DOCUMENT RESUME

ED 475 146 TM 034 821

AUTHOR	Fraas, John W.; Newman, Isadore
TITLE	Ordinary Least Squares Regression, Discriminant Analysis, and Logistic Regression: Questions Researchers and Practitioners Should Address When Selecting an Analytic Technique.
PUB DATE	2003-02-00
NOTE	42p.; Paper presented at the Annual Meeting of the Eastern Educational Research Association (Hilton Head Island, GA, February 26-March 1, 2003).
PUB TYPE	Information Analyses (070) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Discriminant Analysis; *Least Squares Statistics; *Regression (Statistics); *Research Methodology; *Selection

ABSTRACT

The purpose of this paper is to assist researchers, practitioners, and graduate students in identifying and addressing key questions related to the task of choosing among the analytic techniques designed to analyze a dichotomized dependent variable with a set of independent variables. The discussion is limited to (1) the analysis of data by the analytic procedures of ordinary least squares regression, discriminant analysis, or logistic regression; (2) the use of the Statistical Package for the Social Sciences (registered) computer software; and (3) a dependent variable consisting of two groups. The paper states that researchers need to address the adequacy of each technique with respect to two basic questions. What impact do possible violations of underlying assumptions have on the results? Does a given technique readily produce the type of information required to address the research question. An analysis of a data set is provided to illustrate how addressing these issues can assist in the selection process. (Contains 2 tables and 23 references.) (Author)

Running head: ORDINARY LEAST SQUARES REGRESSION

Ordinary Least Squares Regression, Discriminant Analysis, and Logistic Regression: Questions Researchers and Practitioners Should Address When Selecting an Analytic Technique

John W. Fraas

Ashland University

Isadore Newman

The University of Akron

Paper presented at the annual meeting of the Eastern Educational Research Association

Hilton Head Island, February 26 - March 1, 2003

Abstract

The purpose of this paper is to assist researchers, practitioners, and graduate students in identifying and addressing key questions related to the task of choosing among the analytic techniques designed to analyze a dichotomized dependent variable with a set of independent variables. The discussion is limited to (a) the analysis of data by the analytic procedures of OLS regression, discriminant analysis, or logistic regression; (b) the use of the SPSS® computer software; and (c) a dependent variable consisting of two groups. The paper states that researchers need to address the adequacy of each technique with respect to two basic questions. What impact do possible violations of underlying assumptions have on the results? Does a given technique readily produce the type of information required to address the research question? An analysis of a data set is provided to illustrate how addressing these issues can assist in the selection process.

Ordinary Least Squares Regression, Discriminant Analysis, and Logistic Regression: Questions Researchers and Practitioners Should Address When Selecting an Analytic Technique

A number of researchers have noted that many research studies call for the analysis of a dichotomous dependent variable (Cabrera, 1994; Peng, Lee, &Ingersoll, 2002), that is, a variable that consists of two values used to identify two groups of subjects. Peng et al. noted that, traditionally, researchers utilized ordinary least squares (OLS) regression or discriminant analysis to analyze the data in such studies. Cabrera (1994) and Manski and Wise (1983) referred to studies in which the researchers used logistic regression to analyze their dichotomous dependent variables rather than OLS regression or discriminant analysis.

This paper attempts to identify and examine the key questions researchers and practitioners should address when deciding whether to use OLS regression, discriminant analysis, or logistic regression to analyze a dichotomized dependent variable. We have restricted our discussion to research situations in which the dependent variable consists of only two groups and the SPSS® 11.0 computer software is used as the means of data analysis.

In our attempt to identify and examine the key questions we have assumed that researchers who analyze dichotomous dependent variables do so with one or *more* goals in mind. One such goal is to identify the statistically significant independent variables and be able to judge their practical significance. A second possible goal is to accurately classify future subjects as members of the two groups identified in the dependent variable. A third possible goal is to predict probability values for future subjects that will indicate their chances of belonging to the group assigned the value of one in the dependent variable.

The remaining portion of this paper consists of six sections. The first three sections contain brief discussions of OLS regression, discriminant analysis, and logistic regression used in conjunction with a dependent variable designed to represent two groups. The major concerns regarding the application of each technique to a dichotomized dependent variable, which are divided into those that relate to the type of information the technique provides and the underlying assumptions of the technique, are also presented in these three sections. The fourth section contains the results of the application of each technique to a set of Ashland University data, which are used in the fifth section of the paper. The fifth section identifies and discusses key questions researchers should address when deciding whether to use OLS regression, discriminant analysis, or logistic regression. In addition, the results produced in the fourth section are examined in light of these key questions. The fifth section is followed by a summary.

OLS Regression

In a regression model the relationship between a single dependent variable and several independent variables is estimated. The model postulates that the values of the dependent variable equal a linear combination of the independent variables plus an error term. Such a model can be represented as follows:

$$\mathbf{Y} = \beta_0 + \beta_1\mathbf{X}_1 + \beta_2\mathbf{X}_2 + \ldots \beta_k\mathbf{X}_k + \varepsilon \qquad \text{[Equation 1.0]}$$

where:

1. The βs are the regression coefficients.
2. $\mathbf{Y}$ represents a column vector for the dependent variable.
3. The $\mathbf{X}$s are column vectors for the independent variables.
4. The column vector of errors of prediction is represented by ε.

This regression model is linear in the β parameters but it may or may not be linear with respect to **Y** or the **X**s. Models that are not linear with respect to **Y** or the **X**s can be formed in a number of ways including (a) the values contained in **Y** or the **X**s are transformed by a power other than one or (b) the products of **X** column vectors are included in the model. As noted by Chatterjee, Hadi, and Price (2000, p. 13) "all nonlinear functions [with respect to **Y** and the **X**s] that can be transformed into linear functions are called *linearizable* functions. Accordingly, the class of linear models is actually wider than it might appear . . . because it includes all linearizable functions."

The parameters (β values) in Equation 1.0 can be estimated using the OLS method. The model containing the estimated parameters can be represented as follows:

$$\mathbf{Y} = \hat{\beta}_0 + \hat{\beta}_1\mathbf{X}_1 + \hat{\beta}_2\mathbf{X}_2 + \ldots \hat{\beta}_k\mathbf{X}_k + \mathbf{e} \qquad \text{[Equation 1.1]}$$

where:

1. $\hat{\beta}_0$ is the estimate of β_0, $\hat{\beta}_1$ is the estimate of β_1, etc.
2. The symbol **e** denotes the residual term, which conceptually is analogous to ε and can be regarded as an estimate of ε.

The predicted values of **Y**, represented by ($\hat{\mathbf{Y}}$), are obtained by substituting each case's value for each independent variable into the following regression equation:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + \ldots \hat{\beta}_k X_{ik} \qquad \text{[Equation 1.2]}$$

where $\hat{Y}_i$ is the predicted value for the ith case.

When OLS regression is applied to a dichotomized dependent variable, the model is referred to as Linear Probability Model (LPM). The output produced by the analysis of an LPM model by the SPSS® computer software can be used to (a) identify which independent variables' coefficients are statistically significant, (b) classify subjects with respect to group membership, and (c) produce probability values that future subjects will be members of the group assigned the value of one in the dependent variable. Some of these pieces of information are more easily obtained than others.

The statistical testing of the independent variables' coefficients are a straightforward process when using the OLS regression output produced by the SPSS® computer software. A *t* test of each coefficient and the corresponding probability level is listed directly on the output. The classification of subjects, however, is not as easy. To classify subjects, researchers need to dichotomize the predicted probability values, which the SPSS® computer software calculates and lists in the data set. The researchers need to use the "Recode" subroutine to assign a value of one to any probability value of less than .50 and a value of one to any probability value greater than or equal to .50. Using the "Crosstabs" subroutine a classification matrix can be constructed that reveals the number and percentage of subjects correctly and incorrectly classified.

The SPSS® software does calculate and list the predicted probability values for the subjects included in the analysis and the holdout group (subjects not included in the analysis). If practitioners want to calculate probability values for future subjects, they would need to be supplied the coefficient values produced by the study. Assuming the coefficient values are supplied, the practitioners who use the SPSS® computer software would obtain the predicted probability values for the subjects by using the "Compute" subroutine to multiple the values of the independent variables of those subjects, which are stored in a data file, by the corresponding coefficients values. Thus the predicted probability values for future subjects can readily be calculated.

Potential Problems with OLS Regression Models

A model resulting from the application of OLS regression to a binary dependent variable, that is an LPM, poses three potential problems. The first potential problem is related to the issue of the type of information the technique provides. The other two potential problems are related to underlying assumptions of the technique. These potential problems are as follows:

1. A coefficient for a given independent variable indicates the change in the conditional probability of being classified in the group assigned the value of one in the dependent variable for a one-unit change in the independent variable. This change in the conditional probability is linear and unaffected by the initial conditional probability value. This characteristic of OLS produces two problems. First, while the probability value that a given subject belongs to the group assigned the value of 1 in the dependent variable will fall between 0 and 1, the predicted values are not restricted to this range. As noted by Austin, Yaffee, and Hinkle (1994), predicted values that fall below 0

and above 1 are illogical and not interpretable. We believe, however, that although such values may make some researchers uncomfortable, they may not affect the predictability of the model with respect to its classification accuracy. Second, a one-unit change in the independent variable will produce the same change in the conditional probability when the initial conditional probability is .90 as when it is .50.

2. The assumption of normality of the error term (ε) in OLS is not tenable for an LPM because the error term for a given set of independent variables can take on only two values. As noted by Gujarati (1988, p. 469) "although OLS does not require the disturbances [error term values] to be normally distributed, we assumed them to be so distributed for the purpose of statistical inference, that is, hypothesis testing."
3. Gujarati (1988) demonstrates that variance of the error term (ε) is heteroscedastic. Although this condition does not result in biased OLS estimates, they are inefficient. Thus the validity of the statistical tests conducted on the OLS coefficients is questionable.

Researchers who are considering using OLS regression rather than discriminant analysis or logistic regression should attempt to assess the impact each of these concerns may have on their analysis.

Discriminant Analysis

Similar to OLS regression discriminant analysis attempts to estimate the relationship between a dichotomized dependent variable and a set of independent variables. A discriminant analysis derives a linear combination of the independent variables, which is referred to as a discriminant function, that best discriminates between the groups contained in the dependent variable. The discriminant function takes the following form:

$$Z_i = a + W_1 X_{i1} + W_2 X_{i2} + \ldots + W_p X_{ip} \qquad \text{[Equation 1.3]}$$

where:

1. Z_i represents the discriminant scores for the ith subject.
2. The a symbol represents the intercept value.
3. W_p represents the discriminant weight for the p independent variable.
4. X_{ip} represents the value of the p independent variable for the ith subject.

Once the discriminant function's coefficients are estimated, a discriminant score, which is referred to as a discriminant Z score, is calculated for each subject using these estimated coefficients. The mean discriminant Z score is calculated for the members of each of the two groups. A mean discriminant Z score for a group is referred to as its centroid. The discriminant function, the discriminant Z scores, and the group centroids are used as the basis to (a) identify which independent variables that contribute to the difference between the group means are statistically significant, (b) classify people with respect to group membership, and (c) produce probability values that future subjects will be classified as members of the group assigned the value of one in the dependent variable.

Although the SPSS® computer software does not directly calculate whether the estimated coefficient for a given variable is statistically significant, assuming the researchers are not interested in a stepwise procedure, it can be calculated using the following formula:

$$F_{change} = \frac{(n - p - 2)(1 - \lambda_{p+1}/\lambda_p)}{(\lambda_{p+1}/\lambda_p)} \qquad \text{[Equation 1.4]}$$

where:

1. The symbol n is the total number of cases.
2. The symbol p is the number of independent variables in the model.
3. The symbol λ_p is Wilk's lambda before adding the variable to the model.
4. The symbol λ_{p+1} is Wilk's lambda after inclusion of the variable in the model.

Since the values for λ_p and λ_{p+1} can be obtained from the SPSS® computer software by analyzing one model that contains all the independent variables and additional models that delete only one of the independent variables, a statistical test of each independent variable's coefficient can be conducted.

Researchers can easily obtain classifications of subjects based on the discriminant function both for subjects included in the analysis and subjects withheld from the analysis. The discriminant Z score calculated for each subject is compared to the cut score to determine which group the subject is assigned. The cut score is the average of the two group centroids when the prior probability of any subject belonging to the group assigned the value of one is assumed to be .50 (it is a weighted average of the centroids when the probability is not set at .50). The SPSS® computer software computes the percentages of subjects in each group as well as the total correctly classified for subjects included in the analysis and subjects withheld from the analysis.

With respect to the probability values, the SPSS® computer software calculates a probability value for each subject included in the analysis. Norušis (1999) stated:

> "One way to compute these probabilities for each case [the probabilities that indicate the likelihood that each subject belongs to the group assigned a value of 1] is to first compute the Mahalanobis distance (D^2) to each group mean from the case, and then compute the ratio of exp($-D^2$) for the group over the sum of exp($-D^2$) for all the groups" (p. 259).

These probability values are listed in the output for the subjects in the holdout group as well as the subjects included in the analysis. It should be noted, however, that unless a practitioner has the original data set that was used to estimate the discriminant coefficients, calculation of the probability values for future subjects would be, to say the least, a difficult task.

Potential Problems with Discriminant Analysis

Researchers who choose to analyze a dichotomized dependent variable with discriminant analysis should consider three potential problems. The first potential problem deals with the techniques underlying assumptions and the other two relate to the difficulty in obtaining certain types of information. The potential problems are as follows:

1. As noted by Hair, Anderson, Tatham, and Black (1998) "discriminant analysis relies on strictly meeting the assumptions of multivariate normality and equal variance-covariance matrices across groups-- assumptions that are not met in many situations" (p. 276). Truett, Cornfield, and Kannel (1967) noted that the

assumption of multivariate normality is unlikely to be satisfied in actual data sets. If these assumptions are not met, Glessner, Kamakura, Malhotra, and Zmijewski (1988) stated that the coefficient estimates obtained by discriminant analysis are neither efficient nor consistent. Thus, as noted by Press and Wilson (1978), this condition may lead to the erroneous inclusion of meaningless variables in the discriminant function.

2. If the goal of a researcher is to provide estimates that could be used by a practitioner to calculate probability values of group membership for future subjects, the information produced by a discriminant analysis will make that task a daunting one. The practitioner would need to develop a computer program that uses the Mahalanobis distance values produced by the original study and calculate the Mahalanobis distance values for the subjects in which the practitioner is interested.
3. Researchers may find it rather difficult to inform practitioners of the change in the dependent variable associated with a given change in an independent variable in any meaningful way. That is, a coefficient generated by a discriminant analysis will indicate the change in the discriminant score associated with a given change in the independent variable. Practitioners may find it difficult to assess the practical significance of the change in those terms.

Once again, researchers who are considering using discriminant analysis rather than OLS regression or logistic regression should attempt to assess the impact each of these concerns may have on their analysis.

Logistic Regression

In a logistic regression analysis, the researcher directly estimates the probability of an event occurring, such as a subject belonging to the group assigned the value of one in the dependent variable. Specifically, the procedure used to calculate the logistic coefficients compares the probability of an event occurring with the probability of its not occurring for each subject. This ratio of the two probability values, which is referred to as the odds ratio, is transformed by calculating its natural logarithm value. As noted by Cizek and Fitzgerald (1999) "the logarithmic transformation of the odds ratio, called the 'log odds ratio', is used to express the odds on an equal interval scale. The transformation results in a scale with units called 'logits'--a contraction of the terms logistic and units" (p. 227).

A logistic regression model estimates the log odds, logit of p, as a linear combination of the independent variables:

$$\text{Logit }(p) = \beta_0 + \beta_0 + \beta_1\mathbf{X}_1 + \beta_2\mathbf{X}_2 + \ldots \beta_k\mathbf{X}_k \qquad \text{[Equation 1.5]}$$

where:

1. β_0, β_1, β_2, are maximum likelihood estimates of the logistic regression coefficients.
2. $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_k$ are column vectors of the values for the independent variables.

The results produced by the analysis of a logistic regression model can be used to (a) determine which coefficients are statistically significant, (b) estimate the probability that a subject will possess the characteristic represented by the value of one in the dependent variable, and (c) classify subjects with respect to group membership.

With respect to the goal of statistically testing and interpreting the coefficient of each independent variable with logistic regression, the logistic regression analysis produced by the SPSS® computer software produces a Wald test for each coefficient. The Wald test is the square of the ratio of its coefficient to its standard error. Similar to a t test of an OLS regression coefficient, a logistic coefficient is deemed to differ significantly from zero when the Wald test's probability value is less than the established alpha level.

With respect to the goal of estimating a probability of a subject belonging to the group assigned a value of one in the dependent variable, the logit values, can be converted to a probability as follows:

$$P_i = \frac{\exp(\text{logit}_i)}{1 + \exp(\text{logit}_i)} \qquad \text{[Equation 1.6]}$$

where:

1. P_i is the predicted probability that subject i belongs to the group assigned the value of one.
2. The symbol $\exp(\text{logit}_i)$ is the natural logarithm of the predicted logit for subject i.

The estimated probabilities could be used to classify subjects as being members of one group or the other. If the predicted probability value is less than .50, the subject would be classified as a member of the group assigned the value of zero in the dependent variable. If the probability is equal to or greater than .50, the subject would be identified as a member of the group assigned the value of one.

It is important to note that the logistic transformation of the dependent variable causes the coefficients estimated from a logistic regression model to differ from those obtained from an OLS regression model with respect to predicted probabilities. The coefficients estimated from an OLS regression model are linear with respect to the probability values, while the coefficients estimated for a logistic regression model are linear with respect to the *logit* values. The logistic regression coefficients are not, however, linear with respect to the probability values. That is, the change in the probability value of the event occurring associated with a given change in an independent variable is not constant across the range of initial probability values.

Potential Problems with Logistic Regression Analysis

Two characteristics of logistic regression that differ from OLS regression and discriminant analysis eliminate some of the concerns previously expressed for those two methods. First, the estimates obtained from a logistic regression model remain consistent and efficient even when the independent variables do not follow a multivariate normal distribution with equal variance-covariance matrices across groups (Gessner et al., 1988). Thus the statistical test results for the logistic coefficient may be less problematic than those obtained for discriminant analysis. Second, the probability values produced by a logistic regression analysis are bounded by the values of zero and one, which is not the case for probability values estimated by the OLS regression model.

A potential problem researchers who choose to analyze a dichotomized dependent variable with a logistic regression model must address is related to the type of information the logistic analysis produces. When researchers use logistic regression they must consider how to express the logistic coefficients in a form that will have meaning to practitioners. This interpretation problem is twofold:

1. As previously stated, a logistic coefficient for a given independent variable indicates the change in the log odds (logit) associated with a one-unit change in the variable. Few practitioners will find this interpretation to be meaningful with respect to evaluating the practical significance of the variable. We believe most practitioners would find it more meaningful to relate the change in the independent variable to the change in the probability that the subject is a member of the group assigned the value of one in the dependent variable.
2. The problem with converting a logistic coefficient to a probability value is that, as previously discussed, the change in the probability associated with a given change in the independent variable is not constant across the initial probability values. The issue for the researcher becomes: Can the change in the probability associated with a given change in the independent variable be communicated to practitioners and other researchers as one value and, if so, how should that value be calculated?

These two related interpretation problems should be addressed by researchers who contemplate using logistic regression to analyze a dichotomized dependent variable.

Applications of OLS Regression, Discriminant Analysis, and Logistic Regression

To further develop the questions that researchers should address when deciding whether to use results of OLS regression, discriminant analysis, or logistic regression, each method was applied to a set of data in which the dependent variable consisted of two values (a) the value of one indicated that a student stayed at Ashland University [AU] and (b) the value of zero identified a student as leaving AU.

Five independent variables were used. The labels for these variables and the information they included for each student were as follows:

1. High school grade point average (HSGPA)
2. American College Test score (ACT)
3. Gender of the student (GENDER).
4. Amount of AU aid in thousands of dollars (AUAID)
5. Amount of student financial need in thousands of dollars (NEED)

The HSGPA, ACT, AUAID, and NEED variables were metric; while the dichotomized gender variable contained zero and one values to represent females and males, respectively.

Prediction Validation

The sample of 525 students was divided into two data sets. The first set, which consisted of 443 students, was analyzed by each of the three analytic methods, that is, OLS regression, discriminant analysis, and logistic regression. The second set, which consisted of 82 students, was identified as a holdout group. The holdout group was used to evaluate each technique's ability to accurately classify students as either remaining or not remaining at AU.

Shrinkage Estimates

The results produced by each technique were cross validated through the use of shrinkage estimates (McNeil, Newman, & Kelly, 1996). The shrinkage estimates were calculated for each technique as follows:

1. The sample of 525 subjects was divided into two sample groups, which were identified as Sample 1 (n = 262) and Sample 2 (n = 263).
2. Each of the three analytic methods was used to analyze the data in Sample 1. Along with the coefficients generated by each technique the R^2 value produced by

the OLS regression, the chi-square value corresponding to the Wilks' lambda value produced by the discriminant analysis, and the Nagelkerke R^2 value produced by the logistic regression were recorded.

3. The coefficients produced by a given technique for Sample 1 were used to generate a value for each of the subjects in Sample 2. The set of values were generated by multiplying each subject's data by the corresponding coefficient and adding the constant to the sum of these products. Theses values formed a variable, which was labeled NEWVAR.
4. A model was designed with the dichotomized dependent variable, which identified group membership, and the NEWVAR variable serving as the independent variable. Each of the three analytic techniques was used to analyze this model for the data contained in Sample 2. The R^2 value produced by the OLS regression, the chi-square value corresponding to the Wilks' lambda value produced by the discriminant analysis, and the Nagelkerke R^2 value produced by the logistic regression were recorded.
5. The shrinkage estimate was calculated for each technique as follows:
   (a) OLS regression -- The R^2 value of Sample 2 was divided by the R^2 value of Sample 1.
   (b) Discriminant analysis -- a contingency coefficient was calculated from the results produced for Sample 1 by dividing the chi-square value corresponding to its Wilks' lambda value by the sum of the chi-square value and the sample size. A contingency coefficient value calculated for

Sample 2 in the same manner was divided by the contingency coefficient value for Sample 1.

(c) Logistic regression -- The Nagelkerke R^2 value for Sample 2 was divided by the Nagelkerke R^2 value for Sample 1.

The shrinkage estimates were compared to assess the relative abilities of the three techniques to produce stable results.

Violations of Underlying Assumptions

A number of characteristics regarding the assumptions of normality and equal variance-covariance matrices of the data set that contained 525 cases should be noted. First, one of the independent variables consisted of just two values and the univariate distributions of the four metric independent variables differed from normality according to Shapiro-Wilk and Lilliefors test results. In spite of the fact that examinations of the normal probability plots for the four metric independent variables appeared to indicate the departure from normality did not appear to be severe, the data did not strictly adhere to the multivariate normal distribution assumption.

Second, Levene test results indicate the assumption of equal variance for two of the independent variables is violated. In addition, based on the value of Box's M statistic, the null hypothesis of equal group covariance matrices is rejected. Thus the assumption of equal variance-covariance matrices is questionable. Third, the highest Variance Inflation Factor (VIF) for any of the five independent variables was less than 2.00. Thus a high degree of relationship between the independent variables, that is multicollinearity, did not exist.

Data Analysis

As previously stated, researchers analyze a dichotomized dependent variable for at least one of the following goals: (a) to identify the statistically significant independent variables and be able to judge their practical significance, (b) to accurately classify future subjects as members of the two groups identified in the dependent variable, and (c) to predict probability values for future subjects that will indicate their membership in the group assigned the value of one in the dependent variable. To compare the results produced by the three analytical methods (OLS regression, discriminant analysis, and logistic regression), the data set containing 443 cases was analyzed by each method. The results of the analyses are listed in Table 1.

Insert Table 1 about here

Statistical and Practical Significance of the Coefficients

If the researchers' goal was to identify the statistically significant independent variables and to be able to judge their practical significance, the question is: Does it matter which method is used to analyze the data? To address the first portion of this question, that is, the identification of the statistically significant coefficients, the coefficients of the independent variables produced by each analytic method and their corresponding *p* values are listed in Table 1.

It should be noted that the signs of the discriminant analysis coefficients are opposite the signs of the coefficients produced by the other two methods. This is the result of the signs of the group centroids being -.162 for the group remaining at AU (the group designated as Group 1) and +.175 for the group of students who did not remain at AU (the group designated as Group 0). The signs of these group centroids indicate that a one-unit increase in, say, the AUAID variable is associated with a .32 *decrease* in the discriminant score, which moves the discriminant score further from the centroid of Group 0 and thereby increases the probability of being classified as remaining at AU. Thus the direction of the changes in the probabilities of being classified as belonging to the group assigned the value of one are the same for the coefficients regardless of which analytical method was used.

An examination of the *p* values listed in Table 1, which were generated by the statistical tests of the coefficients, reveals that all three methods identified the same two independent variables (AUAID and GENDER) as being statistically significant at the .05 level. In addition, the three analytic methods produced similar *p* values for the corresponding coefficient.

With respect to the practical significance portion of the question, we believe the most meaningful way to relate practical significance of a given variable to practitioners is through the level of change in the probability of being classified as a member of the group assigned the value of one associated with a given change in the variable. The ease of assessing the practical significance of each statistically significant coefficient in this manner is not equivalent for the three analytic methods.

Since the OLS regression used with an LPM reflects a *linear* relationship between a linear combination of independent variables and the dependent variable, its coefficients are the easiest to gauge with respect to practical significance. To illustrate, the coefficient value of .0268 for the AUAID variable indicates that a $1000 increase in AU financial aid is associated with a .0268 increase in the probability of being classified as remaining at AU.

The discriminant coefficient for AUAID (-.327) indicates that a $1000 increase in AU financial aid is associated with a .327 of a point decrease in the discriminant score. We believe that such a value is more difficult for practitioners to judge with respect to its practical significance than the corresponding change in the probability produced by the OLS regression analysis. Changing the discriminant coefficient into a probability value is not an easy task. We are not aware of any computer program or software that produces such a conversion.

A coefficient obtained from a logistic regression analysis indicates the change in the logit (log odds) for a one-unit change in the independent variable. Thus the .111 coefficient value for the AUAID variable indicates the change in the logit associated with a $1000 increase in AU financial aid. We believe that practitioners will find this value to be difficult to judge from a practical standpoint. This value can readily be converted, however, to a probability value.

When converting a change in the logit value to a probability value, it is important to note that a logistic coefficient is linear with respect to the logit values and not the probability values. This fact causes the changes in the probability values to vary across the range of initial probability values. Thus when converting a change in the logit value into a probability value, an initial probability value needs to be specified (Petersen, 1985). Establishing the initial probability at the mean of the dependent variable, which is the proportion of cases in Group 1 where the change in the initial probability corresponds to a one-unit change in the dependent variable, can be calculated as follows:

$$P_c = \frac{e^{\ln(Pi/(1-Pi))+b}}{1+e^{\ln(Pi/(1-Pi)+b}} - Pi \qquad \text{[Equation 1.7]}$$

where:

1. P_c represents the change in the probability value.
2. Pi represents the initial probability value. This value is set equal to the proportion of the sample belonging to the group assigned the value of one in the dependent variable (i.e., the sample mean of the dependent variable).
3. The symbol *exp* represents the base of the natural logarithm.
4. The symbol *b* represents the logistic regression coefficient for the given predictor variable.

The value produced by Equation 1.7 is referred to as the Delta-p statistic. The Delta-p statistic for the AUAID logistic coefficient value of .111 would be calculated as follows for the initial probability value of .519 (proportion of subjects in Group 1):

$$P_c = \frac{e^{\ln(.519/(1-.519))+.111}}{1+e^{\ln(.519/(1-.519))+.111}} - .519$$

$$P_c = .028$$

This Delta-p statistic value indicates that a $1000 change in AUAID is associated with a .028 increase in the probability of being classified as remaining at AU when the initial probability of remaining at AU is .519.

Two points should be noted with respect to this Delta-p statistic value. First, the value of .028 is almost identical to the OLS regression coefficient for the AUAID variable (.027). Second, as previously noted, the change in the initial probability for a given change in the independent variable is not constant across the range of initial probability values.

With respect to the second point, Fraas, Drushal, and Graham (2002) designed a Microsoft Excel® program that can be used to assess the degree of variation in these probability change values over a range of initial probability values. As noted by Fraas, et al., these probability change values should be calculated for the range of predicted probability values generated by the logistic regression analysis. For the study under consideration the minimum and maximum predicted probability values obtained from the logistic regression model were approximately .25 and .75, respectively. The probability change values calculated for this range of predicted probability values reached a minimum of .020 and a maximum of .028. Since these probability values are close to the Delta-p value of .028, a practitioner could appropriately use this value as the one to judge the practical significance of the AUAID variable.

Ability to Accurately Classify

Gessner, et al. (1988) stated "some researchers have found that a range of alternative techniques produce similar abilities to classify observations correctly" (p. 49). A review of the classification accuracy results generated by the three analytic methods for the holdout group, which are contained in Table 2, indicate that the results are in line with this statement.

Insert Table 2 about here

The percentage of total correctly classified cases for the OLS regression (56.1%), discriminant analysis (57.3%), and logistic regression (56.1%) were nearly identical. The methods did differ somewhat, however, with respect to correct classification when considering each group separately. The discriminant analysis was slightly more accurate in classifying students who did not remain at AU (48.7%) than either the OLS regression (41.0%) or logistic regression methods (41.0%). The reverse was true, however, for classifying students who did remain at AU. That is OLS regression (69.8%) and logistic regression (69.8%) results were slightly more accurate than discriminant analysis (65.1%). These results are consistent, at least when comparing discriminant analysis and logistic regression, to the conclusion stated by Press and Wilson (1978): "It is unlikely that the two methods [discriminant analysis and logistic regression] will give markedly different results . . . unless there is a large proportion of observations whose x-values lie in regions of the factor space with linear logistic response probabilities near zero or one" (p. 705).

Probability Values

The probability values generated by each of the three analytic methods were highly related. Correlating the set of probability values generated by each method with the corresponding probability values generated by the other methods produced three correlation coefficients that exceeded .99. These results existed for the group of students used in the analysis as well as the group of students who constituted the holdout group.

With respect to relative levels of the three sets of probability values, the OLS regression and logistic regression methods produced nearly identical mean values of approximately .519; while the mean probability value produced by the discriminant analysis (.501) was nearly .02 lower. It appears that for this set of data the selection of one analytical method over the others would produce small differences in the predicted probabilities.

Stability of the Results

One key issue regarding results produced by an analytic technique is whether the results are stable (Newman, McNeil, & Fraas, 2003). To assess the relative stability of the results produced by the three analytic techniques, shrinkage estimates were calculated for each technique as previously described. A review of the shrinkage estimates for the OLS regression (.704), discriminant analysis (.699), and logistic regression (.694) were nearly identical. Thus for the AU data set each analytic method appeared to produce results with basically the same degree of stability.

Key Questions to Address

We believe researchers and practitioners should address two questions when deciding whether to use OLS regression, discriminant analysis, or logistic regression. What impact do the violations of underlying assumptions have on the results? Does a given technique readily produce the type of information required to address the research question? This section discusses these two questions as they relate to the three possible goals of the research study that uses OLS regression, discriminant analysis, or logistic regression to analyze a dichotomized dependent variable and a set of independent variables.

Key Questions as Related to the Goal of Identifying Significant Coefficients

Each of the three methods can be used to determine which independent variable coefficients are statistically significant with the SPSS® computer software. Such an assessment is easier to accomplish with OLS regression and logistic regression than with discriminant analysis. Although the SPSS® computer software does not statistically test each individual coefficient in the discriminant model, a researcher can test each coefficient by applying Equation 1.4 to the results as previously discussed.

The more important issue regarding the statistical testing of the coefficients, however, is the effect that any violations of assumptions have on the testing results. We believe researchers will find their greatest concern regarding the violations of underlying assumptions centers on the OLS regression and discriminant analysis methods. As previously discussed, Gujarati (1988) demonstrated that the assumption of normality of the error term for OLS regression models used to analyze dichotomized dependent variables (LPM models) is not tenable. Gujarati also demonstrated that homoscedastic variance of the error term assumption cannot be maintained in LPMs.

With respect to the normality violation, Gujarati noted that "as sample size increases indefinitely . . . the OLS estimators tend to be normally distributed generally" (p. 470). Thus for large samples the violations of the assumption of normality of the error term may not pose a substantial problem in the statistical testing of the OLS regression coefficients. To deal with the problem of unequal variance-covariance matrices Goldberger (1964) proposed a two-step weighted least squares approach as a means of dealing with this problem. Austin et al. (1994), however, found this solution somewhat unwieldy.

Regarding the use of discriminant analysis in a situation in which the assumptions of multivariate normal distribution with equal variance-covariance matrices are not met, Press and Wilson (1978) stated that "the discriminant function estimators . . . will not be consistent . . . [and] meaningless variables will tend to be erroneously included in . . . [the] discriminant function" (p. 701). Lachenbruch (1975) reported, however, that discriminant analysis is a rather robust technique that can tolerate some deviation from these assumptions.

As previously noted, the results of the statistical tests of the coefficients produced by the application of the three analytical methods to the AU data set examined in this paper were very similar (see Table 1). Thus the impact of violating the underlying assumptions may not be critical to the identification of statistically significant coefficients in all situations. If researchers are concerned that such violations place into question test results produced by the OLS regression and discriminant analysis methods, however, they may want to consider using logistic regression.

With respect to providing practitioners with coefficients that can be assessed in terms of practical significance, we took the position that this assessment is best done by revealing the change in the probability of being classified in the group assigned a value of one that is associated with a given change in an independent variable. As previously discussed, the coefficients produced by discriminant analysis do not directly reveal such information and they do not lend themselves to such a conversion.

On the other hand, the OLS regression and logistic regression methods can readily provide this type of information. An OLS regression coefficient directly indicates the change in the probability the subject is a member of Group 1 that is associated with a given change in the independent variable. Researchers should be mindful that the change in the initial probability indicated by the OLS regression coefficient is constant. That is, it does not vary along with the initial probability level. Thus a predicted probability value can be less than 0 or greater than 1. If researchers are willing to assume a linear relationship between the independent variables and the probability values of being a member of Group 1, OLS regression may be appropriate to use. We believe the linear assumption is most appropriate when the predicted probability values are located between approximately .25 and .75.

If researchers are not willing to assume a linear relationship, logistic regression is a more appropriate technique to use. As previously discussed, although the logistic regression coefficient does not reveal the change in the probability value directly, it can be converted into such a value. The logistic regression coefficient indicates the change in the logit (log odds value) for a given change in the independent variable. To express this change as a probability, a corresponding Delta-p value can be calculated.

The Delta-p value indicates the change in the *initial* probability that the person would be classified as a member of Group 1 for a given change in the independent variable (where the initial probability is set equal to the mean of the dependent variable). It is important to determine, however, if the changes in the probability values associated with various initial probability levels differ considerable from the Delta-p value. If so, the researcher may want to report a series of changes in the probability values associated with initial probability values that fall within the range of predicted probabilities calculated by the logistic regression model (Fraas et al., 2002).

Key Questions as Related to the Goal of Accurately Classifying Future Cases

As previously noted, possible violations of the underlying assumptions are a greater concern when OLS regression or discriminant analysis is used. The assumptions most likely to be violated when OLS regression is used are the assumptions regarding the normality and the homoscedasticity of the error term. If the OLS model is not determined by statistical testing of the independent variables, we believe less than extreme violations of these assumptions should have a minimal impact on the accuracy of the classification of future subjects.

The impact of the violations of multivariate normal distribution and equal variance-covariance matrices assumptions are discussed by Klecka (1980). He argues that the violation of these assumptions can affect classification accuracy. As previously noted, however, discriminant analysis is a rather robust technique (Lachenbruch, 1975). The similar classification results produced by our application of the three analytic methods may be evidence of support for Lachenbruch's position and the findings reported by Wilensky and Rossiter (1978). If a researcher remains concerned with using discriminant analysis in the presence of these assumption violations, Stevens (1996) suggests that under such conditions "an alternative classification procedure [logistic regression] is desirable" (p.287).

With respect to providing essential classification information, the discriminant analysis and logistic regression results produced by the SPSS® computer software directly provide such information. As previously discussed, the researcher must generate the classification tables for both the group analyzed and the holdout group when using the OLS technique with the SPSS® computer software.

Key Questions as Related to the Goal of Predicting Probability Values for Future Subjects

The degree of impact of the violations of underlying assumptions is not clear with respect to the predicted probabilities produced by the three analytic techniques. Again, we believe that mild violations of assumptions will likely produce similar probability values with the use of any of the three analytic techniques. This was the case for the results produced for the AU data by each of the analytic techniques. Since the predicted probabilities were so highly correlated and their mean values were similar, the relative impact of the violations noted for the AU data set on the predicted probability values produced for the holdout group appears to be minimal for the three analytical techniques. If the data reflect substantial violations in the assumptions, however, researchers may want to use the values produced by logistic regression to predict probabilities for future subjects.

The information that a practitioner needs to be able to forecast the probability for future subjects is best accomplished with either the OLS regression or the logistic regression methods. Practitioners could predict the probability for a future subject as long as the researcher reported either the OLS regression coefficients or the logistic regression coefficients. Multiplying each OLS regression coefficient with its corresponding value for the future subject and adding those values to the constant would provide a predicted probability.

If the practitioner multiplies each logistic regression coefficient by its corresponding value for the future subject and adds these products to the constant, the predicted logit will be obtained. By substituting this value into Equation 1.6, the practitioner will obtain the predicted probability for the subject.

As previously noted, the predicted probability values produced by a logistic model are restricted to the range of 0 and 1. Again, practitioners should remember that when the OLS regression method is used, the predicted probability values are not restricted to the range of 0 and 1. To deal with this problem some have suggested that predicted probabilities outside the 0 and 1 range should be truncated (Anderson, Auguier, Hauck, Oakes, Vandaele, Weisberg, 1980; Kennedy, 1985).

We believe that if predicted probabilities are between approximately .25 and .75, OLS regression and logistic regression will often produce similar predicted probability values due to the fact the relationship between the linear combination of the independent variables and the probability values of group membership will tend to be linear within that range. If the predicted probabilities fall outside the range of approximately .25 and .75, however, logistic regression may be the preferred method for producing predicted probability values for two reasons. First, the relationship between a given independent variable and the probabilities will be less likely to fit a linear function. Second, the likelihood of predicting probability values outside of the range of 0 and 1 increases.

Discriminant analysis, as produced by the SPSS® computer software, does not readily provide the information needed by a practitioner to generate probability values for future subjects. The difficulty lies in the need for the calculation of the Mahalanobis (D^2) distance values for the future subjects. Thus we do not see discriminant analysis as a viable method for meeting the research goal of providing information that practitioners could use to predict probability values for future subjects.

The Issue of Replicability

Studies comparing the replicability of the three analytic methods under various data conditions need to be conducted. The replication estimates for the results produced by the application of the three analytic techniques to the AU data revealed similar degrees of replicability. Additional studies need to be conducted, however, on the stability of the results produced by these three techniques when applied to data with various characteristics.

Regardless of which goal is established by researchers for a given study and which analytic method is used, we believe they should assess and report the stability of their results. Regardless of whether the researchers select OLS regression, discriminant analysis, or logistic regression, such an assessment can be made through the calculation of a shrinkage estimate, as previously described. If the shrinkage estimate indicates the results are unstable, the results produced by the analysis should be used cautiously by practitioners. If researchers routinely reported shrinkage estimates, we believe research practices would be strengthened.

Summary

Researchers engaged in studies that involve dichotomized dependent variables and a set of independent variables can choose from a number of analytic methods. Three such methods are OLS regression, discriminant analysis, and logistic regression. The issue addressed in this paper is what questions should be addressed when selecting from among these analytic methods. We believe that researchers need to address the adequacy of each technique with respect to two basic questions. What impact do possible violations of underlying assumptions have on the results? Does a given technique readily produce the type of information required to address the research question?

We believe that before these questions can be addressed researchers must establish a clear goal or goals for the analysis. Three goals of studies that involve dichotomized dependent variables and a set of independent variables are (a) to identify the statistically significant independent variables and be able to judge their practical significance, (b) to accurately classify future subjects as members of the two groups identified in the dependent variable, and (c) to predict probability values for future subjects that will indicate their chances of belonging to the group assigned a value of one in the dependent variable.

When the goal of a study is to identify statistically significant independent variables, the assumptions most likely to be violated and have an impact on the method of analysis are the ones related to OLS regression and discriminant analysis. Although the OLS regression and discriminant analysis may be somewhat robust to violations of the underlying assumptions, logistic regression may be an appropriate alternative when a researcher is concerned about the impact that such violations may have on the results. If the goal is also to provide information that can be used to judge the practical significance of the statistically significant independent variables, we believe that OLS regression and logistic regression conducted with the SPSS® computer software provides information that can best meet that goal.

With respect to the goal of accurate classification, we believe that if the violations of underlying assumptions are not severe, the three analytic methods produce similar classification results. With respect to the usefulness of the information provided by the three analytic methods used in conjunction with the SPSS® computer software, all three methods provide information that a practitioner can use to accurately classify future subjects.

If the goal of a study is to provide information to practitioners that could be used to predict probability values for future subjects, our review of the methods used in conjunction with the SPSS® computer software suggests that OLS regression and logistic regression can best meet this goal. If the predicted probabilities approach 0 or 1, logistic regression may be the preferred analytic method for supplying such values.

Although the shrinkage estimates reported in this paper revealed little difference in the stability of the results produced by the three analytic techniques, future studies on the relative stability of the results produced by the techniques should be conducted. Regardless of the findings of such studies, however, we believe researchers should report shrinkage estimates. Reporting such estimates would enable other researchers and practitioners to gauge the stability of the results. No matter which analytic method is used, if the results are not stable, researchers and practitioners must be cautious in the use of such results.

This paper attempted to identify and address key questions researchers and practitioners should consider when deciding whether to use OLS regression, discriminant analysis, and logistic regression in studies that involve a dichotomous dependent variable and a set of independent variables. We hope our discussion of these analytic methods and the questions related to the selection process provides researchers, practitioners, and graduate students with a useful framework by which to determine which analytic method may be most appropriate for a given research project.

References

Anderson, S., Auguier, A. Hauck. W. A., Oakes, D., Vandaele, W., & Weisberg, H. I. (1980). *Statistical methods for comparative studies.* New York: Wiley.

Austin, J. T., Yaffee, R. A., & Hinkle, D. E. (1992). Logistic regression for research in higher education. *Higher Education: Handbook of Theory and Research, 8*, 379-410.

Cabrera, A.F. (1994). Logistic regression analysis in higher education: An applied perspective. In J.C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (pp. 225-254). New York: Agathon Press.

Chatterjee, S., Hadi, A. S., & Price, B. (2000). *Regression analysis by example.* New York: John Wiley & Sons, Inc.

Cizek, G.H., & Fitzgerald, S.M. (1999, January). An introduction to logistic regression. *Measurement and Evaluation in Counseling and Development, 31*(4), 223-244.

Fraas, J. W., Drushal, J. M., & Graham, J. (2002, October). *Expressing logistic regression coefficients as changes in initial probability values: Useful information for practitioners.* Paper presented at the meeting of the Mid-Western Educational Research Association, Columbus, OH.

Gessner, G., Kamakura, W. A., Malhotra, N. K., & Zmijewski, M. E. (1988). Estimating models with binary dependent variables: Some theoretical and empirical observations. *Journal of Business Research 16,*(1), 49-65.

Goldberger, A. S. (1964). *Econometric theory. New York: Wiley.*

Gujarati, D. N. (1988). *Basic econometrics.* New York: McGraw-Hill.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Kennedy, J. J. (1985). *Analysis of qualitative data.* New York: Praeger.

Klecka, W. R. (1980). *Discriminant analysis.* Beverly Hills, CA: Sage Publications.

Lachenbruch, P. A. (1975). *Discriminant analysis.* New York: Hafnew Press.

Manski, C. F., & Wise, D. A. (1983). *College choice in America.* Cambridge, MA: Harvard University Press.

McNeil, K., Newman, I., & Kelly, F. J. (1996). *Testing research hypotheses with the general linear model.* Carbondale, IL: Southern Illinois University Press.

Newman, I., McNeil, K., & Fraas, J. W. (2003, April). *Déjà vu: Another call for replications of research, again.* Paper presented at the annual meeting of the American Educational Research Association, Chicago: IL.

Norušis, M.J. (1999). *SPSS regression models™ 10.0.* Chicago, IL: SPSS, Inc.

Peng, C. Y., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research 96* (1), 3-14.

Petersen, T. (1985). A comment on presenting results from logit and probit models. *American Sociological Review, 50*(1), 130-131.

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

Stevens, J. (1996). *Applied multivariate statistics for the social sciences.* Mahwah, NJ: Lawrence Erlbaum Asociates.

Truett, J., Cornfield, J., & Kannel, W. (1967). A multivariate analysis of the risk of coronary heart disease in Framingham. *Journal of Chronic Diseases, 20,* 511-524.

Wilensky, C. R., & Rossiter (1978). OLS and logit study. *American statistical association 1978 proceedings of the social statistics section,* 260-265.

Table 1

OLS Regression, Discriminant Analysis, and Logistic Regression Coefficients[a]

	Coefficient and p values[b]		
Variables	OLS regression	Discriminant analysis[c]	Logistic regression
HSGPA	-.113 (.112)	1.376 (.101)	-.468 (.110)
ACT	.0036 (.697)	-.041 (.503)	.014 (.698)
AUAID	.0268 (.022)	-.327 (.021)	.111 (.023)
NEED	-.0004 (.246)	.0501 (.246)	.0171 (.241)
GENDER	-.150 (.007)	1.827 (.006)	-.612 (.007)
Constant	.715	-2.391	.893

[a]Dependent variable values are 0 and 1 when the students did not remain and did remain at AU, respectively.

[b]$N = 443$; $n_0 = 213$, and $n_1 = 230$

[c] Since the centroids for the Group 0 and Group 1 were .175 and -.162, respectively, the signs of the coefficients for the discriminant analysis will be opposite of the signs of the coefficients for OLS regression and logistic regression.

Table 2

Classification Accuracy for Subjects in the Holdout Sample[a]

Group	Observed number	OLS correctly classified	Discriminant correctly classified	Logistic correctly classified
Did not remain at AU	39	16 (41.0%)	19 (48.7%)	16 (41.0%)
Did remain at AU	43	30 (69.8%)	28 (65.1%)	30 (69.8%)
Total	82	46 (56.1%)	47 (57.3%)	46 (56.1%)

[a] First figure is the number of subjects correctly classified; while the second figure, which is enclosed in the parentheses, is the percent correctly classified.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC
Educational Resources Information Center

TM034821

REPRODUCTION RELEASE
(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title:	Ordinary Least Squares Regression, Discriminant Analysis, and Logistic Regression: Questions Researchers and Practitioners Should Address When Selecting an Analytic Technique
Author(s):	John W. Fraas and Isadore Newman
Corporate Source: Ashland University, Ashland, Ohio, 44805	Publication Date: 2/28/03

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
X		
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature: John W Fraas	Printed Name/Position/Title: John W. Fraas, Trustees' Professor	
Organization/Address: Ashland University, 401 College Avenue, Ashland, Ohio, 44805	Telephone: 419-289-5930	FAX: 419-289-5989
	E-Mail Address: jfraas@ashland.edu	Date: 3/6/03

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse: University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com