raw coefficient	Standardized coefficient	Loading
Constant	-29.013	0	
Log Cu	1.253	0.172	0.334
Log Pb	-0.494	-0.083	0.258
Log Ni	6.690	0.653	0.428
Log Mn	9.308	0.945	0.724

We had no theoretical basis for ordering our variables so we entered them in a step-down analysis in order of their univariate *F*-ratios. Log Mn entered first, then we tested log Ni with log Mn as a covariate, then log Cu with log Mn and log Ni as covariates and finally log Pb with log Mn, log Ni and log Cu as covariates. We were not interested in testing hypotheses about the covariates and adjusted the significance levels for the site effects with a Holm correction (Chapter 3).

Source	df	MS	*F*	Adj *P*
Log Mn				
Site	2	0.244	23.608	<0.001 (0.004)
Residual	9	0.010		
Log Ni				
Site	2	0.034	3.407	0.085 (0.255)
Log Mn	1	0.007		
Residual	8	0.010		
Log Cu				
Site	2	0.011	0.512	0.620 (0.910)
Log Mn	1	<0.001		
Log Ni	1	0.023		
Residual	7	0.021		
Log Pb				
Site	2	0.033	0.901	0.455 (0.910)
Log Mn	1	0.021		
Log Ni	1	0.022		
Log Cu	1	0.002		
Residual	6	0.037		

The step-down analysis suggests that none of the variables contributes significantly to the difference between groups when entered after log Mn, i.e. none of the sitee effects for any variable are significant once log Mn is included as a covariate.

Box 16-2 Worked example of MANOVA: plant functional groups and leaf characters

Reich *et al*. (1999) examined the generality of leaf traits from different species across a range of ecosystems and geographic regions. We will use two of their locations (Colorado and Wisconsin) and two of their functional groups (forbs and shrubs) in a crossed design. There were between three and eleven species in each cell and five response variables were measured: specific leaf area ($\log_{10}$ transformed), leaf nitrogen concentration, mass-based net photosynthetic capacity, area-based net photosynthetic capacity and leaf diffusive conductance at photosynthetic capacity.

There is some concern about the assumption of homogeneity of variances and covariances, especially as Levene's test for homogeneity of variances was statistically significant for three (log specific leaf area, leaf N and G_s) out of the five variables.

There were no significant multivariate test statistics for either main effect or the interaction. Note that since there were only two levels of each factor, the df, the approximate *F*-ratios and the *P* values were identical for each term for all three multivariate statistics:

	Wilk's *l*	Pillai trace	Hotelling-Lawley trace	df	*F*	*P*
Location	0.573	0.427	0.745	5,16	2.384	0.085
Functional group	0.549	0.450	0.819	5,16	2.622	0.065
Interaction	0.836	0.164	0.196	5,16	0.626	0.682

The univariate *F* tests indicate that the only significant effect was that of functional group for nitrogen concentration in leaves, although the effect of functional group for mass-based net photosynethetic capacity and of location for leaf diffusive conductance were marginal:

Source	df	*F*	*P*
Location	1, 19		
Log specific leaf area		0.880	0.359
Leaf N		0.005	0.947
A_{Mass}		1.025	0.323
A_{Area}		0.042	0.841
G_S		3.756	0.069
Functional group	1, 20		
Log specific leaf area		2.299	0.145
Leaf N		5.305	0.032
A_{Mass}		3.254	0.086
A_{Area}		1.148	0.297
G_S		2.645	0.119
Interaction	1, 20		
Log specific leaf area		1.979	0.175
Leaf N		0.774	0.389
A_{Mass}		1.624	0.217
A_{Area}		0.065	0.802

G_S	1.112	0.304

The standardized discriminant function coefficients for each main effect and interaction would not normally be of much interest given that there were no significant effects from the MANOVA. We present them simply to illustrate that there is a separate discriminant function for each effect in the model and we can interpret these coefficients just as we would for single factor MANOVAs.

Variable	Location	Functional group	Interaction
Log specific leaf area	-2.002	-1.721	1.309
Leaf N	1.798	1.499	-1.294
A_{Mass}	0.612	1.409	0.479
A_{Area}	-1.489	-1.615	0.338
G_S	1.436	1.472	-0.709

Box 16-3 Worked example of discriminant function analysis: trace metals in marine sediments

We will illustrate a discriminant function analysis using the data from Haynes *et al*. (1995) - see Box 16-1. The aim here is to try and predict site membership of stations based on the four variables recorded for each station. The variance explained by each discriminant function was:

	Eigenvalue	% of variance
Function 1	9.979	94.6
Function 2	0.570	5.4

The first discriminant function explains most of the between-group (between-site) variance. The MANOVA test showed a significant difference between sites in the first discriminant function (Pillai trace = 1.272, df = 8, 14, *F*-ratio = 3.058, *P* = 0.033; see Box 16-1).

The relative contributions of each of the four trace metals to each discriminant function were:

	Raw coefficient		Standardized coefficient		Loading	
	1	2	1	2	1	2
Constant	-29.013	-0.822	0	0		
Log Cu	1.253	3.030	0.172	0.415	0.334	-0.271
Log Pb	-0.494	-5.042	-0.083	-0.847	0.258	0.845
Log Ni	6.690	-3.126	0.653	-0.305	0.428	0.409
Log Mn	9.308	2.864	0.945	0.291	0.724	-0.159

The general pattern is the same for raw and standardized coefficients and loadings. Manganese and nickel contribute the most to the first function (Box 16-1) whereas lead contributes most to the second function. Note that within a discriminant function, the direction of the sign for each variable is arbitrary, i.e. the positives and negatives could be reversed with no change in interpretation.

The classification functions for each site are:

	Delray	Seaspray	Woodside
Constant	-339.675	-421.174	-534.398
Log Cu	14.723	14.191	22.851
Log Pb	-24.237	-18.519	-27.090
Log Ni	171.225	195.893	216.269
Log Mn	258.338	282.345	320.356

These classification functions were solved for each station and each station classified to the site with the highest value:

	Delray	Seaspray	Woodside	% correct
Delray	4	0	0	100
Seaspray	0	4	0	100
Woodside	0	0	4	100
Total	4	4	4	100

Note that % successful prediction is perfect. The classification matrix produced using a jackknife technique was:

	Delray	Seaspray	Woodside	% correct
Delray	3	1	0	74
Seaspray	1	3	0	75
Woodside	1	1	2	50
Total	5	5	2	67

Note that the jackknifed model results in lower % successful prediction but these % may be a more reliable indicator of classification success because we have excluded each observation when calculating the classification coefficients.

A discriminant function plot using group mean scores showed that the three sites discriminate clearly along function 1 but there is little separation along function 2, not surprisingly since function 1 explained nearly all of the variation between sites (Figure 16-1).

Box 16-4 Worked example of discriminant function analysis: plant functional groups and leaf characters

We examined our ability to discriminate between the four location and functional group combinations for species of plants on which Reich et al. (1999) measured five variables - see Box 16-2. Like the two factor MANOVA earlier, the multivariate tests indicated no significant differences between the four groups (e.g. Pillai trace = 0.902, df = 15, 54, *F*-ratio = 1.548, *P* = 0.121) for the first discriminant function

The following classification functions were solved for each species (object):

	Colorado Forb	Colorado Shrub	Wisconsin Forb	Wisconsin Shrub
Constant	-535.136	-570.858	-576.439	-593.038
Log specific leaf area	557.743	580.616	582.442	592.914
Leaf N	-2.688	-2.981	-3.010	-3.129
A_{Mass}	-1.126	-1.154	-1.134	-1.173
A_{Area}	24.450	25.467	25.356	26.184
G_S	-0.227	-0.249	-0.248	-0.262

Each species was classified to the location and functional group combination with the highest value for the classification function. The classification matrices showed that we could more correctly classify species to some combinations than others:

	Colorado Forb	Colorado Shrub	Wisconsin Forb	Wisconsin Shrub	% correct
Colorado Forb	3	0	0	0	100
Colorado Shrub	0	3	0	1	75
Wisconsin Forb	1	3	6	1	55
Wisconsin Shrub	0	1	0	5	83
Total	4	7	6	7	71

Jackknifed classification matrix:

	Colorado Forb	Colorado Shrub	Wisconsin Forb	Wisconsin Shrub	% correct
Colorado Forb	3	0	0	0	100
Colorado Shrub	2	0	0	2	0
Wisconsin Forb	2	2	5	2	45
Wisconsin Shrub	0	2	0	4	67
Total	7	4	5	8	50

We were most successful at classifying species from the Colorado-forb combination and least from the Colorado-shrub combination.

The plot of the scores for the first two discriminant functions shows that there is considerable overlap between the different groups for both functions (Figure 16-2). Colorado forbs were the tightest group and we were most successful at classifying these species.

(a)

	Log_{10} Cu	Log_{10} Pb	Log_{10} Ni	Log_{10} Mn
Log_{10} Cu	0.196			
Log_{10} Pb	0.152	0.273		
Log_{10} Ni	0.164	0.192	0.165	
Log_{10} Mn	0.306	0.275	0.273	0.487

(b)

	Log_{10} Cu	Log_{10} Pb	Log_{10} Ni	Log_{10} Mn
Log_{10} Cu	0.169			
Log_{10} Pb	0.001	0.254		
Log_{10} Ni	0.045	0.031	0.086	
Log_{10} Mn	-0.011	0.033	-0.026	0.093

(c)

	Log_{10} Cu	Log_{10} Pb	Log_{10} Ni	Log_{10} Mn
Log_{10} Cu	0.369			
Log_{10} Pb	0.153	0.523		
Log_{10} Ni	0.209	0.223	0.251	
Log_{10} Mn	0.295	0.308	0.247	0.579

Table 16-1 Groups (a) and residual (b) and total (c) sums of squares and cross products matrices for data from Haynes et al. *(1995). The main diagonals are sums of squares between groups, within groups and total and the other elements are cross products.*

Source	df	Wilk's l	F	P
Clipping (C)	4, 56	0.467	23.950	<0.001
Emasculation (E)	4, 45	0.936	3.251	0.439
C x E	4, 56	0.826	2.768	0.029

Table 16-2 MANOVA results from Juenger & Bergelson (2000) who tested the effects of clipping, emasculation and their interaction on four response variables (flower, fruit, and seed production, total seed mass) of the perennial wildflower, the scarlet gilia.

Source	df	Pillai's trace	*F*	*P*
Infected or not	3,54	0.58	24.43	<0.001
Marsh zone	3,54	0.38	11.00	<0.001
Patch size	3,54	0.51	18.56	<0.001
Infection x zone	3,54	0.19	4.12	0.010
Infection x size	3,54	0.41	12.60	<0.001
Zone x size	3,54	0.13	2.60	0.062
Infection x zone x size	3,54	0.09	1.84	0.150

Table 16-3 MANOVA results from Pennings & Calloway (1996) who set up an experiment in a saltmarsh with three factors: Cuscuta *infection by the parasitic plant* Cuscuta salina*, zone within saltmarsh and size of patch. They recorded the biomass of three non-parasitic plant species and analyzed these three response variables with a three factor (infection, marsh zone, patch size) crossed MANOVA.*

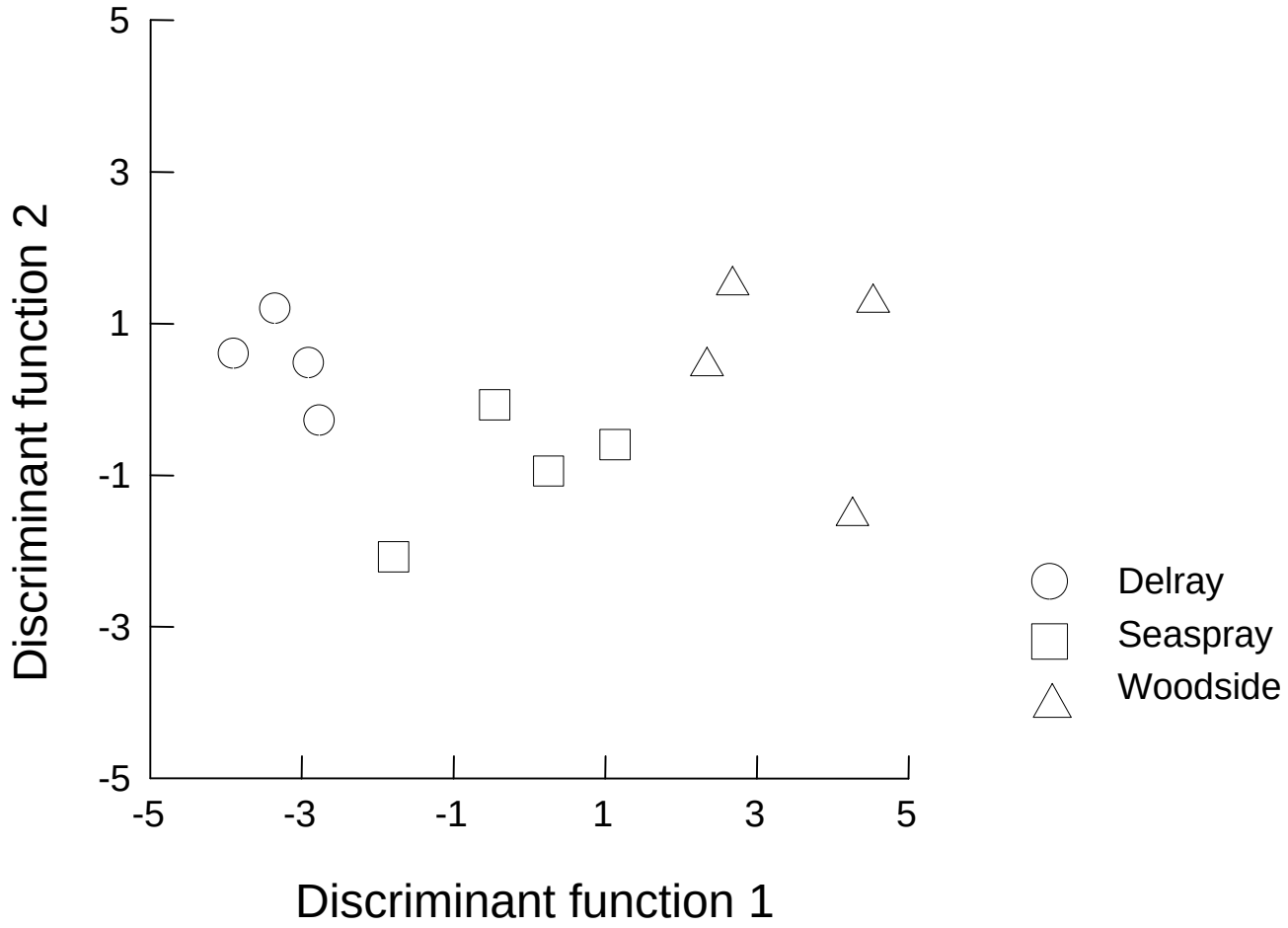
Variable	Standardized coefficient	Loading (correlation)
Survivorship	0.208	0.015
Size	0.634	0.593
Larval period	1.015	0.757

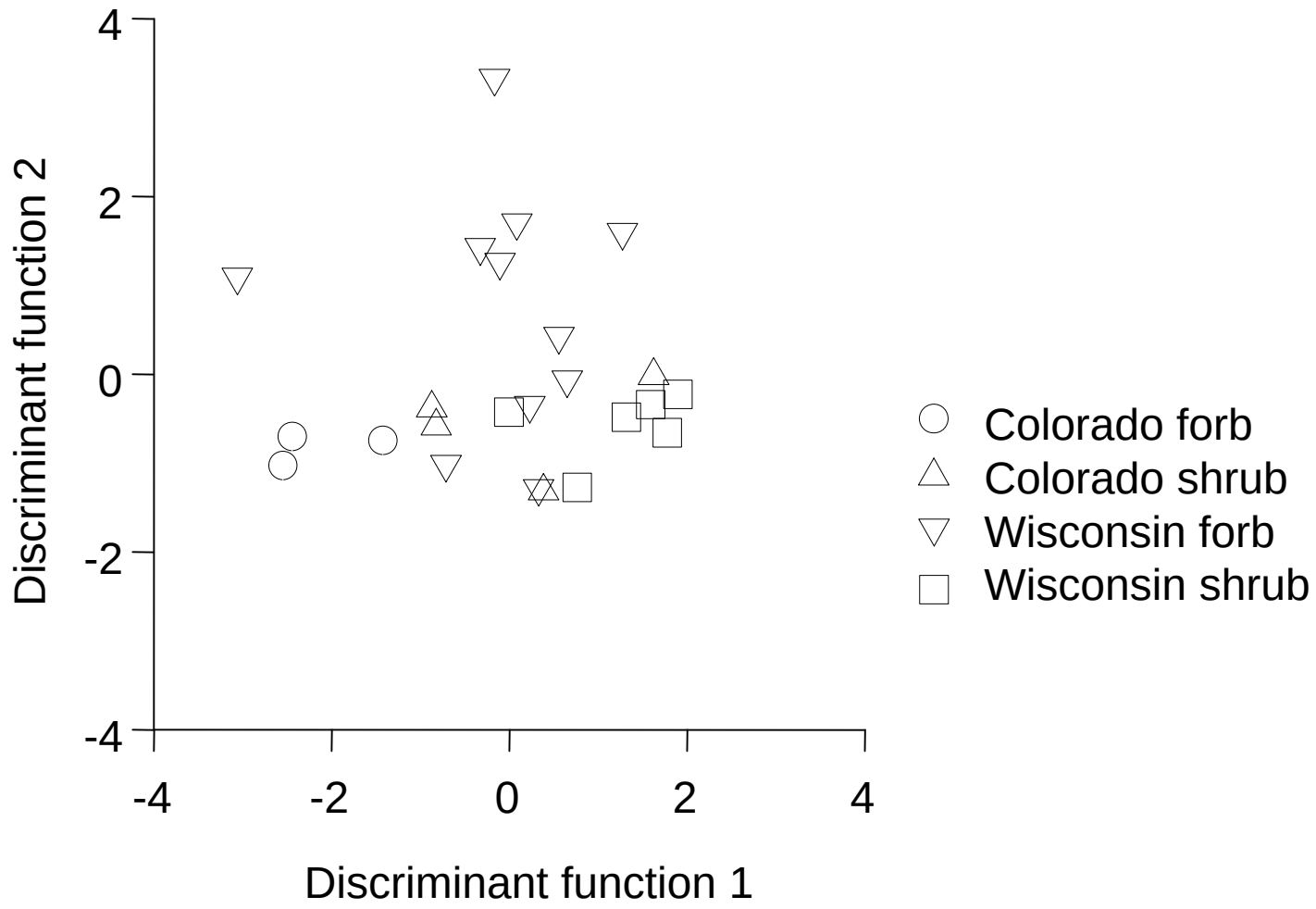
Table 16-4 Discriminant function analysis from Skelly (1995). The grouping variable was frog species (Pseudoacris triseriata and P. crucifer) and standardized coefficients and loadings for the three variables on the first discriminant function are provided. Larval period contributed most to the separation between species.

Captions to Figures

Figure 16-1 Plot of discriminant function scores for each replicate station for the first two functions from discriminant function analysis of data from Haynes et al. *(1995). The four variables were concentrations of the metals Cu, Pb, Ni and Mn in the sediment, all* log_{10} *transformed.*

Figure 16-2 Plot of discriminant function scores for each replicate species for the first two functions from discriminant function analysis of data from Reich et al. *(1999). The five variables were leaf characters: log specific leaf area, Leaf N,* A_{Mass}, A_{Area}, *and* G_S.


5
3
1
-1
-3
-5
-5
-3
-1
1
3
5
Discriminant function 2
Discriminant function 1
Delray
Seaspray
Woodside


Discriminant function 2
4
2
0
-2
-4
-4
-2
0
2
4
Discriminant function 1
Colorado forb
Colorado shrub
Wisconsin forb
Wisconsin shrub