DOCUMENT RESUME

ED 422 391 TM 028 953

AUTHOR	Fan, Xitao; Wang, Lin
TITLE	Comparing Linear Discriminant Function with Logistic Regression for the Two-Group Classification Problem.
PUB DATE	1998-04-00
NOTE	42p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Classification; Comparative Analysis; Monte Carlo Methods; *Probability; Simulation
IDENTIFIERS	*Linear Discriminant Function; *Logistic Regression; Predictive Discriminant Analysis

ABSTRACT

The Monte Carlo study compared the performance of predictive discriminant analysis (PDA) and that of logistic regression (LR) for the two-group classification problem. Prior probabilities were used for classification, but the cost of misclassification was assumed to be equal. The study used a fully crossed three-factor experimental design (with 200 replications in each cell) manipulating sample size, prior probabilities, and equal/unequal covariance matrices. Two data patterns were simulated to provide a replication mechanism within the study. The major findings are: (1) PDA and LR have comparable performance for two groups with equal prior probabilities; and (2) for two groups with unequal prior probabilities, LR minimizes the error rate for the smaller group, and PDA minimizes the error rate of the larger and total sample. Consistency was observed across the two data patterns. The findings reveal a picture of PDA and LR that seems to be more complicated than that typically portrayed in the literature. Limitations of the study are noted, and future directions are suggested. (Contains 2 figures, 5 tables, and 29 references.) (Author/SLD)

PDA_LR.DF1

COMPARING LINEAR DISCRIMINANT FUNCTION WITH LOGISTIC REGRESSION FOR THE TWO-GROUP CLASSIFICATION PROBLEM

Xitao Fan

Utah State University

Xitao Fan
Department of Psychology
Utah State University
Logan, UT 84322-2810
Phone: (801)797-1451
E-Mail: fafan@cc.usu.edu

Lin Wang

American College Testing

Running Head: Classification by PDA and LR

Paper presented at the 1998 Annual Meeting of the American Educational Research Association, April 13-17, San Diego (*Session # 27.51*).

ABSTRACT

This Monte Carlo study compared the performance of predictive discriminant analysis (PDA) and that of logistic regression (LR) for the two-group classification problem. Prior probabilities were used for classification, but the cost of misclassification was assumed to be equal. The study used a fully crossed three-factor experimental design (with 200 replications in each cell): sample size, prior probabilities, and equal/unequal covariance matrices. Two data patterns were simulated to provide a replication mechanism within the study. The major findings are: 1) PDA and LR have comparable performance for two groups with equal prior probabilities; 2)for two groups with unequal prior probabilities, LR minimizes the error rate for the smaller group, and PDA minimizes the error rate of the larger and the total sample. Consistency was observed across the two data patterns. The findings reveal a picture about PDA and LR which seems to be more complicated than typically portrayed in the literature. Limitations of the study were noted, and future directions were suggested.

In social and behavioral sciences in general, and in education (e.g., for the problem of school dropout) and psychology (e.g., for identifying those with certain pathological symptoms) in particular, there is often the need to classify individuals into different groups, or to predict an individual's group membership, based on a battery of measurements. Both discriminant analysis and logistic regression have been the popular statistical tools for this purpose (Yarnold, Hart, & Soltysik, 1994). The relative efficacy of these two statistical methods under different data conditions, however, has been an issue of debate (e.g., Barón, 1991; Dattalo, 1994; Dey & Astin, 1993). Prior to exploring the relevant issues in some detail, some readers may appreciate a brief review of the two statistical methods. Additional details about these methods are provided elsewhere (cf. Hosmer, 1989; Huberty, 1994).

Brief Review of the Two Methods

Predictive Discriminant Analysis for Two Groups

As discussed by Huberty (1994), in social and behavioral science research, discriminant analysis (DA) is often used for two purposes: to describe major group differences (descriptive discriminant analysis, DDA), and to classify subjects into groups, i.e., to predict subjects' group membership (predictive discriminant analysis, PDA). In DDA, the researcher is primarily interested in gaining insights about how the variables explain the group differences. In PDA, the primary interest is in how accurately subjects can be classified into different groups based on a set of measurements. This study focuses on PDA, and its application for the two-group problem.

Suppose multiple measurements ($\underline{X}$: $\underline{X}_1$, $\underline{X}_2$,..., $\underline{X}_p$) are taken for two populations: π_1 and π_2. The multiple measurements are from joint multivariate normal distributions, with population parameters being ($\underline{\mu}_1, \underline{\Sigma}$) and ($\underline{\mu}_2, \underline{\Sigma}$) respectively for the two populations. In other words, the two populations have different population means ($\underline{\mu}_1$, $\underline{\mu}_2$) on the multiple measurements, but they have the common covariance matrix ($\underline{\Sigma}$). For these two populations, a function, $\underline{Y}$, can be formed by linearly combining the original multiple measurements $\underline{X}$ as follows:

$$Y = \mathbf{a}'\mathbf{X} = a_1X_1 + a_2X_2 + ... + a_nX_p$$

If we set the linear coefficients to the following:

$$\mathbf{a}' = (\mu_1 - \mu_2)' \Sigma^{-1} \qquad (1)$$

Then we have the linear composite $\underline{Y}$:

$$Y = \mathbf{a}'X = (\mu_1 - \mu_2)' \Sigma^{-1} X \qquad (2)$$

The linear function (2) above is known as <u>Fisher's linear discriminant function</u>, and the vector $\underline{\mathbf{a}}'$ contains the <u>discriminant function coefficients</u> which combine the original measurements $\underline{X}$ into the linear composite $\underline{Y}$. The most important characteristic of this linear function is that, the ratio of between-group variance to within-group variance on this function $\underline{Y}$ is maximized (Johnson & Wichern, 1988; Kshirsagar, 1972). In essence, the Fisher's linear discriminant function translates the two <u>multivariate</u> populations (π_1 and π_2) into two <u>univariate</u> populations, and the two univariate population means are maximally separated relative to the within-group population variance on the linear composite $\underline{Y}$.

Because maximum separation between the two population means is achieved on $\underline{Y}$, this linear discrimination function can be used for classification. For this purpose, we need to find the midpoint between the two population means on $\underline{Y}$. Because the means of populations π_1 and π_2 on $\underline{Y}$ are:

$$\mu_{1Y} = \mathbf{a}'\boldsymbol{\mu}_1, \quad \mu_{2Y} = \mathbf{a}'\boldsymbol{\mu}_2,$$

The midpoint (m) between the two population means on $\underline{Y}$ is:

$$\begin{aligned} m &= \tfrac{1}{2}(\mu_{1Y} + \mu_{2Y}) \\ &= \tfrac{1}{2}(\mathbf{a}'\boldsymbol{\mu}_1 + \mathbf{a}'\boldsymbol{\mu}_2) \\ &= \tfrac{1}{2}\mathbf{a}'(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \end{aligned} \qquad (3)$$

Once this midpoint is identified, classification for new observations is straightforward and easily implemented as follows (Huberty, 1994; Johnson & Wichern, 1988):

For a new observation with measurements $\mathbf{x}_i$,

Classify $\mathbf{x}_i$ to population π_1, if $y_i = \mathbf{a}'\mathbf{x}_i \geq m$

Classify $\mathbf{x}_i$ to population π_2, if $y_i = \mathbf{a}'\mathbf{x}_i < m$ (4)

where $\underline{\mathbf{a}}'$ is defined in (1). Alternatively, the classification rule above can be expressed as:

Classify $\mathbf{x}_i$ to population π_1, if $y_i = \mathbf{a}'\mathbf{x}_i - m \geq 0$

Classify $\mathbf{x}_i$ to population π_2, if $y_i = \mathbf{a}'\mathbf{x}_i - m < 0$ (5)

This classification rule essentially says that if the linear composite score of the new observation is closer to the mean composite score for Population 1, classify this observation to Group 1; otherwise classify this observation into Group 2 (Huberty, 1994, p. 138). The classification rule in (5), however, assumes both equal prior probabilities (equal proportions) of the two populations π_1 and π_2, and equal cost of misclassification for the two populations. When the prior probabilities of the two populations and the cost of misclassification are not equal for the two populations, a classification rule should take these two factors into account in order to achieve optimal results (Johnson & Wichern, 1988).

In many research situations in education and psychology, prior probabilities are actually far from being equal (e.g., to predict school dropouts vs. graduates; to classify subjects into a normal group vs. a pathological group). In the same vein, there are many situations in which the consequences of misclassification for the two populations is quite different. A linear classification rule which takes into consideration of both unequal prior probabilities and unequal cost of misclassification for the two populations is sometimes known as the Anderson's classification function (Johnson & Wichern, 1988), and this function takes the form:

Classify $\mathbf{x}_i$ to population π_1, if

$$a'x_i - m \geq \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \quad (6)$$

Otherwise, classify $\mathbf{x}_i$ into population π_2.

In the classification function above, $c(1|2)$ is the cost of misclassifying a π_2 member into π_1, and $c(2|1)$ is the cost of misclassifying a π_1 member into π_2. p_2 is the prior probability of π_2, and p_1 is the prior probability of π_1. It is easy to see that, if the cost of misclassification is equal for the two populations, and the prior probability for the two populations is the same, the right side of the equation becomes ln[1]=0, thus (6) becomes (5). In other words, this Anderson classification function is a more general classification function which subsumes the classification rule (5). In the present study, the cost of misclassification for the two populations is assumed to be equal, but the prior probabilities for the two populations may differ. So the classification rule used in the present study of PDA is the following:

Classify $\mathbf{x}_i$ to population π_1, if $a'x_i - m \geq \ln(\frac{p_2}{p_1})$, (7)

otherwise, classify $\mathbf{x}_i$ into population π_2.

Readers may have noticed that all the formulas presented so far involve population parameters only. In real research situations, a researcher only has the sample statistics. For sample classification rules comparable to all those presented above, simply substitute $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$ (sample mean vectors) for $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and substitute $\mathbf{S}_{pooled}$ (pooled sample covariance matrix) for $\boldsymbol{\Sigma}$.

Logistic Regression

Given two populations with group membership as a dichotomous variable, the problem of classification can also be accomplished through logistic regression (LR). While discriminant analysis is part of the general linear model (GLM) (Knapp, 1978; Fan, 1996; Thompson, 1991), logistic regression is not, because it models the nonlinear probabilistic function of the dichotomous variable (Neter, Wasserman, & Kutner, 1989). A graphic example of such a function is presented in Figure 1 for the case of one predictor variable. It should be noted that, because this probabilistic function is nonlinear, the increment on the function $\underline{Y}$ associated with a unit increase in the independent variable $\underline{X}$ will not be constant across all the ranges of $\underline{X}$ (Cleary & Angel, 1984).

Insert Figure 1 about here

Given a binary (dichotomous) outcome variable $\underline{Y}$ (Y=1,2), such as group membership in a two-group situation, and a battary of measurements on the set of continuous variables $\mathbf{X}$ ($\mathbf{X}$: $\mathbf{X}_1$, $\mathbf{X}_2$,..., $\mathbf{X}_p$), the probability of belonging to one group (e.g., Y=2) can be modeled through:

$$Y = \frac{e^{(\beta' X)}}{1 + e^{(\beta' X)}} \tag{7}$$

where $\beta' X = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$, and $\underline{Y}$ is the probability of an observation belonging to Group 2. Alternatively, (7) can be expressed as:

$$\log_e \left(\frac{Y}{1 - Y}\right) = \beta' X = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \tag{8}$$

While the estimation of linear discriminant function parameters ($\hat{\mathbf{a}}'$ in Equation 2)can be accomplished analytically based on ordinary least squares procedures, the estimation of logistic regression model parameters ($\hat{\beta}'$ in Equation 7) cannot be obtained analytically, because there is no closed form solution. Consequently, maximum likelihood estimators for logistic regression model are obtained iteratively, which requires much more intensive computation than the least squares procedure for linear discriminant function. This has been considered as a practical disadvantage of LR by some researchers (e.g., Cleary & Angel, 1984). But with the ever-increasing computing power available to almost all researchers nowadays, the relevance of this argument is probably diminishing rapidly.

Once the logistic regression model (7) is established, i.e., the parameters in the model are properly estimated ($\hat{\underline{\beta}}'$), the model is frequently used for making predictions for new observations. Predicting a binary outcome (e.g., group membership for two groups) for an observation with $\underline{\mathbf{X}}_i$ is straightforward: classify $\underline{\mathbf{X}}_i$ into Group 2 if the predicted probability is large, and classify the observation into Group 1 if the predicted probability is small.

The problem is to determine the cutoff point for the predicted probability above which $\underline{\mathbf{X}}_i$ will be classified into Group 2, and below which $\underline{\mathbf{X}}_i$ will be classified into Group 1. In many situations, when the two groups are approximately equal in terms of their population proportions, 0.5 is often chosen as the cutoff point. When information about the prior probabilities of the groups is available, such information should be used in classification. For example, if in a student population, 20% of students require remedial education for passing the minimum competency test (Group 2), and the other 80% do not (Group 2), this information of prior probabilities can be used to set the cutoff point in the prediction rule (see Neter, Wasserman, & Kutner, 1989, pp. 609-611, for more discussion on this topic).

In addition to the prior probabilities, the cost of misclassification for the two groups should also be considered in the classification rule. In the present study, equal cost of misclassification is assumed for the two groups, and this issue is not explored. Readers interested in this issue should consult other sources (e.g., Johnson & Wichern, 1988).

Issues in Comparing PDA and LR for Classification Accuracy

Since both PDA and LR can be used for predicting or classifying individuals into different groups based on a set of measurements, a logical question often asked is: how do the two techniques compare with each other? In the literature, there has been quite some discussion about the relative merits of these two different techniques (e.g., Dattalo, 1994; Fraser, Jensen, Kiefer, & Popuang, 1994; Wilson & Hardgrave, 1995).

Theoretically, PDA is considered as having more stringent data assumptions. Two prominent assumptions for PDA are multivariate normality of data, and homogeneity of the covariance matrices of the groups (Johnson & Wichern, 1988; Stevens, 1996). However, it is not entirely clear what consequences the violation of these assumptions have on PDA analysis results. LR, on the other hand, is considered relatively free of these stringent data assumptions (Cox & Snell, 1989; Neter, et al., 1989; Tabachnick & Fidell, 1996). Although there is no strong logical reason to expect the superiority of one technique over the other in classification accuracy when the assumptions for PDA hold, it would be reasonable to expect that LR should have the upper hand when some of these assumptions for PDA are not tenable (Neter, et al., 1989; Tabachnick & Fidell, 1996).

Research findings about the relative performance of these two methods appear to be inconsistent. With regard to data normality, Efron (1975) showed that under the optimal data condition of multivariate normality and equal covariance matrices for the groups, linear discriminant function is more economical and more efficient than logistic regression. When the data are not multivariate normal, results from some simulation studies (e.g., Barón, 1991; Bayne, Beauchamp, Kane, & McCabe, 1984) indicated that LR performed better than PDA. This finding, however, has not been unequivocally supported by the studies which compared the two techniques by using extant data sets, because quite a few studies involving actual nonnormal data sets suggested very little practical difference between the two techniques (e.g., Cleary & Angel, 1984; Dey & Astin, 1993; Meshbane & Morris, 1996).

With regard to the condition of equal covariance matrices for PDA, there appears to be a lack of empirical studies to compare the relative performance of PDA and LR when this condition does not hold. Researchers seem to assume that LR should be the method of choice when the two groups do not have equal covariance matrices (Harrell & Lee, 1985; Press & Wilson, 1978). Several studies which involved extant data sets did not suggest that PDA's performance would suffer appreciably because the assumption was violated (Knoke, 1982; Meshbane & Morris, 1996). No one seems to have specifically manipulated this condition in simulation studies to examine its effect on the performance of PDA and LR.

Relative performance of PDA and LR under different sample size conditions is also an issue of interest. Viewed from the perspective of statistical estimation in general, maximum likelihood estimators (as in LR) tend to require larger samples to achieve stable results than ordinary least square estimators (as in PDA). Inconsistent results have been reported about the relative performance of the two techniques with regard to sample size conditions. For example, in a simulation study, Harrell and Lee (1985) implied that PDA performed better under small sample size conditions. The Johnson and Seshia (1992) showed that, when the techniques were applied to real data sets, the findings did not clearly show that this was the case.

In addition to the three issues (data normality, equal covariance matrices, sample size), another issue which has attracted relatively little attention in the literature is the situation when two groups have drastically different proportions in a population, and the effect of this condition on the classification accuracy of PDA and LR. Neter, et al., (1989, p. 582) pointed out that, even for a valid logistic regression model, the middle range of the probabilistic function (say, .25 - .75) is practically <u>linear</u> (see Figure 1). This implies that in situations where the prior probabilities for the two groups are approximately equal, thus the cutoff point is in the middle range of the probabilistic function, it may make very little practical difference whether PDA or LR is used for classification.

On the other hand, when the prior probabilities are drastically different (e.g., .10 vs. .90 for two groups), the probabilistic function becomes more <u>nonlinear</u> in the extreme ranges, and consequently, logistic regression model may be theoretically better than linear discriminant function. This argument was echoed by other researchers (e.g., Cleary & Angel, 1984; Dey & Astin, 1993; Press & Wilson, 1978). The issue that LR should perform better than PDA in situations where prior probabilities for the two groups are drastically different has rarely been investigated empirically.

The inconsistent results in the literature may partially be attributable to the nonsystematic approach used in many studies which used single or a couple extant sample data sets to compare the two techniques (e.g., Angel & Cleary, 1984; Dey & Astin, 1993; Knoke, 1982; Press & Wilson, 1978; Wilson & Hardgrave, 1995;). Unfortunately, the insight such studies could offer about these issues is limited, and the degree of internal and external validity of the findings of these studies is generally not high, for reasons to be discussed momentarily. Even studies which involved multiple extant data sets (e.g., Meshbane & Morris, 1996) did not shed as much light on the issues as they appeared to.

There are several reasons for the limited internal and external validity of these studies. First, using extant data sets gives researchers no control of data characteristics, thus making it impossible to systematically investigate the impact of each individual factor, because in extant data sets, the effects of these relevant factors are often hopelessly confounded with each other. Second, most of these studies did not provide enough information about the data characteristics, making it very difficult to synthesize the results across studies. For these reasons, simulation studies with strong experimental control will be useful to assess the effects of these relevant factors.

Methods

This study considered three of the four issues discussed above: homogeneity of covariance matrices, sample size, and prior probabilities. Both because data normality condition has received the most attention in previous research, and for the reason of keeping the study manageable, data non-normality was not examined in the present study.

Design

A fully crossed three-factor experimental design represented graphically by Figure 2 was implemented for each of the two data structure patterns described in Table 1.

Insert Figure 2 about here
Insert Table 1 about here

In Table 1, the first data structure is arbitrarily determined. The second data structure was adapted from a real data structure presented by Stevens (1996, p. 268, for the first and the fourth groups). The two data structure patterns differ in the number of predictors (3 vs. 8 correlated predictors respectively), and in the correlation pattern among the variables. In Table 1, the degrees of group separation in the multivariate space, as measured by the Mahalanobis distance $[D^2=(\mu_1-\mu_2)'\ \Sigma^{-1}(\mu_1-\mu_2)]$ are also included.

The three factors manipulated under each data pattern were: sample size (4 levels: 60, 100, 200, 400), equality of covariance matrices (2 levels: equal, unequal), and prior probabilities for Group 1 and Group 2 (three levels: 0.50:0.50, 0.25:0.75, and 0.10:0.90). The fully crossed design for the two data structure patterns, with 200 replications in each cell, required the generation and model-fitting of 9600 ([4×2×3×200]×2) samples. The fully crossed design allows for systematic assessment of the impact of the three factors on the classification accuracy of PDA and LR.

Although no theoretical guidelines are available about what is a small or a large sample size for the purpose of classification for the two methods, the review of Meshbane and Morris (1996) of 32 real research data sets used for two-group classification has sample sizes ranging from 100 to 285. Compared with these 32 data sets, the sample size conditions specified in this study (60, 100, 200, 400) could be considered as ranging from relatively small to moderately large.

The degree of inequality of covariance matrices (Σs) between the two groups was specified *a priori* as one group having variances approximately 2-4 times larger than the other group[1]. Also, in this study, when both covariance matrices and group proportions were unequal, the group with smaller proportion has smaller variances on the predictor variables. The specification of unequal Σs in this fashion, however, reduced the degree of group separation (D^2), as indicated by the Mahalanobis distances in Table 1. So this factor (equal or unequal Σs) was confounded by the factor of group separation. Such confounding will be more fully discussed in the *Results and Discussions* section later in the paper.

The three prior probability (*population* proportions) conditions started from equal probabilities for the two groups (0.50:0.50) to the extreme of 0.10:0.90. The specification of these prior probabilities was motivated by the consideration that

---

[1] Note. When variances (σ_i^2) are unequal across the two groups, so will be the covariances (σ_{ij}), since $\sigma_{ij} = r_{ij}\, \sigma_i\, \sigma_j$.

PDA and LR may be minimally different in classification in the middle range of the probabilistic function, but the LR may model the extreme range better than PDA. The respective classification error rates of PDA and LR under these data conditions were obtained from each of the 200 samples within each cell, and their performance was compared based on the classification error rates.

Data Source and Model fitting

Data generation was accomplished by using the SAS normal data generator. Multivariate normal data were simulated through the matrix decomposition procedure (Kaiser & Dickman, 1962) and appropriate linear transformations. For each sample, first a pseudo-population was generated which was 20 times larger than the size of the sample. This pseudo-population had the exact proportions of the two groups under the three prior probability conditions (0.50:0.50, 0.25:0.75, and 0.10:0.90). Once this pseudo-population was generated, a simple random sample of a specified sample size (60, 100, 200, or 400) was drawn from this pseudo-population. In other words, although the population proportions of the groups were exactly specified, the sample proportions were not exact. This procedure models the research reality: sample proportion varies around the population proportion within the limits of sampling error.

Although statistical inference assumes an infinite population from which a sample is drawn, as Glass and Hopkins (1996, p. 224) point out, when the sampling fraction n/N =.05 or less (n: sample size; N: finite population size), the precision of statistical inferences would not be affected. This consideration motivated the decision of generating a pseudo-population 20 times larger than the sample size.

Once a sample was drawn, the sample data were fitted to both the linear discriminant analysis model and the logistic regression model, and the classification error rates from the two models were obtained. For PDA, SAS RPOC DISCRIM was used for model fitting, and the linear classification rule with the appropriate prior probabilities was used in the classification. For LR, SAS PROC LOGISTIC was used for LR model fitting, and the prior probability for the modeled group (logistic regression models the probabilistic function of one of the two groups) was specified for the classification. The classification error rates for the two groups as well as for the total sample under both PDA and LR were collected and saved in a SAS data file for later analyses.

Because both PDA and LR classification contains upward bias, due to the fact that the model estimation and classification are done on the same sample, bias-corrected classification error rates for the two methods were used in the present study. For PDA, the bias correction was achieved through the leave-one-out approach (Huberty, 1994; Lachenbruch, 1967), which is often known as "jackknifing" in the context PDA (Johnson & Wichern, 1988). For LR, due to the intensive computation involved, to fit the model for each observation could be computationally expensive (Knoke, 1982; SAS Institute, 1997, p. 461). Instead of the leave-one-out strategy, the SAS PROC LOGISTIC program implements a less expensive one-step algebraic approximation for correcting the upward bias. Interested readers are referred to the original source for this bias correction (SAS Institute, 1997, pp. 461-468).

The programming of the simulation study was accomplished through a combination of SAS Macro language, SAS PROC IML (interactive matrix language), and SAS statistical procedures, in SAS Window Version 6.12 (SAS Institute, 1997).

Results and Discussions

It turns out that the results are not as straightforward as what our literature review led us to believe. In the following sections, the relevant results will be presented with regard to the effects of the three factors (prior probabilities, un/equal covariance matrices, and sample size) on the classification accuracy of PDA and LR. Wherever appropriate, interpretations and implications of the findings are discussed.

Prior Probabilities for the Two Groups

Table 2 presents the mean classification error rates for Group 1 (the group which is equal or the smaller of the two). As discussed in the literature review section, it is expected that LR would perform better than PDA as this group's proportion becomes smaller, because LR is believed to be better in modeling the probabilistic function at the extreme. The results in Table 2 indicated that this expectation was confirmed by the results from both data structure patterns.

Insert Table 2 about here

In the top half of Table 2 (Data Structure Pattern 1), when covariance matrices are equal, and the two groups have equal population proportions (priors=0.50, equal Σs), PDA and LR have approximately equal classification error rates for this group (10%). But as the Group 1 proportion becomes more extreme, the

PDA's error rate increased rapidly (about 20% for prior=.25; about 35% for prior=.10), while LR's classification error rate is relatively stable (about 11% for prior=.25; about 13% for priors=.10, for larger sample sizes). Even under smaller sample size conditions (n=60, 100), LR still performed considerably better than PDA. The same phenomenon is observed for the second data structure pattern: PDA and LR have comparable classification error rates for prior=.50, but LR performed better than PDA when the prior became smaller.

In Table 2, under the condition of unequal $\underline{\Sigma}$s, both PDA and LR performed worse than they did under equal $\underline{\Sigma}$s. However, the readers are reminded that this condition is confounded with that of group separation to some degree: the specification of unequal Σs in the present study actually reduced the group separation. As a result, it is expected that the performance of both PDA and LR would suffer. A close look at Table 2 reveals that LR's error rates for unequal $\underline{\Sigma}$s are relatively close to those under equal $\underline{\Sigma}$s under larger sample size conditions (n=200, 400), indicating very minor effect of unequal $\underline{\Sigma}$s on LR for this smaller group. On the other hand, PDA's performance in classifying the smaller group members under unequal $\underline{\Sigma}$s becomes substantially worse than under equal $\underline{\Sigma}$s, and than LR, with error rates reaching as high as high as 90%. The only exception is that better PDA performance is seen for unequal $\underline{\Sigma}$s and the priors=0.50. It should be noted that there is a high degree of consistency across the two data structure patterns, making the observation less likely to be a "fluke" caused by a particular data structure.

The findings above indicate that, if prior probabilities are known to be approximately equal, the choice of PDA and LR is probably not that important. But when the prior probabilities are known to be unequal to a considerable degree, and we are concerned about the accurate classification of the members of the smaller group, LR appears to be the method of choice, whether or not the condition of equal Σs is met.

Table 3 presents the classification error rates for Group 2, the group which is equal or larger of the two. The findings observed in Table 2 are reversed here: PDA performed approximately equally well as LR for priors=.50, but performed substantially better than LR for priors>.50, whether or not the equal Σs condition is met. Again, this observation is consistent across the two data structure patterns.

Insert Table 3 about here

The opposite results in Table 2 and Table 3 regarding the efficacy of PDA and LR appear to indicate that, when the groups have unequal proportions, for classification methods such as PDA and LR, one group's loss may often be the other group's gain. In other words, for a given data pattern, choose one technique for minimizing one group's classification error may often mean increasing the classification error for the other group. Which method to choose may have to depend on the consequences of misclassification for the groups involved.

Table 4 presents the total classification error rates for both groups combined. There appear to be two noteworthy observations. First, when the two groups have equal prior probabilities (0.50:0.50), PDA and LR have comparable total classification error rates for both equal and unequal $\underline{\Sigma}$s conditions, except that LR has slightly higher error rates when sample size is small (e.g., n=60). But when the two groups have known prior probabilities unequal to an appreciable degree (0.25:0.75, 0.10:0.90), PDA has lower <u>total</u> classification error rate than LR for all conditions examined in this study, although the difference may be small (e.g., for priors of 0.25:0.75).

Insert Table 4 about here

<u>Sample Size</u>

In general, sample size appears to have minor influence on the classification accuracy for PDA and LR. A close look at Table 2 to Table 4 indicates that LR showed slightly higher classification error rates when the sample was small (e.g., n=60). PDA classification error rates, on the other hand, showed little influence of sample size. This observation agrees with theoretical expectations: PDA requires smaller sample sizes for ordinary least squares solution of PDA function coefficients ($\hat{\underline{a}}'$), while LR requires larger sample sizes for its maximum likelihood solution of regression coefficients ($\hat{\underline{\beta}}'$).

<u>Homogeneity of Covariance Matrices</u>

As discussed previously, it is often difficult to separate the effect of unequal $\underline{\Sigma}$s with that of group separation. Obviously, the specification of unequal $\underline{\Sigma}$s in this study reduced the separation of the two groups in the multivariate space. For this reason, the interpretation for the difference between equal vs. unequal $\underline{\Sigma}$s for PDA and LR is confounded with the factor of group separation.

For example, in Table 2, for priors=0.25 under Data Structure Pattern 1, and for n=100, the error rates are 0.20 (PDA) and 0.12 (LR) respectively for Equal $\underline{\Sigma}$s. For Unequal $\underline{\Sigma}$s and for the same prior=0.25, the error rates are 0.53 (PDA) and 0.16 (LR) respectively. If there were no confounding of group separation, we could conclude that PDA performed much worse than LR for unequal $\underline{\Sigma}$s than it did for equal $\underline{\Sigma}$s. But due to the confounding, we could also say that PDA performed much worse than LR for smaller group separation than it did for larger group separation. For this reason, our interpretation of the effect of unequal $\underline{\Sigma}$s will be qualified as "the effect of unequal $\underline{\Sigma}$s/smaller group separation".

Table 2 for the smaller group classification error rates shows 1) for equal priors for the two groups (priors=.50), PDA performed slightly better than LR for the condition of unequal $\underline{\Sigma}$s/smaller group separation; 2) when prior probability become smaller (priors=.25, .10), PDA's performance rapidly deteriorated under unequal $\underline{\Sigma}$s/smaller group separation, with its classification error reaching unacceptable levels ; 3) although the condition of unequal $\underline{\Sigma}$s/smaller group separation would also affect LR, its performance was much better than that of PDA for the smaller group.

What was observed in Table 2 regarding unequal $\underline{\Sigma}$s/smaller group separation for the smaller group was reversed in Table 3 for the larger group, parallel to the condition of prior probabilities. Here we see the same phenomenon of "one group's loss is another group's gain". For the larger group (priors=.75, .90), PDA's performance turns out to be noticeably better than LR's under the condition of unequal Σs/smaller group separation, contrary to the observation from Table 2 about the smaller group.

If one is concerned about total classification error rate across groups under the condition of unequal Σs/smaller group separation, Table 4 shows that PDA performs better than LR in general, except for equal prior probability (prior=.50). The better performance of PDA as measured by the total classification error rate is more obvious as the prior probabilities for the two groups become more different. Again, consistency is observed for both data structure patterns.

Sources of Variation of the Classification Error Rates

To better understand the extent to which each factor examined in this study has contributed to the variation of classification error, analysis of variance was conducted to partition the variance of the classification error. Table 5 presents the results of this variance partitioning both for the separate error rates of the two groups (smaller and larger groups), and that for the total sample.

For Group 1 (equal or smaller group), for both the data structure patterns, the largest contributor to the variation of the classification error rate is the prior probability, accounting for 22% and 30% of total variance respectively for the two data patterns. Also, both method factor (PDA vs. LR) and covariance factor (equal vs. unequal Σs) caused considerable amounts of variation in the classification error rate. In addition, the sizable interaction term between prior probability and classification method (P*M) indicates that the influence of prior probability is not uniform for the two methods, as was evident in Table 2. As discussed previously (Table 2), for this group, LR performed relatively well for all the conditions of prior probability, while PDA performed poorly when the prior probability became smaller.

For Group 2, as well as for the total error rate, the most prominent source for the variation of error rate is the factor of equal/unequal $\underline{\Sigma}$s condition, and this observation is consistent for both data structure patterns. For Group 2, the prior probability, the method, and the interaction between the two also account for sizable portions of the variance. For the total classification error rate, however, the influence of prior probability and classification method played a relatively minor roles. It is also obvious that sample size does not have any obvious impact on the classification error rate for either group or for the total error rate, although previous discussion about Table 2 to Table 4 revealed that it might be a factor for LR, especially when sample size is relatively small. It should be noted that for the balanced design implemented in this study, the partitioned variance for all the sources (including interaction terms) are orthogonal, i.e., they are additive to a total of 100%.

Practical Implications of the Results

The previous results (Table 2 to Table 5) and the discussion revealed some phenomena not well documented in the literature. As discussed in literature review section, the issue of prior probabilities has rarely been examined, nor has the issue of unequal Σs (although this condition is confounded by group separation to some extent in this study). Almost all previous studies focused on the situation where the two groups have approximately equal proportions, or the two groups' proportions are not drastically different (e.g., Meshbane & Morris, 1996; Press & Wilson, 1978). This, probably, is the major reason that many studies came to the conclusion that the two methods do not have obvious practical differences in their performance (e.g., Dattalo, 1994; Dey & Astin, 1993; Meshbane & Morris, 1996). While this general conclusion is supported by the findings of this study for prior=0.50 condition (equal proportions of the two groups), the picture is far more complicated than that.

When the two groups do not have approximately equal proportions, as is very common in educational and psychological research (e.g., in education, predicting those who may need remedial education, or those who may drop out of school; in psychology, predicting those who may develop some sort of psychological disorder), the conclusion that the performance of two methods is comparable can be quite misleading, as revealed in this study. For the research practitioner, it is important to understand the dynamics of these major variables so that an informed choice between the two methods can be made.

As indicated by many previous studies, and supported by the present study, for situations where the two groups have approximately equal proportions, it may not make much practical difference which of the two methods is chosen for the purpose of classification, since they have similar classification accuracy.

But for situations where the two groups have very unequal proportions, the choice of one over the other can be quite important, since the two methods may have very different performance, depending on which group's classification error we are most interested in minimizing.

If we are most concerned about minimizing the classification error rate for the smaller group, as in a situation where the consequence of misclassification for members of this group is more serious (i.e, the cost of misclassification for this group is higher) than that for the members of the larger group (e.g., to identify subjects who may develop some type of disorder which can be effectively treated, but which may cause long term psychological damage if ignored), LR appears to be the method of choice (see results in Table 2), regardless of whether the assumption of equal Σs can be met. Theoretically, LR is expected to model the probabilistic function near the extreme range better than PDA because of the curvilinear relationship between the predictors and the probabilistic function at this range. The empirical findings of this study confirmed this theoretical expectation.

On the other hand, if we are interested in minimizing the misclassification rate for the larger group, PDA appears to be preferable over LR (see the results in Table 3), for conditions of equal/unequal Σs. In the same vein, if we are interested in minimizing the total classification error rate regardless of those of the larger and smaller groups, PDA appears to be the preferred method, because it has consistently lower overall classification error rate.

Limitations and Future Directions

Like many other studies, this study has its own limitations. As discussed previously, the specification for unequal Σs is, to some extent, confounded by the degree of group separation. This confounding makes it less clear how unequal Σs has impacted the performance of PDA and LR. Future studies may benefit from isolating the effect of unequal Σs on PDA and LR by specifying conditions of unequal Σs which will maintain or minimally change the group separation. On the same note, group separation itself is often considered as a another relevant factor which may affect the performance of PDA and LR, as discussed by Harrell & Lee (1985). This factor was not specifically addressed here. Although two different data structure patterns were simulated, the two data patterns were similar in the separation of the two groups (see Table 1), thus limiting the generalizability of the findings to some degree. Future research may consider data patterns more varied on this and other dimensions.

Summary and Conclusions

This Monte Carlo study compared the performance of predictive discriminant analysis (PDA) and that of logistic regression (LR) for the two-group classification problem. A fully crossed three-factor experimental design was used in this study: sample size (four levels), prior probabilities (three levels), and condition of covariance matrices (two levels). To reduce the likelihood of chance discovery, two different data structure patterns were used: an arbitrarily specified data pattern with three correlated predictors, and a pattern modeled after real research data with eight correlated predictors.

Within each cell condition under each data structure pattern, 200 random samples were generated based on the specified population parameters for the two groups, and PDA and LR were used to classify the sample members into one of the two groups. Prior probabilities for the two groups were used for both PDA and LR, but equal cost of misclassification was assumed for the two groups. Classification error rates for both groups as well as for the total sample were collected ans saved for subsequent analyses. Bias correction measures were implemented for the classification error rates for both techniques (PDA and LR). The design of the experiment required a total of 9600 random samples ([4×3×2×200]×2). The results indicate the following:

1. When the two groups have approximately equal proportions, PDA and LR appear to have comparable performance for the condition of equal $\underline{\Sigma}$s, and their performances differ slightly for the condition of unequal $\underline{\Sigma}$s.
2. When the two groups have very different proportions,
   a) the choice of LR appears to minimize the classification error rate for the <u>smaller</u> group, for both equal/unequal $\underline{\Sigma}$s;
   b) the choice of PDA appears to minimize the classification error rate of the <u>larger</u> group, for both equal/unequal $\underline{\Sigma}$s;
   c) PDA appears to minimize the <u>total</u> classification error rate for both equal/unequal $\underline{\Sigma}$s.
3. Sample size appears to play a very minor role in the classification accuracy of the two methods, except when LR is used under small sample size conditions.

Consistency was observed across the two data structure patterns, making it less likely that the findings are chance discoveries caused by the idiosyncracies of a particular data structure. The results of this study reveals a picture about PDA and LR which is more complicated than what has been typically portrayed in the literature. The results show that the choice of PDA and LR in research practice should be closely related to the proportions of the groups. Furthermore, if possible, the cost of misclassification for each group needs to be considered so as to determine which group's classification error rate, or the total error rate, should be minimized. These considerations are likely to help the research practitioner to make an informed choice between the two methods.

References

Barón, A. E. (1991). Misclassification among methods used for multiple group discrimination - The effects of distributional properties. Statistics in Medicine, 10, 757-766.

Bayne, C. K., Beauchamp, J. J., Kane, V. E., & McCabe, G. P. (1983). Assessment of Fisher and logistic linear and quadratic discrimination models. Computational Statistics and Data Analysis, 1, 257-273.

Cleary, P. D., & Angel, R. (1984). The analysis of relationships involving dichotomous dependent variables. Journal of Health and Social Behavior, 25, 334-348.

Cox, D. R., & Snell, E. J. (1989). The analysis of binary data (2nd ed.). London: Chapman and Hall.

Dattalo, P. (1994). A comparison of discriminant analysis and logistic regression. Journal of Social Service Research, 19, 121-144.

Dey, E. L., & Astin, A. W. (1993). Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression. Research in Higher Education, 34, 569-581.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. Journal of the American Statistical Association, 70, 892-898.

Fan, X. (1996). Canonical correlation analysis as a general analytical model. In B. Thompson (Ed.), Advances in Social Science Methodology: A Research Annual, Vol. 4, pp. 71-94. Greenwich, CT: JAI Press, Inc. Fan, 1996;

Fraser, M. W., Jensen, J. M., Kiefer, D., & Popuang, C.

(1994). Statistical methods for the analysis of critical events. Social Work Research, 18(3), 163-177.

Glass, G. V., & Hopkins, K. D. (1996). Statistical methods in education and psychology (3rd ed.). Boston, MA: Allyn and Bacon.

Harrell, F. E. Jr., & Lee, K. L. (1985). A comparison of the discrimination of discriminant analysis and logistic regression under multivariate normality. In P. K. Sen (Ed.), Biostatistics: statistics in biomedical, public health and environmental sciences (pp. 333-343). Amsterdam: North Holland.

Hosmer, D. W. (1989). Applied logistic regression. New York: Wiley.

Huberty, C. J. (1994). Applied discriminant analysis. New York: Wiley.

Johnson, B. & Seshia, S. S. (1992). Discriminant analysis when all variables are ordered. Statistics in Medicine, 11, 1023-1032.

Johnson R. A., & Wichern, D. W. (1988). Applied multivariate statistical analysis (2nd ed.). Englewood Cliffs, New Jersey: Prentice Hall.

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

Knoke, J. D. (1982). Discriminant analysis with discrete and continuous variables. Biometrics, 38, 191-200.

Kshirsagar, A. M. (1972). Multivariate analysis. New York: Marcel Dekker.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological

Bulletin, 85, 410-416.

Lachenbruch, P. A. (1967). An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. Biometrics, 23, 639-645.

Meshbane, A., & Morris, J. D. (April, 1996). Predictive discriminant analysis versus Logistic regression in two-group classification problems. Paper presented at the annual meeting of the American Educational Research Association, New York, NY. (ERIC Documentation Reproduction Services No. ED 400 280)

Neter, J., Wasserman, W., & Kutner, M. H. (1989). Applied linear regression models (2nd ed.). Boston, MA: Irwin.

Press, S. J., & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association, 73, 699-705.

SAS Institute Inc. (1997). SAS/STAT software: Changes and enhancements through release 6.12. Cary, NC: SAS Institute, Inc.

Stevens, J. (1996). Applied multivariate statistics for the social sciences, (3rd ed.). Mahwah, New Jersey: LEA.

Tabachnick, B. G., & Fidell, L. S. (1996). Using multivariate statistics (3rd ed.). New York, NY: HarperCollins College Publishers.

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

Wilson, R. L., & Hargrave, B. C. (1995). Predicting graduate student success in an MBA program: Regression versus classification. Educational and Psychological Measurement, 55, 186-195.

Yarnold, P. R., Hart, L. A., & Soltysik, R. C. (1994). Optimizing the classification performance of logistic regression and Fisher's discriminant analyses. Educational and Psychological Measurement, 54, 73-85.

Table 1

Two Data Structure Patterns Simulated in the Study

<u>Data Structure 1</u>

X1	1.00		
X2	0.30	1.00	
X3	0.50	0.40	1.00
μ_1	5.00	5.00	5.00[a]
μ_2	9.00	9.00	9.00[b]
σ^2	4.00	4.00	4.00[c]
σ^2	16.00	16.00	16.00[d]

Group Separattion (the Mahalanobis Distance: $D^2=(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)$):

Equal Σs: $D^2 = 6.70$

Unequal Σs: $D^2 = 2.68$ (for priors: 0.50:0.50)

<u>Data Structure 2</u>

X1	1.00							
X2	0.45	1.00						
X3	0.05	0.25	1.00					
X4	0.35	0.05	0.25	1.00				
X5	0.35	0.10	0.35	0.55	1.00			
X6	0.05	0.25	0.50	0.15	0.40	1.00		
X7	-.35	0.05	0.40	0.15	0.30	0.41	1.00	
X8	0.30	0.30	0.50	0.35	0.60	0.50	0.45	1.00
μ_1	12.50	15.00	15.95	12.65	12.15	14.15	18.20	15.20[a]
μ_2	11.40	14.25	15.00	11.30	12.90	15.00	19.20	14.50[b]
σ^2	1.00	2.00	2.00	1.50	1.20	2.00	2.50	2.00[c]
σ^2	4.00	6.00	3.00	4.50	4.80	6.00	7.50	8.00[d]

Group Separattion (the Mahalanobis Distance: $D^2=(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2)$):

Equal Σs: $D^2 = 6.80$

Unequal Σs: $D^2 = 3.26$ (for priors: 0.50:0.50)

a Mean vector for Group 1.
b Mean vector for Group 2.
c For the condition of <u>equal</u> covariance matrices, this set of variances is used for <u>both</u> groups.
d For the condition of <u>unequal</u> covariance matrices, this set of variances is used for Group 2, and the set above is used for Group 1.

Table 2

Classification Error Rates for Group 1 (Equal or Smaller Group)

Priors	Σ	Method	60	100	200	400
Data Structure Pattern 1			Sample Size			
0.50	Equal	PDA	11 (05)	10 (04)	10 (03)	10 (02)
		LR	12 (05)	11 (04)	10 (03)	10 (02)
0.25		PDA	21 (11)	20 (08)	20 (06)	20 (04)
		LR	15 (08)	12 (06)	11 (03)	11 (02)
0.10		PDA	40 (25)	35 (18)	33 (11)	33 (07)
		LR	28 (22)	18 (11)	13 (06)	11 (04)
0.50	Unequal	PDA	11 (06)	11 (05)	10 (03)	10 (02)
		LR	17 (08)	16 (06)	15 (04)	15 (03)
0.25		PDA	54 (20)	53 (16)	53 (12)	53 (08)
		LR	17 (12)	16 (09)	14 (06)	14 (04)
0.10		PDA	92 (13)	94 (11)	95 (06)	96 (04)
		LR	27 (24)	20 (15)	16 (10)	14 (06)
Data Structure Pattern 2						
0.50	Equal	PDA	12 (06)	11 (04)	10 (03)	10 (02)
		LR	15 (05)	12 (03)	11 (03)	10 (02)
0.25		PDA	24 (12)	22 (08)	21 (06)	19 (04)
		LR	23 (09)	16 (06)	12 (04)	11 (03)
0.10		PDA	44 (24)	36 (15)	34 (11)	33 (07)
		LR	40 (21)	29 (14)	17 (06)	12 (04)
0.50	Unequal	PDA	12 (06)	12 (05)	11 (03)	11 (02)
		LR	18 (07)	16 (06)	15 (04)	15 (03)
0.25		PDA	39 (17)	40 (13)	42 (09)	41 (07)
		LR	22 (09)	19 (07)	16 (05)	15 (04)
0.10		PDA	79 (22)	82 (15)	83 (12)	84 (09)
		LR	38 (24)	25 (14)	19 (10)	16 (05)

Note. Each table entry is the mean classification error rate (standard deviation) based on the classification error rates of 200 random samples. Second place decimal point is omitted.

Table 3

Classification Error Rates for Group 2 (Equal or Larger Group)

Data Structure Pattern 1			Sample Size			
Priors	Σ	Method	60	100	200	400
0.50	Equal	PDA	11 (05)	10 (04)	10 (03)	10 (02)
		LR	11 (05)	11 (04)	10 (03)	10 (02)
0.75		PDA	05 (03)	05 (02)	04 (02)	04 (01)
		LR	11 (05)	10 (03)	10 (03)	10 (02)
0.90		PDA	02 (02)	02 (01)	02 (01)	02 (01)
		LR	09 (06)	09 (05)	09 (03)	09 (02)
0.50	Unequal	PDA	28 (07)	27 (05)	26 (04)	26 (02)
		LR	24 (07)	24 (06)	23 (04)	23 (03)
0.75		PDA	11 (03)	10 (02)	10 (02)	09 (01)
		LR	24 (07)	24 (05)	24 (04)	23 (03)
0.90		PDA	03 (02)	02 (01)	02 (01)	01 (01)
		LR	23 (09)	24 (07)	24 (05)	24 (04)
Data Structure Pattern 2						
0.50	Equal	PDA	12 (06)	11 (04)	10 (03)	10 (02)
		LR	15 (05)	12 (04)	11 (03)	10 (02)
0.75		PDA	06 (03)	05 (02)	05 (02)	05 (01)
		LR	12 (04)	10 (04)	10 (02)	10 (02)
0.90		PDA	03 (02)	02 (01)	02 (01)	02 (01)
		LR	10 (04)	08 (04)	09 (03)	09 (02)
0.50	Unequal	PDA	28 (07)	26 (06)	24 (04)	24 (03)
		LR	25 (08)	23 (06)	22 (04)	21 (03)
0.75		PDA	13 (03)	11 (02)	10 (02)	09 (01)
		LR	22 (07)	22 (06)	21 (03)	20 (03)
0.90		PDA	04 (02)	03 (01)	02 (01)	02 (01)
		LR	19 (08)	21 (07)	20 (04)	21 (03)

Note. Each table entry is the mean classification error rate (standard deviation) based on the classification error rates of 200 random samples. Second place decimal point is omitted.

Table 4

Total Classification Error Rates (Both Groups)

Priors	Σ	Method	60	100	200	400
Data Structure Pattern 1			Sample Size			
.50:.50	Equal	PDA	11 (04)	10 (03)	10 (02)	10 (02)
		LR	12 (04)	11 (03)	10 (02)	10 (02)
.25:.75		PDA	09 (04)	09 (03)	08 (02)	08 (01)
		LR	12 (04)	11 (03)	10 (02)	10 (02)
.10:.90		PDA	06 (03)	05 (02)	05 (01)	05 (01)
		LR	10 (05)	10 (05)	10 (03)	10 (02)
.50:.50	Unequal	PDA	19 (05)	19 (04)	18 (03)	18 (02)
		LR	20 (05)	19 (05)	19 (03)	19 (02)
.25:.75		PDA	22 (06)	21 (04)	20 (03)	20 (02)
		LR	22 (06)	22 (04)	21 (03)	21 (02)
.10:.90		PDA	11 (02)	11 (01)	11 (01)	11 (01)
		LR	23 (09)	22 (06)	23 (05)	23 (03)
Data Structure Pattern 2						
.50:.50	Equal	PDA	12 (05)	11 (03)	10 (02)	10 (01)
		LR	15 (04)	12 (03)	11 (02)	10 (01)
.25:.75		PDA	11 (04)	09 (03)	09 (02)	08 (01)
		LR	14 (04)	12 (03)	11 (02)	10 (02)
.10:.90		PDA	07 (04)	06 (02)	05 (01)	05 (01)
		LR	12 (04)	10 (04)	10 (03)	10 (02)
.50:.50	Unequal	PDA	20 (05)	19 (04)	18 (03)	17 (02)
		LR	21 (05)	19 (04)	18 (03)	17 (02)
.25:.75		PDA	20 (05)	18 (04)	18 (03)	17 (02)
		LR	22 (05)	21 (05)	20 (03)	19 (02)
.10:.90		PDA	12 (03)	11 (02)	10 (01)	10 (01)
		LR	20 (07)	21 (06)	20 (04)	21 (03)

Note. Each table entry is the mean classification error rate (standard deviation) based on the classification error rates of 200 random samples. Second place decimal point is omitted.

Table 5

Variance Partitioning for Classification Error Rates

Data Structure Pattern 1

Source	Group 1	Group 2	Total
Total $\underline{R}^2$	**85.13**	**82.49**	**72.95**
Prior (P)	22.14	14.00	3.73
Method (M)	18.22	15.96	5.95
Covariance Σ (C)	10.78	30.42	52.78
Sample Size (N)	.	.	.
P * M	14.92	12.57	6.67
P * C	4.91	2.72	.
M * C	7.04	2.71	.
P * M * C	6.04	3.87	1.83
.			
.			

Data Structure Pattern 2

Source	Group 1	Group 2	Total
Total $\underline{R}^2$	**79.14**	**79.53**	**67.30**
Prior (P)	30.35	19.83	5.76
Method (M)	11.38	13.52	8.18
Covariance Σ (C)	7.20	27.84	43.39
Sample Size (N)	1.72	.	2.80
P * M	12.30	9.75	4.89
P * C	3.63	2.46	.
M * C	4.78	1.21	.
P * M * C	5.15	3.14	.
.			
.			

Note. 1) The tabled entries are the η^2s:

$$\eta^2 = [(\text{Source Sum of Squares})/(\text{Total Sum of Squares})] \times 100$$

2) Those interaction terms which account for less than 1% of the total variance are not listed. For a listed source, a dot is used to indicate that it accounts for less than 1% of the total variance.

Figure Captions

Figure 1 Graphic Representation of the Study Design

Figure 2 Logistic Regression Function with One Predictor

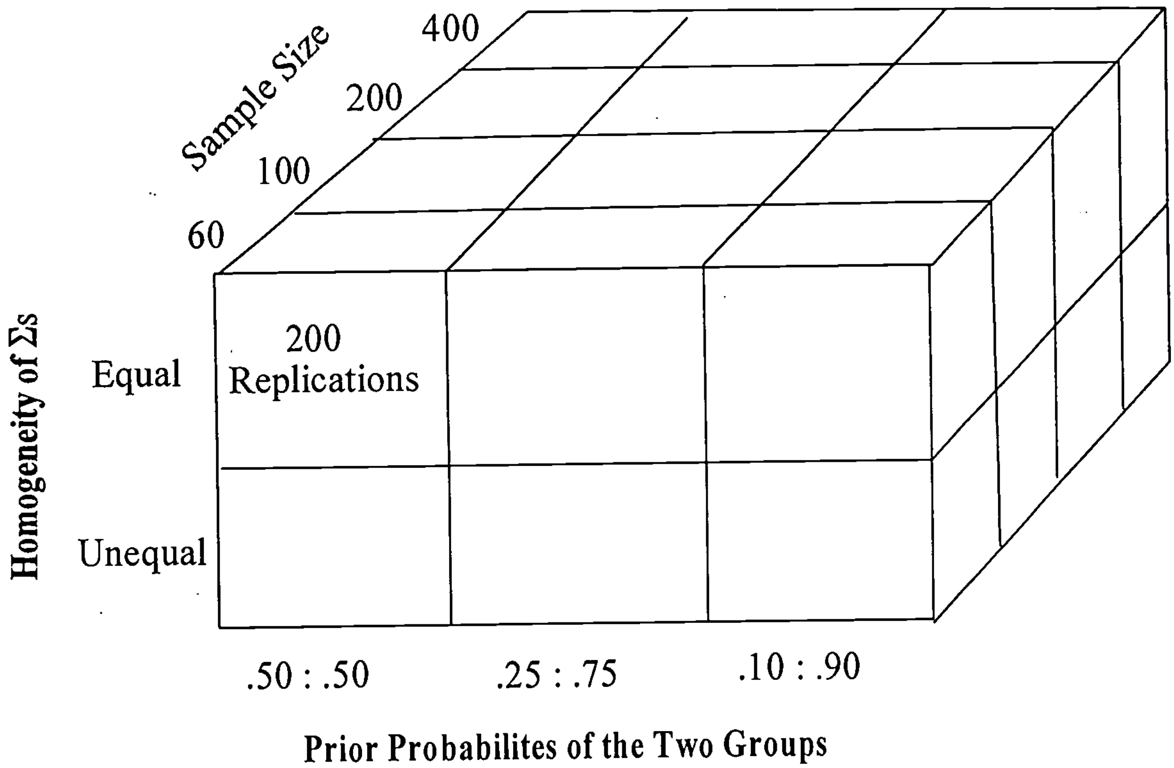
Sample Size
400
200
100
60
Homogeneity of Σs
Equal
Unequal
200
Replications
.50 : .50
.25 : .75
.10 : .90
Prior Probabilites of the Two Groups

Figure 1

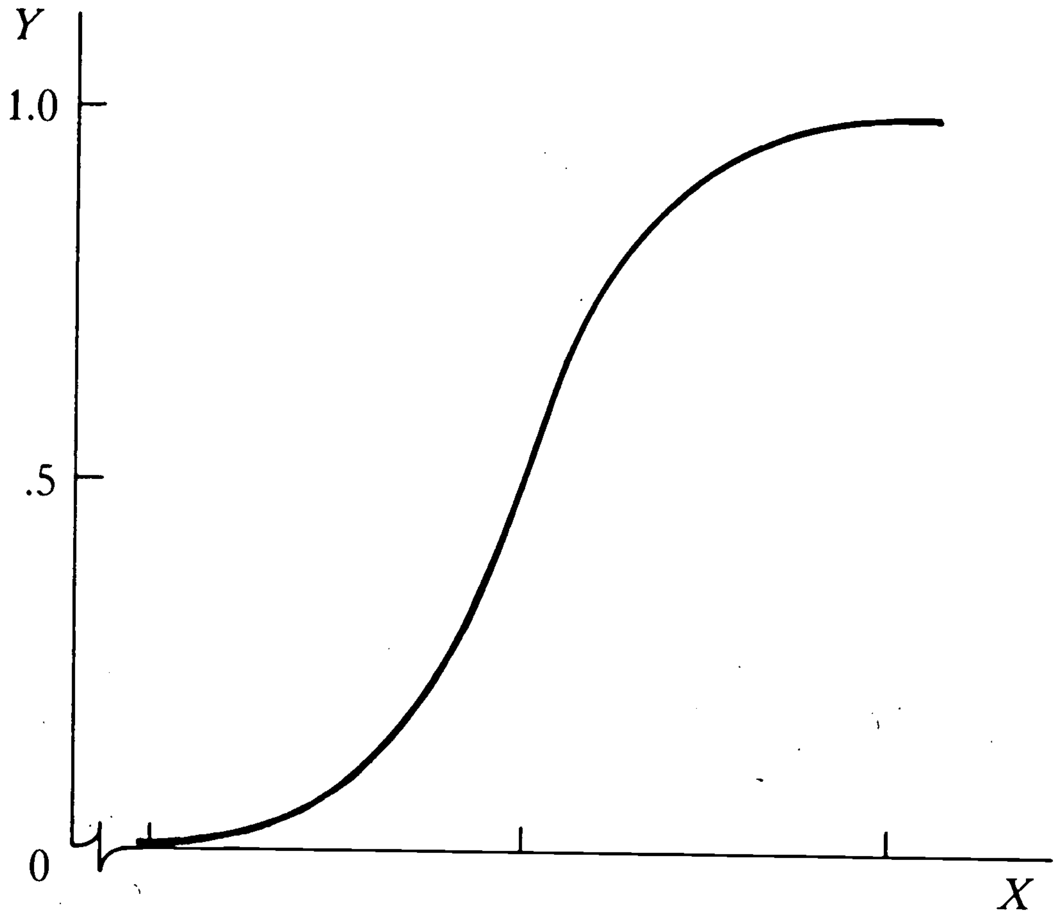

Figure 2

Signature: Xitao Fan	Printed Name/Position/Title: Assistant Professor	
Organization/Address: Dept. of Psychology Utah State University, Logan, UT 84322-2810	Telephone: (435)797-1451	FAX: (435)797-1448
	E-Mail Address: fafan@cc.usu.edu	Date: April 28, 1998

ERIC® Clearinghouse on Assessment and Evaluation

University of Maryland
1129 Shriver Laboratory
College Park, MD 20742-5701

Tel: (800) 464-3742
(301) 405-7449
FAX: (301) 405-8134
ericae@ericae.net
http://ericae.net

March 20, 1998

Dear AERA Presenter,

Congratulations on being a presenter at AERA[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a printed copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at http://ericae.net.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (424)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to: AERA 1998/ERIC Acquisitions
University of Maryland
1129 Shriver Laboratory
College Park, MD 20742

This year ERIC/AE is making a **Searchable Conference Program** available on the AERA web page (http://aera.net). Check it out!

Sincerely,

Lawrence M. Rudner

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1]If you are an AERA chair or discussant, please save this form for future use.