I. INTRODUCTION

Discriminant analysis (also known as discriminant function analysis) is a powerful descriptive and classificatory technique developed by R. A. Fisher in 1936 to (a) describe characteristics that are specific to distinct groups (called descriptive discriminant analysis); and (b) classify cases (i.e., individuals, subjects, participants) into pre-existing groups based on similarities between that case and the other cases belonging to the groups (sometimes called predictive discriminant analysis). The mathematical objective of discriminant analysis is to weight and linearly combine information from a set of \( p \)-dependent variables in a manner that forces the \( k \) groups to be as distinct as possible. A discriminant analysis can be performed using several different statistical programs including DISCRIMINANT (SPSS; Norusis & Norusis, 1998), DISCRIM (SAS; Shin, 1996), and stepwise discriminant analysis (BMDP; Dixon, 1992).

Specific descriptive questions that can be answered through discriminant analysis include the following: (a) In what ways do various groups in a study differ? (b) What differences exist between and among the number of groups on a specific set of variables? (c) Which continuous variables...
best characterize each group, or, which continuous variables are not characteristic of the individual groups? (d) Given the results of a multivariate analysis of variance (MANOVA) indicating that group differences exist in the data, what specific variables best account for these differences? Predictive discriminant analysis addresses the question, Given the characteristics of a case as measured by a set of variables, to what predefined group does the case belong? Predictive discriminant analysis requires the use of classification rules (namely, discriminant functions) derived from a previous descriptive discriminant analysis on a data set for which the group membership of the cases is known. Those classification rules (i.e., derived weights and linear combination resulting from the descriptive discriminant analysis) are used in the classificatory analysis to assign new cases to the predefined groups.

In this chapter we give a thorough and complete discussion of what investigators need to know and do to use discriminant analysis properly. We begin with a brief layout of the specific steps and procedures necessary to conduct a descriptive discriminant analysis. This is followed by a more detailed discussion of each step, and of the information investigators need to complete the steps. Information about how to properly interpret the results of a descriptive discriminant analysis is provided, followed by a discussion of predictive discriminant analysis. Finally, we describe reporting requirements for publishing results of either a descriptive or predictive discriminant analysis. Throughout the discussion we integrate examples of the specific procedures using data drawn from a published discriminant analysis study.

II. ILLUSTRATIVE EXAMPLE

The illustrative study we will use is "Cranial shape in fruit, nectar, and exudate feeders: Implications for interpreting the fossil record" (Dumont, 1997). Dumont was interested in examining whether the morphological characteristics of primate cranial and dentary fossils could be used to distinguish among fruit feeders, nectar feeders, and exudate feeders (primates that gouge bark to feed on the sap). If successful in that effort, the researcher wanted to know in what ways the three types of feeders differed with respect to cranial and dentary morphology. In addition, Dumont wanted to derive classification rules to use in categorizing unclassified primate fossils. With a sample size (n) of 131 fossils and 22 (p) discriminator variables, Dumont concluded that, indeed, it was possible to discriminate among these 3 (k) groups of primates based on cranial and dentary characteristics. Dumont made available to us the complete data for 28 cases to use for illustrative purposes throughout this chapter. We conducted a descriptive discriminant analysis on the three groups (fruit, nectar, and exudate feeders) using three of the eight discriminator variables that evidenced strength in
discriminating between the groups: total skull length, minimum skull width, and dentary depth at canine. We were unable to obtain a data set of cases (similar to the Dumont, 1997, data set) for which group membership was unknown.

Our chosen illustrative study will demonstrate discriminant analysis relevance and use with data most relevant to anthropology and the biological and medical sciences. The interested reader is referred to Betz (1987), Brown and Tinsley (1983), Klecka (1975), and Tankard and Harris (1980) for case illustrations relevant to other areas of the social sciences and humanities (i.e., discrimination among college freshmen who differ in intent to continue mathematics coursework; participants in different leisure activities; political factions in the British Parliament, and television viewers and nonviewers).

III. DESCRIPTIVE DISCRIMINANT ANALYSIS

We recommend the following steps in performing a descriptive discriminant analysis:

1. Determine if discriminant analysis will provide statistical results that answer your research questions (i.e., is descriptive discriminant analysis suitable for answering your question?).
2. Determine the appropriateness of the data set for discriminant analysis.
3. Define the groups that will be used in the analysis.
4. Select the variables that will be used in the analysis.
5. Test the data to assure that the assumptions of discriminant analysis are met. If some assumptions are not met, determine whether discriminant analysis is robust for those assumptions.
6. Perform the analysis.
7. Interpret the results.

A. Data Requirements for Discriminant Analysis

We have already determined that discriminant analysis will help answer Dumont’s (1997) question of whether fossil morphological characteristics can be used to distinguish among primate feeder groups. The next step is to determine if the Dumont data meet the requirements for performing a discriminant analysis.

1. Groups and Discriminant Variables

Discriminant analysis requires a data set that contains two or more mutually exclusive groups and scores on two or more variables for each case
in the group. The groups may be constructed on the basis of demographic characteristics (e.g., sex, ethnicity, and socioeconomic status), personal attributes (e.g., height, psychological disorder, and blood type), or past or present behavior (e.g., whether one watches television, hallucinates, or smokes cigarettes). The groups should be constructed on a dimension that the investigator deems critical to understanding the differences that exist between the groups. In our example, we have three mutually exclusive groups of primates: fruit, nectar, and exudate feeders.

Variables should be chosen that are believed to represent dimensions on which the groups are expected to differ. Labeled discriminator or discriminant variables in discriminant analysis, they are called dependent variables in analyses such as analysis of variance (ANOVA) and MANOVA or predictor variables in analyses such as Multiple Linear Regression (cf. Brown & Tinsley, 1983). Discriminant analysis evaluates the degree to which such variables differentiate the groups, hence the name "discriminator variables." The effectiveness of the discriminant analysis depends on the extent to which the groups differ significantly on these variables. Therefore, the decision to select certain variables as potential discriminator variables is critical to the success of discriminant analysis.

Several guidelines can be followed in selecting discriminator variables. First, relevant theory should be consulted to determine which variables have the greatest potential for distinguishing among the groups of interest. Second, investigators should consider variables that have been shown to be relevant to group discrimination by previous research or theory (Brown & Tinsley, 1983). ANOVA results indicating group differences on single variables and factor analysis results indicating underlying factors among sets of variables are additional kinds of empirical evidence that can be consulted in selecting discriminator variables. Of course, an investigator's professional opinion also can be relied upon when selecting potential discriminator variables. These methods can be used singularly, but we recommend that investigators use them conjointly in choosing the discriminator variables. In our illustrative study, Dumont (1997) recognized that dietary habits are known to distinguish among live primates, and those habits result from or in differences in cranial and dentary morphology. She reasoned, therefore, that the differences in total skull length, minimum skull width, and dentary depth at the canine tooth would be useful in describing and classifying primate fossils.

Investigators should take care to select discriminator variables that are not highly intercorrelated. Intercorrelated variables causes difficulty in accurately determining the precise loadings (viz., partial correlations) of the individual variables on the discriminant functions (a problem discussed later). In short, if the variables are highly correlated with each other, they will likely load on the same function and, thus, not contribute in a unique (nonredundant) way to group discrimination. We recommend that investi-
gators consult published correlation matrices (or compute the complete correlations among the variables if data are available) and select variables that are not correlated significantly to use as discriminator variables. In the illustrative example, the correlations among three discriminator variables ($r$ length-width $= -0.21$, length-depth $= -0.17$, and width-depth $= 0.02$) were not statistically significant (see Table 8.1).

As in the illustrative study, the selected variables should satisfy the requirements for ordinal level measurement, at minimum. Ordinality makes it possible to describe the relations among the groups along the continuum underlying the variables of interest. Specifically, it permits the development of continuous scales on which the differences among the groups on the variables can be denoted. Nondichotomous nominal variables can be used, but they must be dummy coded into dichotomous categories (see Morrison, 1974). Doing so may have the effect of greatly increasing the number of variables used in the analysis. In the illustrative case, the cranial and dentary measurements were made to the nearest 0.1 mm using digital calipers and with a high degree of reliability (see Dumont, 1997).

Although two or more variables are required for the analysis, the number of ($p$) variables should not exceed the number of ($k$) groups. It is important to note that the difficulty of interpreting the results of the analysis increases with the number of discriminator variables selected for analysis (Brown & Tinsley, 1983). This is due to the increasing complexity of interrelations among the variables in accounting for group differences and the increasing subtlety in differences among the groups. In addition, the probability that significant results will be spurious increases as the ratio of discriminator variables to individual cases exceeds 10:1, as discussed in the next section. For these reasons, we suggest that researchers restrict the discriminator variables to those that have major theoretical and empirical relevance.

In summary, the data set should consist of two or more groups having at least ordinal scores on two or more discriminator variables. The discriminator variables should measure dimensions that the investigator views as important to understanding the differences that exist among groups. Importance can be determined on the basis of theory, past research, or another compelling rationale. The discriminator variables should not be intercorrelated with each other. The number of variables should not exceed the number of groups.

2. Sample Size

The investigator needs to determine the sample size needed to perform a discriminant analysis. We suggest that investigators avoid very large sample sizes because statistical tests are more likely to yield significant results for trivial differences under such circumstances. As sample size increases, the variability in the sampling distributions (i.e., the denominator in most correlational statistics) decreases, assuming all other factors are held con-
stant. This will result in a reduction of the overlap between sampling distributions under the null and alternative hypotheses, leading to a greater probability of finding statistical significance (viz., rejecting the null hypothesis). However, small sample sizes are not recommended because the idiosyncrasies in the sample will unduly influence the statistical results, thereby lessening the stability and generalizability of the results (Huberty, 1975, Morrison, 1974).

Small sample sizes also should be avoided because of the need to determine the reliability of the results of a discriminant analysis by confirming their applicability to other samples; this is known as cross-validation. Some cross-validation procedures involve using some part of the total sample to derive the discriminant functions (called a "developmental sample") and evaluating the performance of those functions to correctly classify the remaining part of the sample (called a "cross-validation sample") (see Tinsley & Brown, chapter 1, this volume).

Brown and Tinsley (1983) recommended that the total sample size should be at least ten times the number of discriminator variables. Stevens (1996) argued that the ratio of cases to variables should be more on the order of 20 to 1. We recommend the total sample size should lie within these two recommendations, with care given to ensuring that the sample size for the developmental sample meets the requirements for the number of cases in each group.

Ideally, the cases in the sample would be distributed evenly across the groups (Huberty, 1975; Tatsuoka, 1970), but often this is not possible nor reasonable. When the groups of interest differ in the number of cases, Brown and Tinsley (1983) recommend that the number of cases in each group be at least equal to the number of variables. A more conservative criterion (Huberty, 1975) is that the number of cases in the smallest group be at least three times the number of variables, allowing one-third of the cases to be used for cross-validation purposes. The criteria of Brown and Tinsley (1983) and Huberty (1975) are not as inconsistent as they may appear. The difference in their recommendations is that Brown and Tinsley did not assume, as does Huberty, that the cross-validation and developmental samples would be drawn from one sample.

Descriptive discriminant analysis requires that there are complete data on the discriminator variables and that group membership is known. Those cases that do not have scores on all of the discriminator variables automatically are excluded from the analysis, possibly affecting the adequacy of the data in satisfying sample and group size requirements adversely. Furthermore, cases having missing data may be systematically different from those cases containing complete data, so use of those with incomplete data could result in generalizability problems (Norusis, 1990). We recommend that, upon completion of the analysis, the researcher examine the analysis output and determine whether the number of cases included in the analysis still
meets the sample and group size requirements. If sample size requirements are not met, acknowledge the limitations associated with low sample and group sizes when reporting and interpreting the results. Furthermore, applied researchers should compare the cases with missing data to those with complete data to see if they differ systematically. If they differ, generalizability problems are evident and should be discussed in reporting the results.

Our illustrative data set consisting of three discriminator variables and 28 cases does not satisfy the minimum sample size criterion \( (N = 30) \) of having ten cases per variable recommended by Brown and Tinsley (1983). In situations such as this, researchers must use caution in interpreting the results. In our data set, the fruit, nectar, and exudate groups have 12, 11, and 5 cases, respectively, thereby satisfying the Brown and Tinsley requirement that the number of cases in each group be equal to or exceed the number of variables, but not Huberty's (1975) more conservative criterion of at least 3 cases for every variable in each group. Group 3 does not meet the latter requirement. This is another reason to use caution when interpreting the results. Lastly, it is important to note that there are complete data on the discriminators and group membership variables on all 28 cases.

3. Assumptions

Because the data used in discriminant analysis involve multiple variables, the assumptions required for a discriminant analysis are the same as those required for other multivariate analyses (i.e., MANOVA). Violations of the following assumptions are expected to inflate the Type I error rate: (a) independence of observations (independent measures); (b) multivariate normality (the observations based on the discriminator variables are normally distributed), and (c) homogeneity of covariance matrices (the population covariance matrices based on the discriminator variables are equal). (Huberty and Petoskey discuss these assumptions in detail in chapter 7, this volume). Current evidence suggests that discriminant analysis is robust with respect to violation of assumptions of multivariate normality and of homogeneity of covariance matrices (Stevens, 1996). Nevertheless, conscientious researchers should examine and report whether these assumptions hold.

The assumption of independent observations is critical; the likelihood of making a Type I error increases markedly if this assumption is violated. To assure that the scores for each of the variables (observations) are independent (that is, not highly correlated), the researcher needs to examine the correlation matrix for the discriminator variables. In general, correlations having an absolute value of less than .3 have no interpretive value (Tabachnick & Fidell, 1989) and do not violate the assumption. Correlations having an absolute value equal to or greater than .3 indicate violation of the assumption.
We would like to note some ideas advanced by Stevens (1996) that appear to have merit, but we have found no corroborating empirical or theoretical evidence to warrant an opinion on our part; we offer them to the reader in order to be complete in our treatment of discriminant analysis and because they are interesting. Stevens (1996) argues that correlations greater than .3 are not problematic if they are not statistically significant. In the event that they are, Stevens contends that the researcher could alter the alpha level to a more conservative one and see if they are still significantly correlated. Stevens (1996) also contends that significant correlations within each group are not as problematic to the validity of a discriminant analysis as those that might occur in the pooled within-groups matrix. Therefore, researchers should examine the pooled within-groups correlation matrix and those within each group for the discriminator variables to determine whether the observations (measures) are correlated across groups or only within groups.

No statistical test for multivariate normality is available on SPSS, SAS, or BMDP. However, Stevens (1996) suggests that the data follow a multivariate normal distribution if the distributions of scores on the discriminator variables are normal. Investigators should examine these distributions for indications that the assumption of multivariate normality is likely to have been violated (Norusis, 1990). Keep in mind that this simple alternative is no guarantee that the multivariate distributions do not depart from normality.

Discriminant analysis is especially robust to violations of the assumption of homogeneity of the covariance matrices if the ratio of the largest group n divided by the smallest group n is less than 1.5 (Stevens, 1996). Investigators can test for violations of the homogeneity of covariance assumption using the Box's M statistic in SAS and SPSS. When violated, Klecka (1975) notes that the worst consequence is that cases are more likely to be classified into the group with the greatest dispersion. When the assumption has been violated, some statistical packages permit investigators the option of proceeding with the analysis using separate covariance matrices for each group instead of the pooled within-group covariance matrix (cf. Cocozzelli, 1988). In short, the available literature indicates that violation of the homogeneity of covariances assumption is not of major importance to conducting a valid discriminant analysis and, even when violated, a number of effective options are available to the researcher.

4. Using Prior Probabilities

When calculating the classification function, coefficients used to assign preclassified cases to groups, it is ordinarily assumed that a case has an equal probability of being a member of any of the groups; statistical packages use this assumption by default. Sometimes investigators may be tempted to use unequal prior probabilities (e.g., the proportion of cases in each group);
the commercially available statistics packages allow the investigator to input other probabilities. However, we caution that the use of different prior probabilities can cause the function coefficients to differ radically from those derived using equal prior probabilities. We agree with Stevens (1996) that the decision to assume other than equal prior probabilities is justifiable only when the investigator has a firm theoretical or empirical basis for believing that the differences in group size reflect actual differences in the size of the respective populations, as in the case of racial and ethnic groups.

5. The Statistical Assumptions and Our Illustrative Case

In our example, the assumption, independent observations, was tested by calculating a pooled within-groups correlation matrix. As previously presented, the three discriminator variables were minimally correlated. The assumption of multivariate normality was assessed by inspecting the distribution of scores on the three discriminator variables (see Figure 8.1). These figures show a normal distribution of scores on each group of the three variables. The third assumption of equal population covariance matrices for the dependent variables was tested using Box's M statistic. The statistically significant F ratio \( F(12, 744) = 2.53, p \leq .003 \) suggested that the population covariance matrices were not equal. Nevertheless, we concluded that this violation was not very problematic for discriminant analysis (cf. Klecka, 1975), and we recognized that we should be able to evaluate the likelihood of errors by examining the determinants of the pooled within-groups covariance. Lastly, in the absence of any research documenting that the differences in group sizes are reflective of actual population sizes, we assumed that cases have an equal probability of being a member of any of the groups.

**FIGURE 8.1** Distribution of the dentary depth at canine for the sample cases. \( SD = .54; \) mean = 4.04; \( N = 28.00 \).
B. Components of Discriminant Analysis Results

A proper interpretation of discriminant analysis requires an accurate understanding of four important concepts: discriminant coefficient, discriminant function, discriminant scores, and group centroids.

1. Discriminant Coefficients

Discriminant analysis calculates mathematical weights for scores on each discriminator variable that reflect the degree to which scores on that variable differ among the groups being discriminated. Thus, the discriminator variables on which more groups differ and on which the groups differ most receive the most weight; these weights are referred to as discriminant coefficients.

Most computer programs used to perform discriminant analysis allow investigators to consider both unstandardized and standardized discriminant function coefficients. Unstandardized coefficients, calculated on raw scores for each variable, are of most use when the investigator seeks to cross-validate or replicate the results of a discriminant analysis or to assign previously unclassified subjects or elements to a group. However, the unstandardized coefficients cannot be used to compare variables or to determine what variables play the greatest role in group discrimination because the scaling for each of the discriminator variables (i.e., their means and standard deviations) usually differ. In our illustrative study, for example, the means and standard deviations of the three discriminator variables differ across the three primate feeder groups, making direct comparisons of the variables difficult: total skull length \( (m = 4.04, sd = .54) \), minimum skull width \( (m = 1.06, sd = .32) \), and dentary depth at canine \( (m = .29, sd = .06) \). A comparison of unstandardized coefficients for the discriminator variables will distort the apparent contributions of the variables to the function because of the scaling differences.

Standardized discriminant coefficients are used to determine the comparative relations of discriminator variables to the functions. Standardized discriminant coefficients have been converted to z scores (i.e., \( m = 0, sd = 1 \)) to eliminate scaling differences among the discriminator variables. As a result, investigators can determine the degree to which each discriminator variable is associated with differences among the groups by examining the absolute magnitude of the standardized discriminant coefficients and the relative importance of each discriminator variable to group discrimination.

Table 8.1 presents the standardized and unstandardized discriminant function coefficients for our illustrative study. If we attempted to interpret the unstandardized functions, we would conclude that total skull length and dentary depth at canine were equally important in defining Function 1. However, inspection of the standardized function coefficients reveals that Function 1 is defined primarily by total skull length. The unstandardized
TABLE 8.1 Discriminant Function Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Standardized function</th>
<th>Unstandardized function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Total skull length</td>
<td>1.0</td>
<td>0.16</td>
</tr>
<tr>
<td>Minimum skull width</td>
<td>0.05</td>
<td>0.95</td>
</tr>
<tr>
<td>Dentary depth at canine</td>
<td>0.18</td>
<td>-0.37</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

coefficients suggest that dentary depth at canine receives twice the weight of minimum skull width on Function 2, but the standardized coefficients reveal that the latter actually should be given three times the weight of the former.

2. Discriminant Functions and Discriminant Scores

Discriminant analysis forms one or more weighted linear combinations of discriminator variables called discriminant functions. Each discriminant function has the general form:

$$D = a + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p,$$

where $D$ is the discriminant score, $a$ is the Y-intercept of the regression line, $b$ is the discriminant function coefficient, $x$ is the discriminator variable raw score, and $p$ is the number of discriminator variables. The $Y$-intercept is a constant that adjusts the function to account for the scaling differences in the discriminator variables that are present when raw (unadjusted) scores are used for the analysis; this adjustment is unnecessary when standard scores are analyzed. A discriminant score for each function is calculated for each case in a sample by multiplying each case's discriminator variable raw score by the associated weight (viz., discriminant coefficient). Then, the discriminator scores are used to calculate the average discriminant score of the cases belonging to a group (i.e., group centroid) for each discriminant function. In descriptive discriminant analysis, these scores are used to assign each case to a group. Discriminant scores can be requested as output from most statistical packages. The maximum number of functions that can be produced in a discriminant analysis is equal to the lesser of the number of discriminator variables ($p$) or number of groups minus one ($k - 1$).

In the illustrative study, discriminant analysis produced two unstandardized discriminant functions. Function 1 had unstandardized weights ($b$) of 3.14, .15, and 2.94 and an intercept ($a$) of $-13.73$ (see Table 8.1). Function 2 had unstandardized weights of .49, 3.21, and $-5.99$ and an intercept of $-3.64$. There were three primate feeder groups in our example.
study and three discriminator variables, so our discriminant analysis could produce up to two discriminant functions \((k - 1 = 3 - 1 = 2)\) (see Table 8.2), maximally.

The discriminant functions produced by an analysis will differ in the amount of group discrimination for which they can account and the particular groups they best differentiate. These points will be elaborated upon in our later discussion of the interpretation of discriminant functions.

### 3. Group Centroids

In discriminant analysis, the group centroids represent the mean discriminant score of the members of a group on a given discriminant function. For classification and prediction purposes, the discriminant score of each group case (e.g., each individual) is compared to each group centroid, and the probability of group membership is calculated. The closer a score is to a group centroid, the greater the probability the case belongs to that group. This will be discussed in greater detail in the predictive analysis section.

Group centroids reveal how much and in what ways the groups are differentiated on each function. The absolute magnitude of the group centroids indicates the degree to which a group is differentiated on a function, and the sign of the centroid indicates the direction of the differentiation. Table 8.3 shows that Function 1 discriminates nectar feeders from fruit and exudate feeders. Nectar feeders scored at the positive end on the bipolar function and exudate feeders at the negative end of the function. Function 2 discriminated exudate feeders from fruit and nectar feeders,

---

**TABLE 8.2 Evaluating Discriminant Functions**

<table>
<thead>
<tr>
<th>Function</th>
<th>Eigenvalue</th>
<th>Relative percent</th>
<th>Canonical correlation</th>
<th>Wilks's lambda</th>
<th>Chi-square</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.00</td>
<td>87.0</td>
<td>.82</td>
<td>.26</td>
<td>32.56</td>
<td>6</td>
<td>.00</td>
</tr>
<tr>
<td>2</td>
<td>.30</td>
<td>13.0</td>
<td>.48</td>
<td>.77</td>
<td>6.24</td>
<td>2</td>
<td>.04</td>
</tr>
</tbody>
</table>

**TABLE 8.3 Group Centroids**

<table>
<thead>
<tr>
<th>Group</th>
<th>Function 1</th>
<th>Function 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fruit feeders</td>
<td>-1.03</td>
<td>- .44</td>
</tr>
<tr>
<td>Nectar feeders</td>
<td>1.66</td>
<td>.02</td>
</tr>
<tr>
<td>Exudate feeders</td>
<td>-1.17</td>
<td>1.01</td>
</tr>
</tbody>
</table>
with exudate feeders at the positive end of the function and fruit and nectar feeders at the negative end of the function.

C. Interpreting Discriminant Functions

Interpreting the results of a discriminant analysis depends, in large part, on the interpretation of the discriminant functions. The function is defined by the discriminant coefficients that are used to weight a case's scores on the discriminator variables. The task facing the researcher is to examine the standardized discriminant functions to identify the conceptual dimension (or construct) underlying each function.

First, researchers should identify the discriminator variables that have the highest and lowest weights on a function. The size of the coefficient tells the investigator how much a discriminator variable contributes to group discrimination on a function, and the sign tells the investigator the direction of the relation. For example, in our illustrative study, all of the discriminator variables on Function 1 are positively related to the function, but only total skull length contributes meaningfully to group discrimination on the function (see Table 8.1). Thus, high discriminant scores on Function 1 are associated with greater total skull length. On Function 2, minimum skull width contributes most to group discrimination on the function and in the positive direction; dentary depth at canine is moderately and negatively related to group discrimination on the function. Consequently, high scores on Function 2 are associated with greater minimum skull width and, to a lesser extent, with decreased dentary depth at canine.

Also, researchers should examine the structure matrix coefficients, which show the correlation of discriminator variables with the function. In a manner similar to interpreting factors in factor analysis (Cudeck, chapter 10, this volume), investigators can identify the construct underlying each function by studying the discriminators on a given function. The absolute magnitude indicates the strength of the relation between each variable and the function and the sign indicating the direction of the relation. For example, the structure matrix indicated that total skull length ($r = .98$) is most highly correlated with Function 1, and the other discriminator variables contributed little to group discrimination ($r$'s for minimum skull width and dentary depth at canine were $-.16$ and $.01$, respectively). Minimum skull width ($r = .91$) is most highly correlated with Function 2, but dentary depth at canine also contributed meaningfully, albeit negatively, to group discrimination ($r = -.37$; $r$ for total skull length was $.03$). Structure matrix correlation coefficients less than $.30$ typically are not interpreted because the square of the structure matrix coefficient reveals that such discriminators account for less than 10% of variability in the function. Though labeling and describing the functions includes an element of subjectivity, it is an essential aspect of achieving a parsimonious description of group differ-
ences. This procedure is consistent with the rationale underlying discriminant analysis, that information from many variables considered together is more useful in differentiating groups than information on singular variables in isolation.

In addition, researchers should determine the meaning associated with high or low scores on a function. In our example, we conclude that high scores on Function 1 are associated with higher scores on total skull length. Higher scores on Function 2 are associated with high scores on minimum skull width and moderately lower scores on dentary depth at canine.

Finally, researchers should examine the group centroids for each discriminant function and identify the groups having the highest and lowest scores. Those groups are differentiated the best on that function (see Table 8.3), and the group centroids provide some insight into the unique attributes of each group.

D. Rotating Discriminant Functions

Much like in the case of factor analysis, rotating significant discriminant functions can facilitate the interpretation of the functions. The investigator is cautioned to keep in mind that when rotation (varimax) is performed, the relative standing of the functions is altered so that the first function may not necessarily account for the maximum amount of group differences on the discriminator variables (cf. Stevens, 1996). With regards to the illustrative example, we did not rotate the discriminant functions because the structure matrix provided sufficient information to assist in the interpretation of the functions.

E. Evaluating the Importance of Functions

1. Primacy

   In a manner analogous to factor analysis, the first function generated in a discriminant analysis accounts for the greatest proportion of differences among the groups. Each subsequent function that is generated accounts for a decreasing proportion of the differences among the groups. In our example, two functions were generated (see Table 8.2), with Function 1 accounting for more group differences than Function 2. Furthermore, the proportion of group differences accounted for by each function is statistically independent of that accounted for by the other functions. Discriminant analysis will generate the maximum possible number of discriminant functions, in decreasing order of importance in accounting for group differences, but it is possible that the functions do not differ significantly (either statistically or practically) in the proportion of group differentiation for which they account.
2. Wilks's Lambda and Chi-Square

Wilks's lambda, the ratio of within-group variability to total variability on the discriminator variables, is an inverse measure of the importance of the functions (Betz, 1987; Brown & Tinsley, 1983; Huberty, 1975; Tatsuoka, 1970). Values close to 1 indicate that almost all of the variability in the discriminator variables is due to within-group differences (differences between cases in each group); values close to 0 indicate that almost all of the variability in the discriminator variables is due to group differences. A chi-square test based on lambda indicates whether the variability that is systematically related to group differences is statistically significant. The test is performed over and over again after each function is extracted until a nonsignificant result is obtained or all of the functions have been generated. In our example, the Wilks's lambda value for Function 1 is .26 (see Table 8.2), signifying that most of the variability captured by Function 1 can be attributed to group differences. The Wilks's lambda value for Function 2 is .77, suggesting that little of the variability captured by Function 2 is attributed to between-group differences. Because the purpose of a discriminant analysis is to detect group differences, these lambda statistics suggest that Function 2 is of little importance.

The chi-square values show that, prior to extraction, the variability related to group differences is statistically significant at the .05 level. After extracting Function 1, the variance remaining in the discriminator variables is still significantly related to the group differences. Therefore, a second function was extracted. A reasonable conclusion for the investigator based on Wilks' lambda and the chi-square test is that, although only a small amount of group differences is accounted for by Function 2, the proportion of group differences is statistically significant and worthy of attention.

3. Eigenvalues

Eigenvalues indicate the ratio of between-groups variability to within-groups variability for a function. The larger the eigenvalue, the better at accounting for the group differences are the discriminator variables loading on the function. In our illustrative case, the eigenvalue for Function 1 (2.00) is much larger than the eigenvalue of Function 2 (.30) (see Table 8.2). This is consistent with the Wilks' lambda statistic, which indicated that the ratio of between-group variability to within-group variability detected by Function 1 is much larger than that detected by Function 2. Given our earlier interpretation of the functions and study of the group centroids, we can conclude that total skull length, as opposed to minimum skull width and dentary depth and canine, plays an almost exclusive role in group discrimination on Function 1 and that nectar feeders are the most unique of the three groups with respect to this variable.
4. Relative and Absolute Percent

Unlike Wilks's lambda and the associated chi-square test, the relative and absolute percent statistics are indices of the practical rather than the statistical significance of the functions for group discrimination. Not all functions that account for a statistically significant amount of group variability on the discriminator variables are of conceptual or practical value in distinguishing among the groups. The relative percent statistic, also known as the percent of variance accounted for and percent of variance explained, is obtained by dividing each eigenvalue by the sum of the eigenvalues and is a direct index of the relative importance of each function in accounting for between-group differences. The relative percent shows the proportion of the total amount of between-group variance that is attributable to a particular function. In our example, the relative percent for Functions 1 and 2 are 87% and 13%, respectively (see Table 8.2).

The absolute percent, calculated by hand by dividing each eigenvalue by the number of discriminator variables used in the analysis, indicates the magnitude of between-group variability explained by each function relative to the amount of between-group variation. Thus, with our example, the absolute percent for Function 1 is $2.00/3$, which equals 66.67%. The absolute percent for Function 2 is $.30/3$, which equals 10%. The total percent of between-group variability explained by the functions is 76.67% (i.e., $66.67\% + 10\%$) or the sum of the eigenvalues divided by the number of discriminator variables (i.e., $[2.00 + \ldots + .30]/3 = .77$).

The absolute percent is useful in gauging how effective the discriminator variables were in distinguishing among the groups. It is in this context that the relative percent is interpreted. It is possible for a statistically significant function with a large relative percent to account for little real variability among the groups when the total amount of the variability in the groups that is accounted for by discriminators is small. In such a case, the amount of variance in the discriminator variables that is systematically related to group differentiation is actually less than it appears from an observation of the relative percent alone.

5. Canonical Correlation

Some of the variability represented by a function is unrelated to group differences. This variability may be related to within-group differences, or to various types of errors that occurred in the data collection or data entry process. The canonical correlation, characteristically generated by discriminant analysis, indicates the relation between scores on the function and group differences, and, therefore, provides yet another indication of the practical value of the functions. The canonical correlation indicates the amount of group variability captured by each function, and the canonical correlation squared indicates the proportion of variance in a function that is related to group differences. As with the eigenvalues, larger canonical
correlations indicate more important functions. In our example, the canonical correlations for Functions 1 and 2, respectively, are .82 and .48 (see Table 8.2). The squares of these correlations are .67 and .23, respectively. These values indicate a strong relation between scores on Function 1 and group differences, and a weak relation between scores on Function 2 and group differences.

6. Integrating Information about the Functions

Integrating the information yielded by these statistics requires a solid grasp of the research and theory relevant to the area under investigation. All of the statistics will yield the same rank order of importance of the functions, but they give different clues as to the degree of importance of a function. The most common strategy (cf. Betz, 1987) is to use the chi-square test of significance to determine the maximum number of functions that may be important. Then examine the Wilks' lambda value, the canonical correlations, and the absolute percent values to determine the adequacy of the significant functions in explaining the variability among the groups. Finally, examine the relative percent values to decide the importance of each function.

With respect to our illustrative study, the chi-square test indicates that there are two statistically significant functions. The Wilks' lambda statistic indicates that a large portion of between-group differences is captured by Function 1, and a much smaller portion is captured by Function 2. The canonical correlations indicate a strong relation between scores on Function 1 and group differences, and a weak relation between scores on Function 2 and group differences. Lastly, the relative percents and absolute percents confirm that there is a substantially greater portion of between-group variability captured by Function 1 as compared to Function 2.

F. Evaluating Classification Accuracy

The final determination of the usefulness of a set of functions rests on the ability of the functions to accurately classify members to groups (Betz, 1987). An evaluative classification can be performed as part of the descriptive discriminant analysis to determine the accuracy with which the discriminant functions assign individuals to groups. This evaluative classification includes multiple procedures for determining accuracy (hit) rates and cross-validation.

1. Classification Accuracy

Evaluating classification accuracy (i.e., hit rates) is an important means of determining the statistical and the practical usefulness of a set of computed functions. Investigators often intend to use the functions generated in a discriminant analysis to classify cases that have not yet been classified or
that have been classified by some alternative and possibly inferior method. Before attempting to classify new cases, however, the investigator needs to have some indication of the accuracy of the functions in classifying cases.

The actual procedure involves classifying the original cases in the sample using the functions and evaluating the accuracy of these classifications. When performing a descriptive discriminant analysis, the investigator begins with a sample of cases whose group membership is known. After the discriminant analysis has been performed and the discriminant functions generated and evaluated, discriminant scores are calculated on each selected function, and group centroids are calculated for each function. Next, the group to which each case would be assigned on the basis of their scores on the discriminator variables is determined. Finally, the group to which each case would be assigned is compared to the group to which the case actually belongs, and the percentage of correct assignments is calculated. The procedure results in a percentage indicating the proportion of cases assigned to the groups correctly. Norusis (1990) provides a useful discussion of the mathematical computations involved in determining group assignments.

Table 8.4 presents the classification results for our illustrative study. It shows that the percentage of primate fossils correctly classified using the functions generated in the analysis were 83.3% for the fruit feeder group,

<table>
<thead>
<tr>
<th>Predicted group membership</th>
<th>Fruit feeders</th>
<th>Nectar feeders</th>
<th>Exudate feeders</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Actual group membership</strong></td>
<td><strong>Developmental sample: Classification count</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruit feeders</td>
<td>10</td>
<td>0</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Nectar feeders</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Exudate feeders</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>Developmental sample: Classification percent</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruit feeders</td>
<td>83.3</td>
<td>.0</td>
<td>16.7</td>
<td>100</td>
</tr>
<tr>
<td>Nectar feeders</td>
<td>9.1</td>
<td>90.9</td>
<td>.0</td>
<td>100</td>
</tr>
<tr>
<td>Exudate feeders</td>
<td>1.0</td>
<td>.0</td>
<td>80.0</td>
<td>100</td>
</tr>
<tr>
<td><strong>Cross-validation sample: Classification count</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruit feeders</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Nectar feeders</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Exudate feeders</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td><strong>Cross-validation sample: Classification percent</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fruit feeders</td>
<td>75.0</td>
<td>8.3</td>
<td>16.7</td>
<td>100</td>
</tr>
<tr>
<td>Nectar feeders</td>
<td>9.1</td>
<td>90.9</td>
<td>.0</td>
<td>100</td>
</tr>
<tr>
<td>Exudate feeders</td>
<td>20.0</td>
<td>.0</td>
<td>80.0</td>
<td>100</td>
</tr>
</tbody>
</table>
90.9% for the nectar feeder group, and 80% for the exudate feeder group. The overall hit rate (i.e., percent of correct predictions) for the analysis was 85.7%.

2. Cross-Validation

It is critical to cross-validate the results of a discriminant analysis, especially if the researcher intends to classify other samples into the groups of interest. As stated previously, discriminant analysis is a maximizing procedure in which the discriminant coefficients are derived mathematically to maximize the distinctions among the groups. In doing so, discriminant analysis capitalizes on chance differences among the sample groups that occur because of idiosyncratic characteristics of the sample. Because these idiosyncratic characteristics differ from sample to sample, the usefulness of the classification functions (i.e., discriminant functions) is uncertain. Generalizability is particularly uncertain for the lesser functions generated in an analysis. Thus, the investigator needs to determine how much shrinkage in the hit rate can be expected when classifying cases that were not used to derive the functions.

Methods of cross-validation abound (see Tinsley & Brown, chapter 1, this volume); we will illustrate two that are among the more feasible. In the jackknife method (also referred to as the "leave-one-out" method and the U-method), one sample case is systematically held out and the discriminant analysis is performed on the remaining sample. Then, the excluded observation is classified into one of the groups using the discriminant functions generated by the analysis. This procedure is repeated until each member of the sample has been held out and classified. The information resulting from this procedure can be used to determine the stability of various statistics produced by the discriminant analysis. The proportion of cases correctly classified can be compared to that expected on the basis of chance using the z test of proportions that will be discussed later in the chapter. The jackknife procedure is available on BMDP, SAS, and the newer versions of SPSS.

Cross-validation using the hold-out method involves splitting the sample, randomly, into two parts with two-thirds of the sample belonging to a "developmental" sample and one-third being allocated to a "cross-validation" sample. The discriminant analysis is performed on the developmental sample, and the functions derived are used to classify the members of the smaller cross-validation sample. The accuracy of classification for the smaller sample indicates the hit rate investigators can expect to achieve for newer samples.

Examination of Table 8.4 reveals that the analysis correctly classified 85.7% of the 28 cases in the developmental sample using the derived functions. We performed the jackknife cross-validation procedure to determine shrinkage and found that 82.1% of cross-validated classifications were cor-
rect. This value is lower than the hit rate achieved for the derivation sample, which will almost always be the case. Thus, the use of the discriminant functions derived from the developmental sample to classify independent samples can be expected to result in approximately 82% of the cases being correctly classified. Next, we need to determine the significance of this hit rate.

3. Chance Proportions and Hit Rates

Computer programs do not provide a significance test of the proportion of cases correctly classified in a discriminant analysis. Researchers can evaluate classification accuracy by comparing the proportion of cases correctly classified to the base or chance rate using a z test of proportions (cf. Betz, 1987; Brown & Tinsley, 1983; see Formula 5). The base rate is the proportion of cases expected to be correctly classified on the basis of the best available alternative strategy (i.e., results from current classification practices including the use of previously derived discriminant functions). The base rate is the preferred standard against which to evaluate classification accuracy, but it is often the case in research that no alternate strategy exists. In such circumstances, investigators can test the significance of their classification results by comparing the proportion of correct predictions to the proportion expected on the basis of chance.

There are a number different ways of determining the chance rate (Brown & Tinsley, 1983; Morrison, 1974). When the groups to which cases were classified are of equal size, the proportion of correct classifications expected to occur on the basis of chance alone is

\[ \frac{1}{k}, \]  

(2)

where \( k \) equals the number of groups into which members are being classified. This assumes that each case had an equal chance of being classified into any one of the groups. In our three-group example, Formula 2 yields a chance rate of 1/3 or 33%, signifying that approximately nine cases of the 28 would be expected to be classified correctly on the basis of chance.

Notice, however, that the groups in our example are not of equal size. Therefore, the chances of being classified into each group are not equal. Two strategies for calculating chance proportions exist when the group sizes are unequal. Sometimes, investigators are interested only in maximizing the overall classification accuracy (i.e., they do not care about the particular groups into which cases are being classified nor about differences in the types of classification errors being made). In this situation, classification accuracy would be maximized by classifying all members into the largest group and the chance proportion is calculated using the formula

\[ \frac{n}{N}, \]  

(3)

where \( n \) is the number of cases classified into the largest group and \( N \) is the total sample size.
where \( n \) is the number of persons in the largest group and \( N \) is the total sample size (i.e., the sum of all members of all groups). The three groups in our example, consisted of 12, 11, and 5 cases, so 12/28 (43%) would be classified correctly by chance if the investigator was only interested in maximizing correct classifications.

Typically, researchers are interested in more than simply maximizing the number of correct classifications. In many situations, the different types of assignment errors differ in their theoretical or practical significance; investigators might be interested in correctly classifying cases into all of the groups. Formulas 2 and 3 can grossly overestimate the chance probability of correct predictions in these circumstances. The proper formula for calculating chance probability in such cases is one that considers the proportion of the total sample in each group:

\[
p_1 a_1 + p_2 a_2 + \cdots + p_k a_k,
\]

where \( p \) is the proportion of the total sample actually belonging to a group, \( a \) is the actual proportion of cases classified by discriminant analysis into a particular group, and \( k \) is the number of the groups (Betz, 1987; Brown & Tinsley, 1983). In our example, \( p_1, p_2, \) and \( p_3 \) equal .43, .39, and .18, respectively, \( a_1, a_2, \) and \( a_3 \) equal .43, .38, .21, respectively, and the chance rate equals .43(.43) + .39(.38) + .18(.21) = .19 + .15 + .04 = .38. Thus, if the researcher was interested in classifying the cases into even the smallest group, the likelihood of doing so correctly by chance would be 38%.

Formulas 2, 3, and 4 will yield identical results when the number of cases in each group are equal, but they diverge as a function of the extent to which the group sizes differ. Even small differences in the calculated chance rate can influence the outcome of the \( z \) test of proportions (Brown & Tinsley, 1983, p. 304). In our example, Formulas 1, 2, and 3 yield different chance rates of 33%, 43%, and 38%. Given the actual hit rate observed in a given study, such differences in the chance rate can be of enormous significance.

The chance proportion can be compared to the proportion correctly classified using the \( z \) test of proportions. This test must be performed by hand because it is not provided by the commercially available statistical packages. The formula for the \( z \) test of proportions is

\[
z = \frac{(Np_a - Np_c)}{(Np_c(1 - p_c))},
\]

where \( N \) = the total sample size, \( p_a = \) the proportion of cases correctly classified using discriminant analysis, and \( p_c = \) the proportion of cases expected to be classified correctly on the basis of chance. The obtained \( z \) value then can be compared to the \( z \) value for a one-tailed test from the \( z \) table found in the appendices of most statistics books. For example, if alpha for the test is chosen to be .05, the critical value of \( z \) is 1.65. In our illustrative study, \( N = 28, p_a = .82, \) and \( p_c = .38. \) Applying Formula 5,
\[ z = \frac{(28(.86) - 28(.38))(1 - .38)}{28(.38)} = 1.88, \] which exceeds the critical value of 1.65. Therefore, we conclude that the classification accuracy of our analysis, even after cross-validation, exceeds that expected on the basis of chance at the .05 level of significance.

**IV. PREDICTIVE DISCRIMINANT ANALYSIS**

Up to this point, our discussion has focused primarily on the descriptive uses of discriminant analysis. There are occasions in which professionals encounter cases for which group membership is not known, but it is important or desirable to determine its category membership. It may be important to know whether a newly discovered poem or painting belongs to the body of work of an artist of historical significance. It may be a newly discovered fossil whose phylum is debated by paleontologists. It might be to determine whether an individual who possesses a given set of characteristics is likely to be successful or unsuccessful in an educational or medical/psychological treatment program. Descriptive discriminant analysis will identify the distinct groups that exist within the sample, establish classification rules (viz., discriminant functions) based on the characteristics of the cases whose group membership is known, and assign new cases to these groups using the established classification rules. In this analysis, however, discriminant analysis is classifying cases for which group membership is known. Predictive discriminant analysis involves the use of derived discriminant functions to classify new cases for which group membership is not known.

In order to perform a predictive discriminant analysis, researchers must have the classification rules (viz., discriminant functions) from a previous descriptive discriminant analysis or researchers must perform such an analysis to obtain those rules. The investigator can perform a descriptive discriminant analysis on the cases of known membership and then use the derived discriminant functions to classify cases for which group membership is not known. When the cases of unknown group membership are contained in a sample of cases for which group membership is known, some statistical packages allow the investigator to perform a descriptive discriminant analysis on the cases of known membership and, as an option, use the discriminant functions that were derived in that analysis to classify cases in the data set for which group membership is not known.

Another method of performing classifications is to use the discriminant functions derived from one sample of cases in a descriptive analysis, which have been preassigned to a set of groups, to classify unassigned cases from other samples to those groups. This method of classifying cases makes use of the set of coefficients, called Fisher's linear discriminant functions or classification function coefficients, which are produced for each group by discriminant analysis. These functions can be obtained from one or two
sources: an investigator’s earlier descriptive discriminant analysis or that performed by earlier investigators. Fisher’s linear discriminant function coefficients for our illustrative analysis are presented in Table 8.5.

It is important to note that classification functions differ from the discriminant functions discussed earlier in that classification functions are produced for each group for which classification is sought. The discriminant functions discussed earlier were not group specific; a case would receive a score on a discriminant function and be assigned to the group for which discriminant analysis determines that the statistical probability is highest for group membership. Using classification functions, new cases will receive a discriminant score on each group-specific function. The classification function coefficients are unstandardized and are used to multiply the scores of a case on the appropriate discriminator variables. Next, the products are summed and the constant (a) is added. This yields a discriminant score for each classification function. If we had a case of unknown group membership and were interested in classifying it, the case would have received three discriminant scores based on the three classification functions presented in Table 8.5. The case should be assigned to the group for which it has the largest discriminant score.

It is important to reemphasize that researchers cannot use the standardized discriminant functions and their coefficients in these computations because the means and standard deviations used to develop the standardized coefficients in the descriptive discriminant analysis sample (also called the derivation or developmental sample) most likely will differ from the means and standard deviations of the sample to-be-classified (i.e., the cross-validation, replication, and classification samples).

We had no data for primate fossils for which group membership was unknown. Therefore, consider the following fabricated situation where an anthropologist discovers a primate fossil whose total skull length, minimum skull width, and dentary depth at canine measured 4.59 mm, .99 mm, and .30 mm, respectively. The anthropologist’s objective is to determine the primate classification of the fossil. With the classification functions in Table 8.5, the fossil obtains the following discriminant scores: fruit feeders, 142.44 \[\{(42.50 \times 4.59 \text{ mm}) + (20.53 \times .99 \text{ mm}) + (115.67 \times .30 \text{ mm}) + (-107.66) = \]

<table>
<thead>
<tr>
<th>TABLE 8.5 Fisher's Linear Discriminant Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fruit feeders</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Total skull length</td>
</tr>
<tr>
<td>Minimum skull width</td>
</tr>
<tr>
<td>Dentary depth at canine</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Nectar feeders</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Total skull length</td>
</tr>
<tr>
<td>Minimum skull width</td>
</tr>
<tr>
<td>Dentary depth at canine</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Exudate feeders</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Total skull length</td>
</tr>
<tr>
<td>Minimum skull width</td>
</tr>
<tr>
<td>Dentary depth at canine</td>
</tr>
<tr>
<td>Constant</td>
</tr>
</tbody>
</table>
V. OTHER ISSUES AND CONCERNS

A. Following up Multivariate Analysis of Variance

A common but somewhat unfortunate practice in the social sciences is to follow MANOVA by a series of univariate ANOVAs or *t*-tests. With few exceptions, we argue against such uses of ANOVA or *t* tests. We join a number of researchers who argue, instead, for the use of discriminant analysis (cf. Betz, 1987; Borgen & Seling, 1978; Bray & Maxwell, 1982; Brown & Tinsley, 1983; Huberty, 1975; Tatsuoka, 1970). MANOVA is used to determine whether groups formed on the basis of one or more independent variables differ on two or more dependent variables (see Huberty & Petoskey, chapter 7, this volume). MANOVA controls for the inflation of the Type 1 error rate associated with performing MANOVAs or multiple *t*-tests on correlated dimensions. A statistically significant MANOVA indicates the presence of group differences on the dependent variables or combinations of the dependent variables (viz., multivariates). Typically, subsequent analysis is necessary to determine what groups differ on the dependent variables and how they differ.

Discriminant analysis allows the investigator to examine group distinctiveness on the multivariates present in the set of variables; ANOVA and *t*-tests do not. In discriminant analysis, the variables relate to similar qualities and quantities that distinguish among the groups' "load" on the same functions, thereby revealing the latent dimensions underlying the dependent variables related to the group differences.

It is somewhat paradoxical that this advantage of discriminant analysis, identifying the multidimensionality underlying correlated dependent measures, is also a potential disadvantage. The interpretability of discriminant functions, specifically of the discriminant coefficients, is problematic when dependent variables (viz., discriminator variables) are highly correlated (Betz, 1987; Bray & Maxwell, 1982; Brown & Tinsley, 1983). The discriminatory power of the analysis is unaffected, but the weights assigned to the correlated variables by the discriminant analysis are reduced because they provide statistically redundant information. Though this information is redundant for predictive purposes, it may be essential for interpretive purposes. Thus, the relation of a particular discriminator variable to a particular function may be obscured if that discriminator variable is related to the other discriminator variables.
The disadvantage of discriminant analysis as a follow-up to MANOVA can be nullified by selecting uncorrelated dependent (discriminant) variables or by factor analyzing the variables to obtain orthogonal factors that can be used as the discriminator variables (cf. Brown & Tinsley, 1983; Sanathanan, 1975). An alternate strategy is to examine the correlations (i.e., the structure matrix coefficients) between scores on each discriminator variable and each function to discover whether relations exist that were obscured in the calculation of the discriminant function coefficients.

B. Stepwise Discriminant Analysis

Researchers often use stepwise discriminant analysis in an effort to discover the "best" subset of discriminator variables to use in discriminating groups. Space limitations prohibit a full treatment of how to perform and interpret such an analysis; interested readers are referred to Stevens (1996) and Huberty (1994). However, we caution readers that stepwise discriminant analysis is far less likely to yield replicable results than descriptive discriminant analysis because of its "outrageous" capitalization on chance and the nuances of the particular sample (Thompson, 1995). Using large-sample sizes and a small number of predictor variables reduces this problem only slightly (cf. Thompson, 1995). Furthermore, the significance tests used with stepwise discriminant analysis are positively biased and erroneous (cf. Rencher & Larson, 1980; Stevens, 1996; Thompson, 1995). Therefore, we agree with Thompson (1995) that stepwise discriminant analysis should be eschewed for publication purposes. We recommend instead that investigators interested in the relative importance of variables to discrimination inspect the standardized discriminant function coefficients and the structure matrix coefficients. The reader is also referred to Huberty (1994) for other strategies.

C. Reporting the Results of a Discriminant Analysis

Consumers of research in which discriminant analysis was performed need specific information to understand and evaluate the results. Researchers should report the following:

1. The mean and standard deviation of the sample on the study variables and the correlations among the variables
2. The Wilks' lambda and the chi-square statistics, the relative percents, and the canonical correlation
3. The proportion of cases correctly classified, the chance or base rate, the procedure used in calculating the chance or base rate, and the z test of proportions comparing the proportion of expected and observed classifications
4. The standardized discriminant function coefficients
5. The unstandardized classification functions for each group
6. The unstandardized function coefficients and the group centroids
7. The results of the cross-validation analysis

VI. CONCLUSION

Discriminant analysis is a powerful tool for analyzing and describing group differences and for classifying cases into groups formed on the basis of their similarities and differences on multiple variables. We used an anthropological example to illustrate the applicability of this technique in the physical and biomedical sciences, in addition to the social sciences, where its use is greater. It should be obvious that discriminant analysis also can be useful in the arts and humanities. Discriminant analysis has great utility as a tool for applied researchers confronted with the problems of selection, placement, and taxonomy, regardless of discipline. However, we can not overemphasize the importance of cross-validation and of testing classification hit rates against the base or chance rate.

REFERENCES


