DOCUMENT RESUME

ED 367 705 TM 021 173

AUTHOR Henington, Carlen
TITLE A Primer on the Use of Predictive Discriminant Analysis.
PUB DATE Jan 94
NOTE 18p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (San Antonio, TX, January 27-29, 1994).
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Correlation; Least Squares Statistics; Measurement Techniques; *Multivariate Analysis; *Prediction; Predictor Variables; Research Problems; Statistical Analysis; Validity
IDENTIFIERS *Descriptive Discriminant Analysis; *Predictive Discriminant Analysis

ABSTRACT

It has been increasingly realized that (1) multivariate methods are essential in most quantitative studies (Fish, 1988; Thompson, 1992), and (2) all conventional parametric analytic methods are correlational and invoke least squares weights (e.g., the beta weights in regression) (Knapp, 1978; Thompson, 1991). The present paper reviews one very popular multivariate analytic method that explicitly invokes weighting to optimize one criterion: the analytic method that researchers have come to call predictive discriminant analysis (Huberty and Barton, 1989; Huberty and Wisenbaker, 1992). Predictive discriminant analysis (PDA) is differentiated from descriptive discriminant analysis (DDA) (Dolenz, 1993) by a focus on predicting membership in intact groups. The paper is intended as a primer introducing researchers to the distinction between PDA and DDA and explaining analytic issues related to the PDA application (Van Epps, 1987). For example, adding predictors can actually result in worse prediction in PDA, though in no other analytic methods can adding predictors result in worse effects. Methods for evaluating predictor variable importance using leave-one-out (L-O-O) strategies are also explored. Included are two tables and one figure. (Contains 17 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 367 705

U S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it
Minor changes have been made to improve reproduction quality
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Carlen Henington

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A PRIMER ON THE USE OF PREDICTIVE DISCRIMINANT ANALYSIS

CARLEN HENINGTON

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, January 27, 1994.

TM021172

ERIC

BEST COPY AVAILABLE

Abstract

Two recent realizations have been increasingly reflected in the contemporary analytic practice of behavioral scientists. First, it has been increasingly realized that multivariate methods are essential in most quantitative studies (Fish, 12988; Thompson, 1992). Second, it has been increasingly recognized that *all* conventional parametric analytic methods are *correlational* and invoke least squares *weights* (e.g., the beta weights in regression) (Knapp, 1978; Thompson, 1991).

The present paper reviews one very popular multivariate analytic method that explicitly invokes weighting to optimize a criterion--the analytic method that researchers have come to call predictive discriminant analysis (Huberty & Barton, 1989; Huberty & Wisenbaker, 1992). Predictive discriminant analysis (PDA) is differentiated from descriptive discriminant analysis (DDA) (Dolenz, 1993) by a focus on predicting membership in intact groups.

The paper is intended as a primer introducing researchers to the distinction between PDA and DDA, and exploring analytic issues related to the PDA application (Van Epps, 1987). For example, *adding* predictors can actually result in *worse* prediction in PDA, though in no other analytic methods can adding predictors result in worse effects. Methods for evaluating predictor variable importance using leave-one-out (L-O-O) strategies will also be explored.

A Primer on the Use of Predictive Discriminant Analysis

In the social sciences we often seek to predict values of the dependent or outcome variable from a set of independent or predictor variables. Because there are a large number of variables which affect human behavior it is preferable to study as many variables which affect behavior as possible. Thus, many researchers are interested in multiple outcomes which have multiple effects (Thompson, 1986). When we use two or more dependent variables in a study, we use multivariate statistics to reduce experimentwise error rates and to identify statistically significant results which exist (Fish, 1988; Thompson, 1992; Thompson, 1991).

Multivariate statistical methods include multivariate analysis of variance, multiple regression, canonical correlation, and discriminant analysis. Multivariate analysis of variance (MANOVA) examines data in which both the dependent and independent variables are measured on an interval or ratio scale (Norusis, 1990; Pedhazur, 1982). When the dependent variables are on an interval or ratio scale and the independent variables are either interval, ratio, or ordinal, then it is possible to use multiple regression analysis (Pedhazur, 1990). Canonical correlation analysis examines relations between two sets of variables within a single group (Pedhazur, 1990). Discriminant analysis is used when there are two sets of variables which are examined simultaneously and in which the dependent variables are in ordinal or nominal scale (Huberty & Barton, 1989; Pedhazur, 1982).

This paper will address one of these multivariate methods, discriminant analysis. A brief background and definition of discriminant analysis will be given; then, the differences between predictive discriminant analysis (PDA) and descriptive discriminant analysis (DDA) will be explored with a focus on analytic issues related to the PDA application; and, finally, methods for evaluating predictor variable importance using leave-one-out (L-O-O) strategies will also be explored.

ERIC

Background Information and Definition

Developed by Fisher in 1936, discriminant analysis was intended to classify objects into one of two clearly defined groups (Pedhazur, 1982). In general, discriminant analysis determines a set of weights to assign individual scores so that the ratio of between-groups sums of squares and cross-products of pooled within-group sums of squares will be maximized. This method maximizes discrimination between the groups (Pedhazur, 1982). Although the meaning of discriminant analysis varies somewhat from researcher to researcher and from textbook to textbook, recently it has been used for two purposes, according to Huberty and Barton (1989): prediction of group membership and description of MANOVA results.

The first purpose is known as predictive discriminant analysis (PDA). PDA uses a set of independent or predictor variables, and one dependent or criterion nominally- or ordinally-scaled variable with two or more levels. The criterion variable is used for grouping purposes, also commonly known as classification. Fisher originally suggested that classification should be based on a linear combination of the discriminating variables to minimize variation within the groups and still maximize group differences (Klecka, 1980). The second purpose is known as descriptive discriminant analysis (DDA). As opposed to PDA, DDA involves a set of two or more criterion variables and a set of one or more grouping or dependent variables with two or more levels (Huberty & Barton, 1989). Thus, PDA is used to predict group membership, while DDA is used to explain or describe group differences. The two types of discriminant analysis are distinguishable by the roles of the variable sets. Huberty and Barton (1989) used the following figure to illustrate the variable roles in each of the two types of analysis.

Insert Figure 1 about here

ERIC

Basic Assumptions of Discriminant Analysis

Although discriminant analysis is a robust technique, it is important to consider the following basic assumptions (Klecka, 1980):

1) there must be two or more mutually exclusive groups;

2) there must be two or more cases per group;

3) the variables used to discriminant between the groups must be in either interval or ratio scale (this makes the use of means and variances possible);

4) any number of discriminating variables are possible, however at a bare minimum there must be at least two more cases than there are discriminating variables;

5) no discriminating variable can be a linear combination of other discriminating variables, thus no two perfectly correlated discriminating variables can be used at the same time;

6) the covariance matrices for each group should be approximately equal;

7) each group is drawn from a normal distributed population, allowing for precise computation of tests of significance and probability of group membership.

Although discriminant analysis is a statistical technique robust in its ability to withstand some violation of these assumptions. Many authors caution against violating some of these assumptions. Of specific concern is the use of an internal versus external classification rule. When the classification rule is developed with a set of cases and that same set of cases are reclassified, this is known as <u>internal</u> classification or analysis (Betz, 1987; Hsu, 1989; Huberty & Barton, 1989). In general, it is believed that internal analysis is acceptable if the number of cases in a data set is five times the number of predictor variables; however, many authors (Hsu, 1989; Huberty & Barton, 1987) report that a more desirable practice is to generate a classification rule and then use an <u>external</u> analysis in which the rule is applied to another set of cases, thus avoiding the internal tendency to overestimate the true hit rate.

A second issue is the equality of covariance matrices assumption. If this

ERIC

assumption is met, a so-called "linear' rule may be used. However, when this assumption is not met, a quadratic rule, maximizing the total probability of missclassification, may be used (Huberty & Barton, 1989).

A linear rule calculates the weights by analyzing the "pooled" covariance matrix, i.e., an average of the separate matrices of each group. A quadratic rule calculates the predictor variables' weights from the separate covariance matrices of each group. Deviation from the assumption of equality of covariance assumption can be determined through the use of an F statistic computed through the SPSS command DISCRIMINANT (Huberty & Wisenbaker, 1992). But, this test is also sensitive to deviation from multivariate normality. Discussion of the application of the quadratic form to correct for unequal variance-covariance is beyond the scope of this paper. See Joachimsthaler and Stam (1988) for a review of the procedures in computing the quadratic rule.

Another important assumption, is the third assumption (Dolenz, 1993). It is noted that as the ratio of the number of discriminator variables to the number of individuals increases, there is a likelihood that the accuracy of the discriminators decreases if the weights determined on the first sample are applied to a second sample. The seventh assumption, that of multivariate normal distribution, is not too much of an issue if the group sizes are equal (Lachenbruch, 1975). Additional concerns specific to PDA will be explored following a discussion of the differences between the two types of discriminant analysis.

Differentiating Between the Two Types of Discriminant Analysis

Although both types of discriminant analysis involve multiple response variables as well as multiple groups, the sampling design for PDA and DDA usually are different. In DDA, there will be two or more criterion variables and a set of one or more grouping variables with two or more levels (Huberty & Barton, 1989). In this situation there may be two groups with samples drawn in a quasiexperimental manner for each of the three

ERIC

criterion variables, or there may be one sample with random assignment to each of the groups. In DDA, the results are usually reported in the form of a MANOVA based on the statistical assessment of the effects of the grouping variables (Huberty & Barton, 1987; Klecka, 1980) with an F value and a p value. Additionally, a plot of the linear composite of each outcome variable may be shown. Through examination of the plot, it is possible to evaluate differences in group means.

Conversely, in PDA the response or predictor variables are used to determine group membership or classification. Here group membership is the criterion variable. This variable must have two or more levels. Using PDA, it is possible to determine: 1) which variables are useful in predicting group membership or classification, 2) how the variables might be combined within an equation to predict the most likely group membership, and 3) the accuracy of the derived equation or discriminant function (Betz, 1987; Klecka, 1980). The classification function or formula has been expressed by Klecka (1980) in the following manner:

$$h_k = b_{k0} + b_{k1}X_1 + b_{k2}X_2 + \dots b_{kp}X_p + a$$

where h_k is the score for group k, the b's are coefficients that need to be derived and which are applied to the variables X, and a is an additive constant. Individual cases are classified into the group with the highest h. Klecka refers to this formula has the "simple classification functions" (p. 43). The coefficient is derived through the following formula:

$$b_{ki} = (n.-g) \sum_{j=1}^{p} a_{ij} X_{jk}$$

where b_{ki} is the coefficient for variable i in the equation corresponding to group k, and a_{ij} is an element from the inverse of the within-groups sum of cross-products matrix (a

calculation using matrix algebra and beyond the scope of this paper). Calculation of a constant is also required:

$$b_{k0} = -.5 \sum_{j=1}^{p} b_{kj} X_{jk}$$

Therefore, variables on which the groups are similar receive smaller weights, while variables on which the groups differ more are generally weighted more heavily. Through this formula it is possible to see that group differences are maximized while group similarities are minimized. The coefficients are expressed in the form of a table of classification coefficients similar to Table 1.

Insert Table 1 about here

Through applying the coefficients in Table 1 to the variables' raw data, we would then determine the scores for each of the three groups. For example, suppose the scores obtained through applying the coefficients to variable 1 for the first case raw data were 34.888, 14.227, and 7.437; then for that individual case, the classification might be in Group 1 because the first score is the largest. Thus, we have assigned that case to the closest group--that which is "it has the highest probability of belonging" (Klecka, 1980, p. 45).

Issues in the Interpretation of Predictive Discriminant Analysis

Interpretation of the coefficients (*b*'s) are not usually informative because, like the *b* weights in regression, they are not in standardized form and are thus arbitrary numbers which only have the property that the case resembles most closely that group on which it has the highest score. Klecka (1980) states that a more informative method of classification is to measure the distances from the individual case to each of the group centroids, thus through this technique we classify the case in the closest group based on distance. Klecka

(1980) provides the following formula for this purpose:

$$D^2 (X/G_k) = (n.-g) \sum_{i=1}^{p} \sum_{j=1}^{p} a_{ij} (X_i - X_{ik}) (X_j - X_{jk})$$

where $D^2\ (X/G_k)$ is the squared distance from point (a specific case) to the centroid of group k. Once D^2 has been calculated for each group, then the case would be classified into the group with the smallest D^2. As can be seen in this formula, the smaller the D^2 the better the match or prediction. With this information about distance from the group centroid, it is possible to make inferences about the probabilities for correct classification. For example, if the D^2 are very different then it is not difficult to determine which group the case probably belongs to. However, if the D^2 are approximately the same, the case may have a probability of belonging to more than one group or perhaps none of the groups. In this situation, classification into a group is likely to be meaningless (Klecka, 1980).

It is possible to test the accuracy of the classification procedure by applying the process to cases in which the classification is already known and then applying the classification rule to them. The results are reported in the form of a classification table with actual group membership compared to predicted classification. From the classification table it is possible to estimate prediction "hit rates". The proportion of correct classifications indicates the accuracy of the classification rule and "confirms the degree of group separation" (Klecka, 1980, p. 49). See Table 2 for an example of a classification matrix.

Insert Table 2 about here

It is possible to have prediction rates which are lower than a prediction which may occur by chance. For this reason it is important to examine the percentage of correct predictions <u>based on chance</u>. If the groups are of equal size, the formula to compute

prediction by chance is $1/k$, where k is the number of groups (Betz, 1987; Daniels & Darcy, 1988). Thus if there are three groups, as in our example in Table 1, the chance likelihood of correctly classifying an individual case is 1/3 or .333. In situations where the sample sizes are unequal, Betz (1987) presents two methods of determining correct classification by chance: 1) with the assumption that all correct predictions are equal in value the formula n/N is used--n is the size of the largest group and N is the total sample size (this method assumes the same prediction for all cases and is not useful in predicting in advance those case which will not be correctly classified); and 2) a formula which assumes a comparable rate of error for all groups using the following formula:

$$p_1a_1 + p_2a_2 + p_3a_3 + \ldots p_ka_k,$$

where p values are the proportion of cases in the sample which belong to each group, a values are the proportion actually classified as belonging to that group and k is the number of groups. In explaining the shortcoming of the first formula, Betz states: "assume that we have 300 successes and 100 failures in a job training program. If we make a prediction of success for every individual, we will be correct 75% of the time: that is 300/400 = .75" (p. 396). To indicate the more accurate second formula she states: "Assume that in the earlier example, the discriminant function led to the prediction of 60% success and 40% failures. By inserting these values into the formula we would have a chance rate of correct prediction of (.75) (.60) + (.25) (.40) = .55" (p. 396). Notice that the chance of correct prediction using the second formula is much more conservative.

The final steps necessary to obtain all possible information about the ability of the usefulness of the predictor variables require further exploration. When examining the importance of the predictors it is important to look at the relative percentage (an indication of the proportion of variance due to variance in the discriminators), the canonical correlation (the proportion of variance related to differences among the groups and the

absolute amount of variance or, simply stated, the correlation coefficients between scores on each variable and the scores on each function). In making a decision about whether to include a discriminator or not it is important to look at the canonical correlation. Canonical correlations indicate the relationship between the variance accounted for by group differences and a predictive function (Brown & Tinsley, 1983; Dolenz, 1993). The proportion of variance in a function related to group differences is the squared canonical correlation. In looking at the canonical correlations a general rule is to look at the larger numbers as more important. Additionally, computation of a series of Wilk's lambda and chi-squares are made. The Wilk's lambda and chi-square also indicate the importance of the functions in accounting for group differences. The larger the lambda, the less information is remaining in the discriminator variables (Brown & Tinsley, 1983).

Another issue which should be considered in making probability estimates for classification is the situation in which borderline or questionable assignment occurs. It is important to note that discriminant analysis is a maximization procedure--it capitalizes on sample-specific error (Betz, 1987). To determine the long-term predictive accuracy of the function or classification rule it is necessary to use a method of cross-validation. Many methods have been outlined by various sources (Betz, 1987; Brown & Tinsley, 1983; Taylor, 1991). The four methods which are commonly presented in the literature are: 1) cross-validation using a holdout sample from the original sample; 2) double cross-validation; 3) the bootstrap method; and 4) the jackknife or "leaving-one-out" (L-O-O) method.

In the holdout method of cross-validation, the sample is split into two parts (it is usually desirable to have the first part have a large proportion of the cases because this will give a more stable discriminant function). A discriminant analysis is completed on the first part of the sample, then the weight from this subsample is applied to the second holdout subsample for classification. The major drawback of this method is it requires a large

sample. In double cross-validation, the total sample is divided nearly in half with separate discriminant analysis performed on each half. The discriminant functions are then applied to the opposing half. Again, the drawback of this method is its required large sample. In the third method, the bootstrap method involves creating a mega data set by duplication of the sample over and over, then performing a discriminant analysis on a random sample of this new data set. In the fourth, L-O-O, method one case at a time is held out and the discriminant analysis is computed on the remaining cases. That discriminant function is then applied to the individual cases. Error rates are computed cumulatively. The jackknife method is available on BMDP (Betz, 1987).

Summary

In conclusion, Brown and Tinsley (1983) state that five pieces of information should be reported when using discriminant analysis:

1) the standardized discriminant function coefficients;
2) the group centroids (assists the reader in determining how the groups differ on the functions, also useful in classifying future cases of unknown group membership);
3) the relative percent, absolute percent and canonical correlation (gives information about the relationships between the functions and group differences);
4) the statistical test significance; and
5) the proportion of correct classifications and the statistical significance of the cross-validation method or other replication statistics.

Based on the information provided in this paper, the usefulness of discriminant analysis as a tool to study group differences and for selection, intervention, and placement should be apparent to the reader. The basic explanations and procedures provided in this paper are intended to assist the researcher new to discriminant analysis to understand the procedures and limitations of discriminant analysis.

References

Betz, N. E. (1987). Use of discriminant analysis in counseling psychology research. Journal of Counseling Psychology, 34, 393-403.

Brown, M. T., & Tinsley, H. E. A. (1983). Discriminant analysis. Journal of Leisure Research, 15, 290-310.

Daniels, M. R., & Darcy, R. (1988). Notes on the use and interpretation of discriminant analysis. American Journal of Political Science, 27, 359-381.

Dolenz, B. (1993, January). Descriptive discriminant analysis: An application. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service NO. ED 355 274)

Fish, L. (1988). Why multivariate methods are usually vital. Measurement and Evaluation in Counseling and Development, 22, 158-168.

Huberty, C., & Wisenbaker, J. (1992). Discriminant analysis: Potential improvements in typical practice. In B. Thompson (Ed.), Advances in social science methodology (Vol. 2, pp. 169-208). Greenwich, CT: JAI Press.

Huberty, C. J., & Barton, R. M. (1989). Methods, plainly speaking: An introduction to discriminant analysis. Measurement and Evaluation in Counseling and Development, 22, 158-168.

Hsu, L. M. (1989). Discriminant analysis: A comment. Journal of Counseling Psychology, 36, 244-247.

Joachimsthaler, E. A., & Stam, A. (1988). Four approaches to the classification problem in discriminant analysis: An experimental study. Decision Sciences, 19, 322-333.

Klecka, W. J. (1980). Discriminant analysis. Beverly Hills, CA: Sage.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.

Norusis, M. J. (1990). The SPSS guide to data analysis for release 4. Chicago: SPSS,

Inc.

Pedhazur, E. J. (1982). Multiple Regression in behavioral research: Explanation and prediction (2nd ed.). Fort Worth: Harcourt Brace Jovanovich, Inc.

Taylor, D. L. (1991). Evaluating the sample specificity of discriminant analysis results using the jackknife statistic. Paper presented at the Annual Meeting of the Southwest Educational Research Association, San Antonio. (ERIC Document Reproduction Service NO. ED 320 574)

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in counseling and Development, 24 (2), 80-95.

Thompson, B. (1992). Two and one-half decades of leadership in measurement and evaluation. Journal of Counseling and Development, 70, 434-438.

Van Epps, P. D. (1987, January). Heuristics to facilitate understanding of discriminant analysis. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas. (ERIC Document Reproduction Service NO. ED 284 871)

Figure 1

Variable Roles in Predictive Discriminant Analysis and Descriptive Discriminant Analysis with One Grouping Variable

		Analysis	
		PDA	**DDA**
Variable Type	**Response**	Predictors	Criteria
	Grouping	Criterion	Predictor

(Taken from Huberty & Barton, 1989, p. 159)

ERIC

Table 1

Classification Coefficients

Variable	Group 1	Group 2	Group 3
X_1	9.821	2.193	1.056
X_2	.121	4.793	1.186
X_3	7.221	-2.789	10.888
X_4	1.333	1.563	1.003
Constant	3.789	27.456	10.897

Table 2

Classification Matrix

	Predicted Group			
Actual Group	1	2	3	Sum
1	75*	12	6	93
2	14	25*	3	42
3	10	6	33*	49
Unknown	2	4	1	7
Sum	101	47	43	191

* cases correctly classified and considered "hits"

ERIC