DOCUMENT RESUME

ED 075 496 TM 002 597

AUTHOR Huberty, Carl J.; Blommers, Paul J.
TITLE An Empirical Comparison of the Accuracy of Selected Multivariate Classification Rules.
PUB DATE 73
NOTE 42p.; Paper presented at American Educational Research Association Meeting (New Orleans, Louisiana, February 25-March 1, 1973)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Classification; *Comparative Analysis; Computer Science; Group Membership; Multiple Regression Analysis; *Predictor Variables; *Probability; Statistical Analysis

ABSTRACT

This study involved two phases: first when classification was based on the calibration sample, and second in a cross-validation setting. Computer-generated data were used. Results obtained from rules based on probabilities of group membership were compared for accuracy when classifying in the discriminant space and in the predictor variable spaces. In the first phase accuracy was greater in the predictor variable spaces, while the reverse was true in the second phase. In general, rules based on probabilities of group membership were approximately equally accurate, and more accurate than a rule related to a multiple regression analysis. Other findings are also discussed. (Author)

AN EMPIRICAL COMPARISON OF THE ACCURACY OF SELECTED MULTIVARIATE CLASSIFICATION RULES

Carl J Huberty
University of Georgia

Paul J. Blommers
University of Iowa

Abstract

This study involved two phases: first when classification was based on the calibration sample, and second in a cross-validation setting. Computer generated data were used. Results obtained from rules based on probabilities of group membership were compared for accuracy when classifying in the discriminant space and in the predictor variable spaces. In the first phase accuracy was greater in the predictor variable spaces, while the reverse was true in the second phase. In general, rules based on probabilities of group membership were approximately equally accurate, and more accurate than a rule related to a multiple regression analysis. Other findings are also discussed.

AN EMPIRICAL COMPARISON OF THE ACCURACY OF SELECTED MULTIVARIATE CLASSIFICATION RULES

It is sometimes of interest to a researcher to classify individuals or objects into one of several categories or classes on the basis of a set of observed measures. In so doing the researcher must choose an appropriate multivariate classification rule. From the investigator's viewpoint, appropriateness may be determined in terms of 1) having a sound theoretical basis, 2) being capable of accommodating more than two criterion populations, and 3) having been adapted, or readily adaptable, to computer programming. Many such multivariate classification decision rules have been proposed. This study dealt with five selected rules of identifying the population to which an individual may be assigned.

In a multivariate classification problem we begin with, say, k subsamples (or groups) of individuals, independently drawn from k well defined populations. There are, say, p measures available for each of the N individuals that comprise the total sample. The criterion employed in classifying an individual on the basis of the p measures depends upon the particular decision rule used. For each of the rules investigated, it is necessary to calculate k numerical values--one corresponding to each of the k groups--for every individual and then to classify according to the criterion of the rule employed.

In this study, two basic assumptions were made: 1) the k populations are p-variate normal, and 2) the k population covariance matrices are

identical.[1] A brief discussion of the rules under scrutiny in this investigation follows.

Rule 1. Assuming equal costs of misclassification and making the above two assumptions, a (linear) discriminant score of individual n for the gth population is (Rao, 1965, p. 488)

$$[1] \quad -\tfrac{1}{2} \log_e |\Sigma| - \tfrac{1}{2} (\underline{X}_n - \underline{\mu}_g)' \Sigma^{-1} (\underline{X}_n - \underline{\mu}_g) + \log_e \pi_g ,$$

where

Σ = the common population (p x p) covariance matrix,

$\underline{X}_n$ = the (p x 1) predictor score vector for individual n,

$\underline{\mu}_g$ = the (p x 1) centroid for population g, and

π_g = the probability that individual n (selected at random) is from population g.

Since the expression

$$-\tfrac{1}{2} \log_e |\Sigma| - \tfrac{1}{2} \underline{X}_n' \Sigma^{-1} \underline{X}_n$$

would be common to all k scores for a given individual, the equivalent of [1], for classification purposes, is

$$\underline{\mu}_g' \Sigma^{-1} \underline{X}_n - \tfrac{1}{2} \underline{\mu}_g' \Sigma^{-1} \underline{\mu}_g + \log_e \pi_g .$$

Substituting (maximum likelihood) estimates of Σ and $\underline{\mu}_g$ yields a discriminant score of individual n for group g:

$$[2] \qquad L_{ng} = \bar{\underline{X}}_g' C^{-1} \underline{X}_n - \tfrac{1}{2} \bar{\underline{X}}_g' C^{-1} \bar{\underline{X}}_g + \log_e \pi_g ,$$

where

$\bar{\underline{X}}_g$ = (p x 1) vector of observed mean scores for group g, and

C = the (p x p) pooled within-groups covariance matrix based on the sample of N individuals.

Here the elements of the matrix C are given by

$$c_{ij} = \frac{e_{ij}}{(N - k)}, \quad i, j = 1, \cdots, p,$$

where

$$e_{ij} = \sum_{g=1}^{k} \sum_{n=1}^{N_g} (X_{ign} - \bar{X}_{ig})(X_{jgn} - \bar{X}_{jg}),$$

with

N_g = the number of individuals in group g,

X_{ign} = the score on predictor variable i for individual n of group g, and

$\bar{X}_{ig}$ = the mean of predictor variable i for group g.

Of course, since sample estimates are used, an optimum solution cannot be claimed.

The first rule investigated may be stated[2] as

[3] $$R_1 : L_{ng} \geq L_{ng'}, \quad g, g' = 1, \ldots, k \; ; \; g \neq g' .$$

This is a maximum likelihood method of classification in the sense that it is equivalent (when population values are known) to the rule which assigns the individual with measurements $\underline{X}$ to that population for which the posterior probability of population membership has the highest value.

Rule 2. Another indirect application of the criterion of highest probability is as follows (see Cooley and Lohnes, 1971, p. 264). For

each individual n the values of k quadratic forms are determined; these quadratic forms are given by

$$[4] \qquad \chi^2_{ng} = \underset{\sim}{x}'_n \, D_g^{-1} \, \underset{\sim}{x}_n ,$$

where

D_g = the (p x p) covariance matrix for group g, and

$$\underset{\sim}{x}'_n = (X_{1n} - \bar{X}_{1g}, X_{2n} - \bar{X}_{2g}, \cdots, X_{pn} - \bar{X}_{pg}).$$

If sampling is from a p-variate normal population, the resulting values of [4] follow a chi-square distribution with p degrees of freedom (Rao, 1952, p. 55). Geometrically, the value of [4] is a measure of the closeness of individual n's profile point, $(X_{1n}, \ldots, X_{pn})$, in the p-space of group g to the centroid of group g. Since the tabled probability of a given chi-square value is the likelihood of obtaining a larger value, this probability may be considered the proportion of individual points that would be expected (in the long run) to lie beyond the locus of density (or "centour") on which this particular individual's point lies. The "centour score" for an individual, then, yields an estimate of the percent of individuals in a population that are further from the centroid. The higher the group "centour score" of an individual, the greater is his similarity to that group. Hence, this decision rule involves assigning individual n to that population for which his centour score is highest, or equivalently, to that population which yields the lowest chi-square sample value given by [4]. Symbolically, this rule may be stated as

[5] $$R_2 : \chi^2_{ng} \lesssim \chi^2_{ng'}, \quad g, g' = 1, \ldots, k \; ; \; g \neq g'.$$

Heterogeneity of dispersion enters into [5] in a somewhat undesirable way in that the greater the dispersion for a particular sample the greater the likelihood that individuals of unknown classification will be identified with it. Nor does [5] take into account prior probabilities, π_g, of population membership. Only if the sample dispersion matrices are identical and the prior probabilities are the same does [4] result in a minimum number of misclassifications. One way of adjusting for heterogeneity of dispersion and prior probability consists of modifying [5] (see Tatsuoka, 1971, pp. 222-225 for other modifications of [5]):

[6] $$\chi^2_{ng} \lesssim \chi^2_{ng'} - \log_e \frac{|D_g|}{|D_{g'}|} + 2 \log_e \frac{\pi_g}{\pi_{g'}} .$$

Rule 3. A more direct application of the classification criterion of highest posterior probability of population membership is the basis for the third decision rule (see Cooley and Lohnes, 1971, p. 267). Here a set of hypotheses regarding population membership of individual n is involved. One of the hypotheses is to be retained and the others rejected. The following notation for determining the likelihood of such an hypothesis is used:

[7] $$P(H_g \mid \underline{X}_n) , \quad g = 1, \ldots, k.$$

This denotes the probability of hypothesis g, given the score vector of individual n. Hypothesis g, H_g, states that individual n belongs to population g. For each individual there would be k such hypotheses, and that

hypothesis for which the likelihood is a maximum is selected.

The conditional probabilities of [7] can be computed from Bayes' formula, provided the prior probabilities are known or can be estimated. Let $P(\underset{\sim}{X}_n \mid H_g)$ denote the conditional probability of observing the score vector $\underset{\sim}{X}_n$, given that H_g is true; i.e., the probability that individual n selected at random from population g will have the particular combination of predictor scores, $\underset{\sim}{X}_n$ (technically, within a "small" neighborhood of $\underset{\sim}{X}_n$). Then the conditional probability that H_g is true, given that observation vector $\underset{\sim}{X}_n$ for individual n was obtained is

$$[8] \qquad P_{ng}(H_g \mid \underset{\sim}{X}_n) = \frac{\pi_g \, P(\underset{\sim}{X}_n \mid H_g)}{\sum_{g'=1}^{k} \pi_{g'} \, P(\underset{\sim}{X}_n \mid H_{g'})}, \quad g = 1, \cdots, k.$$

Under the assumption that the predictor variable vectors follow a p-variate normal distribution in each of the k populations, [8] may be expressed as

$$[9] \qquad P_{ng}(H_g \mid \underset{\sim}{X}_n) = \frac{\pi_g |D_g|^{-\frac{1}{2}} \exp(-\frac{1}{2}\chi^2_{ng})}{\sum_{g'=1}^{k} \pi_{g'} |D_{g'}|^{-\frac{1}{2}} \exp(-\frac{1}{2}\chi^2_{ng'})}, \quad g = 1, \ldots, k,$$

where $\chi^2_{ng'}$ is defined by [4]. Formula [9] gives the (posterior) probability that individual n with the score vector $\underset{\sim}{X}_n$, selected at random from the entire sample, will be a member of population g. Thus, a third classification rule is to assign an individual to that population for which his posterior probability of population membership is largest.

Such a rule may be expressed as

[10] $$R_3 : P_{ng} \geq P_{ng'} , \quad g, g' = 1 \cdots, k ; g \neq g'.$$

Since [9] utilizes information regarding both prior probabilities and differences in dispersions, [10] produces the same results as those obtained by using [6] (Huberty, 1969, p. 79).

Since the denominators of P_{ng} and $P_{ng'}$ for a given individual are identical, classification according to Rule 3 may be equivalently performed by assigning individual n to that population g for which the value of

$$\pi_g |D_g|^{-\frac{1}{2}} \exp(-\tfrac{1}{2} \chi^2_{ng})$$

is a maximum. And since the logarithm function is monotonic increasing, this is equivalent to maximizing

$$\log_e \pi_g - \tfrac{1}{2} \log_e |D_g| - \tfrac{1}{2}(\underline{X}_n - \bar{\underline{X}}_g)' D_g^{-1} (\underline{X}_n - \bar{\underline{X}}_g).$$

If D_g is replaced by Σ and $\bar{\underline{X}}_g$ by $\underline{\mu}_g$, then this is the same as maximizing [1]. That is, Rule 1 and Rule 3 are the same except for the sample covariance matrix used in computing the k values of each classification statistic.

It is of interest to note that according to Rule 2 and 3, classification is performed in the p-dimensional predictor variable spaces. To classify in the reduced or discriminant space, it is necessary to determine linear composites of the original scores, i.e., discriminant function values (see Cooley and Lohnes, 1971, Ch. 9), and to separate the new

space defined by these functions into k mutually exclusive regions. (These function values are not to be confused with the discriminant scores mentioned in relation to Rule 1.) To determine the dimension of the reduced space either the eigenvalues of the usual determinantal equation may be subjected to a significance test, or a subset of the non-zero eigenvalues that accounts for a large percent, say 90, of the discriminating power of the predictor variable may be chosen. It was decided, for this study, to employ Bartlett's test of significance (Rao, 1952, p. 373). (An α-level of .10 was used.) As Tatsuoka (1971, p. 233) points out, it is reasonable to confine attention to only those discriminant functions that are statistically significant so as to decrease the reliance on apparent differences among the criterion groups due to sampling error.

Rule 4. The formulation of the classificatory problem as conceived by Knutsen (1955) and Horst (1956a) involves finding separate regression equations contrasting each criterion group in turn with all others. In finding the regression equation corresponding to group g, the dichotomous criterion variable assumes the value 1 for individuals in group g and 0 otherwise. To determine the weights used in Horst's "least squares" multiple classification method, the standard procedure is followed. The least squares estimates of one set of population regression coefficients are given by

[11] $$\underline{b} = T^{-1} \underline{v} ,$$

where

$\underline{b}$ = the (p x 1) column vector of sample weights,

T = the (p x p) total sample deviation score cross-products matrix, and

$\underline{v}$ = the (p x 1) column vector of deviation score cross-products of the predictors and the (dichotomous) criterion, y -- the deviations being taken from the grand mean.

Here, the elements of T are given by

$$t_{ij} = \sum_{g=1}^{k} \sum_{n=1}^{N_g} (X_{ign} - \bar{X}_i)(X_{jgn} - \bar{X}_j), \; i,j = 1, \cdots, p,$$

where

$\bar{X}_i$ = the mean of the scores on predictor variable i for all N individuals.

The jth element of $\underline{v}$ for group g contrasted with the remaining groups is given by

$$\sum_{n=1}^{N} x_{jgn} y_{gn} = \sum_{n=1}^{N_g} X_{jgn} - \frac{N_g}{N} \sum_{n=1}^{N} X_{jgn}.$$

To generalize [11] to represent the k sets of weights, we write the (p x k) matrix, B, where

[12] $$B = T^{-1}V.$$

The gth column of B is the set of weights corresponding to the gth group, and the (p x k) matrix V is a similar extension of $\underline{v}$. The weights obtained from [12] are appropriate for use with deviation scores. If individuals are to be classified into one of several populations using raw score data, a correction term must be applied in order to achieve group-to-group compar-

..se of raw scores instead of deviation scores does not change the "slope" of the hyperplane but only the point of origin. Therefore, the regression weights would remain the same. The raw score formula for predicting membership of individual n in population g from the score vector

$$\underline{X}'_n = (X_{1n}, \cdots, X_{pn})$$

is

[13] $$Y_{ng} = \underline{b}'_g \underline{X}_n + c_g ,$$

where

$\underline{b}_g$ = the gth column of B, and

$c_g = \bar{Y}_g - \underline{b}'_g \bar{\underline{X}}.$

In the latter expression,

$\bar{Y}_g$ = the mean of the criterion measure which, when the gth group is the one considered, is simply $\frac{N_g}{N}$, and

$\bar{\underline{X}}$ = the (p x 1) vector of means of the p predictor variables for the total sample.

Classification according to this decision rule requires that for each individual, k composite (regression) scores be obtained (using equation [13]) which are his k predicted criterion values. Since an individual's actual criterion score is 1 or 0, depending upon his group membership, it follows that an individual n may logically be identified with that population for which his composite weighted score is nearest

to unity. Thus, we have

[14] $$R_4 : | Y_{ng} - 1.0 | \leq | Y_{ng'} - 1.0 |, \; g, g' = 1, \cdots, k; \; g \neq g'.$$

Rule 5. The last rule investigated is based upon posterior probability of population membership, as in Rule 3. However, the methods of calculating the posterior probabilities for the two rules differ. The formula used to compute the posterior probabilities involved in Rule 5 is based on "Case E. $\Sigma_g = \Sigma$ but unknown, $\underline{\mu}_g$ unknown," presented by Geisser (1966, p. 155). (See also, Cooley and Lohnes, 1971, p. 269.) Under the assumption of p-variate normality, Geisser obtains the "predictive density of a future observation (vector) given the available data" via a Bayesian approach. Geisser's density function is

[15] $$h(\underline{X}_n \mid H_g) \propto \left[\frac{N_g}{N_g + 1}\right]^{\frac{p}{2}} \left[1 + \frac{N_g \, (\underline{x}_n' \, C^{-1} \, \underline{x}_n)}{(N_g + 1)(N - k)}\right]^{\frac{-(N - k + 1)}{2}},$$

where all symbols have been previously defined. Thus, the probability that an individual n belongs to population g, given that he has a score vector $\underline{X}_n$ is by Bayes' formula,

[16] $$Q_{ng} \, (H_g \mid \underline{X}_n) = \frac{\pi_g \, h(\underline{X}_n | H_g)}{\sum_{g'=1}^{k} \pi_{g'} \, h \, (\underline{X}_n | H_{g'})}, \quad g = 1, \cdots, k.$$

Hence, the last rule used in this study may be stated as

[17] $$R_5 : \; Q_{ng} \geq Q_{ng'}, \; g, g' = 1, \cdots, k; \; g \neq g.$$

Of course, since for a given individual, the k values of the denominator of [16] are identical, individual n may be assigned to that population corresponding to the largest value of $\pi_g \; h \; (\underline{X}_n | H_g)$.

All five of these decision rules were judged to be appropriate according to the three criteria mentioned at the outset.

Procedure

To effect the simulation of drawing random samples of size N from k p-variate normal populations with a known common covariance matrix a highspeed electronic computer--IBM System 360, Model 65--was used. In this study the number of predictor variables considered was p = 10, and the numbers of criterion groups were k = 3 and k = 5.

To obtain the population covariance matrix, Σ, and the population mean matrix, M_{pop}, we proceeded as follows. The classical factor analysis model (Harman, 1967, p. 15) may be represented by

[18] $$z_j = \sum_{h=1}^{m} a_{jh} F_h + d_j U_j, \qquad j = 1, \cdots, 10,$$

where z_j, F_h, and U_j are standardized normal variates, $m < 10$, and a_{jh} and d_j are real-valued constants. Further assumptions underlying this model are

$$\text{Cov}\,(F_h, F_{h'}) = \text{Cov}\,(F_h, U_j) = \text{Cov}\,(U_j, U_{j'}) = 0,$$

where these covariances refer to population values. Hence,

[19] $$1 = \sigma_j^2 = E\,(z_j^2) = \sum_{h=1}^{m} a_{jh}^2 + d_j^2, \qquad j = 1, \cdots, 10,$$

and

[20] $$\rho_{jj'} = E(z_j z_{j'}) = \sum_{h=1}^{m} a_{jh} a_{j'h}, \quad j, j' = 1, \cdots, 10; j \neq j'.$$

Thus, Σ (10 x 10) is obtained using the relationship

[21] $$\Sigma = A_{pop} A'_{pop} + D^2_{pop},$$

where

A_{pop} = a (10 x m) matrix of elements a_{jh}, and

D_{pop} = a (10 x 10) diagonal matrix with diagonal elements d_j.

Equation [21] is the matrix expression for equations [19] and [20].

The communality ($\sum_{h=1}^{m} a^2_{jh}$) of each of the predictor variables was arbitrarily set at .75, thus making the reliability of variable z_j at least .75.[3] This condition gives a D_{pop} matrix with all diagonal elements equal to .50. Any A_{pop} matrix which is consistent with equations [19] and [20], under the constraint that $\sum_{h=1}^{m} a^2_{jh} = .75$, will suffice.

Separation between the k populations was accomplished by prescribing a (10 x k) population weight matrix:[4]

$$W_{pop} = \left[\begin{array}{ccc|cc} w_{11} & w_{12} & w_{13} & w_{14} & w_{15} \\ w_{21} & w_{22} & w_{23} & w_{24} & w_{25} \\ w_{31} & w_{32} & 0 & w_{34} & w_{35} \\ w_{41} & 0 & w_{43} & w_{44} & 0 \\ 0 & w_{52} & w_{53} & 0 & w_{55} \\ w_{61} & 0 & 0 & w_{64} & w_{65} \\ 0 & w_{72} & 0 & w_{74} & 0 \\ 0 & 0 & w_{83} & 0 & w_{85} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] .$$

From W_{pop} a (10 x k) population mean matrix, M_{pop}, was then obtained (see Footnote 4) by the relationship

$$M_{pop} = \Sigma W_{pop} .$$

The non-zero weights were chosen so that the population values, $\Delta^2_{gg'}$, of Mahalanobis' generalized distance between (the centroids of) two populations g and g', were "significant"; i.e., the significance test presented by Rao (1965, p. 480) would find the sample counterpart, $D^2_{gg'}$, significant[5] for the smallest value of N used in this study. These population distances are given by

[22] $\Delta^2_{gg'} = (\underline{\mu}_g - \underline{\mu}_{g'})' \Sigma^{-1} (\underline{\mu}_g - \underline{\mu}_{g'})$.

Corresponding to each total sample size, sample score matrices of size $(10 \times N_g)$ were generated from each of the k p-variate normal populations having the common covariance matrix, Σ. To generate these sample score matrices a procedure similar to that suggested by Kaiser and Dickman (1962) was employed. A number was selected from a uniform (0,1) distribution using a subroutine called RANDU, corresponding to which a number from a normal (0,1) "continuous" distribution was located. This technique was used to produce the elements of both an $(m \times N_g)$ matrix $\hat{F}$ and a $(10 \times N_g)$ matrix $\hat{U}$. The subsample score matrix corresponding to group g was then obtained (see equation [18]) using

$$X_g = A\hat{F} + D\hat{U} + M_g ,$$

where

X_g = the $(10 \times N_g)$ matrix of "observed" scores, and

M_g = the $(10 \times N_g)$ matrix, the ith row of which contains the (constant) value μ_{ig}.

Thus, in essence, random samples were selected on the orthogonal F and U matrices, and the observed scores were obtained by the above transformation.

There were two phases in the present study. In addition to the two assumptions mentioned at the outset, Phase I of the study was carried out under the following three restrictions: 3) the a priori probabilities of group membership, π_g, g = 1, ...k, are identical; 4) the number of individuals drawn from each population is the same; and 5) each sample of

individuals is classified on the basis of data for that sample. For both k-values the total sample sizes considered were N = 90, 150, 300 and 450. Because of the fourth restriction in this phase a random sample of size N was composed of k subsamples of size N/k each. In this phase sampling from each of the k populations was repeated 100 times for each value of N. Restriction 3) lead us to the "uniform ignorance" assumption of $\pi_g = 1/k$, $g = 1, \ldots, k$ (see Tatsuoka, 1971, pp. 225-226); hence the π-values called for in computing the probabilities in Rule 3 were deleted. Rule 5 was not included in Phase I because under conditions 3) and 4) it yields results that are identical to those of Rule 1 (Huberty, 1971).

The purpose of Phase II was to empirically investigate the accuracy of the five selected classification rules in a situation where conditions 3), 4), and 5) were either removed or modified in such a way as to make them more compatible with the "real world." In practical applications of multivariate classification theory, the number of individuals in each of the subsamples that represent the k populations of interest often is not the same. Hence, the k prior probabilities, estimates of which are usually based on the subsample sizes, can not be taken to be identical. Further, the investigator often desires to classify individuals whose proper classification has not been determined at the time when sampling took place. Thus, he seeks a relatively accurate classification rule to employ with individuals other than those in the "calibration sample." The proportion of misclassifications obtained by using the calibration or norming sample--the "apparent" error rate--tends to underestimate the "actual" error rate--the long run frequency of misclassifications using replication samples (Geisser, 1970).

Intuitively this seems reasonable since classificati[illegible] individuals in the norming sample [illegible] is based on a rule which is not only in some sense optimal for these particular individuals but is likely to capitalize favorably on chance sampling fluctuations. It is for these reasons that the situation of Phase I was altered by the removal or modification of conditions 3), 4), and 5).

In Phase II of the study the total sample sizes considered were N = 150, 300, 450, and 600 when three criterion groups were involved and N = 300, 450, 600 when k = 5. A random sample of size N was composed of k subsamples of size $N_1, \ldots, N_k$. For k = 3, the ratios of subsample sizes were arbitrarily set at 9:5:1. For k = 5, the ratios selected were 12:9:5:3:1.[6] The assignment of subsample size to each respective group was made by employing a table of random numbers. For example, with k = 5 and N = 450, the assignment was: $N_1 = 75$, $N_2 = 15$, $N_3 = 45$, $N_4 = 180$, and $N_5 = 135$.

Rather than apply the selected rules to the sample of "individuals" on which the classification statistics are determined, the necessary matrix calculations were performed on the first sample and classification according to each of the five[7] selected rules was carried out on the succeeding 100 samples. Thus the comparative accuracy of the rules was determined in a "cross-validation" setting.

Data Analysis

The criterion used to judge the accuracy of each of the five decision rules is the number of correct classifications over k groups. In analyzing the accuracy of population identification provided by the rules, a descriptive analysis was performed on the number of correct classifications over

all k groups for each value of N and each value of k. Frequency distributions of the number of correct classifications for each run of the experiment were obtained and comparisons were made of 1) the means and standard deviations of these distributions, and 2) the intercorrelations among the numbers of correct classifications obtained from each of the rules investigated.

An analysis of variance was employed in comparing the results of the rules statistically. The "treatments-by-subjects" design described by Lindquist (1953, Ch. 6) was applied[8] to the numbers of correct classifications to compare the results of Rules 2 and 3 in both the predictor variable spaces and the discriminant space. The "treatments" were the rules being investigated, while the "subjects" were the (100) runs. The following five hypotheses were tested:

(i) $H_0 : \mu_{2V} = \mu_{3V} = \mu_{3D}$,

(ii) $H_0 : \mu_{2V} = \mu_{2D}$,

(iii) $H_0 : \mu_{3V} = \mu_{3D}$,

(iv) $H_0 : \mu_{2?} = \mu_{3?}$, and

(v) $H_0 : \mu_1 = \mu_{2.5} = \mu_4 = \mu_5$,

where

μ_{iV} = the population mean of the numbers of correct classifications using Rule i (i = 2 or 3) in the predictor variable spaces,

μ_{iD} = the population mean of the numbers of correct classifications using Rule i (i = 2 or 3) in the discriminant space,

$\mu_{i?}$ = the population mean of the numbers of correct classifications using Rule i (i = 2 or 3) in the space determined by the results of the tests of hypotheses (ii) and (iii), and

$\mu_{2.5}$ = the population mean of the numbers of correct classifications for the rule and space that is expected to be "best" in accord with the results of the test of hypothesis (iv).

Hypotheses (ii) and (iii) were tested to determine if discriminating power is different when classifying in the discriminant space instead of in the predictor variable spaces. Of course, the test of hypothesis (iv) was only made if the first hypothesis was rejected. The procedure used to test hypotheses (ii), (iii), and (iv) is that of Scheffé (1959, p. 66). (As mentioned previously Rule 5 was not considered in Phase I of the study.)

Results

Many of the reported studies in the behavioral sciences that employ "multiple group discriminant analysis" were done for the purpose of assigning an individual to one of a finite number of populations to which he may belong. This assignment was made on the basis of a set of characteristics observed on the individual. Sometimes, especially in personnel work and career planning studies, discriminant analysis is used primarily as a means of analyzing group and variable differences; and the classification aspect of the analysis is considered as a by-product. The primary concern of the present study, however, was one of identification; that is, how accurately do different decision rules identify an individual as a member of one of

several criterion populations? Of course, one way to answer this question and, hence, compare the efficiency of these rules, is to determine for each rule the proportion of correct classifications of a fixed sample of subjects. For a given set of rules being investigated the "best" one may be defined as that one which consistently, over repeated sampling, produces the highest proportion of correct classifications.

To aid in the interpretation of the results the percents, means, and standard deviations based on chance expectations are included in the appropriate tables. The resulting values are based on the multinomial situation present here. The expected proportion of correct classifications for a given N-value across all k groups is

$$p = \frac{1}{N} \sum_{g=1}^{k} \pi_g N_g .$$

In this study π_g was taken to be N_g/N. Of course, in Phase I, $p = 1/k$. The expected number of correct classifications is given by Np; and the expected standard deviation of the distribution of correct classifications is $\sqrt{Np(1-p)}$.

The results of this study will be discussed separately for each phase and each k-value.

Phase I -- k = 3. The percents, means, and standard deviations of the numbers of correct classifications, as reported in Table 1, reflect the

- - - - - - - - - - - -

Insert Table 1 here

- - - - - - - - - - - -

efficiency of each of the four rules studied. The means given in Table 1 indicate that Rules 2V and 3V (classifying in the predictor variable spaces

according to Rules 2 and 3, respectively) identify population membership most accurately. This result was also evident from a comparison of the frequency distributions of the numbers of correct classifications for each rule. The distributions resulting from classification using Rules 2V and 3V were quite similar. The corresponding distributions for Rules 1, 2D, 3D, and 4 were also similar though markedly lower. This similarity was also apparent from the intercorrelations among the numbers of correct classifications according to the different rules. Hypothesis (i) was clearly rejected for all four values of N; the value of each mean square ratio (MSR) was at least 500. The differences indicated in hypotheses (ii) and (iii) were both highly significant--as judged by the Scheffé procedure--with the use of the predictor variable spaces being favored over the discriminant space. Since both hypotheses (ii) and (iii) were rejected for all N-values, the results of this phase of the study do not agree with those reported by Lohnes (1961). That is, a significant loss in discriminating power was observed when classifying according to Rules 2 and 3 in the discriminant space rather than in the predictor variable spaces.

The outcomes of the tests of hypotheses (ii) and (iii) indicate that hypothesis (iv) could be stated as

$$\text{(iv)} \quad H_0 : \mu_{2V} = \mu_{3V} .$$

In no case could this hypothesis be rejected. However, because in each case the observed mean of Rule 3V was greater than that of Rule 2V, the fifth hypothesis was stated as

$$\text{(v)} \quad H_0 : \mu_1 = \mu_{3V} = \mu_4 .$$

The values of the four MSRs in the testing of (v) were in no case less than 600. Hence, in each case hypothesis (v) was rejected; and Scheffé's method was again employed in making follow-up pairwise comparisons. Differences between Rule 3V means on the one hand and Rule 1 and Rule 4 means on the other were highly significant indicating that Rule 3V could be expected to produce (in the long run) the highest number of correct classifications. Further, it was found that the classification accuracies of Rules 1 and 4 were not significantly different. This result is not consistent with that found by Knutsen (1955), who concluded, from his single sample, that Rule 4 was more accurate than Rule 1.

Phase I -- k = 5. Except for the fact that the numbers of correct classifications were considerably larger, the results for five groups were very similar to those for the three-group case. As seen in Table 2 the reflected trends (from rule to rule and along N-values) are parallel to those results when k = 3. The outcomes of the tests of the five hypotheses of interest were also the same for the five-group situation as for the case of the three groups.

- - - - - - - - - - - -

Insert Table 2 here

- - - - - - - - - - - -

Phase II -- k = 3. The percents, means and standard deviations of the numbers of correct classifications are given in Table 3. These results indicate the superiority of Rules 1, 3D, and 5 in terms of accuracy of identification of population membership. That these three rules tend to produce similar degrees of classification accuracy is apparent from tables (not presented here) of the intercorrelations and frequency distributions

of the numbers of correct classification. The correlations among these rules ranged from .85 to .99. It was pointed out earlier that in a "true" sense Rules 1 and 3 are equivalent. That is, the classification statistics used are the same except for the sample covariance matrix used in computing the k values for each individual. Furthermore, it has been shown that when the N_g-values are the same, Rules 1 and 5 are equivalent. Thus, descrepancies among the results yielded by Rules 1, 3, and 5 are merely a function of the "goodness" of the sample estimates of Σ and of π_g , $g = 1, \ldots, k$.

Based on the analysis of variance results (no MSR was less than 200) hypothesis (i) can not be considered tenable in any of the four sample size cases. As judged by Scheffé's procedure, the differences indicated by hypotheses (ii) and (iii) were highly significant in all cases. Rule 2

- - - - - - - - - - - - -

Insert Table 3 here

- - - - - - - - - - - - -

in the predictor variable spaces produced a greater number of correct classifications than in the reduced or discriminant space for all N-values. Rule 3, on the other hand, appears to be more accurate for all N-values considered in this study if it is applied in the reduced space. This latter result is not consistent with that found in Phase I.

The outcomes of the tests of hypotheses (ii) and (iii) indicate that hypothesis (iv) can be stated as

$$\text{(iv)} \quad H_0 \ : \mu_{3D} = \mu_{2V} \ .$$

This hypothesis was rejected in all four situations, with Rule 3 (in the discriminant space) being favored for each sample size. This result lead to the following statement for hypothesis (v):

$$\text{(v)} \quad H_0 : \mu_1 = \mu_{3D} = \mu_4 = \mu_5 \ .$$

Since the MSRs resulting from the analysis of variance in the testing of this hypothesis were at least 700 for any N-value, these four rules were not judged to be equally accurate. The method of Scheffé was again employed in making the follow-up pairwise comparisons. Differences between Rule 4 means on the one hand and means of Rules 1, 3D, and 5 on the other were highly significant. In all situations the significant results indicated that Rule 4 would be expected to produce (in the long run) fewer correct classifications. Further, it was found that the classification accuracies of Rules 1 and 3D, 1 and 5, and 3D and 5 were, in each case, not significantly different. The results of the tests following hypothesis (v) are summarized in Table 4.

- - - - - - - - - - - -

Insert Table 4 here

- - - - - - - - - - - -

Phase II -- k = 5. As in the three-group case, Rules 1, 3D, and 5 produced the greatest number of correct classifications (see Table 5). Again hypothesis (i) was rejected in each case (all MSRs were at least 300), and of the four rules considered in this hypothesis results of applying Scheffé's procedure indicated Rule 3D as being the most accurate. Table 6 summarizes the results of the application of Scheffé's individual

comparison test following significant MSRs (all ratios were at least 1000) of the tests of hypothesis (v). Note that for N = 450 Rules 1 and 5 are more accurate than Rule 3 (in the discriminant space). For the other two N-values, the comparisons of Rule 1 versus Rule 3D and Rule 5 versus Rule 3D indicated no significant differences in classification accuracy; and Rules 1 and 5 were equally accurate for all N-values. Again, as in the case of k = 3, Rule 4 was found to be less accurate than Rules 1, 3D, and 5.

- - - - - - - - - - - - - - -

Insert Tables 5 & 6 here

- - - - - - - - - - - - - - -

It should be noted that sampling performed with different orderings of subsample sizes produced very similar results; that is, nearly the same rankings of the rules (with respect to classification accuracy) resulted. Also, it is of interest to note that the results of applying Rule 5 in the reduced space were almost identical to its application in the predictor variable spaces; the mean of the seven differences of mean number of correct classifications over both values of k (four for k = 3, three for k = 5) was .79 and the mean correlation between classification accuracy in the different spaces over both values of k was .95.

Discussion

Any remarks concerning the results of Phase I must necessarily be made in light of a situation characterized by the following restrictions: 1) the k criterion populations are p-variate normal, 2) the k population covariance matrices are identical, 3) the a priori probabilities of population membership are identical, 4) the number of individuals drawn from

each population is the same, and 5) each sample of individuals is classified on the basis of the sample data. It was judged that the population covariance (i.e., correlation) matrix used in this study was not atypical of those found in practice (Cochran, 1964, p. 186).

Perhaps one of the most striking findings of Phase I was the decreasing accuracy of classification with increasing N. One possible reason for this is the potential bias that may have been brought into the situation because of the use of each observed group covariance matrix rather than the pooled matrix and/or the sampling process itself. If a bias was present the larger sample sizes would reflect this more distinctly causing lesser accuracy. Another explanation which accounts in part for the outcome at hand is as follows. Let us consider Rule 2V in which the classification of individual n is based on the distance of his profile point to the centroid of each of the k groups. In the case of group g, this distance is equal to the mean distance between n's profile point and each of the profile points for the N_g individuals in the group.[9] In the event that individual n is a member of group g, one of the distances entering into the mean necessarily has a null value. The effect of this would be expected to be inversely related to the size of N_g. In a two-group case the "true" proportion of correct

classifications is easily estimable, and in the situation of this study is obtainable directly using $\Phi(\Delta/2)$ where Φ is the standard normal distribution function and Δ^2 is defined by [22] (Dunn, 1971). The simulation procedure was repeated exactly as in Phase I except for the consideration of only two criterion populations ($k = 2$). The proportion of correct classifications also decreased with increasing total sample size, and the proportion seemingly approached $\Phi(\Delta/2)$, which was .755 in this case. As pointed out by Glick (1972), computing the probability of correct classification for $k > 2$ involves evaluating an integral over a p-dimensional region, the form of which may vary with each rule.

That, for the given situation, all of the rules yielded results that were better than if the classifications were made at random is clear from a comparison with the chance expectations. A statistical test of "discriminatory power" is available (Press, 1972, pp. 381-383) but, because of the obvious outcomes, is not reported here.

The fact that the numbers of correct classifications were larger for $k = 5$ than for $k = 3$ does not necessarily imply that classification accuracy for comparable N-values improves as k increases. One reason for this result in this phase of the study may be that the separation of the populations (as measured by Mahalanobis' generalized distance function, Δ^2) in the $k = 5$ case was greater than in the case of $k = 3$. For $k = 3$ the mean Δ^2 was 1.55; for $k = 5$ the mean was 3.86. A closer examination of Tables 1 and 2 reveals that under the conditions of the present study the accuracy of classification from use of most of the rules appears to depend upon N_g rather than k. For example, the application of Rule 2 in the variable spaces yielded 81.1% accuracy for $k = 3$ and $N_g = 30$ versus 80.1% for $k = 5$ and $N_g = 30$; 74.0% for $k = 3$ and $N_g = 50$ versus 71.1% for $k = 5$ and $N_g = 60$; 67.7% for $k = 3$ and $N_g = 100$ versus 67.8% for $k = 5$ and $N_g = 90$. Such a conclusion would be valid if the distances between the population centroids were the same for both $k = 3$ and $k = 5$, but they were not.

The final remarks pertaining to the results of Phase I have to do with classification in the discriminant space as compared to classification based directly upon the original p variables. It is a fact that the results of classification based on Rules 2 and 3 (assuming identical population covariance matrices) are the same regardless of whether the original p variables or the k-1 discriminant functions are used in computing the chi-square values. This is true because the differences among the group centroids are exhausted by the k-1 discriminant functions; thus the distance between an individual and a group centroid would be the same

in the p-dimensional spaces of the total sample as in the discriminant space. So, first, differences in classification accuracy may be expected because of sampling fluctuations. It is important to note that with the use of each group covariance matrix, rather than the pooled covariance matrix, such fluctuations are indeed taken into consideration. Secondly, since some of the discriminant functions were considered "nonsignificant" the dimensionality of the discriminant space was sometimes less than k-1, particularly in the five-group case.

In Phase II of the study restrictions 3), 4), and 5) mentioned at the outset of this section were removed. The marked weakness of Rule 2D compared to Rule 3D in Phase II deserves some comment. The same chi-square values, based on the same number of discriminant functions, were operated on by both of these rules. Upon examining the results according to subsample size it was found that Rule 2D was as accurate or slightly more so than Rule 3D for smaller N_g-values, whereas Rule 3D yielded considerably better accuracy for larger N_g-values. That is, by taking into consideration the prior probabilities, as does Rule 3D, it was often found that the largest of the posterior probabilities was associated with the largest subsample size. That this may result is suggested by Tatsuoka (1971, p. 226).

A result of Phase II which differed from that of Phase I involves the accuracy of Rule 3 when classifying in the discriminant space versus classifying in the predictor variable spaces. In Phase II of the study classifying in the discriminant space yielded better results. It was conjectured that this superiority of accuracy is a matter of the

stability of the discriminant functions (over repeated sampling). By "stability" is meant that since a discriminant function is a linear composite of the predictor variables it would exhibit less sampling fluctuation over repeated sampling than the single variables which enter into the composite. However, an important consideration here is that the classification statistics used in Rule 3V are based only on the N_g individuals in group g (to determine D_g^{-1}) whereas the statistics used in Rule 3D are based on the total sample of individuals (to determine the discriminant functions).

To summarize, then, in terms of classification accuracy it can be expected that, under the conditions of either Phase I or Phase II, Rule 2 applied in the discriminant space and Rule 4 will not in the long run do as well as the other rules. Since a true classification problem is one which involves assigning as yet "unlabeled" individuals to one of a number of well-defined criterion populations it may be well to only consider classification accuracy of rules applied in a cross-validation setting. (However, a proportion of correct classifications when based on a "large" calibration sample may be used as an index of discriminatory power of a set of predictors.) In this sense it was found in Phase II that Rule 1, 3D, and 5 can be expected to yield the largest, and nearly the same, number of correct classifications.

REFERENCES

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and a correlation between errors in the two-way classification. Annals of Mathematical Statistics, 1954, 25, 484-498.

Cochran, W. G. On the performance of the linear discriminant function. Technometrics, 1964, 6, 179-190.

Cooley, W. W., and Lohnes, P. R. Multivariate data analysis. New York: Wiley, 1971.

Dunn, O. J. Some expected values for probabilities of correct classifications in discriminant analysis. Technometrics, 1971, 13, 345-353.

Geisser, S. Predictive discrimination, in Krishnaiah, P. R. (Ed.), Proceedings of the international symposium on multivariate analysis. New York: Academic Press, 1966, 149-163.

Geisser, S. Discriminatory practices, in Meyer, D. L. and Collier, R. O. (Eds.), Bayesian statistics. Itasca, Illinois: Peacock, 1970, 57-70.

Glick, N. Sample-based classification procedures derived from density estimates. Journal of American Statistical Association, 1972, 67, 116-122.

Harman, H. H. Modern factor analysis. Chicago: University of Chicago Press, 1967.

Horst, P. Multiple classification by the method of least squares. Journal of Clinical Psychology, 1956a, 12, 3-16.

Horst, P. Least square multiple classification for unequal subgroups, Journal of Clinical Psychology, 1956b, 12, 309-315.

Huberty, C. J An empirical comparison of selected classification rules in multiple group discriminant analysis. Unpublished doctoral dissertation, University of Iowa, 1969.

Huberty, C. J The equivalence of two multivariate classification schemes. Multivariate Behavioral Research, 1971, 6, 243-246.

Kaiser, H. F., and Dickman, K. Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. Psychometrika, 1962, 27, 179-182.

Knutsen, M. An empirical comparison of the linear discriminant function and multiple regression techniques in the classifying of subjects into three categories. Unpublished doctoral dissertation, University of Washington, 1955.

Linquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.

Lohnes, P. R. Test space and discriminant space classification models and related significance tests. Educational and Psychological Measurement, 1961, 12, 322-327.

Press, S. J. Applied multivariate analysis. New York: Holt, Rinehart and Winston, 1972.

Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley, 1952.

Rao, C. R. Linear statistical inference and its application. New York: Wiley, 1965.

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

Tatsuoka, M. M. Multivariate analysis. New York: Wiley, 1971.

FOOTNOTES

[1]It is recognized that an alternative multivariate model is sometimes considered; when a "total population" concept makes sense, we are dealing with k "subpopulations." The difference may bear on interpretation of results.

[2]This is read as follows: "Assign individual n to population g if his discriminant score, L_{ng}, corresponding to population g is largest."

[3]Although the underlying statistical model of discriminant analysis assumes no errors of measurement, in this study the predictor variables are allowed to be fallible in that each variable has a maximum error variance of .25. Allowing for non-zero unreliability is in agreement with most studies involving psychometric measurement.

[4]The choice of this array of zero and non-zero weights was made in view of a study which treats the variable selection problem. The A_{pop} and Σ matrices, the routine employed in arriving at A_{pop}, and the W_{pop} and M_{pop} matrices have been deposited with the National Auxiliary Publications Service. Order Document No. from National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc. 22 West 34th Street, New York, New York 10001. Remit in Advance $ for photocopies or $ for microfiche and make checks payable to: Research and Microfilm Publications, Inc.

[5]A low significance level (α = .10) was employed so as to subject the classification procedures to a relatively severe test of efficiency. If the separation of the populations is great, many classificatory schemes would appear to be equivalently efficient. Also, it was of interest to keep the probability of making a Type II error--claiming the groups are not significantly different when they actually are--low; hence, a large α-value was used.

[6]The smallest total sample size that could be used with k = 3 was 150 and with k = 5 was 300 since the given ratios require one subsample of size 10; with p = 10 a smaller number of "individuals" would produce a dispersion matrix, D, whose determinant is zero (since the rank of D would be less than 10).

[7] A modification of the least squares technique used in arriving at the classification statistic for Rule 4 was required because of the unequal subsample sizes employed in Phase II (see Horst, 1956b).

[8] The conservative test suggested by Box (1954) was employed in all situations.

[9] This stems from the fact that

$$\underline{X}_n - \bar{\underline{X}}_g = \frac{1}{N_g}\left[(\underline{X}_n - \underline{X}_{1g}) + (\underline{X}_n - \underline{X}_{2g}) + \cdots + (\underline{X}_n - \underline{X}_{N_g g})\right].$$

TABLE 1

Phase I

Percents, Means, Standard Deviations of the Number of Correct Classifications for k = 3[a]

	Rule						
	1	2V	2D	3V	3D	4	Chance Expectations
N = 90	66.6% 59.96 (4.44)	81.1% 73.04 (3.57)	66.0% 59.44 (5.31)	83.0% 74.70 (3.54)	66.9% 60.19 (4.98)	66.2% 59.60 (4.46)	33.3% 30 (4.47)
N = 150	63.4% 95.06 (5.87)	74.0% 110.93 (6.28)	63.5% 95.19 (5.48)	75.0% 112.46 (5.92)	63.9% 95.79 (6.17)	63.3% 94.94 (5.73)	33.3% 50 (5.77)
N = 300	61.7% 185.16 (7.74)	67.7% 203.21 (8.69)	61.9% 185.66 (7.86)	68.4% 205.34 (9.02)	61.8% 185.31 (7.67)	61.6% 184.90 (7.50)	33.3% 100 (8.16)
N = 450	61.2% 275.46 (10.03)	65.2% 293.59 (11.11)	61.2% 275.35 (10.12)	65.8% 296.31 (10.59)	61.4% 276.26 (10.16)	61.2% 275.40 (10.09)	33.3% 150 (10.00)

[a]Standard deviations are given in parentheses.

TABLE 2

Phase I

Percents, Means, and Standard Deviations of the Number of Correct Classifications for k = 5[a]

	Rule						
	1	2V	2D	3V	3D	4	Chance Expectations
N = 90	69.1% 62.20 (4.53)	89.4% 80.50 (3.25)	71.0% 63.93 (5.22)	91.0% 81.86 (2.93)	72.4% 65.16 (4.79)	68.4% 61.53 (4.50)	20.0% 18 (3.79)
N = 150	66.9% 100.38 (5.28)	80.1% 120.18 (4.73)	68.5% 102.69 (5.68)	81.2% 121.87 (4.40)	68.9% 103.41 (5.18)	66.1% 99.19 (5.12)	20.0% 30 (4.90)
N = 300	63.6% 190.67 (8.26)	71.1% 213.41 (7.98)	64.2% 192.64 (8.34)	71.8% 215.36 (7.66)	64.6% 193.72 (8.34)	63.2% 189.52 (8.66)	20.0% 60 (6.00)
N = 450	62.8% 282.67 (10.17)	67.8% 304.91 (10.43)	63.4% 285.40 (10.20)	68.3% 307.43 (9.97)	63.7% 286.49 (10.16)	62.5% 281.05 (9.71)	20.0% 90 (8.48)

[a]Standard Deviations are given in parentheses.

TABLE 3

Phase II

Percents, Means, and Standard Deviations of the Number of Correct Classifications for k = 3[a]

	Rule							
	1	2V	2D	3V	3D	4	5	Chance Expectations
N = 150	66.2% 99.35 (5.36)	62.3% 93.50 (5.29)	57.1% 85.59 (6.81)	63.5% 95.26 (4.84)	65.8% 98.72 (5.12)	56.0% 84.00 (6.48)	66.6% 99.90 (4.99)	47.6% 71 (6.12)
N = 300	68.3% 204.79 (6.67)	64.0% 192.01 (8.13)	56.8% 170.35 (9.08)	64.9% 194.81 (7.75)	68.2% 204.71 (6.39)	56.9% 170.77 (8.64)	68.2% 204.58 (6.51)	47.6% 143 (8.65)
N = 450	69.6% 313.17 (9.01)	63.8% 287.04 (9.64)	58.5% 263.38 (10.85)	67.8% 304.93 (7.45)	69.6% 313.11 (8.72)	59.7% 268.83 (10.26)	69.6% 313.16 (8.85)	47.6% 214 (10.59)
N = 600	69.8% 418.65 (10.23)	64.2% 385.02 (11.06)	62.0% 372.22 (11.72)	67.8% 406.65 (10.89)	69.8% 418.75 (10.34)	59.3% 355.74 (12.50)	69.8% 418.95 (10.13)	47.6% 285 (12.23)

[a]Standard deviations are given in parentheses.

TABLE 4

Mean Differences and Scheffé

Critical Values for k = 3

	$\bar{X}_1-\bar{X}_{3D}$	$\bar{X}_1-\bar{X}_4$	$\bar{X}_5-\bar{X}_1$	$\bar{X}_{3D}-\bar{X}_4$	$\bar{X}_5-\bar{X}_{3D}$	$\bar{X}_5-\bar{X}_4$	Critical Value[a]
N = 150	.63	15.35	.55	14.72	1.18	15.90	1.84
N = 300	.08	34.02	-.21	33.94	-.13	33.81	3.05
N = 450	.06	44.34	-.01	44.28	.05	44.33	3.02
N = 600	-.10	62.91	.30	63.01	.20	63.21	3.83

[a]Based on $_{.99}F_{1,99}$

TABLE 5

Phase II

Percents, Means, and Standard Deviations of the Number of Correct Classifications for k = 5[a]

	Rule							
	1	2V	2D	3V	3D	4	5	Chance Expectations
N = 300	65.9% 197.60 (7.65)	60.4% 181.26 (7.16)	59.0% 176.93 (6.73)	60.8% 182.33 (7.05)	65.3% 195.95 (7.15)	54.6% 163.77 (8.38)	66.0% 197.88 (7.45)	28.8% 87 (7.85)
N = 450	65.5% 294.82 (10.34)	58.4% 262.77 (10.86)	56.6% 254.84 (11.56)	61.1% 274.92 (8.76)	64.8% 291.44 (9.92)	58.4% 262.90 (10.84)	65.6% 295.05 (10.13)	28.8% 130 (9.61)
N = 600	65.0% 390.04 (11.51)	58.6% 351.57 (13.38)	59.9% 359.34 (11.78)	61.8% 370.80 (13.07)	64.6% 387.34 (11.47)	54.5% 326.83 (11.68)	65.0% 389.99 (11.50)	28.8% 173 (11.10)

[a]Standard deviations are given in parentheses.

TABLE 6

Mean Differences and Scheffé´
Critical Values for k = 5

	$\bar{X}_1-\bar{X}_{3D}$	$\bar{X}_1-\bar{X}_4$	$\bar{X}_5-\bar{X}_1$	$\bar{X}_{3D}-\bar{X}_4$	$\bar{X}_5-\bar{X}_{3D}$	$\bar{X}_5-\bar{X}_4$	Critical Value[a]
N = 300	1.65	33.83	.28	32.18	1.93	34.11	2.63
N = 450	3.38	31.92	.23	28.54	3.61	32.15	3.14
N = 600	2.70	63.21	-.05	60.51	2.65	63.16	3.78

[a]Based on $_{.99}F_{1,99}$