Using Multivariate Statistics

Third Edition

Barbara G. Tabachnick
Linda S. Fidell

California State University, Northridge

HarperCollinsCollegePublishers

Acquisitions Editor: Catherine Woods
Project Coordination: Interactive Composition Corporation
Text and Cover Design: John Massey
Cover Photograph: Rick Strange / The Picture Cube
Art Studio: Interactive Composition Corporation
Electronic Production Manager: Eric Jorgensen
Manufacturing Manager: Hilda Koparanian
Electronic Page Makeup: Interactive Composition Corporation
Printer and Binder: RR Donnelley & Sons Company
Cover Printer: New England Book Components, Inc.

Using Multivariate Statistics, Third Edition

Library of Congress Cataloging in Publication Data

Tabachnick, Barbara G.
Using multivariate statistics / Barbara G. Tabachnick, Linda S. Fidell.—3rd ed.
p. cm.
Includes bibliographical references (p. 851–860) and index.
ISBN 0-673-99414-7
1. Multivariate analysis. I. Fidell, Linda S. II. Title.
QA278.T3 1996
519.5'35—dc20 95-16499
CIP

96 97 98 9 8 7 6 5 4 3

CHAPTER 11
Discriminant Function Analysis

11.1 GENERAL PURPOSE AND DESCRIPTION

The goal of discriminant function analysis is to predict group membership from a set of predictors. For example, can a differential diagnosis between a group of normal children, a group of children with learning disability, and a group with emotional disorder be made reliably from a set of psychological test scores? The three groups are normal children, children with learning disability, and children with emotional disorder. The predictors are a set of psychological test scores such as the Illinois Test of Psycholinguistic Ability, subtests of the Wide Range Achievement Test, Figure Drawing tests, and the Wechsler Intelligence Scale for Children.

Discriminant function analysis (DISCRIM) is MANOVA turned around. In MANOVA, we ask whether group membership is associated with reliable mean differences on a combination of DVs. If the answer to that question is yes, then the combination of variables can be used to predict group membership—the DISCRIM perspective. In univariate terms, a significant difference between groups implies that, given a score, you can predict (imperfectly, no doubt) which group it comes from.

Semantically, however, confusion arises between MANOVA and DISCRIM because in MANOVA the IVs are the groups and the DVs predictors while in DISCRIM the IVs are the predictors and the DVs groups. We have tried to avoid confusion here by always referring to IVs as "predictors" and to DVs as "groups" or "grouping variables."[1]

Mathematically, MANOVA and DISCRIM are the same, although the emphases often differ. The major question in MANOVA is whether group membership is associated with reliable mean differences in combined DV scores, analogous in DISCRIM to the question of whether predictors can be combined to predict group membership reliably. In many cases, DISCRIM is carried to the point of actually putting cases into groups in a process called classification.

[1] Many texts also refer to IVs or predictors as discriminating variables and to DVs or groups as classification variables. However, there are also discriminant functions and classification functions to contend with, so the terminology becomes quite confusing. We have tried to simplify it by using only the terms predictors and groups.

Classification is a major extension of DISCRIM over MANOVA. Most computer programs for DISCRIM evaluate the adequacy of classification. How well does the classification procedure do? How many learning-disabled kids in the original sample, or a cross-validation sample, are classified correctly? When errors occur, what is their nature? Are learning-disabled kids more often confused with normal kids or with kids suffering emotional disorder?

A second difference involves interpretation of differences among the predictors. In MANOVA, there is frequently an effort to decide which DVs are associated with group differences, but rarely an effort to interpret the pattern of differences among the DVs as a whole. In DISCRIM, there is often an effort to interpret the pattern of differences among the predictors as a whole in an attempt to understand the dimensions along which groups differ.

Complexity arises with this attempt, however, because with more than two groups there may be more than one way to combine the predictors to differentiate among groups. There may, in fact, be as many dimensions that discriminate among groups as there are degrees of freedom for groups. For instance, if there are three groups, there may be two dimensions that discriminate among groups: a dimension that separates the first group from the second and third groups, and a dimension that separates the second group from the third group, for example. This process is conceptually similar to contrasts in ANOVA except that in DISCRIM the divisions among groups are almost never as neat and clean.

In our example of three groups of children (normal, learning-disabled, and emotionally disordered) given a variety of psychological measures, one way of combining the psychological test scores may tend to separate the normal group from the two groups with disorders, while a second way of combining the test scores may tend to separate the group with learning disability from the group with emotional disorder. The researcher attempts to understand the "message" in the two ways of combining test scores to separate groups differently. What is the meaning of the combination of scores that separates normal from disordered kids, and what is the meaning of the different combination of scores that separates kids with one kind of disorder from kids with another? This attempt is facilitated by statistics available in many of the canned computer programs for DISCRIM that are not printed in some programs for MANOVA.

Thus there are two facets of DISCRIM, and one or both may be emphasized in any given research application. The researcher may simply be interested in a decision rule for classifying cases where the number of dimensions and their meaning is irrelevant. Or, the emphasis may be on interpreting the results of DISCRIM in terms of the combinations of predictors—called discriminant functions—that separate various groups from each other.

A DISCRIM version of covariates analysis (MANCOVA) is available, because DISCRIM can be set up in a sequential manner. When sequential DISCRIM is used, the covariate is simply a predictor that is given top priority. For example, a researcher might consider the score on the Wechsler Intelligence Scale for Children a covariate and ask how well the Wide Range Achievement Test, the Illinois Test of Psycholinguistic Ability, and Figure Drawings differentiate between normal, learning-disabled, and emotionally disordered children after differences in IQ are accounted for.

If groups are arranged in a factorial design, it is frequently best to rephrase research questions so that they are answered within the framework of MANOVA. (However, DISCRIM can in some circumstances be directly applied to factorial designs as discussed in Section 11.6.5.) Similarly, DISCRIM programs make no provision for within-subjects variables. If a within-subjects analysis is desired, the question is also rephrased in terms of MANOVA or Profile Analysis. For this reason the emphasis in this chapter is on one-way between-subjects DISCRIM.

11.2 ■ KINDS OF RESEARCH QUESTIONS

The primary goals of DISCRIM are to find the dimension or dimensions along which groups differ and to find classification functions to predict group membership. The degree to which these goals are met depends, of course, on choice of predictors. Typically, the choice is made either on the basis of theory about which variables should provide information about group membership, or on the basis of pragmatic considerations such as expense, convenience, or unobtrusiveness.

It should be emphasized that the same data are profitably analyzed through either MANOVA or DISCRIM programs, and frequently both, depending on the kinds of questions you want to ask. If group sizes are very unequal, and/or distributional assumptions are untenable, logistic regression also answers most of the same questions. In any event, statistical procedures are readily available within canned computer programs for answering the following types of questions generally associated with DISCRIM.

11.2.1 ■ Significance of Prediction

Can group membership be predicted reliably from the set of predictors? For example, can we do better than chance in predicting whether children are learning-disabled, emotionally disordered, or normal on the basis of the set of psychological test scores? This is the major question of DISCRIM that the statistical procedures described in Section 11.6.1 are designed to answer. The question is identical to the question about "main effects of IVs" for a one-way MANOVA.

11.2.2 ■ Number of Significant Discriminant Functions

Along how many dimensions do groups differ reliably? For the three groups of children in our example, two discriminant functions are possible, and neither, one, or both may be statistically reliable. For example, the first function may separate the normal group from the other two while the second, which would separate the group with learning disability from the group with emotional disorder, is not reliable. This pattern of results indicates the predictors can differentiate normal from abnormal kids, but cannot separate learning disabled kids from kids with emotional disorder.

In DISCRIM, like canonical correlation (Chapter 6), the first discriminant function provides the best separation among groups. Then a second discriminant function, orthogonal to the first, is found that best separates groups on the basis of associations not used in the first discriminant function. This procedure of finding successive orthogonal discriminant functions continues until all possible dimensions are evaluated. The number of possible dimensions is either one fewer than the number of groups or equal to the number of predictor variables, whichever is smaller. Typically, only the first one or two discriminant functions reliably discriminate among groups; remaining functions provide no additional information about group membership and are better ignored. Tests of significance for discriminant functions are discussed in Section 11.6.2.

11.2.3 ■ Dimensions of Discrimination

How can the dimensions along which groups are separated be interpreted? Where are groups located along the discriminant functions and how do predictors correlate with the discriminant functions? In our example, if two significant discriminant functions are found, which predictors correlate highly with each function? What pattern of test scores discriminates between normal children and the other two groups (first discriminant function)? And what pattern of scores discriminates between children with learning disability and children with emotional disorder (second discriminant function)? These questions are discussed in Section 11.6.3.

11.2.4 ■ Classification Functions

What linear equation(s) can be used to classify new cases into groups? For example, suppose we have the battery of psychological test scores for a group of new, undiagnosed children. How can we combine (weight) their scores to achieve the most reliable diagnosis? Procedures for deriving and using classification functions are discussed in Sections 11.4.2 and 11.6.6.[2]

11.2.5 ■ Adequacy of Classification

Given classification functions, what proportion of cases is correctly classified? When errors occur, how are cases misclassified? For instance, what proportion of learning-disabled children is correctly classified as learning-disabled, and, among those who are incorrectly classified, are they more often put into the group of normal children or into the group of emotionally disordered children?

Classification functions are used to predict group membership for new cases and to check the adequacy of classification for cases in the same sample through cross-validation. If the researcher knows that some groups are more likely to occur, or if some kinds of misclassification are especially undesirable, the classification procedure can be modified. Procedures for deriving classification functions and modifying them are discussed in Section 11.4.2; procedures for testing them are discussed in Section 11.6.6.

11.2.6 ■ Strength of Association

What is the degree of relationship between group membership and the set of predictors? If the first discriminant function separates the normal group from the other two groups, how much does the variance for groups overlap the variance in combined test scores? If the second discriminant function separates learning-disabled from emotionally disordered children, how much does the variance for these groups overlap the combined test scores for this discriminant function? This is basically a question of percent of variance accounted for and, as seen in Section 11.4.1, is answered through canonical correlation. A canonical correlation is found for each discriminant function that, when squared, indicates the proportion of variance shared between groups and predictors on that function.

11.2.7 ■ Importance of Predictor Variables

Which predictors are most important in predicting group membership? Which test scores are helpful for separating normal children from children with disorders, and which are helpful for separating learning-disabled from emotionally disordered children?

[2] Discriminant function analysis provides classification of cases into groups where group membership is known, at least for the sample from whom the classification equations are derived. Cluster analysis is a similar procedure except that group membership is not known. Instead, the analysis develops groups on the basis of similarities among cases.

Questions about importance of predictors are analogous to those of importance of DVs in MANOVA, to those of IVs in multiple regression, and to those of IVs and DVs in canonical correlation. One procedure in DISCRIM is to interpret the correlations between the predictors and the discriminant functions, as discussed in Section 11.6.3.2. A second procedure is to evaluate predictors by how well they separate each group from all the others, as discussed in Section 11.6.4. (Or importance can be evaluated as in MANOVA, Section 9.5.2.)

11.2.8 Significance of Prediction with Covariates

After statistically removing the effects of one or more covariates, can one reliably predict group membership from a set of predictors? In DISCRIM, as in MANOVA, the ability of some predictors to promote group separation can be assessed after adjustment for prior variables. If scores on the Wechsler Intelligence Scale for Children (WISC) are considered the covariate and given first entry in DISCRIM, do scores on the Illinois Test of Psycholinguistic Ability (ITPA), the Wide Range Achievement Test, and Figure Drawings contribute to prediction of group membership when they are added to the equation?

Rephrased in terms of sequential discriminant function analysis, the question becomes, Do scores on the ITPA, the Wide Range Achievement Test, and Figure Drawings provide significantly better classification among the three groups than that afforded by scores on the WISC alone? Sequential DISCRIM is discussed in Section 11.5.2. A test for contribution of added predictors is given in Section 11.6.6.3.

11.2.9 Estimation of Group Means

If predictors discriminate among groups, it is important to report just how the groups differ on those variables. The best estimate of central tendency in a population is the sample mean. If, for example, the ITPA discriminates between groups with learning disability and emotional disorder, it is worthwhile to compare and report the mean ITPA score for learning-disabled children and the mean ITPA score for emotionally disordered children.

11.3 LIMITS TO DISCRIMINANT FUNCTION ANALYSIS

11.3.1 Theoretical Issues

Because DISCRIM is typically used to predict membership in naturally occurring groups rather than groups formed by random assignment, questions such as why we can reliably predict group membership, or what causes differential membership are often not asked. If, however, group membership has occurred by random assignment, inferences of causality are justifiable as long as proper experimental controls have been instituted. The DISCRIM question then becomes, Does treatment following random assignment to groups produce enough difference in the predictors that we can now reliably separate groups on the basis of those variables?

As implied, limitations to DISCRIM are the same as limitations to MANOVA. The usual difficulties of generalizability apply to DISCRIM. But the cross-validation procedure described in Section 11.6.6.1 gives some indication of the generalizability of a solution.

11.3.2 Practical Issues

Practical issues for DISCRIM are basically the same as for MANOVA. Therefore, they are discussed here only to the extent of identifying the similarities between MANOVA and DISCRIM and identifying the situations in which assumptions for MANOVA and DISCRIM differ.

Classification makes fewer statistical demands than does inference. If classification is the primary goal, then, most of the following requirements (except for outliers and homogeneity of variance-covariance matrices) are relaxed. If, for example, you achieve 95% accuracy in classification, you hardly worry about the shape of distributions. Nevertheless, DISCRIM is optimal under the same conditions where MANOVA is optimal; and, if the classification rate is unsatisfactory, it may be because of violation of assumptions or limitations.

11.3.2.1 Unequal Sample Sizes and Missing Data

As DISCRIM is typically a one-way analysis, *no special problems are posed by unequal sample sizes in groups.*[3] In classification, however, a decision is required as to whether you want the a priori probabilities of assignment to groups to be influenced by sample size. That is, do you want the probability with which a case is assigned to a group to reflect the fact that the group itself is more (or less) probable in the sample? Section 11.4.2 discusses this issue, and use of unequal a priori probabilities is demonstrated in Section 11.8.

Regarding missing data (absence of scores on predictors for some cases), consult Section 8.3.2.1 and Chapter 4 for a review of problems and potential solutions.

As discussed in Section 9.3.2.1, *the sample size of the smallest group should exceed the number of predictor variables.* Although sequential and stepwise DISCRIM avoid the problems of multicollinearity and singularity by a tolerance test at each step, overfitting (producing results so close to the sample they don't generalize to other samples) occurs with all forms of DISCRIM if the number of cases does not notably exceed the number of predictors in the smallest group.

11.3.2.2 Multivariate Normality

When using statistical inference in DISCRIM, the assumption of multivariate normality is that scores on predictors are independently and randomly sampled from a population, and that the sampling distribution of any linear combination of predictors is normally distributed. No tests are currently feasible for testing the normality of all linear combinations of sampling distributions of means of predictors.

However, DISCRIM, like MANOVA, is robust to failures of normality if violation is caused by skewness rather than outliers. *A sample size that would produce 20 df for error in the univariate ANOVA case should ensure robustness with respect to multivariate normality, as long as sample sizes are equal and two-tailed tests are used.* (Calculation of df for error in the univariate case is discussed in Section 3.2.1.)

Because tests for DISCRIM typically are two-tailed, this requirement poses no difficulty. Sample sizes, however, are often not equal for applications of DISCRIM because naturally occurring groups rarely occur or are sampled with equal numbers of cases in groups. As differences in sample size

[3] Actually a problem does occur because discriminant functions may be nonorthogonal with unequal n (cf. Chapter 13), but rotation of axes is uncommon in discriminant function analysis. Also, highly unequal sample sizes are better handled by logistic regression (Chapter 12) than by discriminant function analysis.

among groups increase, larger overall sample sizes are necessary to assure robustness. As a conservative recommendation, robustness is expected with 20 cases in the smallest group if there are only a few predictors (say, five or fewer).

If samples are both small and unequal in size, assessment of normality is a matter of judgment. Are predictors expected to have normal sampling distributions in the population being sampled? If not, transformation of one or more predictors (cf. Chapter 4) may be worthwhile.

11.3.2.3 ■ Outliers

DISCRIM, like MANOVA, is highly sensitive to inclusion of outliers. Therefore, *run a test for univariate and multivariate outliers for each group separately, and transform or eliminate significant outliers before DISCRIM* (see Chapter 4).

11.3.2.4 ■ Homogeneity of Variance-Covariance Matrices

In inference, when sample sizes are equal or large, DISCRIM, like MANOVA (Section 9.3.2.4) is robust to violation of the assumption of equality of within-group variance-covariance (dispersion) matrices. However, when sample sizes are unequal and small, results of significance testing may be misleading if there is heterogeneity of the variance-covariance matrices.

Although inference is usually robust with respect to heterogeneity of variance-covariance matrices with decently sized samples, classification is not. Cases tend to be overclassified into groups with greater dispersion. If classification is an important goal of analysis, test for homogeneity of variance-covariance matrices.

Homogeneity of variance-covariance matrices is assessed through procedures of Section 9.3.2.4, or by *inspection of scatterplots of scores on the first two canonical discriminant functions produced separately for each group.* These scatterplots are available through BMDP7M or SPSS DISCRIMINANT. *Rough equality in overall size of the scatterplots is evidence of homogeneity of variance-covariance matrices (cf. Section 11.8.1.5).*

Anderson's test, available in BMDP5M or SAS DISCRIM (POOL=TEST) assesses homogeneity of variance-covariance matrices, but is also sensitive to nonnormality. Another overly sensitive test, Box's M, is available in SPSS MANOVA and DISCRIMINANT.

If heterogeneity is found, one can transform predictors, use separate covariance matrices during classification, use quadratic discriminant function analysis, or use nonparametric classification. Transformation of predictors follows procedures of Chapter 4. Classification on the basis of separate covariance matrices, the second remedy, is available through SPSS DISCRIMINANT and SAS DISCRIM. Because this procedure often leads to overfitting, it should be used only when the sample is large enough to permit cross-validation (Section 11.6.6.1). Quadratic discrimination function analysis, the third remedy, is available in BMDP5M, SYSTAT DISCRIM, and SAS DISCRIM. This procedure avoids overclassification into groups with greater dispersion, but performs poorly with small samples (Norušis, 1990).

SAS DISCRIM uses separate matrices and computes quadratic discriminant functions with the instruction POOL=NO. With the instruction POOL=TEST, SAS DISCRIM uses the pooled variance-covariance matrix only if heterogeneity of variance-covariance matrices is not significant. With small samples, nonnormal predictors, and heterogeneity of variance-covariance matrices, SAS DISCRIM offers a fourth remedy, nonparametric classification methods, which avoid overclassification into groups with greater dispersion and are robust to nonnormality.

Therefore, *transform variables if there is significant departure from homogeneity, samples are small and unequal, and inference is the major goal. If the emphasis is on classification and dispersions are unequal, use (1) separate covariance matrices and/or quadratic discriminant analysis if samples are normal and large and (2) nonparametric classification methods if variables are nonnormal and/or samples are small.*

11.3.2.5 Linearity

The DISCRIM model assumes linear relationships among all pairs of predictors within each group. The assumption is less serious (from some points of view) than others, however, in that violation leads to reduced power rather than increased Type I error. The procedures in Section 8.3.2.6 may be applied to test for and improve linearity and to increase power.

11.3.2.6 Multicollinearity and Singularity

Multicollinearity or singularity may occur with highly redundant predictors, making matrix inversion unreliable. Fortunately, most computer programs for DISCRIM protect against this possibility by testing tolerance. Predictors with insufficient tolerance are excluded.

Guidelines for assessing multicollinearity and singularity for programs that do not include tolerance tests, and for dealing with multicollinearity or singularity when it occurs, are in Section 9.3.2.8. Note that analysis is done on predictors, not "DVs" in DISCRIM.

11.4 FUNDAMENTAL EQUATIONS FOR DISCRIMINANT FUNCTION ANALYSIS

Hypothetical scores on four predictors are given for three groups of learning-disabled children for demonstration of DISCRIM. Scores for three cases in each of the three groups are shown in Table 11.1.

The three groups are MEMORY (children whose major difficulty seems to be with tasks related to memory), PERCEPTION (children who show difficulty in visual perception), and COMMUNICATION (children with language difficulty). The four predictors are PERF (Performance Scale IQ of the WISC), INFO (Information subtest of the WISC), VERBEXP (Verbal Expression subtest of the ITPA), and AGE (chronological age in years). The grouping variable, then, is type of learning disability, and the predictors are selected scores from psychodiagnostic instruments and age.

Fundamental equations are presented for two major parts of DISCRIM: discriminant functions and classification equations. Setup and selected output for this example appear in Section 11.4.3 for BMDP7M, SPSS DISCRIMINANT, SAS DISCRIM, and SYSTAT DISCRIM.

11.4.1 Derivation and Test of Discriminant Functions

The fundamental equations for testing the significance of a set of discriminant functions are the same as for MANOVA, discussed in Chapter 9. Variance in the set of predictors is partitioned into two sources: variance attributable to differences between groups and variance attributable to differences within groups. Through procedures shown in Equations 9.1 to 9.3, cross-products matrices are formed.

TABLE 11.1 HYPOTHETICAL SMALL DATA FOR ILLUSTRATION OF DISCRIMINANT FUNCTION ANALYSIS

	Predictors			
Group	PERF	INFO	VERBEXP	AGE
	87	5	31	6.4
MEMORY	97	7	36	8.3
	112	9	42	7.2
	102	16	45	7.0
PERCEPTION	85	10	38	7.6
	76	9	32	6.2
	120	12	30	8.4
COMMUNICATION	85	8	28	6.3
	99	9	27	8.2

$$\mathbf{S}_{\text{total}} = \mathbf{S}_{bg} + \mathbf{S}_{wg} \tag{11.1}$$

The total cross-products matrix ($\mathbf{S}_{\text{total}}$) is partitioned into a cross-products matrix associated with differences between groups ($\mathbf{S}_{bg}$) and a cross-products matrix of differences within groups ($\mathbf{S}_{wg}$).

For the example in Table 11.1, the resulting cross-products matrices are

$$\mathbf{S}_{bg} = \begin{bmatrix} 314.89 & -71.56 & -180.00 & 14.49 \\ -71.56 & 32.89 & 8.00 & -2.22 \\ -180.00 & 8.00 & 168.00 & -10.40 \\ 14.49 & -2.22 & -10.40 & 0.74 \end{bmatrix}$$

$$\mathbf{S}_{wg} = \begin{bmatrix} 1286.00 & 220.00 & 348.33 & 50.00 \\ 220.00 & 45.33 & 73.67 & 6.37 \\ 348.33 & 73.67 & 150.00 & 9.73 \\ 50.00 & 6.37 & 9.73 & 5.49 \end{bmatrix}$$

Determinants[4] for these matrices are

$$|\mathbf{S}_{wg}| = 4.70034789 \times 10^{13}$$

$$|\mathbf{S}_{bg} - \mathbf{S}_{wg}| = 448.63489 \times 10^{13}$$

Following procedures in Equation 9.4, Wilks' Lambda[5] for these matrices is

$$\Lambda = \frac{|\mathbf{S}_{wg}|}{|\mathbf{S}_{bg} + \mathbf{S}_{wg}|} = .010477$$

[4] A determinant, as described in Appendix A, can be viewed as a measure of generalized variance of a matrix.

[5] Alternative statistical criteria are discussed in Section 11.6.1.1.

To find the approximate F ratio, as per Equation 9.5, the following values are used:

$p = 4$ the number of predictor variables

$df_{effect} = 2$ the number of groups minus one, or $k - 1$

$df_{error} = 6$ the number of groups times the quantity $n - 1$, where n is the number of cases per group. Because n is often not equal for all groups in DISCRIM, an alternative equation for df_{error} is $N - k$, where N is the total number of cases in all groups—9 in this case.

Thus we obtain

$$s = \sqrt{\frac{(4)^2(2)^2 - 4}{2(4)^2 + (2)^2 - 5}} = 2$$

$$y = (.010477)^{\frac{1}{2}} = .102357$$

$$df_1 = 4(2) = 8$$

$$df_2 = (2)\left[6 - \frac{4 - 2 + 1}{2}\right] - \left[\frac{4(2) - 2}{2}\right] = 6$$

$$\text{Approximate } F(8, 6) = \left(\frac{1 - .102357}{.102357}\right)\left(\frac{6}{8}\right) = 6.58$$

Critical F with 8 and 6 df at $\alpha = 0.05$ is 4.15. Because obtained F exceeds critical F, we conclude that the three groups of children can be distinguished on the basis of the combination of the four predictors.

This is a test of overall relationship between groups and predictors. It is the same as the overall test of a main effect in MANOVA. In MANOVA, this result is followed by an assessment of the importance of the various DVs to the main effect. In DISCRIM, however, when an overall relationship is found between groups and predictors, the next step is to examine the discriminant functions that compose the overall relationship.

The maximum number of discriminant functions is either (1) the number of predictors or (2) the degrees of freedom for groups, whichever is smaller. Because there are three groups (and four predictors) in this example, there are potentially two discriminant functions contributing to the overall relationship. And, because the overall relationship is reliable, at least the first discriminant function is very likely to be reliable, and both may be reliable.

Discriminant functions are like regression equations; a discriminant function score for a case is predicted from the sum of the series of predictors, each weighted by a coefficient. There is one set of discriminant function coefficients for the first discriminant function, a second set of coefficients for the second discriminant function, and so forth. Subjects get separate discriminant function scores for each discriminant function when their own scores on predictors are inserted into the equations.

To solve for the (standardized) discriminant function score for the ith function, Equation 11.2 is used.

$$D_i = d_{i1}z_1 + d_{i2}z_2 + \ldots + d_{ip}z_p \tag{11.2}$$

> A child's standardized score on the ith discriminant function (D_i) is found by multiplying the standardized score on each predictor (z) by its standardized discriminant function coefficient (d_i) and then adding the products for all predictors.

Discriminant function coefficients are found in the same manner as are coefficients for canonical variates (Section 6.4.2). In fact, DISCRIM is basically a problem in canonical correlation with group membership on one side of the equation and predictors on the other, where successive canonical variates (here called discriminant functions) are computed. In DISCRIM, d_i are chosen to maximize differences between groups relative to differences within groups.

Just as in multiple regression, Equation 11.2 can be written either for raw scores or for standardized scores. A discriminant function score for a case, then, can also be produced by multiplying the raw score on each predictor by its associated unstandardized discriminant function coefficient, adding the products over all predictors, and adding a constant to adjust for the means. The score produced in this way is the same D_i as produced in Equation 11.2.

The mean of each discriminant function over all cases is zero, because the mean of each predictor, when standardized, is zero. The standard deviation of each D_i is 1.

Just as D_i can be calculated for each case, a mean value of D_i can be calculated for each group. The members of each group considered together have a mean score on a discriminant function that is the distance of the group, in standard deviation units, from the zero mean of the discriminant function. Group means on D_i are typically called centroids in reduced space, the space having been reduced from that of the p predictors to a single dimension, or discriminant function.

A canonical correlation is found for each discriminant function following procedures in Chapter 6. Successive discriminant functions are evaluated for significance, as discussed in Section 11.6.2. Also discussed in subsequent sections are loading matrices and group centroids.

If there are only two groups, discriminant function scores can be used to classify cases into groups. A case is classified into one group if its D_i score is above zero, and into the other group if the D_i score is below zero. With numerous groups, classification is possible from the discriminant functions, but it is simpler to use the procedure in the following section.

11.4.2 ■ Classification

To assign cases into groups, a classification equation is developed for each group. Three classification equations are developed for the example in Table 11.1, where there are three groups. Data for each case are inserted into each classification equation to develop a classification score for each group for the case. The case is assigned to the group for which it has the highest classification score.

In its simplest form, the basic classification equation for the jth group ($j = 1, 2, \ldots, k$) is

$$C_j = c_{j0} + c_{j1}X_1 + c_{j2}X_2 + \ldots c_{jp}X_p \tag{11.3}$$

A score on the classification function for group j (C_j) is found by multiplying the raw score on each predictor (X) by its associated classification function coefficient (c_j), summing over all predictors, and adding a constant c_{j0}.

Classification coefficients, c_j, are found from the means of the p predictors and the pooled within-group variance-covariance matrix, **W**. The within-group covariance matrix is produced by dividing each element in the cross-products matrix, $\mathbf{S}_{wg}$, by the within-group degrees of freedom, $N - k$. In matrix form,

$$\mathbf{C}_j = \mathbf{W}^{-1}\mathbf{M}_j \tag{11.4}$$

The row matrix of classification coefficients for group j ($\mathbf{C}_j = c_{j1}, c_{j2}, \ldots, c_{jp}$) is found by multiplying the inverse of the within-group variance-covariance matrix ($\mathbf{W}^{-1}$) by a column matrix of means for group j on the p variables ($\mathbf{M}_j = \overline{X}_{j1}, \overline{X}_{j2}, \ldots \overline{X}_{jp}$).

The constant for group j, c_{j0}, is found as follows:

$$c_{j0} = (-\frac{1}{2})\mathbf{C}_j\mathbf{M}_j \tag{11.5}$$

The constant for the classification function for group j (c_{j0}) is formed by multiplying the row matrix of classification coefficients for group j ($\mathbf{C}_j$) by the column matrix of means for group j ($\mathbf{M}_j$).

For the sample data, each element in the $\mathbf{S}_{wg}$ matrix from Section 11.4.1 is divided by $df_{wg} = df_{error} = 6$ to produce the within-group variance-covariance matrix:

$$\mathbf{W}_{bg} = \begin{bmatrix} 214.33 & 36.67 & 58.06 & 8.33 \\ 36.67 & 7.56 & 12.28 & 1.06 \\ 58.06 & 12.28 & 25.00 & 1.62 \\ 8.33 & 1.06 & 1.62 & 0.92 \end{bmatrix}$$

The inverse of the within-group variance-covariance matrix is

$$\mathbf{W}^{-1} = \begin{bmatrix} 0.04362 & -0.20195 & 0.00956 & -0.17990 \\ -0.21095 & 1.62970 & -0.37037 & 0.60623 \\ 0.00956 & -0.37037 & 0.20071 & -0.01299 \\ -0.17990 & 0.60623 & -0.01299 & 2.05006 \end{bmatrix}$$

Multiplying $\mathbf{W}^{-1}$ by the column matrix of means for the first group gives the matrix of classification coefficients for that group, as per Equation 11.4.

$$\mathbf{C}_1 = \mathbf{W}^{-1}\begin{bmatrix} 98.67 \\ 7.00 \\ 36.33 \\ 7.30 \end{bmatrix} = [1.92, -17.56, 5.55, 0.99]$$

The constant for group 1, then, according to Equation 11.5, is

$$c_{1,0} = (-\frac{1}{2})\,[1.92, -17.56, 5.55, 0.99]\begin{bmatrix} 98.67 \\ 7.00 \\ 36.33 \\ 7.30 \end{bmatrix} = -137.83$$

(Values used in these calculations were carried to several decimal places before rounding.) When these procedures are repeated for groups 2 and 3, the full set of classification equations is produced, as shown in Table 11.2.

In its simplest form, classification proceeds as follows for the first case in group 1. Three classification scores, one for each group, are calculated for the case by applying Equation 11.3:

$$C_1 = -137.83 + (1.92)(87) + (-17.56)(5) + (5.55)(31) + (0.99)(6.4) = 119.80$$

$$C_2 = -71.29 + (0.59)(87) + (-8.70)(5) + (4.12)(31) + (5.02)(6.4) = 96.39$$

$$C_3 = -71.24 + (1.37)(87) + (-10.59)(5) + (2.97)(31) + (2.91)(6.4) = 105.69$$

Because this child has the highest classification score in group 1, the child is assigned to group 1, a correct classification in this case.

This simple classification scheme is most appropriate when equal group sizes are expected in the population. If unequal group sizes are expected, the classification procedure can be modified by setting a priori probabilities to group size. Although a number of highly sophisticated classification schemes have been suggested (e.g., Tatsuoka, 1975), the most straightforward involves adding to each classification equation a term that adjusts for group size.[6] The classification equation for group j (C_j) then becomes

$$C_j = c_{j0} + \sum_{i=1}^{p} c_{ji}X_i + \ln(n_j/N) \tag{11.6}$$

where n_j = size of group j and N = total sample size.

TABLE 11.2 CLASSIFICATION FUNCTION COEFFICIENTS FOR SAMPLE DATA OF TABLE 11.1

	Group 1: MEMORY	Group 2: PERCEP	Group 3: COMMUN
PERF	1.92420	0.58704	1.36552
INFO	-17.56221	-8.69921	-10.58700
VERBEXP	5.54585	4.11679	2.97278
AGE	0.98723	5.01749	2.91135
(CONSTANT)	-137.82892	-71.28563	-71.24188

[6] Output from computer programs reflects this adjustment.

It should be reemphasized that the classification procedures are highly sensitive to heterogeneity of variance-covariance matrices. Cases are more likely to be classified into the group with the greatest dispersion—that is, into the group for which the determinant of the within-group covariance matrix is greatest. Section 11.3.2.4 provides suggestions for dealing with this problem.

Uses of classification procedures are discussed more fully in Section 11.6.6.

11.4.3 Computer Analyses of Small Sample Example

Setup and selected output for computer analyses of the data in Table 11.1, using the simplest methods, are in Tables 11.3 through 11.6, for BMDP7M, SPSS DISCRIMINANT, SAS DISCRIM, and SYSTAT DISCRIM, respectively. There is less consensus than usual in these programs regarding the statistics that are reported owing, in part, to the complexity of discriminant function analysis.

Table 11.3 shows a BMDP7M run with all but the most relevant output suppressed or omitted. Direct DISCRIM is specified in the LEVEL instruction. The first two zeros delete the first two predictors named in the VARIABLE paragraph from the problem. The remaining four predictors, with level 1, are to be used as predictors. FORCE=1 forces all predictors with level 1 into the equation. Because all predictors to be included have the same level, this is a direct rather than a stepwise or sequential analysis.

In the output, PRIOR PROBABILITIES of group membership are set equal by default. At STEP NUMBER 0 univariate ANOVAs for each predictor considered separately, with df = 2 and 6, are given; F ratios for each predictor appear in the column labeled F TO ENTER. TOLERNCE is 1-SMC for each predictor already in the equation. Because there are no predictors in the equation at this point, SMCs are zero. The NO STEP instruction suppresses printing of steps 2 and 3.

At STEP NUMBER 4 all predictors have entered the equation. F TO REMOVE (with 2 and 3 df) tests the significance of the reduction in prediction if a predictor is removed from the equation. Both Wilks' Lambda and the APPROXIMATE F-STATISTIC for the multivariate effect (cf. Sections 9.4.1 and 11.4.1) are provided as a test of the discriminant functions considered together. The F-MATRIX, with 4 and 3 df, shows multivariate pairwise comparisons among the three groups. For example, multivariate $F(4, 3) = 9.70$ for the difference between groups 1 and 2.

In the CLASSIFICATION FUNCTIONS section, the CLASSIFICATION MATRIX shows all cases to be perfectly classified. Rows represent actual group membership; columns show the group into which cases are classified, following procedures in Section 11.4.2. With JACKKNIFED CLASSIFICATION, in which classification equations are developed for each case with that case deleted (cf. Section 11.6.6.2), one case in the third group is misclassified into the second group, reducing the PERCENT CORRECT for group 3 to 66.7 and the overall percentage to 88.9.

The EIGENVALUES show the relative proportion of variance contributed by the two discriminant functions, followed by the CUMULATIVE PROPORTION OF TOTAL DISPERSION (variance) in the solution accounted for by the functions—the first discriminant function accounts for 70.699 percent of the variance in the solution while the second accounts for the remaining variance. This is followed by the results of four multivariate tests, described in Section 9.5.1.

The remaining output describes the two discriminant functions (CANONICAL VARIABLES) produced for the three groups. CANONICAL CORRELATIONS are the multiple correlations between the predictors and the discriminant functions. COEFFICIENTS FOR CANONICAL VARIABLES are weighting coefficients for canonical correlation as described in Chapter 6, and STANDARDIZED .... COEFFICIENTS FOR CANONICAL VARIABLES are raw score discriminant function coefficients computed using the pooled within-cell variance-covariance matrix.

TABLE 11.3 SETUP AND SELECTED BMDP7M OUTPUT FOR DIRECT DISCRIMINANT FUNCTION ANALYSIS ON SAMPLE DATA IN TABLE 11.1

```
/INPUT      VARIABLES ARE 6.  FORMAT IS FREE.  FILE = 'TAPE41.DAT'.
/VARIABLE   NAMES ARE SUBJNO,GROUP,PERF,INFO,VERBEXP,AGE.
            LABEL IS SUBJNO.
/GROUP     VAR IS GROUP.
           CODES(GROUP) ARE 1, 2, 3.
/DISCRIM   LEVEL=2*0, 4*1.  FORCE=1.
/PRINT     NO STEP.
/END

PRIOR PROBABILITIES. . . .   0.33333   0.33333   0.33333

****************************************************************************

STEP NUMBER   0

 VARIABLE      F TO    FORCE TOLERNCE *   VARIABLE      F TO    FORCE TOLERNCE
              REMOVE   LEVEL          *                ENTER    LEVEL
       DF =  2     7                  *        DF =  2     6
                                      *   3 PERF        0.73      1   1.00000
                                      *   4 INFO        2.18      1   1.00000
                                      *   5 VERBEXP     3.36      1   1.00000
                                      *   6 AGE         0.40      1   1.00000

****************************************************************************

STEP NUMBER   4
VARIABLE ENTERED    6 AGE

 VARIABLE      F TO    FORCE TOLERNCE *   VARIABLE      F TO    FORCE TOLERNCE
              REMOVE   LEVEL          *                ENTER    LEVEL
       DF =  2     3                  *        DF =  2     2
 3 PERF         6.89     1   0.10696  *
 4 INFO        18.84     1   0.08121  *
 5 VERBEXP      9.74     1   0.19930  *
 6 AGE          0.27     1   0.53278  *

U-STATISTIC(WILKS' LAMBDA) 0.0104766   DEGREES OF FREEDOM    4    2      6
APPROXIMATE F-STATISTIC        6.577   DEGREES OF FREEDOM    8.00      6.00

 F - MATRIX          DEGREES OF FREEDOM =     4     3

            *1        *2
*2          9.70
*3          7.19      4.57

CLASSIFICATION FUNCTIONS

        GROUP =    *1              *2             *3
VARIABLE
 3 PERF            1.92420        0.58704        1.36552
 4 INFO          -17.56221       -8.69921      -10.58700
 5 VERBEXP         5.54585        4.11679        2.97278
 6 AGE             0.98723        5.01749        2.91135

CONSTANT        -138.91107      -72.38436      -72.34032

CLASSIFICATION MATRIX

GROUP      PERCENT    NUMBER OF CASES CLASSIFIED INTO GROUP
           CORRECT
                        *1        *2         *3
 *1        100.0        3         0          0
 *2        100.0        0         3          0
 *3        100.0        0         0          3
TOTAL      100.0        3         3          3
```

JACKKNIFED CLASSIFICATION

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP		
		*1	*2	*3
*1	100.0	3	0	0
*2	100.0	0	3	0
*3	66.7	0	1	2
TOTAL	88.9	3	4	2

EIGENVALUES

13.48590 5.58923

CUMULATIVE PROPORTION OF TOTAL DISPERSION

0.70699 1.00000

MULTIVARIATE TESTS

STATISTICS	VALUE	F APPROXIMATION F	D.F.	D.F.	PVALUE
WILKS' LAMBDA	0.01048	6.57742	8	6.00	0.0169
D.F. 4 2 6					
PILLAI'S TRACE	1.77920	8.05816	8	8	0.0039
HOTELLING-LAWLEY TRACE	19.07512	APPROXIMATION NOT CALCULABLE			
ROY'S MAXIMUM ROOT	13.48590				

TABLES ARE AVAILABLE FOR ROY'S TEST: MORRISON (1976)

CANONICAL CORRELATIONS

0.96487 0.92100

AVERAGE SQUARED CANONICAL CORRELATION 0.88960

VARIABLE	COEFFICIENTS FOR CANONICAL VARIABLES	
3 PERF	0.17100	-0.10069
4 INFO	-1.26953	-0.10325
5 VERBEXP	0.26493	0.35776
6 AGE	-0.52541	0.24690

VARIABLE	STANDARDIZED (BY POOLED WITHINGROUP VARIANCES) COEFFICIENTS FOR CANONICAL VARIABLES	
3 PERF	2.50352	-1.47406
4 INFO	-3.48961	-0.28380
5 VERBEXP	1.32466	1.78881
6 AGE	-0.50273	0.23625
CONSTANT	-9.67374	-3.45293

GROUP	CANONICAL VARIABLES EVALUATED AT GROUP MEANS	
*1	4.10234	0.69097
*2	-2.98068	1.94169
*3	-1.12166	-2.63265

TABLE 11.4 SETUP AND SELECTED SPSS DISCRIMINANT OUTPUT FOR DISCRIMINANT FUNCTION ANALYSIS OF SAMPLE DATA IN TABLE 11.1

```
DATA LIST     FILE='TAPE41.DAT'    FREE
              /SUBJNO GROUP PERF INFO VERBEXP AGE.
DISCRIMINANT  GROUPS = GROUP(1,3)/
              VARIABLES=PERF TO AGE/
              ANALYSIS=PERF TO AGE/
              METHOD = DIRECT/
              STATISTICS = TABLE.
```

Prior probability for each group is .33333

Canonical Discriminant Functions

Fcn	Eigenvalue	Pct of Variance	Cum Pct	Canonical Corr	:	After Fcn	Wilks' Lambda	Chisquare	DF	Sig
					:	0	.0105	20.514	8	.0086
1*	13.4859	70.70	70.70	.9649	:	1	.1518	8.484	3	.0370
2*	5.5892	29.30	100.00	.9210	:					

* marks the 2 canonical discriminant functions remaining in the analysis.

Standardized Canonical Discriminant Function Coefficients

	FUNC 1	FUNC 2
PERF	-2.50352	-1.47406
INFO	3.48961	-.28380
VERBEXP	-1.32466	1.78881
AGE	.50273	.23625

Structure Matrix:

Pooled-withingroups correlations between discriminating variables
and canonical discriminant functions
(Variables ordered by size of correlation within function)

	FUNC 1	FUNC 2
INFO	.22796*	.06642
VERBEXP	-.02233	.44630*
PERF	-.07546	-.17341*
AGE	-.02786	-.14861*

Canonical Discriminant Functions evaluated at Group Means (Group Centroids)

Group	FUNC 1	FUNC 2
1	-4.10234	.69097
2	2.98068	1.94169
3	1.12166	-2.63265

Classification Results -

Actual Group		No. of Cases	Predicted Group Membership 1	2	3
Group	1	3	3	0	0
			100.0%	.0%	.0%
Group	2	3	0	3	0
			.0%	100.0%	.0%
Group	3	3	0	0	3
			.0%	.0%	100.0%

Percent of "grouped" cases correctly classified: 100.00%

Classification Processing Summary
- 9 Cases were processed.
- 0 Cases were excluded for missing or outofrange group codes.
- 0 Cases had at least one missing discriminating variable.
- 9 Cases were used for printed output.

TABLE 11.5 SETUP AND SAS DISCRIM OUTPUT FOR DISCRIMINANT FUNCTION ANALYSIS OF SMALL SAMPLE DATA OF TABLE 11.1

```
DATA SSAMPLE;
INFILE 'TAPE41.DAT';
INPUT SUBJNO GROUP PERF INFO VERBEXP AGE;
PROC DISCRIM;
     CLASS GROUP;
     VAR PERF INFO VERBEXP AGE;
```

Discriminant Analysis

9 Observations	8 DF Total
4 Variables	6 DF Within Classes
3 Classes	2 DF Between Classes

Class Level Information

GROUP	Frequency	Weight	Proportion	Prior Probability
1	3	3.0000	0.333333	0.333333
2	3	3.0000	0.333333	0.333333
3	3	3.0000	0.333333	0.333333

Discriminant Analysis   Pooled Covariance Matrix Information

Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
4	6.50509421

Discriminant Analysis   Pairwise Generalized Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \, COV^{-1} \, (\bar{X}_i - \bar{X}_j)$$

Generalized Squared Distance to GROUP

From GROUP	1	2	3
1	0	51.73351	38.33673
2	51.73351	0	24.38053
3	38.33673	24.38053	0

Discriminant Analysis   Linear Discriminant Function

$$\text{Constant} = -.5\,\bar{X}_j' \, COV^{-1} \, \bar{X}_j \qquad \text{Coefficient Vector} = COV^{-1} \, \bar{X}_j$$

GROUP

	1	2	3
CONSTANT	-137.81247	-71.28575	-71.24170
PERF	1.92420	0.58704	1.36552
INFO	-17.56221	-8.69921	-10.58700
VERBEXP	5.54585	4.11679	2.97278
AGE	0.98723	5.01749	2.91135

TABLE 11.5 *(CONTINUED)*

```
Discriminant Analysis   Classification Summary for Calibration Data: WORK.SSAMPLE

        Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function:  Posterior Probability of Membership in each GROUP:

 2          _         -1  _                               2                      2
D (X) = (X-X )' COV   (X-X )          Pr(j|X) = exp(-.5 D (X)) / SUM exp(-.5 D (X))
 j          j             j                               j       k              k
```

Number of Observations and Percent Classified into GROUP:

From GROUP	1	2	3	Total
1	3 100.00	0 0.00	0 0.00	3 100.00
2	0 0.00	3 100.00	0 0.00	3 100.00
3	0 0.00	0 0.00	3 100.00	3 100.00
Total Percent	3 33.33	3 33.33	3 33.33	9 100.00
Priors	0.3333	0.3333	0.3333	

Error Count Estimates for GROUP:

	1	2	3	Total
Rate	0.0000	0.0000	0.0000	0.0000
Priors	0.3333	0.3333	0.3333	

(Note that the CONSTANT values are for the standardized equations.) Finally, the section labeled CANONICAL VARIABLES EVALUATED AT GROUP MEANS shows the group centroids, average discriminant function score for each group for each function.

SPSS DISCRIMINANT[7] (Table 11.4) also assigns equal prior probability for each group by default. The output summarizing the Canonical Discriminant Functions appears in two parallel tables separated by colons(:). At the left are shown Eigenvalue, Pct of Variance and Cum Pct of variance accounted for by each function, and Canonical Corr for each discriminant function. At the right are the "peel off" significance tests of successive discriminant functions. When After Fcn is 0, (Chisquare $= 20.514$) all discriminant functions are tested together (0 removed). After Fcn 1 is removed, Chisquare is still statistically significant at $\alpha = .05$ because Sig $= .0370$. This means that the second discriminant function is significant as well as the first. If not, the second would not have been marked as one of the discriminated functions remaining in the analysis.

[7] Note that in the PC+ Implementation of SPSS, the keyword is DSCRIMINANT (drop the first I).

TABLE 11.6 SETUP AND SELECTED SYSTAT DISCRIM OUTPUT FOR DISCRIMINANT FUNCTION ANALYSIS OF SMALL SAMPLE DATA IN TABLE 11.1

```
USE TAPE41
OUTPUT SSDISC.OUT
MODEL  GROUP = PERF INFO VERBEXP AGE
PRINT MEDIUM
ESTIMATE
```

Group frequencies

	1	2	3
Frequencies	3	3	3

Group means

PERF	98.667	87.667	101.333
INFO	7.000	11.667	9.667
VERBEXP	36.333	38.333	28.333
AGE	7.300	6.933	7.633

Between groups Fmatrix - df = 4 3

	1	2	3
1	0.0		
2	9.700	0.0	
3	7.188	4.571	0.0

Wilk's lambda

Lambda = 0.0105 df = 4 2 6

Approx. F= 6.5774 df = 8 6 prob = 0.0169

Classification functions

	1	2	3
Constant	-138.911	-72.384	-72.340
PERF	1.924	0.587	1.366
INFO	-17.562	-8.699	-10.587
VERBEXP	5.546	4.117	2.973
AGE	0.987	5.017	2.911

Variable	F-to-remove	Tolerance	Variable	F-to-enter	Tolerance
3 PERF	6.89	0.106964			
4 INFO	18.84	0.081213			
5 VERBEXP	9.74	0.199297			
6 AGE	0.27	0.532781			

Classification matrix (cases in row categories classified into columns)

	1	2	3	%correct
1	3	0	0	100
2	0	3	0	100
3	0	0	3	100
Total	3	3	3	100

Jackknifed classification matrix

	1	2	3	%correct
1	3	0	0	100
2	0	3	0	100
3	0	1	2	67
Total	3	4	2	89

TABLE 11.6 (CONTINUED)

```
         Eigen      Canonical     Cumulative proportion
         values    correlations    of total dispersion
       ---------   ------------   ---------------------
         13.486        0.965              0.707
          5.589        0.921              1.000

      Wilk's lambda=      0.010
          Approx.F=       6.577  DF=    8,       6  p-tail=  0.0169

     Pillai's trace=      1.779
          Approx.F=       8.058  DF=    8,       8  p-tail=  0.0039

Lawley-Hotelling trace=  19.075
          Approx.F=       4.769  DF=    8,       4  p-tail=  0.0742

Canonical discriminant functions
--------------------------------
                        1          2
 Constant          -9.674      3.453
 PERF               0.171      0.101
 INFO              -1.270      0.103
 VERBEXP            0.265     -0.358
 AGE               -0.525     -0.247

Canonical discriminant functions -- standardized by within variances
--------------------------------------------------------------------
                        1          2
 PERF               2.504      1.474
 INFO              -3.490      0.284
 VERBEXP            1.325     -1.789
 AGE               -0.503     -0.236

Canonical scores of group means
-------------------------------
 1             4.102   -.691
 2            -2.981  -1.942
 3            -1.122   2.633
```

Standardized Canonical Discriminant Function Coefficients (Equation 11.2) are given for deriving discriminant function scores from standardized predictors. Correlations (loadings) between predictors and discriminant functions are given in the Structure Matrix. These are ordered so that predictors loading on the first discriminant function are listed first, and those loading on the second discriminant function second. Then Group Centroids are shown, indicating the average discriminant score for each group on each function.

In the Classification Results table, produced by the TABLE instruction in the STATISTICS paragraph (STATISTICS 13 is used for the PC version), rows represent actual group membership and columns represent predicted group membership. Within each cell, the number and percent of cases correctly classified are shown. For this example, all of the diagonal cells show 100% correction classification (100.0%).

SAS DISCRIM output (Table 11.5) begins with a summary of input and degrees of freedom, followed by a summary of GROUPs, their Frequency (number of cases), Weight (sample sizes in this case), Proportion of cases in each group, and Prior Probability (set equal by default). Information is then provided about the Pooled Covariance Matrix, which can signal problems in multicollinearity/singularity if the rank of the matrix is not equal to the number of predictors. An equation is provided by which distances between groups are derived and then presented in the Mahalanobis D^2 matrix labeled Pairwise Generalized Squared Distances Between Groups. For example the greatest distance is between groups 1 and 2 (51.73351) while the smallest distance is between groups 2 and 3 (24.38053). Equations for classification functions and classification coefficients (Equation 11.6) are given in the following matrix, labeled Linear Discriminant Function. Finally, results of classification are presented in the table labeled Number of Observations and Percent Classified into GROUP where, as usual, rows represent actual group and columns represent predicted group. Cell values show number of cases classified and percentage correct. Number of erroneous classifications for each group are presented, and Prior probabilities are repeated at the bottom of this table.

SYSTAT DISCRIM (Table 11.6), after requesting PRINT MEDIUM, begins with the frequencies and the means for each of the predictors. The Between groups Fmatrix then shows the F-ratio for the multivariate test comparing each group with each other group. Wilks' Lambda is then provided, followed by Classification functions for classifying new cases into groups. This is followed by the F-to-remove, showing the consequences of dropping each predictor, by itself, from the prediction equation.

The next two matrices show classification results, with and without jackknifing, followed by Eigenvalues (relative proportion of variance contributed by each predictor), Canonical correlations (multiple correlations between predictors and groups), and Cumulative proportion of total dispersion in the solution accounted for by the functions, as per BMDP7M. Wilks' Lambda is then repeated, along with Pillai's trace and Lawley-Hotelling trace, as alternative multivariate criteria for evaluated significance of prediction of groups by predictor variables. Canonical discriminant functions (raw discriminant function coefficients) are then shown, followed by coefficients standardized by pooled within-groups variances. The Canonical scores of group means are the centroids for the three groups on the two discriminant functions. Plots (not shown) of discriminant functions, with centroids and cases for all groups are then produced by SYSTAT DISCRIM.

11.5 ■ TYPES OF DISCRIMINANT FUNCTION ANALYSIS

The three types of discriminant function analysis—standard (direct), sequential, and statistical (stepwise)—are analogous to the three types of multiple regression discussed in Section 5.5. Criteria for choosing among the three strategies are the same as those discussed in Section 5.5.4 for multiple regression.

11.5.1 ■ Direct Discriminant Function Analysis

In standard (direct) DISCRIM, like standard multiple regression, all predictors enter the equations at once and each predictor is assigned only the unique association it has with groups. Variance shared among predictors contributes to the total relationship, but not to any one predictor.

The overall test of relationship between predictors and groups in direct DISCRIM is the same as the test of main effect in MANOVA where all discriminant functions are combined and DVs are considered simultaneously. Direct DISCRIM is the model demonstrated in Section 11.4.1.

All the computer programs described in Table 11.11 perform direct DISCRIM; the use of some of them for that purpose is shown in Tables 11.3 through 11.6.

11.5.2 ■ Sequential Discriminant Function Analysis

Sequential (or, as some prefer to call it, hierarchical) DISCRIM is used to evaluate contributions to prediction of group membership by predictors as they enter the equations in an order determined by the researcher. The researcher assesses improvement in classification when a new predictor is added to a set of prior predictors. Does classification of cases into groups improve reliably when the new predictor or predictors are added (cf. Section 11.6.6.3)?

If predictors with early entry are viewed as covariates and an added predictor is viewed as a DV, DISCRIM is used for analysis of covariance. Indeed, sequential DISCRIM can be used to perform stepdown analysis following MANOVA (cf. Section 9.5.2.2) because stepdown analysis is a sequence of ANCOVAs.

Sequential DISCRIM also is useful when a reduced set of predictors is desired and there is some basis for establishing a *priority* order among them. If, for example, some predictors are easy or inexpensive to obtain and they are given early entry, a useful, cost-effective set of predictors may be found through the sequential procedure.

Sequential DISCRIM through BMDP7M for the data of Section 11.4 is shown in Table 11.7. Age is given first priority because it is one of the easiest pieces of information to obtain about a child, and, furthermore, may be useful as a covariate. The two WISC scores are given the same priority 2 entry level and they compete with each other in a stepwise fashion for entry (see Section 11.5.3). Third priority is assigned the VERBEXP score because it is the least likely to be available without special testing.

Significance of discrimination is assessed at each step, and classification functions are provided. Adequacy of classification is shown automatically at the last step but can be requested after any step.

Classification information in SPSS DISCRIMINANT is provided only at the last step, however other useful information, unavailable in other programs, is given at each step. Particularly handy for stepdown analysis is the "change in Rao's V" at every step, as discussed in Section 11.6.1.1.[8]

In SPSS, sequential DISCRIM differs from sequential multiple regression in the way predictors with the same priority enter. In DISCRIM, predictors with the same priority compete with one another stepwise, and only one predictor enters at each step. In multiple regression, several IVs can enter in a single step. Further, the test for the significance of improvement in prediction is tedious in the absence of very large samples (Section 11.6.6.3). This makes DISCRIM less flexible than multiple regression. If you have only two groups and sample sizes are approximately equal, you might consider performing DISCRIM through regression where the DV is a dichotomous variable representing group membership, with groups coded 0 and 1. If classification is desired, preliminary multiple regression analysis with fully flexible entry of predictors could be followed by DISCRIM to provide classification.

In BMDP, the less flexible method (where predictors with the same priority compete for entry stepwise and only one enters at a time) is used for sequential DISCRIM. SYSTAT DISCRIM provides

[8] In using DISCRIM for stepdown analysis following MANOVA, change in Rao's *V* can replace stepdown *F* as the criterion for evaluating significance of successive DVs. Rao's *V* statistics are available through an option in SPSS DISCRIMINANT.

TABLE 11.7 SETUP AND SELECTED BMDP7M OUTPUT FOR SEQUENTIAL DISCRIMINANT FUNCTION ANALYSIS OF SAMPLE DATA IN TABLE 11.1

```
/INPUT     VARIABLES ARE 6.  FORMAT IS FREE.  FILE = 'TAPE41.DAT'.
/VARIABLE  NAMES ARE SUBJNO,GROUP,PERF,INFO,VERBEXP,AGE.
             LABEL IS SUBJNO.
/GROUP     VAR IS GROUP.
           CODES(GROUP) ARE 1, 2, 3.
           NAMES(GROUP) ARE MEMORY, PERCEPT, COMMUN.
/DISCRIM   LEVEL=2*0, 2, 2, 3, 1.  FORCE=3.
/END

****************************************************************************

STEP NUMBER    1
VARIABLE ENTERED    6 AGE

  VARIABLE        F TO    FORCE TOLERNCE *   VARIABLE       F TO    FORCE TOLERNCE
                 REMOVE   LEVEL          *                  ENTER   LEVEL
          DF =   2    6                  *          DF =   2    5
   6 AGE           0.40      1   1.00000 *   3 PERF          0.31     2    0.64611
                                         *   4 INFO          2.47     2    0.83723
                                         *   5 VERBEXP       3.49     3    0.88503

U-STATISTIC(WILKS' LAMBDA) 0.8819122   DEGREES OF FREEDOM    1    2      6
APPROXIMATE F-STATISTIC        0.402   DEGREES OF FREEDOM     2.00     6.00

 F - MATRIX          DEGREES OF FREEDOM =     1     6

             MEMORY    PERCEPT
PERCEPT        0.22
COMMUN         0.18       0.80

CLASSIFICATION FUNCTIONS

             GROUP =    MEMORY        PERCEPT        COMMUN
VARIABLE
  6 AGE                 7.97330       7.57282       8.33738

CONSTANT              -30.20117     -27.35104     -32.91961

****************************************************************************

STEP NUMBER    2
VARIABLE ENTERED    4 INFO

  VARIABLE        F TO    FORCE TOLERNCE *   VARIABLE       F TO    FORCE TOLERNCE
                 REMOVE   LEVEL          *                  ENTER   LEVEL
          DF =   2    5                  *          DF =   2    4
   4 INFO          2.47      2   0.83723 *   3 PERF          9.30     2    0.10809
   6 AGE           0.77      1   0.83723 *   5 VERBEXP      13.13     3    0.20140

U-STATISTIC(WILKS' LAMBDA) 0.4435523   DEGREES OF FREEDOM    2    2      6
APPROXIMATE F-STATISTIC        1.254   DEGREES OF FREEDOM     4.00    10.00

 F - MATRIX          DEGREES OF FREEDOM =     2     5

             MEMORY    PERCEPT
PERCEPT        2.65
COMMUN         0.59       1.12

CLASSIFICATION FUNCTIONS

             GROUP =    MEMORY        PERCEPT        COMMUN
VARIABLE
  4 INFO               -0.23089       0.57401       0.12959
  6 AGE                 8.24090       6.90755       8.18718

CONSTANT              -30.36978     -28.39318     -32.97273
```

TABLE 11.7 *(CONTINUED)*

```
**********************************************************************

STEP NUMBER    3
VARIABLE ENTERED    3 PERF

  VARIABLE        F TO    FORCE  TOLERNCE  *   VARIABLE      F TO    FORCE  TOLERNCE
                 REMOVE   LEVEL            *                 ENTER   LEVEL
        DF =   2    4                      *           DF =   2    3
  3 PERF          9.30      2    0.10809   *   5 VERBEXP     9.74      3    0.19930
  4 INFO         18.01      2    0.14007   *
  6 AGE           0.44      1    0.53300   *

U-STATISTIC(WILKS' LAMBDA) 0.0784812    DEGREES OF FREEDOM    3    2      6
APPROXIMATE F-STATISTIC        3.426    DEGREES OF FREEDOM     6.00      8.00

 F - MATRIX          DEGREES OF FREEDOM =    3     4

           MEMORY    PERCEPT
PERCEPT     13.85
COMMUN       1.78      5.95

CLASSIFICATION FUNCTIONS

         GROUP =    MEMORY        PERCEPT        COMMUN
VARIABLE
  3 PERF           1.65998        0.39091        1.22389
  4 INFO          -7.31838       -1.09502       -5.09593
  6 AGE            1.34609        5.28388        3.10372

CONSTANT         -62.29002      -30.16335      -50.32441
**********************************************************************

STEP NUMBER    4
VARIABLE ENTERED    5 VERBEXP

  VARIABLE        F TO    FORCE  TOLERNCE  *   VARIABLE      F TO    FORCE  TOLERNCE
                 REMOVE   LEVEL            *                 ENTER   LEVEL
        DF =   2    3                      *           DF = 2      2
  3 PERF          6.89      2    0.10696   *
  4 INFO         18.84      2    0.08121   *
  5 VERBEXP       9.74      3    0.19930   *
  6 AGE           0.27      1    0.53278   *

U-STATISTIC(WILKS' LAMBDA) 0.0104766    DEGREES OF FREEDOM    4    2      6
APPROXIMATE F-STATISTIC        6.577    DEGREES OF FREEDOM     8.00      6.00

 F - MATRIX          DEGREES OF FREEDOM =    4     3

           MEMORY    PERCEPT
PERCEPT      9.70
COMMUN       7.19      4.57

CLASSIFICATION FUNCTIONS

         GROUP =    MEMORY        PERCEPT        COMMUN
VARIABLE
  3 PERF           1.92420        0.58704        1.36552
  4 INFO         -17.56221       -8.69921      -10.58700
  5 VERBEXP        5.54585        4.11679        2.97278
  6 AGE            0.98723        5.01749        2.91135

CONSTANT        -138.91107      -72.38436      -72.34032
```

TABLE 11.7 *(CONTINUED)*

CLASSIFICATION MATRIX

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP - MEMORY	PERCEPT	COMMUN
MEMORY	100.0	3	0	0
PERCEPT	100.0	0	3	0
COMMUN	100.0	0	0	3
TOTAL	100.0	3	3	3

JACKKNIFED CLASSIFICATION

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP - MEMORY	PERCEPT	COMMUN
MEMORY	100.0	3	0	0
PERCEPT	100.0	0	3	0
COMMUN	66.7	0	1	2
TOTAL	88.9	3	4	2

SUMMARY TABLE

STEP NO.	VARIABLE ENTERED REMOVED	F VALUE TO ENTER REMOVE	NO. OF VARIAB. INCLUDED	U-STATISTIC	APPROXIMATE F-STATISTIC	DEGREES OF FREEDOM	
1	6 AGE	0.402	1	0.8819	0.402	2.0	6.0
2	4 INFO	2.471	2	0.4436	1.254	4.0	10.0
3	3 PERF	9.303	3	0.0785	3.426	6.0	8.0
4	5 VERBEXP	9.737	4	0.0105	6.577	8.0	6.0

interactive entry of variables for sequential analysis. One or more variables are entered at each interactive step. Sequential DISCRIM is accomplished through a series of runs in SAS DISCRIM.

If you have more than two groups or your group sizes are very unequal, sequential logistic regression is the procedure of choice, and, as seen in Chapter 12, most programs classify cases.

11.5.3 Stepwise (Statistical) Discriminant Function Analysis

When the researcher has no reasons for assigning some predictors higher priority than others, statistical criteria can be used to determine order of entry. That is, if a researcher wants a reduced set of predictors but has no preferences among them, stepwise DISCRIM can be used to produce the reduced set. Entry of predictors is determined by user-specified statistical criteria, of which several are available as discussed in Section 11.6.1.2.

Stepwise DISCRIM has the same controversial aspects as stepwise procedures in general (see Section 5.5.3). Order of entry may be dependent on trivial differences in relationships among predictors in the sample that do not reflect population differences. However, this bias is reduced if cross-validation is used (cf. Section 11.8). Costanza and Afifi (1979) recommend a probability to enter criterion more liberal than .05. They suggest a choice in the range of .15 to .20 to ensure entry of important variables.

Application of stepwise analysis to our small sample example through SPSS DISCRIMINANT is illustrated in Table 11.8. One of five statistical criteria for entry of predictors was selected for this example. Notice that AGE is dropped as a predictor by the statistical criterion and the analysis has only three steps. Compared with output in Table 11.7 where AGE is included, F for discrimination has improved ($F = 6.58$ with AGE included, $F = 10.67$ without AGE). That is, AGE is worthless as a predictor in this example; its inclusion only decreases within-groups degrees of freedom while not increasing between-groups sums of squares. If this result replicates, AGE should be dropped from consideration when classifying future cases.

Progression of the stepwise analysis is summarized in the Summary Table section of Table 11.8. The summary table is followed by discriminant function coefficients, the loading matrix, and group centroids. Moreover, a great deal of classification information is available, not shown in this segment of the output but discussed in Section 11.6.6.

Stepwise as well as sequential DISCRIM are produced by BMDP7M. Format of output for stepwise analysis is identical to that of sequential analysis, as shown in Table 11.7. The user can select among four statistical criteria for entry of predictors. In SAS, stepwise discriminant analysis is provided through a separate program—STEPDISC. Three entry methods are available (cf. Section 11.6.1.2), as well as additional statistical criteria for two of them. SYSTAT provides forward and backward stepwise DISCRIM, with statistical criteria for entry and removal.

11.6 ■ SOME IMPORTANT ISSUES

11.6.1 ■ Statistical Inference

Section 11.6.1.1 contains a discussion of criteria for evaluating the overall statistical significance of a set of predictors for predicting group membership. Section 11.6.1.2 summarizes methods for directing the progression of stepwise DISCRIM and statistical criteria for entry of predictors.

11.6.1.1 ■ Criteria for Overall Statistical Significance

Criteria for evaluating overall statistical reliability in DISCRIM are the same as those in MANOVA. The choice between Wilks' Lambda, Roy's gcr, Hotelling's trace, and Pillai's criterion is based on the same considerations as discussed in Section 9.5.1. Different statistics are available in different programs, as noted in Section 11.7.

Two additional statistical criteria, Mahalanobis' D^2 and Rao's V, are especially relevant to stepwise DISCRIM. Mahalanobis' D^2 is based on distance between pairs of group centroids which is then generalizable to distances over multiple pairs of groups. Rao's V is another generalized distance measure that attains its largest value when there is greatest overall separation among groups.

These two criteria are available both to direct the progression of stepwise DISCRIM and to evaluate the reliability of a set of predictors to predict group membership. Like Wilks' Lambda, Mahalanobis' D^2 and Rao's V are based on all discriminant functions rather than one. Note that Lambda, D^2, and V are descriptive statistics; they are not, themselves, inferential statistics, although inferential statistics are applied to them.

11.6.1.2 ■ Stepping Methods

Related to criteria for statistical inference is the choice among methods to direct the progression of entry of predictors in stepwise DISCRIM. Different methods of progression maximize group differences

TABLE 11.8 SETUP AND SELECTED SPSS DISCRIMINANT OUTPUT FOR STEPWISE DISCRIMINANT FUNCTION ANALYSIS OF DATA IN TABLE 11.1

```
DATA LIST          FILE='TAPE41.DAT'    FREE
                    /SUBJNO TYPE PERF INFO VERBEXP AGE.
DISCRIMINANT        GROUPS = TYPE (1,3)/
                    VARIABLES=PERF TO AGE/
                    ANALYSIS=PERF TO AGE/
                    METHOD = WILKS.

Prior probability for each group is  .33333
---------------- Variables not in the analysis after step   0 ----------------

                              Minimum
Variable   Tolerance   Tolerance   F to enter    Wilks' Lambda
PERF       1.0000000   1.0000000    .73458          .80330
INFO       1.0000000   1.0000000   2.1765           .57955
VERBEXP    1.0000000   1.0000000   3.3600           .47170
AGE        1.0000000   1.0000000    .40170          .88191

At step   1, VERBEXP  was included in the analysis.

                              Degrees of Freedom  Signif.   Between Groups
Wilks' Lambda           .47170     1    2       6.0
Equivalent F           3.36000          2       6.0    .1050

---------------- Variables in the analysis after step   1 ----------------

Variable   Tolerance   F to remove   Wilks' Lambda
VERBEXP    1.0000000    3.3600

---------------- Variables not in the analysis after step   1 ----------------

                              Minimum
Variable   Tolerance   Tolerance   F to enter    Wilks' Lambda
PERF        .3709896    .3709896   5.4218           .14886
INFO        .2019444    .2019444   13.134           .07543
AGE         .8850270    .8850270    .70229          .36825

At step   2, INFO     was included in the analysis.

                              Degrees of Freedom  Signif.   Between Groups
Wilks' Lambda           .07543     2    2       6.0
Equivalent F           6.60264          4      10.0    .0072

---------------- Variables in the analysis after step   2 ----------------

Variable   Tolerance   F to remove   Wilks' Lambda
INFO        .2019444    13.134          .47170
VERBEXP     .2019444    16.708          .57955

---------------- Variables not in the analysis after step   2 ----------------

                              Minimum
Variable   Tolerance   Tolerance   F to enter    Wilks' Lambda
PERF        .1676353    .0912506   10.231           .01233
AGE         .8349769    .1905241    .57354          .05862

At step   3, PERF     was included in the analysis.

                              Degrees of Freedom  Signif.   Between Groups
Wilks' Lambda           .01233     3    2       6.0
Equivalent F          10.6724           6       8.0    .0019
```

TABLE 11.8 *(CONTINUED)*

---------------- Variables in the analysis after step 3 ---------------

Variable	Tolerance	F to remove	Wilks' Lambda			
			.07543INFO	.0912506	22.138	.14886
PERF	.1676353	10.231	.09593			
VERBEXP	.1993789	13.555				

------------ Variables not in the analysis after step 3 ----------------

Variable	Tolerance	Minimum Tolerance	F to enter	Wilks' Lambda
AGE	.5327805	.0812130	.26592	.01048

F level or tolerance or VIN insufficient for further computation.

Summary Table

Step	Action Entered	Action Removed	Vars In	Wilks' Lambda	Sig.	Label
1	VERBEXP		1	.47170	.1050	
2	INFO		2	.07543	.0072	
3	PERF		3	.01233	.0019	

Canonical Discriminant Functions

Fcn	Eigenvalue	Pct of Variance	Cum Pct	Canonical Corr		After Fcn	Wilks' Lambda	Chisquare	DF	Sig
					:	0	.0123	21.977	6	.0012
1*	11.7179	68.55	68.55	.9599	:	1	.1569	9.262	2	.0097
2*	5.3751	31.45	100.00	.9182	:					

* marks the 2 canonical discriminant functions remaining in the analysis.

Standardized Canonical Discriminant Function Coefficients

	FUNC 1	FUNC 2
PERF	1.88474	-1.42704
INFO	-3.30213	-.07042
VERBEXP	1.50580	1.64500

Structure Matrix:

Pooled-withingroups correlations between discriminating variables and canonical discriminant functions

(Variables ordered by size of correlation within function)

	FUNC 1	FUNC 2
INFO	-.23964*	.09887
VERBEXP	.05067	.45030*
AGE	.29955	-.31955*
PERF	.07023	-.18655*

Canonical Discriminant Functions evaluated at Group Means (Group Centroids)

Group	FUNC 1	FUNC 2
1	3.89650	.44986
2	-2.52348	2.06052
3	-1.37301	-2.51039

along different statistical criteria, as indicated in the Purpose column of Table 11.9. The variety of stepping methods, and the computer programs in which they are available, are presented in Table 11.9.

Selection of stepping method depends on the availability of programs and choice of statistical criterion. If, for example, the statistical criterion is Wilks' Lambda, it is beneficial to choose the stepping method that minimizes Lambda. (In SPSS DISCRIMINANT Lambda is the least expensive method, and is recommended in the absence of contrary reasons.) Or, if the statistical criterion is "change in Rao's V," the obvious choice of stepping method is RAO.

Statistical criteria also can be used to modify stepping. For example, the user can modify minimum F for a predictor to enter, minimum F to avoid removal, and so on. SAS allows forward, backward and "stepwise" stepping (cf. Section 5.5.3). Either partial R^2 or significance level is chosen for variables or enter (forward stepping) or stay (backward stepping) in the model. Tolerance (the proportion of variance for a potential predictor that is not already accounted for by predictors in the equation) can be modified in BMDP, SAS, and SPSS stepwise programs. Comparison of programs with respect to these stepwise statistical criteria is provided in Table 11.12.

11.6.2 Number of Discriminant Functions

In DISCRIM with more than two groups, a number of discriminant functions are extracted. The maximum number of functions is the lesser of either degrees of freedom for groups or, as in canonical correlation, principal components analysis and factor analysis, equal to the number of predictors. As in these other analyses, some functions often carry no worthwhile information. It is frequently the case that the first one or two discriminant functions account for the lion's share of discriminating power, with no additional information forthcoming from the remaining functions.

Many of the programs evaluate successive discriminant functions. For the SPSS DISCRIMINANT example of Table 11.4, note that eigenvalues, percents of variance, and canonical correlations are given for each discriminant function for the small sample data of Table 11.1. To the right, with both functions included, the $\chi^2(8)$ of 20.514 indicates a highly reliable relationship between groups and predictors. With the first discriminant function removed, there is still a reliable relationship between groups and predictors as indicated by the $\chi^2(3) = 8.484$, $p = .037$. This finding indicates that the second discriminant function is also reliable.

How much between-group variability is accounted for by each discriminant function? The eigenvalues associated with discriminant functions indicate the relative proportion of between-group variability accounted for by each function. In the small sample example of Table 11.4, 70.70% of the between-group variability is accounted for by the first discriminant function and 29.30% by the second.

SPSS DISCRIMINANT offers the most flexibility with regard to number of discriminant functions. The user can choose the number of functions, the critical value for proportion of variance accounted for (with succeeding discriminant functions dropped once that value is exceeded), or the significance level of additional functions. SYSTAT GLM (but not DISCRIM) provides tests of successive functions, as do SAS DISCRIM and CANDISC (but not STEPDISC). Neither of the BMDP programs provide tests of successive functions.

11.6.3 Interpreting Discriminant Functions

If a primary goal of analysis is to discover and interpret the combinations of predictors (the discriminant functions) that separate groups in various ways, then the next two sections are relevant. Section 11.6.3.1 reveals how groups are spaced out along the various discriminant functions. Section 11.6.3.2 discusses correlations between predictors and the discriminant functions.

TABLE 11.9 METHODS FOR DIRECTING STEPWISE DISCRIMINANT FUNCTION ANALYSIS

Label	Purpose	Program and Option
WILKS	Produces smallest value of Wilks' Lambda (therefore largest multivariate *F*)	SPSS:WILKS BMDP:METHOD=1.
MAHAL	Produces largest distance (D^2) for two closest groups	SPSS:MAHAL
MAXMINF	Maximizes the smallest *F* between pairs of groups	SPSS:MAXMINF
MINRESID	Produces smallest average residual variance ($1 - R^2$) between variables and pairs of groups	SPSS:MINRESI
RAO	Produces at each step largest increase in distance between groups as measured by Rao's *V*	SPSS:RAO
F TO ENTER	At each step picks variables with largest F TO ENTER	BMDP:METHOD=2.
Partial R^2	Specifies the partial R^2 (*p*) to enter and/or stay (remove) at each step (forward/backward stepping)	SAS:PR2ENTRY=*p*, PR2STAY=*p*
Significance	Specifies the significance level (*p*) to enter and/or stay (remove) at each step (forward/backward stepping)	SAS:SLENTRY=*p*, SLSTAY=*p* SYSTAT:Enter=*p*, Remove=*p* SPSS:PIN=*p*, POUT=*p*
F-enter/remove limit	Specifies the F-to-enter and/or remove at each step (forward/backward stepping)	SYSTAT:FEnter=*f*, FRemove=*f* BMDP:ENTER=*f*, REMOVE=*f* SPSS:FIN=*f*, FOUT=*f*
Maximum steps	Specifies the maximum number of steps	SAS:MAXSTEP=*n* SPSS:MAXSTEPS=*n* BMDP:STEP=*n*
Number of variables	Specifies the number of variables in the final model	SAS:STOP=*n*
CONTRAST	Stepping procedure based on groups as defined by contrasts	BMDP7M:CONTRAST. SYSTAT:CONTRAST

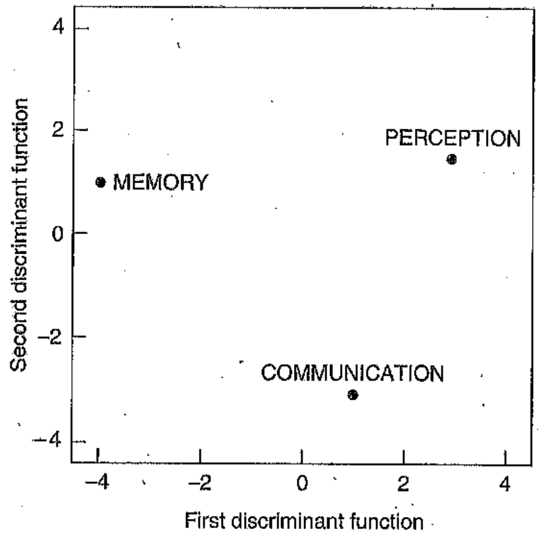


FIGURE 11.1 CENTROIDS OF THREE LEARNING DISABILITY GROUPS ON THE TWO DISCRIMINANT FUNCTIONS DERIVED FROM SAMPLE DATA OF TABLE 11.1.

11.6.3.1 Discriminant Function Plots

Groups are spaced along the various discriminant functions according to their centroids. Recall from Section 11.4.1 that centroids are mean discriminant scores for each group on a function. Discriminant functions form axes and the centroids of the groups are plotted along the axes. If there is a big difference between the centroid of one group and the centroid of another along a discriminant function axis, the discriminant function separates the two groups. If there is not a big distance, the discriminant function does not separate the two groups. Many groups can be plotted along a single axis.

An example of a discriminant function plot is illustrated in Figure 11.1 for the data of Section 11.4. Centroids are obtained from the section called CANONICAL DISCRIMINANT FUNCTIONS EVALUATED AT GROUP MEANS in Table 11.3.

The plot emphasizes the utility of both discriminant functions in separating the three groups. On the first discriminant function (*X* axis), the MEMORY group is some distance from the other two groups, but the COMMUNICATION and PERCEPTION groups are close together. On the second function (*Y* axis) the COMMUNICATION group is far from the MEMORY and PERCEPTION groups. It takes both discriminant functions, then, to separate the three groups from each other.

If there are four or more groups and, therefore, more than two reliable discriminant functions, then pairwise plots of axes are used. One discriminant function is the *X* axis and another is the *Y* axis. Each group has a centroid for each discriminant function; paired centroids are plotted with respect to their values on the *X* and *Y* axes. Because centroids are only plotted pairwise, three significant discriminant functions require three plots (function 1 vs. function 2; function 1 vs. function 3; and function 2 vs. function 3) and so on.

BMDP7M, SYSTAT DISCRIM and SPSS DISCRIMINANT provide a plot of group centroids for the first pair of discriminant functions (called canonical variates in BMDP7M and canonical scores in SYSTAT DISCRIM). Cases as well as means are plotted in SPSS DISCRIMINANT and

SYSTAT DISCRIM, making separations among groups harder to see than with simpler plots, but facilitating evaluation of classification. Simplified plots of group means as well as plots including cases are available in BMDP7M.

Plots of centroids on additional pairs of reliable discriminant functions have to be prepared by hand, or discriminant scores can be passed to a "plotting" program such as BMDP6D. SAS and SYSTAT pass the discriminant scores to plotting programs.

With factorial designs (Section 11.6.5) separate sets of plots are required for each significant main effect and interaction. Main effect plots have the same format as Figure 11.1, with one centroid per group per margin. Interaction plots have as many centroids as cells in the design.

11.6.3.2 ■ Loading Matrices

Plots of centroids tell you how groups are separated by a discriminant function, but they do not reveal the meaning of the discriminant function. The meaning of the function is inferred by a researcher from the pattern of correlations between the function and the predictors.[9] Correlations between predictors and functions are called loadings in both discriminant function analysis and factor analysis (see Chapter 13). If predictors X_1, X_2, and X_3 load (correlate) highly with the function but predictors X_4 and X_5 do not, the researcher attempts to understand what X_1, X_2, and X_3 have in common with each other that is different from X_4 and X_5; the meaning of the function is determined by this understanding. (Read Section 13.6.5 for further insights into the art of interpreting loadings.)

Mathematically, the matrix of loadings is the pooled within-group correlation matrix multiplied by the matrix of standardized discriminant function coefficients.[10]

$$\mathbf{A} = \mathbf{R}_w\mathbf{D} \tag{11.7}$$

The loading matrix of correlations between predictors and discriminant functions, **A**, is found by multiplying the matrix of within-group correlations among predictors, $\mathbf{R}_w$, by a matrix of standardized discriminant function coefficients, **D** (standardized using pooled within-group standard deviations).

For the example of Table 11.1, the loading matrix, called POOLED WITHIN-GROUPS CORRELATIONS BETWEEN DISCRIMINATING VARIABLES AND CANONICAL DISCRIMINANT FUNCTIONS by SPSS DISCRIMINANT, appears as the middle matrix in Table 11.4.

Loading matrices are read in columns; the column is the discriminant function (FUNC 1 and FUNC 2), the rows are predictors (INFO to AGE), and the entries in the column are correlations. For this example, the first discriminant function correlates most highly with INFO (WISC Information scores, $r = .22796$), while the second function correlates most highly with VERBEXP (ITPA Verbal Expression scale, $r = .44630$).

These findings are related to discriminant function plots (e.g., Figure 11.1) for full interpretation. The first discriminant function is largely a measure of INFOrmation, and it separates the group with MEMORY problems from the groups with PERCEPTION and COMMUNICATION problems. The second discriminant function is largely a measure of VERBEXP (verbal expression) and it separates the group with COMMUNICATION problems from the groups with PERCEPTION and MEMORY

[9] Some researchers interpret standardized discriminant function coefficients; however they suffer from the same difficulties in interpretation as standardized regression coefficients, discussed in Section 5.6.1.

[10] Some texts (e.g., Cooley & Lohnes, 1971) and early versions of SPSS use the total correlation matrix to find standardized coefficients rather than the within-group matrix.

problems. Interpretation in this example is reasonably straightforward because only one predictor is highly correlated with each discriminant function; interpretation is much more interesting when several predictors correlate with a discriminant function.

Consensus is lacking regarding how high correlations in a loading matrix must be to be interpreted. By convention, correlations in excess of .33 (10% of variance) may be considered eligible while lower ones are not. Guidelines suggested by Comrey and Lee (1992) are included in Section 13.6.5. However, the size of loadings depends both on the value of the correlation in the population and on the homogeneity of scores in the sample taken from it. If the sample is unusually homogeneous with respect to a predictor, the loadings for the predictor are lower and it may be wise to lower the criterion for determining whether or not to interpret the predictor as part of a discriminant function.

Caution is always necessary in interpreting loadings, however, because they are full, not partial or semipartial, correlations. The loading could be substantially lower if correlations with other predictors were partialed out. For a review of this material, read Section 5.6.1. Section 11.6.4 deals with methods for interpreting predictors after variance associated with other predictors is removed, if that is desired.

In some cases, rotation of the loading matrix may facilitate interpretation, as discussed in Chapter 13. SPSS DISCRIMINANT, SYSTAT GLM, and SPSS MANOVA allow rotation of discriminant functions. But rotation of discriminant function loading matrices is still considered problematic and not recommended for the novice.

11.6.4 ■ Evaluating Predictor Variables

Another tool for evaluating contribution of predictors to separation of groups is available through the CONTRAST procedure of BMDP7M and SYSTAT DISCRIM. Means for predictors for each group are contrasted with means for other groups pooled. For instance, if there are three groups, means on predictors for group 1 are contrasted with pooled means from groups 2 and 3; then means for group 2 are contrasted with pooled means from groups 1 and 3; finally means for group 3 are contrasted with pooled means from groups 1 and 2. This procedure is used to determine which predictors are important for isolating one group from the rest.

One BMDP7M or SYSTAT DISCRIM run is required to isolate the means for each group and contrast them with means for other groups. Because there are only two groups (one group versus pooled other groups), there is only one discriminant function per run. F TO ENTER for each predictor at step 0 is univariate F for testing the reliability of the mean difference between the group singled out and the other groups. Therefore, at step 0, F TO ENTER shows how important a predictor is, by itself, in separating the members of a particular group.

At the last step of each contrast run, with all predictors forced into the analysis, F TO REMOVE reflects the reduction in prediction that would result if a predictor were removed from the equation. It is the unique contribution the predictor makes, in this particular set of predictors, to separation of groups. F TO REMOVE is the importance of a predictor after adjustment for all other predictors in the set.

In order to avoid overinterpretation, it is probably best to consider only predictors with F ratios "significant" after adjusting error for the number of predictors in the set. The adjustment is made on the basis of

$$\alpha = 1 - (1 - \alpha_i)^p \tag{11.8}$$

Type I error rate (α) for evaluating contribution of p predictors to between-group contrasts is based on the error rate for evaluating each predictor and the number of predictors, $i = 1, 2, \ldots, p$.

F TO REMOVE for each predictor at the last step, then, is evaluated at α_i. Predictors that meet this criterion contribute reliable unique variance to separation of one group from the others. Prediction of group membership is reliably reduced if the predictor is deleted from the equation.

Even with this adjustment, there is danger of inflation of Type I error rate because multiple nonorthogonal contrasts are performed. If there are numerous groups, further adjustment might be considered such as multiplication of critical F by $k - 1$, where k = number of groups. Or interpretation can proceed very cautiously, de-emphasizing statistical justification.

As an example of use of F TO ENTER and F TO REMOVE for interpretation, partial output from one BMDP7M CONTRAST run for the data of Table 11.1, between the group with MEMORY problems and the other two groups, is shown in Table 11.10.

With 1 and 3 df at the last step, it is not reasonable to evaluate predictors with respect to statistical significance. However, the pattern of F TO ENTER and F TO REMOVE reveals the predictors most likely to separate each group from the other two. The group with MEMORY deficits is characterized by low scores on WISC INFOrmation, as corroborated by the group means in Table 11.10.

A form of squared semipartial correlation is useful as a measure of strength of association between each predictor and dichotomized group membership. The percent of variance contributed at the last step by reliable predictors is

$$sr_i^2 = \frac{F_i}{\mathrm{df}_{\mathrm{res}}}\,(1 - r_c^2) \tag{11.9}$$

The squared semipartial correlation (sr_i^2) between predictor I and the discriminant function for the difference between two groups (one group vs. all others) is calculated from F TO REMOVE for the ith predictor at the last step with all predictors forced, degrees of freedom for error ($\mathrm{df}_{\mathrm{res}}$) at the last step, and r_c^2 (CANONICAL CORRELATIONS, squared).

For example, strength of association between INFO and the group with MEMORY problems vs. the other two groups (assuming statistical reliability) is

$$sr_i^2 = \frac{33.37}{3}\,(1 - .96357^2) = .7957$$

Almost 80% of the variance in INFO scores overlaps that of the discriminant function that separates the MEMORY group from the PERCEPT and COMMUN groups.

The procedures detailed in this section are most useful when the number of groups is small and the separations among groups are fairly uniform on the discriminant function plot for the first two functions. With numerous groups, some closely clustered, other kinds of contrasts might be suggested by the discriminant function plot (e.g., groups 1 and 2 might be pooled and contrasted with pooled groups 3, 4, and 5). Or, with a very large number of groups, the procedures of Section 11.6.3 may suffice.

If there is logical basis for assigning priorities to predictors, a sequential rather than standard approach to contrasts can be used. Instead of evaluating each predictor after adjustment for all other predictors, it is evaluated after adjustment by only higher priority predictors. This strategy is accomplished through a series of SPSS MANOVA runs, in which Roy-Bargmann stepdown F's (cf. Chapter 9) are evaluated for each contrast.

All the procedures for evaluation of DVs in MANOVA apply to evaluation of predictor variables in DISCRIM. Interpretation of stepdown analysis, univariate F, pooled within-group correlations among predictors, or standardized discriminant function coefficients is as appropriate (or inappropriate) for DISCRIM as for MANOVA. These procedures are summarized in Section 9.5.2.

TABLE 11.10 SETUP AND PARTIAL OUTPUT FOR BMDP7M CONTRAST RUN BETWEEN MEMORY AND OTHER TWO GROUPS FOR SMALL SAMPLE EXAMPLE

```
/INPUT     VARIABLES ARE 6.  FORMAT IS FREE.  FILE = 'TAPE41.DAT'.
/VARIABLE NAMES ARE SUBJNO,GROUP,PERF,INFO,VERBEXP,AGE.
            LABEL IS SUBJNO.
/GROUP      VAR IS GROUP.
            CODES(GROUP) ARE 1 TO 3.
            NAMES(GROUP) ARE MEMORY, PERCEPT, COMMUN.
/DISCRIM    LEVEL=2*0, 4*1.  FORCE=1.
            CONTRAST = 2, -1, -1.
/PRINT      NO STEP.
/PLOT               NO CANON.   CONTRAST.
/END


    MEANS
       GROUP =    MEMORY        PERCEPT       COMMUN         ALL GPS.
VARIABLE
  3 PERF         98.66666      87.66666     101.33334      95.88889
  4 INFO          7.00000      11.66667       9.66667       9.44444
  5 VERBEXP      36.33333      38.33333      28.33333      34.33333
  6 AGE           7.30000       6.93333       7.63333       7.28889
COUNTS                 3.            3.            3.            9.

*************************************************************************

STEP NUMBER    0
  VARIABLE     F TO    FORCE TOLERNCE *   VARIABLE     F TO    FORCE TOLERNCE
               REMOVE  LEVEL          *                ENTER   LEVEL
       DF =    1    7                 *        DF =    1    6
                                      *   3 PERF          0.16    1   1.00000
                                      *   4 INFO          3.56    1   1.00000
                                      *   5 VERBEXP       0.72    1   1.00000
                                      *   6 AGE           0.00    1   1.00000
*************************************************************************

STEP NUMBER    4
VARIABLE ENTERED    6 AGE
  VARIABLE     F TO    FORCE TOLERNCE *   VARIABLE     F TO    FORCE TOLERNCE
               REMOVE  LEVEL          *                ENTER   LEVEL
       DF =    1    3                 *        DF =    1    2
  3 PERF          2.90    1   0.10696 *
  4 INFO         33.37    1   0.08121 *
  5 VERBEXP       2.72    1   0.19930 *
  6 AGE           0.34    1   0.53278 *

CANONICAL CORRELATIONS
                 0.96357
```

11.6.5 Design Complexity: Factorial Designs

The notion of placing cases into groups is easily extended to situations where groups are formed by differences on more than one dimension. An illustration of factorial arrangement of groups is the large sample example of Section 9.7, where women are classified by femininity (high or low) and also by masculinity (high or low) on the basis of scores on the Bem Sex Role Inventory (BSRI). Dimensions of femininity and masculinity (each with two levels) are factorially combined to form four groups: high-high, high-low, low-high, low-low. Unless you want to classify cases, factorial designs are best analyzed through MANOVA. If classification is your goal, however, some issues require attention.

As long as sample sizes are equal in all cells, factorial designs are fairly easily analyzed through a factorial DISCRIM program such as BMDP7M or SYSTAT DISCRIM. A separate run is required for each main effect and interaction wherein the main effect or interaction is specified by contrast coding. Discriminant and classification functions are produced for the effect during the run.

When sample sizes are unequal in cells, a two-stage analysis is often best. First, questions about reliability of separation of groups by predictors are answered through MANOVA. Second, if classification is desired after MANOVA, it is found through DISCRIM programs.

Formation of groups for DISCRIM depends on the outcome of MANOVA. If the interaction is statistically significant, groups are formed for the cells of the design. That is, in a two-by-two design, four groups are formed and used as the grouping variable in DISCRIM. Note that main effects as well as interactions influence group means (cell means) in this procedure, but for most purposes classification of cases into cells seems reasonable.

If an interaction(s) is not statistically reliable, classification is based on significant main effects. For example, interaction is not reliable, in the data of Section 9.7, but both main effect of masculinity and main effect of femininity are reliable. One DISCRIM run is used to produce the classification equations for main effect of masculinity and a second run is used to produce the classification equations for main effect of femininity. That is, classification of main effects is based on marginal groups.

SYSTAT GLM makes no distinction between MANOVA and DISCRIM, therefore factorial DISCRIM is produced by the instruction PRINT=LONG with the setup for a factorial design (cf. Table 9.5). In SYSTAT GLM each main effect and interaction is tested separately and classification information is written to a file. Therefore a two-step procedure is still useful, in which the first analysis determines whether classification results are to be saved for the interaction or main effect(s) steps.

11.6.6 ■ Use of Classification Procedures

The basic technique for classifying cases into groups is outlined in Section 11.4.2. Results of classification are presented in tables such as the CLASSIFICATION MATRIX of BMDP7M (Table 11.3) and SYSTAT DISCRIM (Table 11.6), the CLASSIFICATION RESULTS of SPSS (Table 11.4), or NUMBER OF OBSERVATIONS AND PERCENTS CLASSIFIED INTO GROUP of SAS (Table 11.5) where actual group membership is compared to predicted group membership. From these tables, one finds the percent of cases correctly classified and the number and nature of errors of classification.

But how good is the classification? When there are equal numbers of cases in every group, it is easy to determine the percent of cases that should be correctly classified by chance alone to compare to the percent correctly classified by the classification procedure. If there are two equally sized groups, 50% of the cases should be correctly classified by chance alone (cases are randomly assigned into two groups and half of the assignments in each group are correct), while three equally sized groups should produce 33% correct classification by chance, and so forth. However, when there are unequal numbers of cases in the groups, computation of the percent of cases that should be correctly classified by chance alone is a bit more complicated.

The easier way to find it[11] is to first compute the number of cases in each group that should be correct by chance alone and then add across the groups to find the overall expected percent correct. Consider an example where there are 60 cases, 10 in Group 1, 20 in Group 2, and 30 in Group 3. If prior probabilities are specified as .17, .33, and .50, respectively, the programs will assign 10, 20, and 30 cases to the groups. If 10 cases are assigned at random to Group 1, .17 of them (or 1.7) should be correct by chance alone. If 20 cases are randomly assigned to Group 2, .33 (or 6.6) of them should be correct by chance alone, and if 30 cases are assigned to Group 3, .50 of them (or 15) should be correct by chance alone. Adding together 1.7, 6.6, and 15 gives 23.3 cases correct by chance alone, 39% of the total. The percent correct using classification equations has to be substantially larger than the percent expected correct by chance alone if the equations are to be useful.

Some of the computer programs offer sophisticated additional features that are helpful in many classification situations.

11.6.6.1 Cross-Validation and New Cases

Classification is based on classification coefficients derived from samples and they usually work too well for the sample from which they were derived. Because the coefficients are only estimates of population classification coefficients, it is often most desirable to know how well the coefficients generalize to a new sample of cases. Testing the utility of coefficients on a new sample is called cross-validation. One form of cross-validation involves dividing a single large sample randomly in two parts, deriving classification functions on one part, and testing them on the other. A second form of cross-validation involves deriving classification functions from a sample measured at one time, and testing them on a sample measured at a later time. In either case, cross-validation techniques are especially well developed in the BMDP7M and 5M, SAS DISCRIM, SYSTAT DISCRIM, and SPSS DISCRIMINANT programs.

For a large sample randomly divided into parts, you simply omit information about actual group membership for some cases (hide it in the program) as shown in Section 11.8.2. SPSS DISCRIMINANT, SYSTAT DISCRIM, BMDP7M and BMDP5M do not include these cases in the derivation of classification functions, but do include them in the classification phase. In SAS DISCRIM, the withheld cases are put in a separate data file. The accuracy with which the classification functions predict group membership for cases in this data file is then examined. The SAS manual (SAS Institute, 1990, p. 698) provides details for classifying cases in the data file by including "calibration" information for their classification in yet another file.

When the new cases are measured at a later time, classifying them is somewhat more complicated unless you use SAS DISCRIM (in the same way that you would for cross-validation). This is because none of the other computer programs for DISCRIM allows classification of new cases without repeated entry of the original cases to derive the classification functions. You "hide" the new cases, derive the classification functions from the old cases, and test classification on all cases. Or, you can input the classification coefficients along with raw data for the new cases and run the data only through the classification phase. Or, it may be easiest to write your own program based on the classification coefficients to classify cases as shown in Section 11.4.2.

[11] The harder way to find it is to expand the multinomial distribution, a procedure that is more technically correct but produces identical results to those of the simpler method presented here.

11.6.6.2 ■ Jackknifed Classification

Bias enters classification if the coefficients used to assign a case to a group are derived, in part, from the case. In jackknifed classification, the data from the case are left out when the coefficients used to assign it to a group are computed. Each case has a set of coefficients that are developed from all other cases. Jackknifed classification gives a more realistic estimate of the ability of predictors to separate groups.

BMDP7M and SYSTAT DISCRIM provide for jackknifed classification. When the procedure is used with all predictors forced into the equation, bias in classification is eliminated. When it is used with stepwise entry of predictors (where they may not all enter), bias is reduced. An application of jackknifed classification is shown in Section 11.8.

11.6.6.3 ■ Evaluating Improvement in Classification

In sequential DISCRIM, it is useful to determine if classification improves as a new set of predictors is added to the analysis. McNemar's repeated-measures chi square provides a simple, straightforward (but tedious) test of improvement. Cases are tabulated one by one, by hand, as to whether they are correctly or incorrectly classified before the step and after the step where the predictors are added.

		Early step classification	
		Correct	Incorrect
Later step	Correct	(*A*)	*B*
classification	Incorrect	*C*	(*D*)

Cases that have the same result at both steps (either correctly classified—cell *A*—or incorrectly classified—cell *D*) are ignored because they do not change. Therefore, χ^2 for change is

$$\chi^2 = \frac{(B - C)^2}{B + C} \qquad \text{df} = 1 \tag{11.10}$$

Ordinarily, the researcher is only interested in improvement in χ^2, that is, in situations where $B > C$ because more cases are correctly classified after addition of predictors. When $B > C$ and χ^2 is greater than 3.84 (critical value of χ^2 with 1 df at $\alpha = .05$), the added predictors reliably improve classification.

With very large samples hand tabulation of cases is not reasonable. An alternative, but possibly less desirable, procedure is to test the significance of the difference between two lambdas, as suggested by Frane (personal communication). Wilks' Lambda from the step with the larger number of predictors (Λ_2) is divided by Lambda from the step with fewer predictors (Λ_1) to produce (Λ_D).

$$\Lambda_D = \frac{\Lambda_2}{\Lambda_1} \tag{11.11}$$

Wilks' Lambda for testing the significance of the difference between two Lambdas (Λ_D) is calculated by dividing the smaller Lambda (Λ_2) by the larger Lambda (Λ_1).

Λ_D is evaluated with three degree of freedom parameters: p, the number of predictors after addition of predictors; $\text{df}_{\text{effect}}$, the number of groups minus 1; and the df_{error} at the step with the added predictors. Approximate F is found according to procedures in Section 11.4.1.

For the sequential example of Table 11.7, one can test whether addition of INFO, PERF, and VERBEXP at the last step ($\Lambda_2 = .0105$) reliably improves classification of cases over that achieved at step 1 with AGE in the equation ($\Lambda_1 = .8819$).

$$\Lambda_D = \frac{.0105}{.8819} = .0119$$

where

$$\text{df}_p = p = 4$$

$$\text{df}_{\text{effect}} = k - 1 = 2$$

$$\text{df}_{\text{error}} = N - 1 = 6$$

From Section 11.4.1,

$$s = \sqrt{\frac{(4)^2(2)^2 - 4}{(4)^2 + (2)^2 - 5}} = 2$$

$$y = .0119^{\frac{1}{2}} = .1091$$

$$\text{df}_1 = (4)(2) = 8$$

$$\text{df}_2 = (2)\left[6 - \frac{4 - 2 + 1}{2}\right] - \left[\frac{4(2) - 2}{2}\right] = 6$$

$$\text{Approximate } F(8, 6) = \left(\frac{1 - .1091}{.1091}\right)\left(\frac{6}{8}\right) = 6.124$$

Because critical $F(8, 6)$ is 4.15 at $\alpha = .05$, there is reliable improvement in classification into the three groups when INFO, PERF and VERBEXP scores are added to AGE scores.

11.7 ■ COMPARISON OF PROGRAMS

There are numerous programs for discriminant function analysis in the four packages, some general and some special purpose. Both SPSS and BMDP have a general purpose discriminant function analysis program that performs direct, sequential, or stepwise DISCRIM with classification. In addition, SPSS MANOVA performs DISCRIM, but not classification. BMDP also has a program for quadratic discriminant function analysis. SYSTAT GLM provides discriminant function analysis as part of the general linear model program, but now SYSTAT also has a specialized DISCRIM program, similar to BMDP7M. SAS has three programs, one for inference and discriminant function coefficients, one which adds classification, and a third for stepwise analysis. Finally, if the only question is reliability of predictors to separate groups, any of the MANOVA programs discussed in Chapter 9 is appropriate. Table 11.11 compares features of direct discriminant function programs. Features for stepwise discriminant function are compared in Table 11.12.

TABLE 11.11 COMPARISON OF PROGRAMS FOR DIRECT DISCRIMINANT FUNCTION ANALYSIS

Feature	SAS DISCRIM	SAS CANDISC	SPSS DISCRIMINANT	SPSS MANOVA[a]	SYSTAT DISCRIM	SYSTAT GLM	BMDP7M	BMDP5M
Input								
Optional matrix input	Yes	Yes	Yes	Yes	No	Yes	No	No
Missing data options	No	No	Yes	No	No	No	No	No
Restrict number of discriminant functions	NCAN	NCAN	Yes	No	No	No	No	No
Specify cumulative % of sum of eigenvalues	No	No	Yes	No	No	No	No	No
Specify significance level of functions to retain	No	No	Yes	ALPHA	No	No	No	No
Factorial arrangement of groups	No	No	No	Yes	CONTRASTS	Yes	CONTRASTS	No
Specify tolerance	SINGULAR	SINGULAR	Yes	No	Yes	Yes	Yes	No
Rotation of discriminant functions	No	No	Yes	Yes	No	Yes	No	No
Quadratic discriminant function analysis	POOL=NO	No	No	No	Yes	No	No	Yes
Optional prior probabilities	Yes	N.A.[b]	Yes	N.A.[b]	No	Yes	Yes	Yes
Specify separate covariance matrices for classification	POOL=NO	N.A.	Yes	N.A.	No	No	No	No
Threshold for classification	Yes	N.A.	No	N.A.	No	No	N.A.	No
Nonparametric classification method	Yes	N.A.	No	N.A.	No	No	No	No
Output								
Wilks' Lambda with approx. F	Yes	Yes	Yes	Yes	PRINT MEDIUM	Yes	Yes	Yes
χ^2	No	No	Yes	No	No	Yes	No	No
Generalized distance between groups (Mahalanobis D^2)	Yes	Yes	Yes	No	No	No	No	Yes
Hotelling's trace criterion	Yes	Yes	No	Yes	PRINT MEDIUM	Yes	Yes	No
Roy's gcr (maximum root)	Yes	Yes	No	Yes	No	THETA	Yes	No
Pillai's criterion	Yes	Yes	No	Yes	PRINT MEDIUM	Yes	Yes	No
Tests of successive dimensions (roots)	Yes	Yes	Yes	Yes	No	RESIDUAL ROOTS	No	No
Univariate F ratios	Yes	Yes	Yes	Yes	No	Yes	Yes[c]	No
Group means	Yes	Yes	Yes	Yes	PRINT MEDIUM	Yes	Yes	Yes

TABLE 11.11 *(CONTINUED)*

Feature	SAS DISCRIM	SAS CANDISC	SPSS DISCRIMINANT	SPSS MANOVA[a]	SYSTAT DISCRIM	SYSTAT GLM	BMDP7M	BMDP5M
Total and within-group standardized group means	Yes	Yes	No	No	No	No	No	No
Group standard deviations	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Total, within-group and between-group standard deviations	Yes	Yes	No	No	No	No	No	No
Coefficient of variation	No	No	No	No	No	No	Yes	Yes
Standardized discriminant function (canonical) coefficients	Yes	Yes	Yes	Yes	PRINT MEDIUM	Yes	Yes	No
Unstandardized (raw) discriminant function (canonical) coefficients	Yes	Yes	Yes	Yes	PRINT MEDIUM	No	Yes	No
Group centroids	Yes	Yes	Yes	No	Yes	No	Yes	No
Pooled within-groups (residual) SSCP matrix	Yes	Yes	No	Yes	No	Yes	No	No
Between-groups SSCP matrix	Yes	Yes	No	No	No	No	No	No
Hypothesis SSCP matrix	No	No	No	Yes	No	Yes	No	No
Total SSCP matrix	Yes	Yes	No	No	No	Yes	No	No
Group SSCP matrices	Yes	Yes	No	No	No	No	No	No
Pooled within-groups (residual) correlation matrix	Yes	Yes	Yes	Yes	PRINT LONG	Yes	Yes	Yes
Determinant of within-group correlation matrix	No	No	No	Yes	No	No	No	No.
Between-groups correlation matrix	Yes	Yes	No	No	No	No	No	Yes
Group correlation matrices	Yes	Yes	No	No	PRINT LONG	No	No	Yes
Total correlation matrix	Yes	Yes	No	No	PRINT LONG	No	No	Yes
Total covariance matrix	Yes	Yes	Yes	No	PRINT LONG	No	No	Yes
Pooled within-groups (residual) covariance matrix	Yes	Yes	Yes	Yes	PRINT LONG	Yes	Yes	Yes
Group covariance matrices	Yes	Yes	Yes	No	No	No	No	Yes
Between-group covariance matrix	Yes	Yes	No	No	No	No	No	Yes

TABLE 11.11 *(CONTINUED)*

Feature	SAS DISCRIM	SAS CANDISC	SPSS DISCRIMINANT	SPSS MANOVA[a]	SYSTAT DISCRIM	SYSTAT GLM	BMDP7M	BMDP5M
Determinants of group covariance matrices	Yes	No	No	Yes	No	No	No	No
Homogeneity of variance-covariance matrices	Yes	No	Yes	Yes	Yes	No	Yes[d]	Yes
F matrix, pairwise group comparison	No	No	Yes	No[e]	Yes	No	Yes	Yes
Canonical correlations	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Average squared canonical correlation	No	No	No	No	No	No	Yes	No
Adjusted canonical correlations	Yes	Yes	No	No	No	No	No	No
Standard errors of canonical correlations	Yes	Yes	No	No	No	No	No	No
Eigenvalues	Yes	Yes	Yes	Yes	Yes	No	Yes	No
SMCs for each variable	R-Squared	R-Squared	No	No	No	Yes	No	Yes[f]
SMC divided by tolerance for each variable	RSQ/ (1−RSQ)	RSQ/ (1−RSQ)	No	No	No	No	No	No
Loading (structure) matrix (pooled within-groups)	Yes	Yes	Yes	No	No	Yes	No	No
Total structure matrix	Yes	Yes	No	No	No	No	No	No
Between structure matrix	Yes	Yes	No	No	No	No	No	No
Individual discriminant (canonical variate) scores	No	No	Yes	No	Yes	Data file	Yes	No
Classification features								
Classification of cases	Yes	N.A[b]	Yes	N.A[b]	Yes	No[g]	Yes	Yes
Classification function coefficients	Yes[h]	N.A.	Yes	N.A.	PRINT MEDIUM	No	Yes	Yes

TABLE 11.11 *(CONTINUED)*

Feature	SAS DISCRIM	SAS CANDISC	SPSS DISCRIMINANT	SPSS MANOVA[a]	SYSTAT DISCRIM	SYSTAT GLM	BMDP7M	BMDP5M
Classification matrix	Yes	N.A.	Yes	N.A.	Yes	No	Yes	Yes
Posterior probabilities for classification	Data file	N.A.	Yes	N.A.	PRINT LONG	Data file	Yes	Yes
Mahalanobis' D^2 for cases (outliers)	No	N.A.	No	N.A.	PRINT LONG	Data file[i]	Yes	No
Jackknifed classification matrix	Cross-validate	N.A.	No	N.A.	Yes	No	Yes	No
Classification with a cross-validation sample	Yes	N.A.	Yes	N.A.	No	Yes	Yes	Yes
Plots								
Plot of group centroids alone	No	N.A.	No	N.A.	No	No	Yes	No
All groups scatterplot	No	N.A.	Yes	N.A.	Yes	No	Yes	No
Separate scatterplots by group	No	N.A.	Yes	N.A.	No	No	Yes	No
Territorial map	No	N.A.	Yes	N.A.	No	No	No	No

[a] Additional features reviewed in Section 8.9

[b] SAS CANDISC and SPSS MANOVA do not classify cases

[c] STEP NUMBER 0, F TO ENTER

[d] Group plots of scores on first two discriminant functions

[e] Can be obtained through CONTRAST procedure

[f] Separately for each group

[g] Classification requires SYSTAT TABLES or SYSTAT DISCRIM

[h] Labeled Linear Discriminant Function

TABLE 11.12 COMPARISON OF PROGRAMS FOR STEPWISE AND SEQUENTIAL DISCRIMINANT FUNCTION ANALYSIS

Feature	SPSS DISCRIMINANT	BMDP7M	SAS STEPDISC	SYSTAT DISCRIM
Input				
Optional matrix input	Yes	No	Yes	Yes
Missing data options	Yes	No	No	No
Specify contrast	No	Yes	No	Yes
Factorial arrangement of groups	No	CONTRASTS	No	CONTRAST
Suppress intermediate steps	No	NO STEP	No	No
Suppress all but summary table	NOSTEP	No	SHORT	No
Optional methods for order of entry/removal	3	3	5	2
Forced entry by level (sequential)	Yes	Yes	No	Yes
Force some variables into model	Yes	Yes	INCLUDE	FORCE
Specify tolerance	Yes	Yes	SINGULAR	Yes
Specify maximum number of steps	Yes	Yes	Yes	No
Specify number of variables in final stepwise model	No	No	STOP=	No
Specify F to enter/remove	FIN/FOUT	ENTER/ REMOVE	No	FEnter/ FRemove
Specify significance of F to enter/remove	PIN/POUT	No	SLE/SLS	Enter/Remove
Specify partial R^2 to enter/remove	No	No	PR2E/PR2S	No
Restrict number of discriminant functions	Yes	No	No	No
Specify cumulative % of sum of eigenvalues	Yes	No	No	No
Specify significance level of functions to retain	Yes	No	No	No
Rotation of discriminant functions	Yes	No	No	No
Prior probabilities optional	Yes	Yes	N.A.[a]	No
Specify separate covariance matrices for classification	Yes	No	N.A.	No
Output				
Wilks' Lambda with approximate F	Yes	Yes	Yes	PRINT MEDIUM
χ^2	Yes	No	No	No
Mahalanobis' D^2 (between groups)	Yes	No	No	No
Rao's V	Yes	No	No	No
Pillai's criterion	No	Yes	Yes	PRINT MEDIUM
Tests of successive dimensions (roots)	Yes	No	No	No
Univariate F ratios	Yes	Yes[b]	STEP 1 F	Yes[b]
Group means	Yes	Yes	Yes	PRINT MEDIUM
Within-group and total standardized group means	No	No	Yes	No
Group standard deviations	Yes	Yes	Yes	No
Total and pooled within-group standard deviations	No	No	Yes	No
Coefficients of variation	No	Yes	No	No
Standardized discriminant function (canonical) coefficients	Yes	Yes	No	PRINT MEDIUM
Unstandardized discriminant function (canonical) coefficients	Yes	Yes	No	PRINT MEDIUM
Group centroids	Yes	Yes	No	Yes[c]

TABLE 11.12 *(CONTINUED)*

Feature	SPSS DISCRIMINANT	BMDP7M	SAS STEPDISC	SYSTAT DISCRIM
Pooled within-group correlation matrix	Yes	Yes	Yes	PRINT LONG
Total correlation matrix	No	No	Yes	PRINT LONG
Total covariance matrix	Yes	No	Yes	PRINT LONG
Total SSCP matrix	No	No	Yes	No
Pooled within-group covariance matrix	Yes	Yes	Yes	PRINT LONG
Pooled within-group SSCP matrix	No	No	Yes	No
Group covariance matrices	Yes	No	Yes	PRINT LONG
Group correlation matrices	No	No	Yes	PRINT LONG
Group SSCP matrices	No	No	Yes	No
Between-group correlation matrix	No	No	Yes	No
Between-group covariance matrix	No	No	Yes	No
Between-group SSCP matrix	No	No	Yes	No
Homogeneity of variance-covariance matrices	Yes	Yes[d]	No	Yes
F matrix, pairwise group comparison	Yes	Yes	No	Yes
Canonical correlations, each discriminant function	Yes	Yes	No	Yes
Canonical correlations, average	No	Yes[e]	Yes	No
Eigenvalues	Yes	Yes	No	Yes
Loading (structure) matrix	Yes	No	No	No
Partial R^2 (or tolerance) to enter/remove, each step	Yes	Yes	Yes	Yes
F to enter/remove, each step	Yes	Yes	Yes	Yes
Classification features				
Classification of cases	Yes	Yes	N.A.[a]	Yes
Classification function coefficients	Yes	Yes	N.A.	PRINT MEDIUM
Classification matrix	Yes	Yes	N.A.	Yes
Individual discriminant (canonical variate) scores	Yes	Yes	N.A.	PRINT LONG
Posterior probabilities for classification	Yes	Yes	N.A.	PRINT LONG
Mahalanobis' D^2 for cases (outliers)	No	Yes	N.A.	PRINT LONG
Jackknifed classification matrix	No	Yes	N.A.	Yes
Classification with a cross-validation sample	Yes	Yes	N.A.	Yes
Classification information at each step	No	Yes	N.A.	No
Plots				
Plot of group centroids alone	No	Yes	N.A.	No
All groups scatterplot	Yes	Yes	N.A.	Yes
Separate scatterplots by group	Yes	Yes	N.A.	No
Territorial map	Yes	No	N.A.	No

[a] SAS STEPDISC does not classify cases (see SAS DISCRIM, Table 11.11)
[b] STEP 0, F TO ENTER
[c] Canonical scores of group means
[d] Group plots of scores on first two discriminant functions
[e] Squared

11.7.1 ■ SPSS Package

SPSS DISCRIMINANT[12], features of which are described in both Tables 11.11 and 11.12, is the basic program in this package for DISCRIM. The program provides direct (standard), sequential, or stepwise entry of predictors with numerous options. Strong points include several types of plots and plenty of information about classification. Territorial maps are handy for classification using discriminant function scores if there are only a few cases to classify. In addition, a test of homogeneity of variance-covariance matrices is provided through plots and, should heterogeneity be found, classification may be based on separate matrices. Other useful features are evaluation of successive discriminant functions and availability of loading matrices.

SPSS MANOVA can also be used for DISCRIM and has some features unobtainable in any of the other DISCRIM programs. SPSS MANOVA is described rather fully in Table 9.11, but some aspects especially pertinent to DISCRIM are featured in Table 11.11. MANOVA offers a variety of statistical criteria for testing the significance of the set of predictors (cf. Section 11.6.1) and routinely prints loading matrices. Many other matrices can be printed out, and these, along with determinants, are useful for the more sophisticated researcher. Successive discriminant functions (roots) are evaluated, as in SPSS DISCRIMINANT.

SPSS MANOVA provides discriminant functions for more complex designs such as factorial arrangements with unequal sample sizes. The program is limited, however, in that it includes no classification phase. Further, only standard DISCRIM is available, with no provision for stepwise or sequential analysis other than Roy-Bargmann stepdown analysis as described in Chapter 9.

11.7.2 ■ BMDP Series

BMDP7M deals well with all varieties of DISCRIM. Although designed as a stepwise program, direct DISCRIM is easily produced by forcing entry of all predictors and suppressing all steps except the last. Several inferential multivariate tests are available, along with automatic printing of pairwise multivariate F ratios between groups.

BMDP7M's main advantage over most other DISCRIM programs is identification of outliers through the case classification procedure. Further, the JACKKNIFED CLASSIFICATION feature in BMDP7M assigns a case to a group without using the case in developing the classification coefficients for it. This eliminates or reduces bias in classification, as discussed in Section 11.6.6.2. However, the program lacks a dimension reduction analysis for successive roots and does not provide a loading matrix.

The BMDP7M provision of stepping by contrasts allows analysis of factorial designs as long as sample sizes are equal in cells (Section 11.6.5). Because this program allows for classification as well as statistical inference, it is the program of choice for equal-n factorial designs. (Unequal-n factorial designs require contrast coefficients that reflect choice of adjustment procedure and some considerable sophistication on the part of the user.)

BMDP5M is designed to do quadratic discriminant function analysis and emphasizes classification and various matrices as a result. The program offers a test for homogeneity of variance-covariance matrices, and can be used for linear (but not stepwise) discriminant function analysis as well. Cross-validation is available, but not jackknifed classification. None of the standard canonical analyses or coefficients are available, nor is there dimension reduction analysis or a loading matrix.

[12] DSCRIMINANT in the PC+ version.

11.7.3 ■ SYSTAT System

SYSTAT DISCRIM now is the major discriminant function analysis program, replacing MGLH. The new program deals with all varieties of DISCRIM. Automatic (forward and backward) and interactive stepping are available, as well as a contrast procedure to control entry of variables. Jackknifed classification is produced by default, and the manual shows how to perform cross-validation. Dimension reduction analysis is no longer available, but can be obtained by rephrasing the problem as MANOVA and running it through GLM. Such a strategy also is well suited to factorial arrangements of unequal-n groups. The SYSTAT manual (Wilkinson & Hill, 1994a, pp. 370–373) demonstrates an example of discriminant function analysis through GLM.

The manual shows how to use scatterplot matrices (SYSTAT SPLOM) to evaluate homogeneity of variance-covariance matrices; quadratic discrimination function analysis is available through DISCRIM should the assumption be violated. Several univariate and multivariate inferential tests also are available. SYSTAT DISCRIM can be used to assess outliers through Mahalanobis distance of each case to each group centroid.

11.7.4 ■ SAS System

In SAS, there are three separate programs to deal with different aspects of discriminant analysis, with surprisingly little overlap between the stepwise and direct programs. All of the SAS programs for discriminant function analysis are especially rich in output of SSCP, correlation, and covariance matrices.

The most comprehensive program is DISCRIM. This program now does just about everything that the CANDISC program does, although earlier versions lacked inferential tests. It does not perform stepwise or sequential analysis. This program is especially handy for classifying new cases or performing cross-validation (Section 11.6.6.1). With the expansion of DISCRIM, there seems little advantage to using CANDISC. Both programs offer alternate inferential tests, dimension reduction analysis, and all of the standard matrices of discriminant function results. DISCRIM classifies cases while CANDISC does not, so that some time savings may be gained by using CANDISC for direct DISCRIM when classification is not wanted.

Finally, stepwise (but not sequential) analysis is accomplished through STEPDISC. As seen in Table 11.12, very few additional amenities are available in this program. There is no classification, nor is there information about the discriminant functions. On the other hand, this program offers plenty of options for entry and removal of predictors.

11.8 ■ COMPLETE EXAMPLE OF DISCRIMINANT FUNCTION ANALYSIS

The example of direct discriminant function analysis[13] in this section explores how role-dissatisfied housewives, role-satisfied housewives, and employed women differ in attitudes. The sample of 465 women is described in Appendix B, Section B.1. The grouping variable is role-

[13] The example of hierarchical discriminant analysis from earlier editions of this book (e.g., Tabachnick & Fidell, 1989) has been dropped in favor of sequential logistic regression analysis of the same variables (cf. Section 12.8).

dissatisfied housewives (UNHOUSE), role-satisfied housewives (HAPHOUSE), and working women (WORKING).

Predictors are internal vs. external locus of control (CONTROL), satisfaction with current marital status (ATTMAR), attitude toward women's role (ATTROLE), and attitude toward homemaking (ATTHOUSE). A fifth attitudinal variable, attitude toward paid work, was dropped from analysis because data were available only for women who had been employed within the past five years and use of this predictor would have involved nonrandom missing values (cf. Chapter 4). The example of DISCRIM, then, involves prediction of group membership from the four attitudinal variables.

The direct discriminant function analysis allows us to evaluate the distinctions among the three groups on the basis of attitudes. We explore the dimensions on which the groups differ, the predictors contributing to differences among groups on these dimensions, and the degree to which we can accurately classify members into their own groups. We also evaluate efficiency of classification with a cross-validation sample.

11.8.1 Evaluation of Assumptions

The data are first evaluated with respect to practical limitations of DISCRIM.

11.8.1.1 Unequal Sample Sizes and Missing Data

In a screening run through BMDP7D (cf. Chapters 4 or 8), seven cases had missing values among the four attitudinal predictors. Missing data were scattered over predictors and groups in apparently random fashion, so that deletion of the cases was deemed appropriate.[14] The full data set includes 458 cases once cases with missing values are deleted.

During classification, unequal sample sizes are used to modify the probabilities with which cases are classified into groups. Because the sample is randomly drawn from the population of interest, sample sizes in groups are believed to represent some real process in the population that should be reflected in classification. For example, knowledge that over half the women are employed implies that greater weight should be given the WORKING group.

11.8.1.2 Multivariate Normality

After deletion of cases with missing data, there are still over 80 cases per group. Although a BMDP7D run (not shown) reveals skewness in ATTMAR, sample sizes are large enough to suggest normality of sampling distributions of means. Therefore there is no reason to expect distortion of results due to failure of multivariate normality.

11.8.1.3 Linearity

Although ATTMAR and RACE are skewed, there is no expectation of curvilinearity between these two and the remaining predictors. At worst, ATTMAR in conjunction with the remaining continuous, well-behaved predictors may contribute to a mild reduction in association.

[14] Alternative strategies for dealing with missing data are discussed in Chapter 4.

11.8.1.4 ■ Outliers

To identify univariate outliers, z scores associated with minimum and maximum values on each of the four predictors are investigated through BMDP7D for each group separately. There are some questionable values on ATTHOUSE, with a few exceptionally positive scores. These values are about 4.5 standard deviations from their group means, making them candidates for deletion or alteration. However, the cases are retained for the search for multivariate outliers.

BMDP7M is used to search for multivariate outliers in each group separately. A portion of BMDP7M output for the WORKING group is shown in Table 11.13. Outliers are identified as cases with too large a Mahalanobis D^2, evaluated as χ^2 with degrees of freedom equal to the number of predictors. Critical χ^2 with 4 df at $\alpha = .001$ is 18.467; any case with $D^2 > 18.467$ is an outlier. In Table 11.13, cases 261 ($D^2 = 25.3$) and 299 ($D^2 = 24.8$) are identified as outliers in the group of WORKING women. No additional outliers were found.

The multivariate outliers are the same cases that have extreme univariate scores on ATTHOUSE. Because transformation is questionable for ATTHOUSE (where it seems unreasonable to transform the predictor for only two cases) it is decided to delete multivariate outliers.

Therefore, of the original 465 cases, 7 are lost due to missing values and 2 are multivariate outliers, leaving a total of 456 cases for analysis.

11.8.1.5 ■ Homogeneity of Variance-Covariance Matrices

A second BMDP7M run, Table 11.14, deletes the outliers in order to evaluate homogeneity of variance-covariance matrices. Examination of sample variances for the 4 predictors reveals no gross discrepancies among the groups. ATTMAR shows the largest ratio of variances, where the ratio is less than 3 ($10.29752^2/6.62350^2 = 2.73$) for the HAPHOUSE vs. UNHOUSE groups. There is somewhat more variance in attitude toward marital status in role-dissatisfied than in role-satisfied housewives. Sample sizes are fairly discrepant, with almost three times as many working women as role-dissatisfied housewives, but with two-tailed tests and reasonable homogeneity of variance, DISCRIM is robust enough to handle the discrepancies.

Homogeneity of variance-covariance matrices is also examined through plots of the first two discriminant functions (called canonical variables) produced by BMDP7M using the setup of Table 11.14. As seen in Figure 11.2, the spread of cases for the three groups is relatively equal. Therefore no further test of homogeneity of variance-covariance matrices is necessary.

11.8.1.6 ■ Multicollinearity and Singularity

Because BMDP7M, used for the major analysis, protects against multicollinearity through checks of tolerance, no formal evaluation is necessary (cf. Chapter 4).

11.8.2 ■ Direct Discriminant Function Analysis

Direct DISCRIM is performed through BMDP7M with the 4 attitudinal predictors all forced into the equation. The program instructions and a segment of the output appear in Table 11.15. Because this is a direct analysis, stepwise output is suppressed through NO STEP.

In Table 11.15 the F TO ENTER values at step 0 are univariate F ratios with 2 and 453 df for the individual predictors. Three of the four predictors, all except CONTROL, show univariate F

TABLE 11.13 IDENTIFICATION OF OUTLIERS. SETUP AND SELECTED OUTPUT FROM BMDP7M

```
/INPUT        VAR=13.    FILE='DISCRIM.DAT'.   FORMAT IS  FREE.
/VARIABLE     NAMES ARE CASESEQ, WORKSTAT, MARITAL, CHILDREN,
                   RELIGION, RACE, CONTROL, ATTMAR, ATTROLE, SEL,
                   ATTHOUSE, AGE, EDUC.
              MISSING=2*0, 3*9, 5*0, 1, 2*0.
              LABEL = CASESEQ.
              USE = WORKSTAT, CONTROL, ATTMAR, ATTROLE, ATTHOUSE.
/GROUP        VAR = WORKSTAT.
              CODES(WORKSTAT) ARE 1 TO 3.
              NAMES(WORKSTAT) ARE WORKING, HAPHOUSE, UNHOUSE.
/DISC         LEVEL=6*0, 3*1, 0, 1, 2*0.
              FORCE = 1.
/PRINT        NO STEP.
/END
```

GROUP WORKING		INCORRECT CLASSIFICATIONS	MAHALANOBIS DSQUARE FROM AND POSTERIOR PROBABILITY FOR GROUP		
			WORKING	HAPHOUSE	UNHOUSE
CASE					
252	337		2.7 0.527	4.0 0.263	4.5 0.210
256	341	HAPHOUSE	1.1 0.325	0.5 0.446	1.8 0.229
259	344	HAPHOUSE	0.7 0.329	0.3 0.405	1.2 0.266
260	345	HAPHOUSE	2.8 0.312	2.2 0.421	3.1 0.267
261	346	HAPHOUSE	25.3 0.328	24.1 0.609	28.6 0.063
262	347	UNHOUSE	1.0 0.365	1.7 0.261	1.0 0.374
263	348		1.6 0.481	3.0 0.236	2.7 0.283
264	349		4.7 0.390	5.6 0.244	4.8 0.366
265	355		3.1 0.427	4.6 0.201	3.4 0.372
266	357	HAPHOUSE	1.1 0.314	0.4 0.447	1.6 0.239
267	358	UNHOUSE	5.6 0.280	7.0 0.141	4.2 0.579
268	359		4.6 0.481	6.6 0.182	5.3 0.337
270	362	UNHOUSE	1.1 0.328	1.3 0.291	0.8 0.380
272	365	HAPHOUSE	3.4 0.353	2.7 0.502	5.1 0.145
274	369	UNHOUSE	6.1 0.359	7.5 0.175	5.6 0.466
276	372	HAPHOUSE	7.1 0.244	5.3 0.586	7.8 0.169
278	378		5.8 0.593	7.8 0.219	8.1 0.188
281	381		4.7 0.486	6.3 0.216	5.7 0.297
283	383	UNHOUSE	3.5 0.229	2.9 0.308	2.1 0.463
284	384	UNHOUSE	2.4 0.246	1.9 0.315	1.3 0.439
286	386		0.7 0.406	1.4 0.290	1.3 0.303
287	387		1.1 0.451	2.0 0.277	2.1 0.272
289	397	HAPHOUSE	5.1 0.386	4.6 0.488	7.3 0.126
290	398		6.5 0.444	8.9 0.130	6.5 0.426
291	399		8.5 0.443	9.3 0.294	9.5 0.264
292	400		3.3 0.475	5.4 0.170	3.9 0.355
293	401		9.2 0.423	10.6 0.209	9.5 0.368
295	403		4.0 0.429	6.2 0.144	4.0 0.427
296	404		1.2 0.436	1.8 0.321	2.4 0.243
298	406		7.0 0.410	8.2 0.228	7.3 0.362
299	407	HAPHOUSE	24.8 0.311	23.4 0.616	27.7 0.073
307	425		1.5 0.417	2.3 0.268	2.0 0.315

ratios that are significant at the .01 level. When all 4 predictors are used, the approximate F of 6.274 (with 8 and 900 df based on Wilks' Lambda) is highly significant. That is, there is reliable separation of the three groups based on all four predictors combined, as discussed in Section 11.6.1.1.

TABLE 11.14 SETUP AND SELECTED OUTPUT FROM BMDP7M RUN TO CHECK HOMOGENEITY OF VARIANCE-COVARIANCE MATRICES

```
/INPUT        VAR=13.    FILE='DISCRIM.DAT'.    FORMAT IS  FREE.
/VARIABLE     NAMES ARE CASESEQ, WORKSTAT, MARITAL, CHILDREN,
                     RELIGION, RACE, CONTROL, ATTMAR, ATTROLE, SEL,
                     ATTHOUSE, AGE, EDUC.
              MISSING=2*0, 3*9, 5*0, 1, 2*0.
              LABEL = CASESEQ.
              USE = WORKSTAT, CONTROL, ATTMAR, ATTROLE, ATTHOUSE.
/GROUP        VAR = WORKSTAT.
              CODES(WORKSTAT) ARE 1 TO 3.
              NAMES(WORKSTAT) ARE WORKING, HAPHOUSE, UNHOUSE.
/TRANSFORM      DELETE = 261, 299.
/DISC         LEVEL=6*0, 3*1, 0, 1, 2*0.
              FORCE = 1.
/PRINT        NO STEP.
/PLOT         GROUP=1. GROUP=2. GROUP=3.
/END
```

STANDARD DEVIATIONS

GROUP =	WORKING	HAPHOUSE	UNHOUSE	POOLED
VARIABLE				
7 CONTROL	1.23780	1.30984	1.25401	1.26253
8 ATTMAR	8.53003	6.62350	10.29753	8.36830
9 ATTROLE	6.95618	6.45843	5.75977	6.61150
11 ATTHOUSE	4.45544	3.88348	3.95846	4.20608

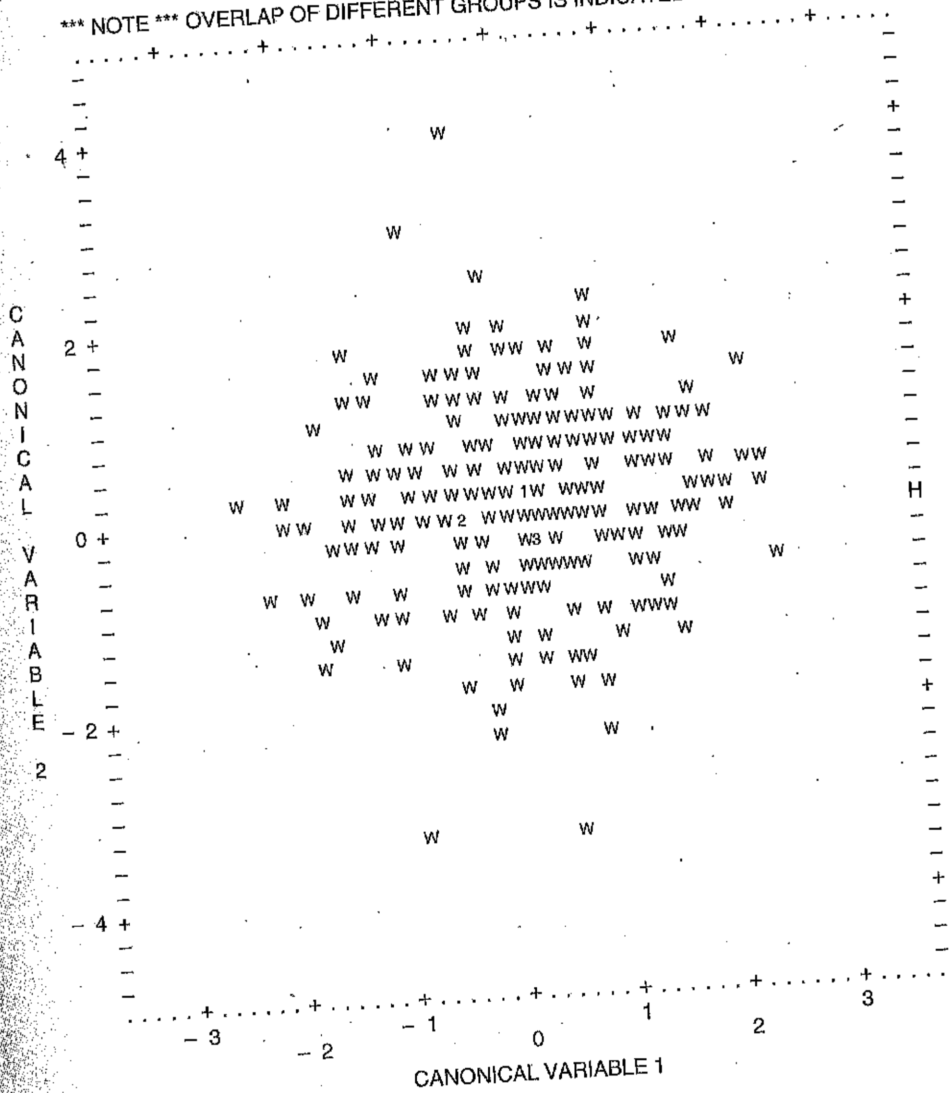
Table 11.15 also shows the classification functions used to classify cases into the three groups (see Equation 11.3) and the results of that classification, with and without jackknifing (see Section 11.6.6). In this case, classification is made on the basis of a modified equation in which unequal prior probabilities are used to reflect unequal group sizes by the use of the PRIOR convention in the setup.

A total of 54.4% of cases are correctly classified by normal procedures, 53.5% by jackknifed procedures. How do these compare with random assignment? Prior probabilities, specified as .52 (WORKING), .30 (HAPHOUSE), and .18 (UNHOUSE), put 237 cases (.52 × 456) in the WORKING group, 137 in the HAPHOUSE group, and 82 in the UNHOUSE group. Of those randomly assigned to the WORKING group, 123 (.52 × 237) should be correct, while 41.1 (.30 × 137) and 14.8 (.18 × 82) should be correct by chance in the HAPHOUSE and UNHOUSE groups, respectively. Over all three groups 178.9 out of the 456 cases or 39% should be correct by chance alone. Both classification procedures correctly classify substantially more than that.

Although BMDP7M does not provide tests of the discriminant functions, it does provide considerable information about them as seen near the end of Table 11.15. Canonical correlations for each discriminant function (.26699 and .18437), although small, are relatively equal for the two discriminant functions. CANONICAL VARIABLES EVALUATED AT GROUP MEANS are centroids on the discriminant functions for the groups, discussed in Sections 11.4.1 and 11.6.3.1.

An additional BMDP7M run for cross-validation is shown in Table 11.16. Approximately 25% of the original cases are randomly selected out in the TRANSFORM paragraph to use for cross-validation. The 111 cases randomly selected out are used to assess the results of classification. New group membership is designated in the GROUP paragraph for these cases: NEWWORK (for WORKING), NEWHAP (for HAPHOUSE), and NEWUN (for UNHOUSE). The other 345 cases

*** NOTE *** OVERLAP OF DIFFERENT GROUPS IS INDICATED BY AN ASTERISK (*).

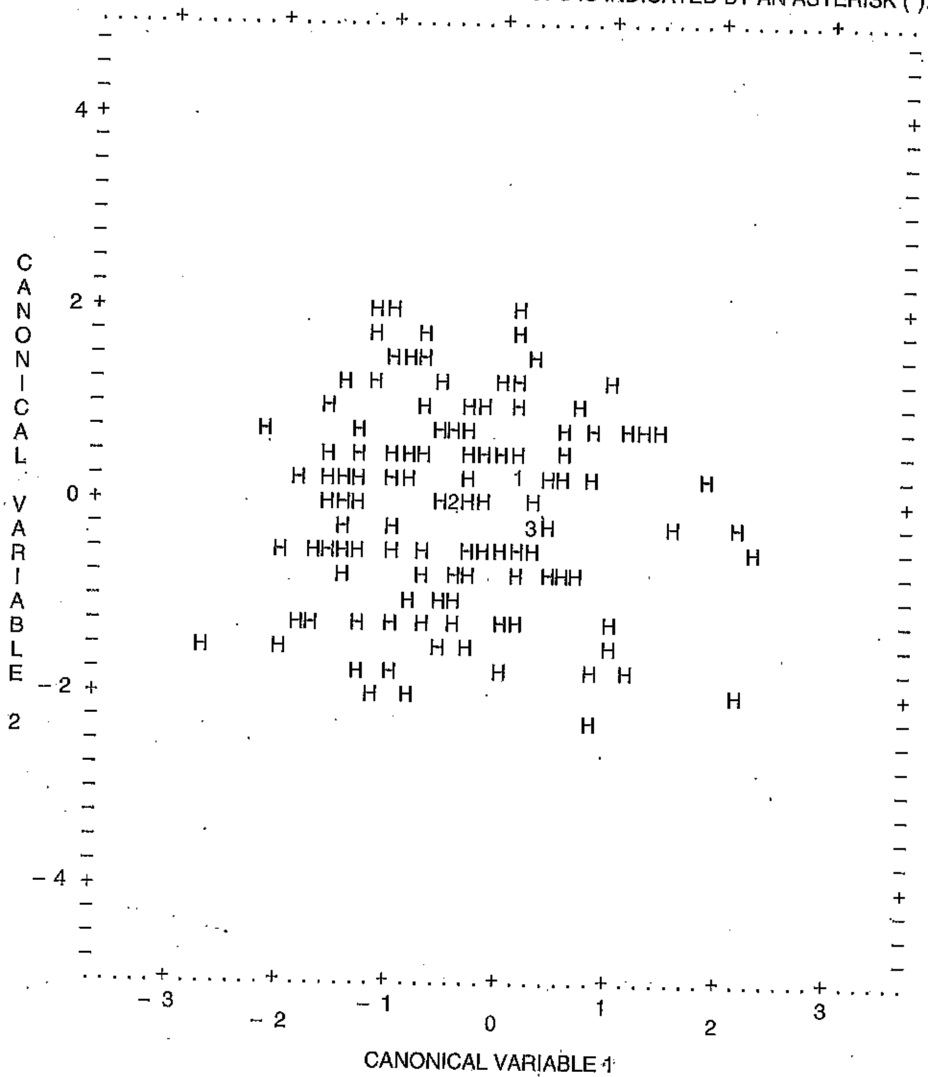


(a)

FIGURE 11.2 SCATTERPLOTS OF CASES ON FIRST TWO CANONICAL VARIATES FOR (a) WORKING WOMEN, (b) ROLE-SATISFIED HOUSEWIVES, AND (c) ROLE-DISSATISFIED HOUSEWIVES. SETUP APPEARS IN TABLE 11.14.

are used, as specified in the GROUP paragraph, to develop discriminant functions and classification equations. Table 11.16 shows results of classification and jackknifed classification for original and cross-validation cases. Unexpectedly, classification improves for the cross-validation sample, although ordinarily it is worse.

*** NOTE *** OVERLAP OF DIFFERENT GROUPS IS INDICATED BY AN ASTERISK (*).

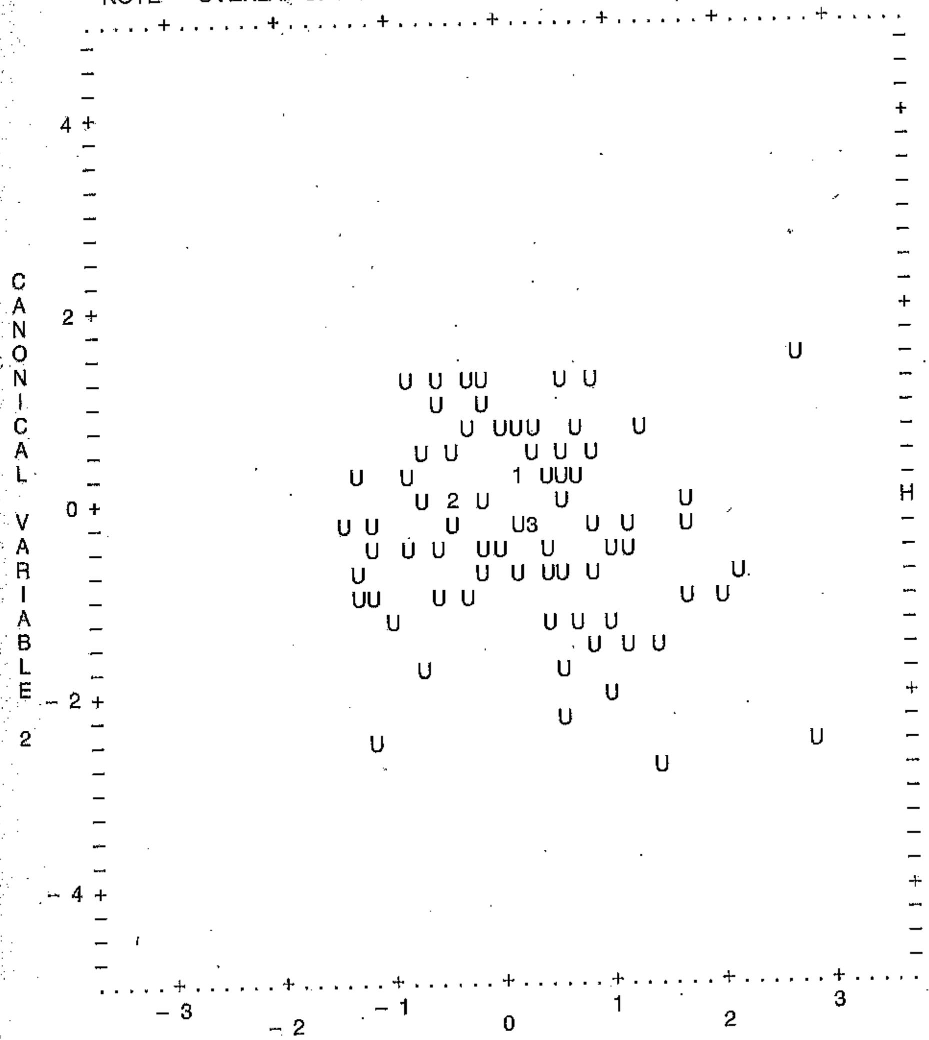


(b)

FIGURE 11.2 *(CONTINUED)*

SPSS DISCRIMINANT is used to test the reliability of the discriminant functions and to provide the loadings for them, as shown in Table 11.17. The total overlap between groups and predictors (AFTER FUNCTION 0) is tested as CHI-SQUARE with 8 df (49.002) and found to be reliable. After the first discriminant function is removed (AFTER FUNCTION 1) CHI-SQUARE with 3 df is 15.614, and still reliable at $p = 0.0014$. This second chi square is a test of the overlap between groups and predictors after the first function is removed. Because there are only two possible discriminant functions with three groups, this is a test of the second discriminant function.

*** NOTE *** OVERLAP OF DIFFERENT GROUPS IS INDICATED BY AN ASTERISK (*).



(c)

FIGURE 11.2 *(CONTINUED)*

SPSS DISCRIMINANT, unlike BMDP7M, also provides the loadings for the predictors on the discriminant functions. Labeled Pooled-within groups correlations between discriminating variables, the first column is the correlations between the predictors and the first discriminant function; the second column is the correlations between the predictors and the second discriminant function. The loadings could be calculated by hand according to Equation 11.7, but they are more conveniently found through SPSS, as shown in Table 11.17. For interpretation of this heterogeneous example, loadings less than .50 are not considered.

TABLE 11.15 SETUP AND PARTIAL OUTPUT FROM BMDP7M DISCRIMINANT FUNCTION ANALYSIS OF FOUR ATTITUDINAL VARIABLES

```
/INPUT        VAR=13.    FILE='DISCRIM.DAT'.    FORMAT IS  FREE.
/VARIABLE     NAMES ARE CASESEQ, WORKSTAT, MARITAL, CHILDREN,
                      RELIGION, RACE, CONTROL, ATTMAR, ATTROLE, SEL,
                      ATTHOUSE, AGE, EDUC.
              MISSING=2*0, 3*9, 5*0, 1, 2*0.
              LABEL = CASESEQ.
              USE = WORKSTAT, CONTROL, ATTMAR, ATTROLE, ATTHOUSE.
/GROUP        VAR = WORKSTAT.
              CODES(WORKSTAT) ARE 1 TO 3.
              NAMES(WORKSTAT) ARE WORKING, HAPHOUSE, UNHOUSE.
              PRIOR = 0.52, 0.30, 0.18.
/TRANSFORM    DELETE = 261, 299.
/DISC         LEVEL=6*0, 3*1, 0, 1, 2*0.
              FORCE = 1.
/PLOT         NO CANON.
/PRINT        NO STEP.  CORR.
/END

PRIOR PROBABILITIES. . . .   0.52000   0.30000   0.18000

WITHIN CORRELATION MATRIX

                  CONTROL      ATTMAR       ATTROLE      ATTHOUSE
                      7            8            9            11

CONTROL     7     1.00000
ATTMAR      8     0.17169      1.00000
ATTROLE     9     0.00912     -0.07010      1.00000
ATTHOUSE   11     0.15500      0.28229     -0.29145      1.00000

****************************************************************************

STEP NUMBER    0

  VARIABLE     F TO   FORCE TOLERNCE *   VARIABLE      F TO    FORCE TOLERNCE
              REMOVE  LEVEL          *                ENTER    LEVEL
        DF =  2  454                 *         DF =  2  453
                                     *    7 CONTROL     2.96     1    1.00000
                                     *    8 ATTMAR      9.81     1    1.00000
                                     *    9 ATTROLE    11.26     1    1.00000
                                     *   11 ATTHOUSE    8.91     1    1.00000

****************************************************************************

STEP NUMBER    4
VARIABLE ENTERED    7 CONTROL

  VARIABLE     F TO   FORCE TOLERNCE *   VARIABLE      F TO    FORCE TOLERNCE
              REMOVE  LEVEL          *                ENTER    LEVEL
        DF =  2  450                 *         DF =  2  449
  7 CONTROL     1.08     1   0.95517 *
  8 ATTMAR      4.90     1   0.90351 *
  9 ATTROLE     9.31     1   0.91201 *
 11 ATTHOUSE    3.22     1   0.83315 *

U-STATISTIC(WILKS' LAMBDA) 0.8971503   DEGREES OF FREEDOM    4   2    453
APPROXIMATE FSTATISTIC         6.274   DEGREES OF FREEDOM    8.00    900.00

 F - MATRIX         DEGREES OF FREEDOM =    4   450

             WORKING  HAPHOUSE
HAPHOUSE      7.57
UNHOUSE       4.12      7.30
```

TABLE 11.15 *(CONTINUED)*

CLASSIFICATION FUNCTIONS

GROUP =	WORKING	HAPHOUSE	UNHOUSE
VARIABLE			
7 CONTROL	3.22293	3.21674	3.36966
8 ATTMAR	0.07664	0.04406	0.09771
9 ATTROLE	1.08148	1.15040	1.13735
11 ATTHOUSE	1.64843	1.62485	1.71834
CONSTANT	-50.30873	-52.00294	-56.54160

CLASSIFICATION MATRIX

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP		
		WORKING	HAPHOUSE	UNHOUSE
WORKING	86.2	206	31	2
HAPHOUSE	27.9	96	38	2
UNHOUSE	4.9	66	11	4
TOTAL	54.4	368	80	8

JACKKNIFED CLASSIFICATION

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP		
		WORKING	HAPHOUSE	UNHOUSE
WORKING	84.9	203	33	3
HAPHOUSE	27.2	97	37	2
UNHOUSE	4.9	66	11	4
TOTAL	53.5	366	81	9

EIGENVALUES

0.07675 0.03519

CUMULATIVE PROPORTION OF TOTAL DISPERSION

0.68566 1.00000

CANONICAL CORRELATIONS

0.26699 0.18437

AVERAGE SQUARED CANONICAL CORRELATION 0.05264

GROUP	CANONICAL VARIABLES EVALUATED AT GROUP MEANS	
WORKING	0.14072	0.15053
HAPHOUSE	-0.41601	-0.05393
UNHOUSE	0.28328	-0.35361

TABLE 11.16 CROSS-VALIDATION OF CLASSIFICATION OF CASES BY FOUR ATTITUDINAL VARIABLES. SETUP AND SELECTED OUTPUT FROM BMDP7M

```
/INPUT        VAR=13.    FILE='DISCRIM.DAT'.    FORMAT IS  FREE.
/VARIABLE     NAMES ARE CASESEQ, WORKSTAT, MARITAL, CHILDREN,
                     RELIGION, RACE, CONTROL, ATTMAR, ATTROLE, SEL,
                     ATTHOUSE, AGE, EDUC.
              MISSING=2*0, 3*9, 5*0, 1, 2*0.
              LABEL = CASESEQ.
              USE = WORKSTAT, CONTROL, ATTMAR, ATTROLE, ATTHOUSE.
/GROUP        VAR = WORKSTAT.
              CODES(WORKSTAT) ARE 1 TO 6.
              NAMES(WORKSTAT) ARE WORKING, HAPHOUSE, UNHOUSE,
                                 NEWWORK, NEWHAP, NEWUN.
              PRIOR = 0.39, 0.225, 0.135, 0.13, 0.075, 0.045.
              USE = 1 TO 3.
/TRANSFORM    DELETE = 261, 299.
              IF(RNDU(11738) LE .25) THEN WORKSTAT = WORKSTAT + 3.
/DISC         LEVEL=6*0, 3*1, 0, 1, 2*0.
              FORCE = 1.
/PLOT         NO CANON.
/PRINT        NO STEP.  NO POST.  NO POINT.
/END
```

CLASSIFICATION FUNCTIONS

GROUP =	WORKING	HAPHOUSE	UNHOUSE
VARIABLE			
7 CONTROL	2.84787	2.80040	3.03103
8 ATTMAR	0.09174	0.05442	0.11964
9 ATTROLE	0.98078	1.04430	1.02213
11 ATTHOUSE	1.42151	1.40166	1.47110
CONSTANT	-45.18960	-46.41790	-50.85888

CLASSIFICATION MATRIX

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP		
		WORKING	HAPHOUSE	UNHOUSE
WORKING	86.3	158	24	1
HAPHOUSE	25.0	70	24	2
UNHOUSE	7.6	54	7	5
NEWWORK	0.0	49	7	0
NEWHAP	0.0	28	12	0
NEWUN	0.0	12	3	0
TOTAL	54.2	371	77	8

JACKKNIFED CLASSIFICATION

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP		
		WORKING	HAPHOUSE	UNHOUSE
WORKING	85.8	157	25	1
HAPHOUSE	22.9	72	22	2
UNHOUSE	4.5	56	7	3
NEWWORK	0.0	49	7	0
NEWHAP	0.0	28	12	0
NEWUN	0.0	12	3	0
TOTAL	52.8	374	76	6

TABLE 11.17 SETUP AND PARTIAL OUTPUT FROM SPSS DISCRIMINANT FOR PREDICTION OF MEMBERSHIP IN THREE GROUPS ON THE BASIS OF FOUR ATTITUDINAL VARIABLES

```
DATA LIST          FILE='DISCRIM.DAT'      FREE
                     /CASESEQ,WORKSTAT,MARITAL,CHILDREN,RELIGION,RACE,
                     CONTROL,ATTMAR,ATTROLE,SEL,ATTHOUSE,AGE,EDUC.
MISSING VALUES     CASESEQ, WORKSTAT, RACE, CONTROL, ATTMAR,
                   ATTROLE, SEL, AGE, EDUC(0)  ATTHOUSE(1)  MARITAL TO
                   RELIGION(9).
SELECT IF          (CASESEQ NE 346).
SELECT IF          (CASESEQ NE 407).
DISCRIMINANT       GROUPS=WORKSTAT(1,3)/
                    VARIABLES=CONTROL ATTMAR ATTROLE ATTHOUSE/
                    ANALYSIS=CONTROL TO ATTHOUSE/
                    PRIORS = SIZE.


Prior Probabilities
   Group    Prior     Label
     1     .52412
     2     .29825
     3     .17763
   Total  1.00000

                         Canonical Discriminant Functions

               Pct of    Cum  Canonical   After  Wilks'
Fcn Eigenvalue Variance  Pct      Corr      Fcn  Lambda  Chisquare    DF   Sig
                                        :    0   .8972     49.002     8  .0000
  1*     .0768   68.57   68.57    .2670 :    1   .9660     15.614     3  .0014
  2*     .0352   31.43  100.00    .1844 :

  * marks the  2 canonical discriminant functions remaining in the analysis.


Structure Matrix:

Pooled-withingroups correlations between discriminating variables
                                    and canonical discriminant functions
(Variables ordered by size of correlation within function).

              FUNC  1    FUNC  2
ATTMAR        .71846*    .32299
ATTHOUSE      .67945*    .33332
ATTROLE      -.63925     .72223*
CONTROL       .28168     .44494*
```

A plot of the placement of the centroids for the three group on the two discriminant functions (canonical variables) as axes appears in Figure 11.3. The points that are plotted are given in Table 11.15 as CANONICAL VARIABLES EVALUATED AT GROUP MEANS.

A summary of information appropriate for publication appears in Table 11.18. In the table are the loadings, univariate F for each predictor, and pooled within-group correlations among predictors (as found through SPSS DISCRIMINANT).

Contrasts run through BMDP7M are shown in Tables 11.19 to 11.21. In Table 11.19, means on predictors for WORKING women are contrasted with the pooled means for HAPHOUSE and

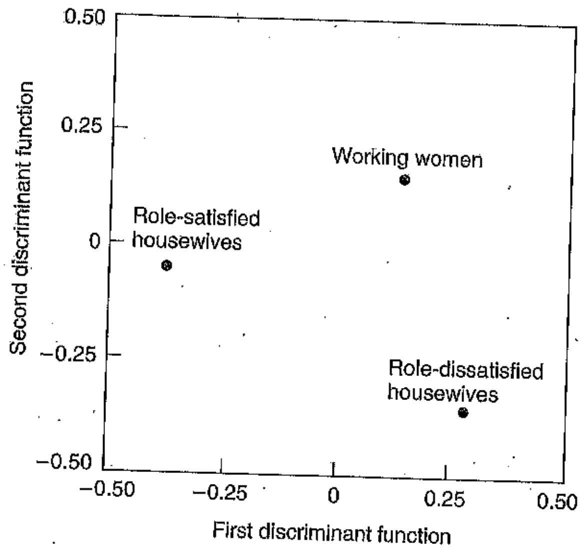


FIGURE 11.3 PLOTS OF THREE GROUPS CENTROIDS ON TWO DISCRIMINANT FUNCTIONS DERIVED FROM FOUR ATTITUDINAL VARIABLES.

TABLE 11.18 RESULTS OF DISCRIMINANT FUNCTION ANALYSIS OF ATTITUDINAL VARIABLES

Predictor variable	Correlations of predictor variables with discriminant functions 1	2	Univariate $F(2, 453)$	Pooled within-group correlations among predictors ATTMAR	ATTROLE	ATTHOUSE
CONTROL	.28	.44	2.96	.17	.01	.16
ATTMAR	.72	.32	9.81		−.07	.28
ATTROLE	−.64	.72	11.26			−.29
ATTHOUSE	.68	.33	8.91			
Canonical R	.27	.18				
Eigenvalue	.08	.04				

UNHOUSE to determine which predictors distinguish WORKING women from others. Table 11.20 has the HAPHOUSE group contrasted with the other two groups; Table 11.21 shows the UNHOUSE group contrasted with the other two groups. F TO ENTER is the univariate F ratio as in ANOVA where each predictor is evaluated separately. F TO REMOVE is the contribution each predictor makes to the contrast after adjustment for all other predictors. Critical F TO REMOVE for the four predictors at $\alpha_i = .01$ (cf. Section 11.6.4) with 1 and 450 df is approximately 6.8.

On the basis of both F TO ENTER and F TO REMOVE, the predictor that most clearly distinguishes the WORKING group from the other two is ATTROLE. The HAPHOUSE group differs from the other two groups on the basis of ATTMAR after adjustment for the remaining predictors; ATTMAR, ATTROLE, and ATTHOUSE reliably separate this group from the other two in ANOVA.

TABLE 11.19 SETUP AND PARTIAL OUTPUT OF BMDP7M CONTRASTING THE WORKING GROUP WITH THE OTHER TWO GROUPS

```
/INPUT          VAR=13.    FILE='DISCRIM.DAT'.   FORMAT IS  FREE.
/VARIABLE       NAMES ARE CASESEQ, WORKSTAT, MARITAL, CHILDREN,
                      RELIGION, RACE, CONTROL, ATTMAR, ATTROLE, SEL,
                      ATTHOUSE, AGE, EDUC.
                MISSING=2*0, 3*9, 5*0, 1, 2*0.
                LABEL = CASESEQ.
                USE = WORKSTAT, CONTROL, ATTMAR, ATTROLE, ATTHOUSE.
/GROUP          VAR = WORKSTAT.
                CODES(WORKSTAT) ARE 1 TO 3.
                NAMES(WORKSTAT) ARE WORKING, HAPHOUSE, UNHOUSE.
/TRANSFORM      DELETE = 261, 299.
/DISC           LEVEL=6*0, 3*1, 0, 1, 2*0.
                FORCE = 1.
                CONTRAST = -2, 1, 1.
/PLOT           NO CANON.
/PRINT          NO STEP.  NO POST.  NO POINT.
/END

*************************************************************************

STEP NUMBER   0

  VARIABLE      F TO    FORCE TOLERNCE *   VARIABLE      F TO    FORCE TOLERNCE
               REMOVE   LEVEL          *                ENTER    LEVEL
        DF =  1  454                   *         DF =  1  453
                                       *   7 CONTROL     1.08     1    1.00000
                                       *   8 ATTMAR      0.13     1    1.00000
                                       *   9 ATTROLE    16.55     1    1.00000
                                       *  11 ATTHOUSE    0.06     1    1.00000

*************************************************************************

STEP NUMBER    4
VARIABLE ENTERED    8 ATTMAR

  VARIABLE      F TO    FORCE TOLERNCE *   VARIABLE      F TO    FORCE TOLERNCE
               REMOVE   LEVEL          *                ENTER    LEVEL
        DF =  1  450                   *        DF =  1  449
  7 CONTROL     0.79     1    0.95517  *
  8 ATTMAR      0.22     1    0.90351  *
  9 ATTROLE    16.87     1    0.91201  *
 11 ATTHOUSE    0.83     1    0.83315  *

EIGENVALUES

               0.07675      0.03519

CANONICAL CORRELATIONS

               0.26699      0.18437
```

The UNHOUSE group does not differ from the other two when each predictor is adjusted for all others but does differ on ATTHOUSE and ATTMAR without adjustment.

A checklist for a direct discriminant function analysis appears in Table 11.22. It is followed by an example of a Results section, in journal format, for the analysis just described.

TABLE 11.20 SETUP AND PARTIAL OUTPUT OF BMDP7M CONTRASTING THE HAPHOUSE GROUP WITH THE OTHER TWO GROUPS

```
/INPUT          VAR=13.    FILE='DISCRIM.DAT'.    FORMAT IS  FREE.
/VARIABLE       NAMES ARE CASESEQ, WORKSTAT, MARITAL, CHILDREN,
                        RELIGION, RACE, CONTROL, ATTMAR, ATTROLE, SEL,
                        ATTHOUSE, AGE, EDUC.
                MISSING=2*0, 3*9, 5*0, 1, 2*0.
                LABEL = CASESEQ.
                USE = WORKSTAT, CONTROL, ATTMAR, ATTROLE, ATTHOUSE.
/GROUP          VAR = WORKSTAT.
                CODES(WORKSTAT) ARE 1 TO 3.
                NAMES(WORKSTAT) ARE WORKING, HAPHOUSE, UNHOUSE.
/TRANSFORM      DELETE = 261, 299.
/DISC           LEVEL=6*0, 3*1, 0, 1, 2*0.
                FORCE = 1.
                CONTRAST = 1, -2, 1.
/PLOT           NO CANON.
/PRINT          NO STEP.  NO POST.  NO POINT.
/END
```

```
**************************************************************************

STEP NUMBER   0

  VARIABLE     F TO   FORCE TOLERNCE *   VARIABLE     F TO   FORCE TOLERNCE
              REMOVE  LEVEL          *                ENTER  LEVEL
        DF =  1  454                 *         DF =  1  453
                                     *   7 CONTROL     3.42    1   1.00000
                                     *   8 ATTMAR     18.95    1   1.00000
                                     *   9 ATTROLE    11.73    1   1.00000
                                     *  11 ATTHOUSE   17.05    1   1.00000

**************************************************************************

STEP NUMBER   4
VARIABLE ENTERED    7 CONTROL

  VARIABLE     F TO   FORCE TOLERNCE *   VARIABLE     F TO   FORCE TOLERNCE
              REMOVE  LEVEL          *                ENTER  LEVEL
        DF =  1  450                 *         DF =  1  449
  7 CONTROL      0.78    1   0.95517 *
  8 ATTMAR       9.66    1   0.90351 *
  9 ATTROLE      5.45    1   0.91201 *
 11 ATTHOUSE     4.09    1   0.83315 *

EIGENVALUES

                0.07675      0.03519

CANONICAL CORRELATIONS

                0.26699      0.18437
```

TABLE 11.21 SET AND PARTIAL OUTPUT OF BMDP7M CONTRASTING THE UNHOUSE GROUP WITH THE OTHER TWO GROUPS

```
/INPUT        VAR=13.    FILE='DISCRIM.DAT'.    FORMAT IS  FREE.
/VARIABLE     NAMES ARE CASESEQ, WORKSTAT, MARITAL, CHILDREN,
                      RELIGION, RACE, CONTROL, ATTMAR, ATTROLE, SEL,
                      ATTHOUSE, AGE, EDUC.
              MISSING=2*0, 3*9, 5*0, 1, 2*0.
              LABEL = CASESEQ.
              USE = WORKSTAT, CONTROL, ATTMAR, ATTROLE, ATTHOUSE.
/GROUP        VAR = WORKSTAT.
              CODES(WORKSTAT) ARE 1 TO 3.
              NAMES(WORKSTAT) ARE WORKING, HAPHOUSE, UNHOUSE.
/TRANSFORM    DELETE = 261, 299.
/DISC         LEVEL=6*0, 3*1, 0, 1, 2*0.
              FORCE = 1.
              CONTRAST = 1, 1, -2.
/PLOT         NO CANON.
/PRINT        NO STEP.  NO POST.  NO POINT.
/END
*********************************************************************************

STEP NUMBER   0

 VARIABLE      F TO    FORCE TOLERNCE *    VARIABLE       F TO    FORCE TOLERNCE
               REMOVE  LEVEL          *                   ENTER   LEVEL
                                      *          DF =  1  453
       DF =  1  454                   *    7 CONTROL       5.81     1    1.00000
                                      *    8 ATTMAR       12.27     1    1.00000
                                      *    9 ATTROLE       0.03     1    1.00000
                                      *   11 ATTHOUSE     11.58     1    1.00000

*********************************************************************************

STEP NUMBER    4
VARIABLE ENTERED    9 ATTROLE

 VARIABLE      F TO    FORCE TOLERNCE *    VARIABLE       F TO    FORCE TOLERNCE
               REMOVE  LEVEL          *                   ENTER   LEVEL
                                      *          DF =  1  449
       DF =  1  450
 7 CONTROL       2.13     1   0.95517 *
 8 ATTMAR        5.56     1   0.90351 *
 9 ATTROLE       1.14     1   0.91201 *
11 ATTHOUSE      6.20     1   0.83315 *

EIGENVALUES

                 0.07675     0.03519

CANONICAL CORRELATIONS

                 0.26699     0.18437
```

TABLE 11.22 CHECKLIST FOR DIRECT DISCRIMINANT FUNCTION ANALYSIS

1. Issues
   a. Unequal sample sizes and missing data
   b. Normality of sampling distributions
   c. Outliers
   d. Linearity
   e. Homogeneity of variance-covariance matrices
   f. Multicollinearity and singularity
2. Major analysis
   a. Significance of discriminant functions. If significant:
      (1) Variance accounted for
      (2) Plot(s) of discriminant functions
      (3) Loading matrix
   b. Variables separating each group
3. Additional analyses
   a. Group means for high-loading variables
   b. Pooled within-group correlations among predictor variables
   c. Classification results
      (1) Jackknifed classification
      (2) Cross-validation
   d. Change in Rao's *V* (or stepdown *F*) plus univariate *F* for predictors

Results

A direct discriminant function analysis was performed using four attitudinal variables as predictors of membership in three groups. Predictors were locus of control, attitude toward marital status, attitude toward role of women, and attitude toward homemaking. Groups were working women, role-satisfied housewives, and role-dissatisfied housewives.

Of the original 465 cases, seven were dropped from analysis because of missing data. Missing data appeared to be randomly scattered throughout groups and predictors. Two additional cases were identified as multivariate outliers with $\underline{p} < .001$, and were also deleted. Both of the outlying cases were in the working group; they were women with extraordinarily favorable attitudes toward housework. For the remaining 456 cases (239 working women, 136 role-satisfied housewives, and 81 role-dissatisfied housewives), evaluation of assumptions of linearity, normality, multicollinearity or singularity, and homogeneity of variance-covariance matrices revealed no threat to multivariate analysis.

Two discriminant functions were calculated, with a combined $\chi^2(8) = 49.00$, $\underline{p} < .01$. After removal of the first function, there was still strong association between groups and predictors, $\chi^2(3) = 15.61$, $\underline{p} < .01$. The two discriminant functions accounted for 69% and 31%, respectively, of the between-group variability. [*χ^2 values and percent of variance are from Table 11.17; cf. Section 11.6.2.*] As shown in

Figure 11.3, the first discriminant function maximally separates role-satisfied housewives from the other two groups. The second discriminant function discriminates role-dissatisfied housewives from working women, with role satisfied housewives falling between these two groups.

The loading matrix of correlations between predictors and discriminant functions, as seen in Table 11.18, suggests that the best predictors for distinguishing between role-satisfied housewives and the other two groups (first function) are attitudes toward current marital status, toward women's role, and toward homemaking. Role-satisfied housewives have more favorable attitudes toward marital status (mean = 20.60) than working women (mean = 23.40) or role-dissatisfied housewives (mean = 25.62), and more conservative attitudes toward women's role (mean = 37.19) than working women (mean = 33.86) or dissatisfied housewives (mean = 35.67). Role-satisfied women are more favorable toward homemaking (mean = 22.51) than either working women (mean = 23.81) or role-dissatisfied housewives (mean = 24.93). [*Group means are shown in Table 11.15.*] Loadings less than .50 are not interpreted.

One predictor, attitudes toward women's role, has a loading in excess of .50 on the second discriminant function, which separates role-dissatisfied housewives from working women. Role-dissatisfied housewives have more conservative attitudes toward the role of women than working women (means have already been cited).

Three contrasts were performed where each group, in turn, was contrasted with the other two groups, pooled, to determine which predictors reliably separate each group from the other two groups. When working women were contrasted with the pooled groups of housewives, after adjustment for all other predictors, and keeping overall $\alpha < .05$ for the four predictors, only attitude toward women's role significantly separates working women from the other two groups, $\underline{F}(1, 450) = 16.87$. The squared semipartial correlation between working women vs. the pooled groups of housewives and attitudes toward women's role is .04. [*Equation 11.9 applied to output in Table 11.19.*] (Attitude toward women's role also separates working women from the other two groups without adjustment for the other predictors.)

Role-satisfied housewives differ from the other two groups on attitudes toward marital status, $\underline{F}(1, 450) = 9.66$, $\underline{p} < .05$. The squared semipartial correlation between this predictor and the role-satisfied group vs. the other two groups pooled is .02. (Attitudes toward marital status, toward the role of women, and toward housework separate role-satisfied housewives from the other two groups without adjustment for the other predictors.)

The group of role-dissatisfied housewives does not differ from the other two groups on any predictor after adjustment for all other predictors. Without adjustment, however, the group differs on attitudes toward marital status (less favorable) and attitudes toward homemaking (less favorable).

Pooled within-group correlations among the four predictors are shown in Table 11.18. [*Produced by run of Table 11.15 in section of output not shown called* WITHIN CORRELATION MATRIX.] Of the six correlations, four would show statistical significance at $\alpha = .01$ if tested individually. There is a small positive relationship between locus of control and attitude toward marital status, with $\underline{r}(454) = .17$, $\underline{p} < .01$, indicating that women who are more satisfied with their current marital status are less likely to attribute control of reinforcements to external sources. Attitude toward homemaking is positively correlated with locus of control, $\underline{r}(454) = .16$, $\underline{p} < .01$, and attitude toward marital status, $\underline{r}(454) = .28$, $\underline{p} < .01$, and negatively correlated with attitude toward women's role, $\underline{r}(454) = -.29$, $\underline{p} < .01$. This indicates that women with negative attitudes toward homemaking are likely to attribute control to external sources, to be dissatisfied with their current marital status, and to have more liberal attitudes toward women's role.

With the use of a jackknifed classification procedure for the total usable sample of 456 women, 244 (53.5%) were classified correctly, compared to 178.9 (39%) who would be correctly classified by chance alone. The 53.5% classification rate was achieved by classifying a disproportionate number of cases as working women. Although 52% of the women actually were employed, the classification scheme, using sample proportions as prior probabilities, classified 80.3% of the women as employed [*366/456 from jackknifed classification matrix in Table 11.15*]. This means that the working women were more likely to be correctly classified (84.9% correct classifications) than either the role-satisfied housewives (27.2% correct classifications) or the role-dissatisfied housewives (4.9% correct classifications).

The stability of the classification procedure was checked by a cross-validation run. Approximately 25% of the cases were withheld from calculation of the classification functions in this run. For the 75% of the cases from whom the functions were derived, there was a 52.8% correct classification rate. For the cross-validation cases, classification actually improved to 55%. This indicates a high degree of consistency in the classification scheme, and an unusual random division of cases into the cross-validation sample.

11.9 ■ SOME EXAMPLES FROM THE LITERATURE

Curtis and Simpson (1977) explored predictors for three types of drug users: daily opioid users, less-than-daily opioid users, and nonusers of opioids. A preliminary discriminant function analysis used 31 predictors related to demographic characteristics, alcohol use, and drug history to predict group membership. Both discriminant functions were statistically reliable and virtually all the predictors showed significant univariate *F*s for group differences, not surprising with the sample size greater than 23,000. For the second analysis, only the 13 predictors with univariate $F > 100$ were selected.

The first discriminant function accounted for 20% of the variance between groups, the second for 1% of the variance. The first discriminant function separated daily opioid users from the other two groups, presumably based on comparison of group centroids, although they were not reported. Correlations between predictors and the discriminant function were used to interpret the function. Daily opioid users were likely to be black, to be older, to have used illicit drugs for a longer period of time, to be responsible for slightly more family members, and to have begun daily use of illicit drugs with a drug other than marijuana.

A series of two-group stepwise discriminant functions was run by Strober and Weinberg (1977) on families that did or did not purchase various merchandise. Each item (e.g., dishwasher, hobby and recreational items, college education) was analyzed in a separate DISCRIM. The hypothesis was that predictors based on wife's employment status would be unrelated to purchase of the items considered in the study. Other predictors were family income, net family assets, stage of family life cycle, whether or not the family recently moved, and whether or not the family was in the market for the item being considered. In none of the analyses did predictors related to wife's employment status show significant *F* to enter the stepwise discriminant function. Different predictors were important for various expenditures. For example, for college education, color TV, and washer, life-cycle stage of the family was the first predictor to enter. For purchases of furniture, net family assets was the first predictor to enter.

For those analyses in which sample size was large enough, part of the sample was reserved for cross-validation. Percent correct classification for original sample and cross-validation sample ranged from 53% to 73%. In all except one analysis in which cross-validation was possible, the difference in correct prediction between the two samples was less than 2%. For hobby and recreational items, however, the classification coefficients did not generalize to the cross-validation sample.

Men involved in midcareer changes were studied by Wiener and Vaitenas (1977). Two personality inventories were used to distinguish midcareer changers from vocationally stable men. Separate discriminant function analyses were run for each inventory. The Edwards Personal Preference Schedule generated 15 predictors. The Gordon Personal Profile and Gordon Personal Inventory generated 8 predictors. For each analysis, Wilks' Lambda showed statistically significant association between groups and predictors, with discriminant functions accounting for 28% and 42% of the variance in the Edwards and Gordon analyses, respectively.

Interpretation was based on standardized discriminant function coefficients (rather than the less biased correlations of predictors with discriminant functions). Midcareer changers scored lower on endurance, dominance, and order scales on the Edwards inventory, and showed less responsibility and ascendancy as measured by the Gordon inventory.

Caffrey and Lile (1976) explored differences between psychology and physics students in attitudes toward science. Predictors were based on a scientific attitude questionnaire, in which the respondent expressed degree of agreement with 15 quotations from various writers on science.

Stepwise discriminant function analysis on upper-division psychology vs. physics majors followed preliminary analyses on a random sample of humanities, social science, and nature science majors.

Three items entered the stepwise analysis. Entering first was a statement suggesting that the results of scientific knowledge about behavior should be used to change behavior. Psychology majors were more likely to agree with this item than physics majors. Second, physics majors were more likely to agree with a statement suggesting that humans are free agents, not amenable to scientific prediction, and third, more likely to agree with a statement that Shakespeare conveys more truth about human nature than results of questionnaires.

Distinction between menopausal and nonmenopausal women in Hawaii was studied by Goodman, Steward, and Gilbert (1977). Predictors included medical, gynecological, and obstetrical history; age and age squared; physical measurements; and blood test results. Separate stepwise discriminant function analyses were run for Caucasian women, Japanese women, and then both groups pooled. Because only two groups were compared at a time, analyses were run through a stepwise multiple regression program. Preliminary runs with 35 predictors allowed elimination of nonsignificant predictors; then the stepwise analyses were rerun. In all analyses, age and age squared were forced into the stepwise series first because of known relationships with the remaining predictors.

Percent of variance accounted for in separating menopausal from nonmenopausal women was 45% and 36% for Caucasian and Japanese women, respectively. After adjustment for effects of age, the only predictor retained in the stepwise analysis for both races was "surgery related to a female disorder." For Caucasian women only, one additional predictor, medication, was selected in the stepwise analysis. For age, age squared, surgery, and medication, discriminant function coefficients and their standard errors were given.[15] Finally, predictors common to both races were selected and heterogeneity of coefficients between Japanese and Caucasian samples were tested.[16] No significant heterogeneity was found—that is, the two races did not produce significantly different discriminant functions. An alternative strategy for this test is a factorial discriminant function analysis, with race as one IV and menopausal vs. nonmenopausal as the other.

[15] Because this two-group analysis was run through a multiple regression program, discriminant function coefficients were produced as beta weights (cf. Chapter 5) and standard errors were available.

[16] The test for heterogeneity reported was that of Rao (1952) as applied by Goodman et al. (1974).