The Workshop

JOHN ALDRICH
CHARLES F. CNUDDE
Michigan State University

*Probing the Bounds of Conventional Wisdom: A Comparison of Regression, Probit, and Discriminant Analysis**

The level of measurement of the dependent variable (nominal, ordinal, interval) crucially affects the selection of statistical techniques. Conventional wisdom further restricts the choice of an appropriate technique. In this Workshop paper, we compare three powerful statistical techniques appropriate for each of the levels of measurement and sharing, insofar as possible, a similar set of assumptions. The three techniques are ordinary least squares regression for intervally measured, probit for ordinally measured, and discriminant analysis for nominally measured dependent variables. The assumptions and uses of each technique are reviewed, and an example of the use of each in political research is presented.

Discussions of the levels of measurement problem in political science frequently lead to one of two conclusions. On the one hand, "radicals" argue that the increased leverage obtained from using statistical procedures which assume that the dependent variable is measured on an interval scale outweighs the consequences of their application to nonintervally measured variables. On

*These statistical evaluation procedures were initially explored in order to solve substantive problems that are only partly referred to in this paper. Further applications will be developed in subsequent reports. Part of the research reported herein was performed pursuant to a grant contract with the National Institute of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

the other hand, "purists" argue that the consequences of such a misuse are too serious to ignore, and that nominal or even ordinal dependent variables can only be analyzed through the use of the relatively less powerful techniques, such as cross-tabulation or ordinal measures of association and correlation. In this paper we discuss three techniques, all of which are useful for testing multivariate and quite sophisticated hypotheses. These three alternative statistical procedures–linear regression, probit analysis, and discriminant analysis–are suitable for the analysis of interval, ordinal and nominal dependent variables, respectively. Underlying our presentation is our belief that decisions concerning the choice of an appropriate statistical technique must consider not only the nature of the measurement of the dependent variable, but also the assumptions contained in the model underlying each procedure. Further, we would argue that it is of equal importance to ensure that these assumptions are compatible with the substantive assumptions of the theory/hypotheses under investigation.

Linear Regression

To begin this explication, let us first consider the linear regression model in a context of an application to political data.

One way to think of a linear function is to consider the identity between the predictions of a "true" theoretical model and the actual observations that the model addresses.[1] If the model were "true," then when we plot its predictions and the actual scores the intersections between these two sets of points would all fall on the main diagonal; i.e., they would form a particular type of a linear function. If there were some error in the predictions, then the validity of the model would depend upon whether the scatter plot of intersections fell about the main diagonal in a "random" way. This example is a special case of the general linear model,

$$y = a + bx + e,$$

which would apply whenever we can so account for actual observations with values predicted from a theoretical model. In this model, a "true" identity would mean that the data would all fall on the main diagonal, and the slope would intersect at the origin. These considerations would mean a slope coefficient (or b value) of 1.0, and an intercept of 0, respectively.

[1] The term "theoretical model" here means some theory which leads to predictions independent of the "statistical theory" (i.e., linear regression) which can be used to *test* the relationship between the theoretical predictions and reality.

A linear model of this type may be developed to predict U.S. House of Representatives seat gains and losses in "off-year" elections. The model's basic assumption is of stability in partisan preferences. Losses in the mid-term elections, then, are simply due to fluctuations back to a normal or expected level after a party gained more than that amount in the preceding presidential election. Therefore we calculate how many seats were gained due to the swing away from the expected partisan split in the presidential year by any party. That gain is what it should lose in the following midterm election, assuming a stable two party system and no systematic "short run forces," such as recessions or Watergate in the off year.

With this logic we obtain a set of predicted losses. If the model is correct, the actual losses should line up with these predictions in a roughly linear way, and in fact should approximate the special case of an identity.

The scatter plot of Democratic losses for the elections from 1938 to 1970 is shown in Figure 1. There are two major deviant cases, 1946 and 1958. In both years, the assumption of no systematic "short run forces" probably is invalid, and those cases are dropped.

To illustrate the theoretical model, we can briefly consider one particular case. In the Roosevelt landslide of 1936 the Democrats made considerable gains in House seats. Some portion of the total Democratic majority that year undoubtedly came from districts Democrats would "normally" win because of the differential affiliation of the electorate to the major parties. This normal or expected proportion of seats for a given party is assumed constant in the model. Our best evidence leads to the inference of relative stability in this proportion since the Great Depression, and so as a means of making a first approximation we may assume it to be a constant. However, refinements could and should be attempted by relaxing this assumption.[2] If this proportion is set as a constant, then the difference between such a value and the percent of the seats the Democrats received in 1936 would be due to factors other than that normal expectation due to party affiliation.[3] In general, we can conceptualize this difference as due to the peculiarities of the particular

[2] The assumed constant expected number of seats for a party would change markedly over time if the partisan split of the electorate markedly changed, or if apportionment rules changed in ways greatly uncorrelated with the partisan division in affiliation, or if electoral outcomes fell into extreme ranges where the "cube rule" curve is not well approximated by a linear function.

[3] The value adopted for the analysis is 55% of the seats in the House of Representatives to the Democratic Party. The value is approximately the same as the "normal vote" for the Democrats in the electorate, as estimated for the period 1952 to 1960 from survey data (Converse, 1966).

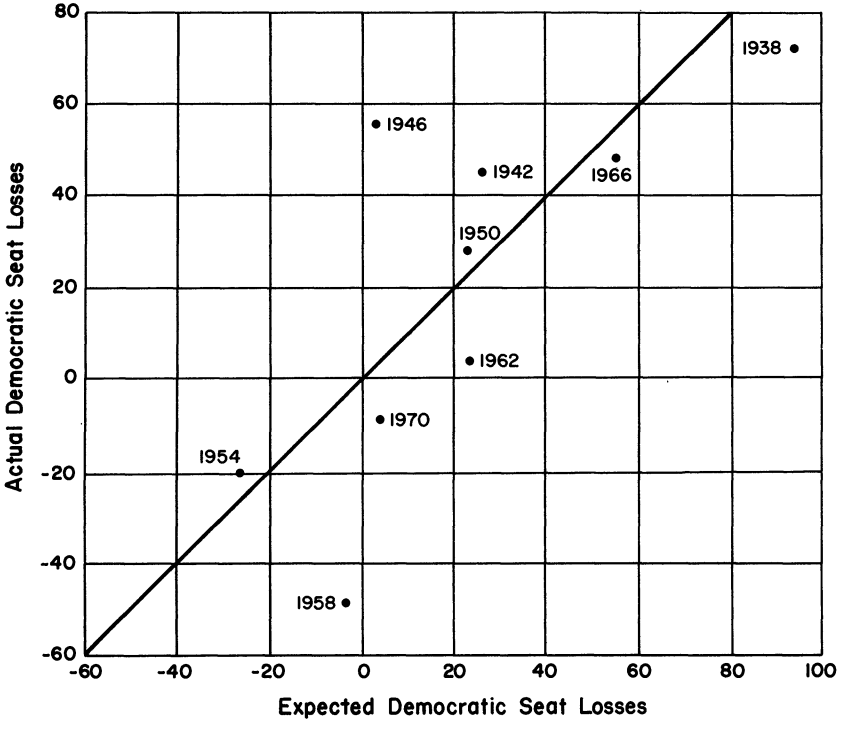

FIGURE 1
Expected and Actual Seat Losses

election of 1936; i.e., all those factors that combined to make up the Roosevelt landslide. Our calculations indicate that, given the number of seats in the House at that time, the number due to the "short run" effects of 1936 was about an 86-seat gain to the Democrats over and above what they would have normally received in the current partisan era.

In the long run, a gain of this sort would be lost in another election where the peculiar factors are apt to be different. As a second approximating assumption, we could hold that it would be entirely coincidental for two elections in a row to have short run factors that would operate to give the same outcome. Considering the gains in a given election, we can make a prediction about the losses in the next under this assumption: our best prediction, knowing nothing else about the next election, is that the outcome will return to the normal expectation. Thus in the long run the normal

expectation is our best predictor of outcomes. The difference between the normal expectation and the actual outcome in a given election is therefore our best prediction of what would be lost or gained in the next election.

In 1936, if the Democrats gained 86 more seats than they should have, given the normal expectation, and if the normal expectation is our best prediction of the 1938 election, then they would tend to lose those 86 seats in 1938. In fact they lost only 62 seats in 1938. However, to examine the utility of this approach to predicting off-year elections, we would have to compare the losses given by the model with the actual losses for a whole series of elections. If we consider the actual losses as the values to be predicted, and the values given by the model as the predictor variable, we can use the linear regression model as a device for making this examination.

The linear regression model would link the actual losses (Y) to the losses given by the model of election outcomes (X), some linear coefficients (α and β), and an error term (e) which stands for all other influences on the actual outcomes. Thus, we have a version of the standard linear equation:

$$Y = \alpha + \beta X + e \tag{1}$$

If in the long run our normal expectation is the best predictor of actual outcomes, then the sum of all other influences across a large number of elections would equal zero. Formally,

$$\Sigma e = 0 \tag{2}$$

If in the long run the deviation from the normal expectation for a given election would only coincidentally be related to that deviation in the next election, then the covariation between e and X would sum to zero in a large number of elections. Formally,

$$\Sigma Xe = 0 \tag{3}$$

Denoting estimated values by hats, writing the linear equation in terms of estimates and rewriting it to set all values equal to the estimated residuals gives

$$e = Y - \hat{\alpha} - \hat{\beta} X \tag{4}$$

Substituting (4) into the versions of (2) and (3) which correspond to estimated values gives

$$n\hat{\alpha} + \hat{\beta} \Sigma X = \Sigma Y \tag{5}$$

$$\hat{\alpha}\Sigma X + \hat{\beta} \Sigma X^2 = \Sigma XY \tag{6}$$

Except for the estimated values of the regression coefficients, all values in the equations can be calculated from the data obtained for the election years in the scatter plot. Since we have two equations and two unknowns, solving for the two unknowns gives the values of the two linear regression coefficients:

$$\hat{\alpha} = 0.9$$

$$\hat{\beta} = 0.8$$

These values come reasonably close to our theoretically derived values; a value of 0.9 for $\hat{\alpha}$ is reasonably close to 0.0 (given a range of −20.0 to +62.0 for actual seat losses), and a value of $\hat{\beta}$ of 0.8 is reasonably close to 1.0. We can tentatively conclude that the linear regression model corresponds to the model of off-year congressional seat losses.[4]

The success of a regression such as this example can be judged by examining the error in fit. We expect some error, of course, in fitting predicted dependent variable values to the actual observations. If we square such error and sum the squared errors, we can examine this "residual variance" as a proportion of the total original variance in Y. That is, we have divided the original variance into "explained" variance and "unexplained" or error variance. The proportion of variance explained is the familiar R^2. This can be used as a measure of "goodness of fit" of the theoretical model to the data. In this case, the R^2 is a very high .830, giving us greater confidence in the adequacy of the estimation.[5]

The procedure used to obtain estimates of the coefficients in regression problems such as this is "ordinary least squares" (hereafter OLS). One crucial assumption of OLS is that the "error terms" have properties that result in zero outcomes under certain conditions. Without these conditions we cannot make the substitution of equations to derive the coefficient values. Another way of saying the same thing is that without this assumption we do not have enough information to solve for the unknown coefficients.

What are the conditions necessary to specify that the sum of the left-out factors and the covariation between the left-out factors and the measured independent variable are both zero? In order to obtain proper OLS estimates, the unknown causes must have values which are randomly selected from a

[4] For example, the estimated coefficients are well within the expected values of $\hat{\alpha} = 0$ and $\hat{\beta} = 1$ at any level of significance desired.

[5] In this case, we have a small number of observations (7). Therefore, it is appropriate to "correct" the R^2 for the relatively few degrees of freedom in the data set. This "corrected R^2" (also denoted as $\bar{R}^2$) is .796, thus not changing our basic conclusion. For the derivation of $\bar{R}^2$ see Johnston (1972, pp. 129–130).

population of such errors which is normally distributed and has a constant variance and a mean of zero.[6] Often in political science we do not have direct observations at the interval level for the dependent variable. When we observe some transformation of the interval level variable, we cannot insure that we have a population of errors that are normally distributed and that meet the other assumptions. In fact, we frequently measure dependent variables in ways that lead us to *suspect* these assumptions.

If our measure of the dependent variable is such that it may take only a limited range of values, we might question the possibility of normally distributed errors. For example, if our dependent variable is dichotomous, as when the vote is measured as "Democratic/Republican" (or as when participation is measured as "Vote/Didn't Vote"), we have the extreme case in which the range is limited to only two values. These dependent variables are linked, for our sample, to an equation which adds the weighted independent variable scores to a constant and the error term, to give these two values. Only under very peculiar circumstances could we have the kind of continuous error term values necessary for even an approximately normal distribution, and yet have such a limited range of variation in the dependent variable. As we move to trichotomous dependent variables, we can easily see that the problem is almost as severe. In general, as our dependent variable takes on a greater range in scores–approaches a continuous variable–OLS becomes more appropriate.

Put another way, a normally distributed error term implies an unrestricted range of variation in the dependent variable. A severely restricted range of variation in the dependent variable tends to undermine the assumption of normally distributed error.

If the assumptions concerning the distribution of unmeasured influences such as normality are called into question, then OLS is an inappropriate estimation procedure. For dichotomous and other categorically measured dependent variables, alternative procedures should be employed.

To illustrate the implication of limited variability in the dependent variable for the other assumptions, we can examine Figure 2. In Figure 2 the scores on the dependent variable are limited to 0.0 and 1.0. Therefore, when

[6] The assumption that the error is normally distributed is important when dealing with small samples such as our example. In large samples, many of the properties of OLS regression estimates, statistical tests, and the like do not require the normality assumption. This follows since in large samples, the sampling distributions of the estimates will be approximately normal regardless of the form of the distribution of the error term. There are many good sources which deal with the regression model and its assumptions (e.g., Johnston, 1972).

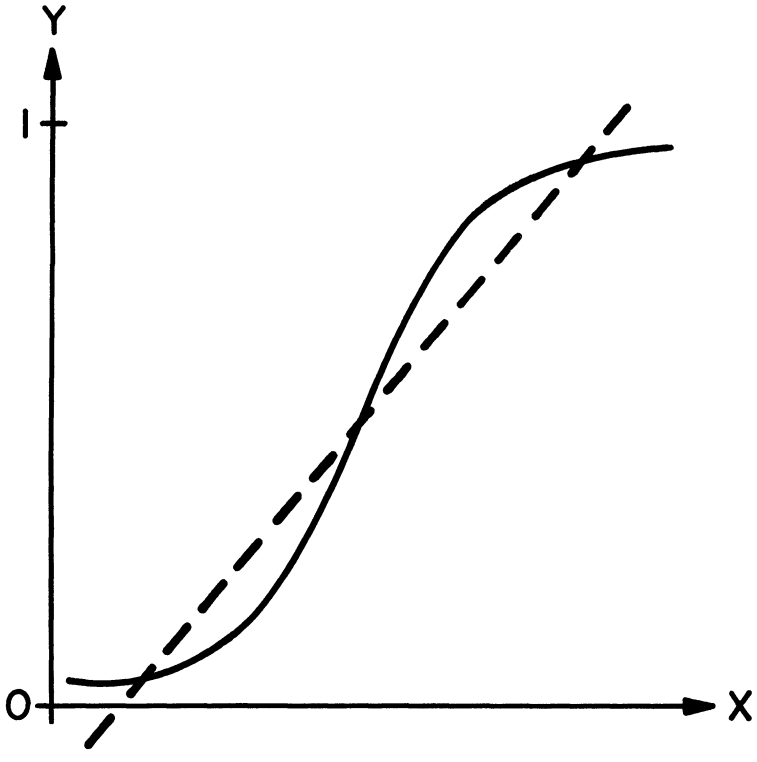

FIGURE 2
Comparison of OLS Estimate and "Typical" S-Shaped Relationship with a Dichotomous Dependent Variable

an independent variable is plotted against a dichotomous dependent variable, the underlying relationship would appear better approximated by an "S" shape than a straight line. If we were to attempt to analyze these data with OLS procedures, we would obtain an estimate of a straight line such as the dotted line in the figure. The estimated slope would appear much like that drawn, because OLS will attempt to "average out" the vertical distances of the actual scores from the estimate, resulting in as many observations below the estimate as above.

Given this typical shape of the actual observations and the characteristic departure from it by the OLS estimate, we can draw conclusions about two assumptions:

1) The residuals (the estimated error) obtained from the OLS estimate will be correlated with the independent variable.

This conclusion can be easily seen from the illustration. The departures from the OLS estimate (the gap between the S-shaped curve and the straight line) are primarily above the line when the values of X are large. The departures from the estimate are primarily below the line when the values of X are small. Thus (assuming a positive X-Y relation) the residuals will be positively correlated with X.

2) The variance of the residuals will vary systematically with the independent variable.

The illustration shows that the variance of the residuals will tend to decrease as we move away from moderate values of X and toward the extreme values. The reason for this conclusion is that the positive and negative residuals—as has been pointed out above—are correlated with X.

In general, dependent variables with restricted variability—such as we find with ordinal and certainly categorical variables—will tend to produce "clumpiness" in actual relationships with independent variables. As OLS attempts to fit a straight line through these patterns, the result will be sets of residuals which are inconsistent with the assumptions of uncorrelated error and constant variance.

Probit Analysis

Thus, we have seen that the failure of the assumption that the dependent variable is intervally measured leads to the violation of several of the OLS assumptions about the error term. Consequently, nonintervally measured dependent variables imply that the OLS regression estimating procedure breaks down. The inapplicability of OLS in this situation is of serious consequence for political science, since we are so often faced with variables that are no more than ordinally measured. There is, however, an alternative model and associated estimating technique that is designed for just this situation, one that retains much of the power of OLS.

One interpretation of the *probit model* emphasizes the similarity of probit and OLS regression. Suppose that we assume a theoretical—but unmeasured—dependent variable which is related to a set of independent variables as in the OLS model. Similarly, the theoretical dependent variable has associated with it a stochastic term meeting the same assumptions as before. In this case, we further assume that the observations of the dependent variable are measured

on only an ordinal scale, a collapsing of the "true" interval scale. In the extreme, only a dichotomy may be observed. Indeed, the most extreme dichotomous case was that for which probit was originally developed (Finney, 1971). Probit has, however, been extended to the n-chotomous case by Aitchison and Silvey (1957), and by McKelvey and Zavoina (1969). The latter citation provides the first application of the technique in political science, so far as we know.[7] They, therefore, discuss probit at some length, exemplifying its uses for our purposes and in our terms, as well as providing an important extension of the theory of probit with numerous useful refinements.

Estimates for the probit model are developed by the method of maximum likelihood. This method capitalizes on the assumed normality of the error term. With this assumption it is possible to determine the probability (or likelihood) of having observed the particular sample data for any given set of values that the parameters might assume. The maximum likelihood criterion is invoked by selecting, as estimates of the true parameters, those values which have associated with them the highest probability of having obtained the observed sample data. Even having imposed the normality assumption, the equations that must be solved are very complex and can only be approximated. However, the estimates thus obtained have a wide variety of important and useful properties, not the least of which include unbiasedness (in the limit or "consistency"), normal sampling distributions (again in the limit), and in this class of estimates, minimum variance ("best" or most "efficient"). A very similar set of properties hold for the OLS estimates if the assumptions of regression hold.

Estimated parameters appear much like those obtained from OLS regression, since the models are very similar. However, the fact that only an ordinal dependent variable is observed limits and weakens the straightforward interpretation of the coefficients. For example, since the scale or "unit of measurement" of the dependent variable is unknown, "slope" or "b" coefficients cannot be interpreted (as in OLS regression) as the amount of change in the dependent variable for a one-unit change in an independent variable. Or, since the concept of variance is undefined for an ordinally measured variable such as our dependent variable, there is no analogue to the "standardized beta" of OLS regression. In fact, since the dependent variable is only

[7] The basic presentation of n-chotomous probit can be found in McKelvey and Zavoina (1969). Also see Cnudde (1971, p. 104). The initial extension of probit for the case of a single independent variable is in Aitchison and Silvey (1957).

ordinally measured, estimates are only unique up to positive linear transformations. (Technically, all of these comments can be summarized by noting that the model is "underidentified.") Nonetheless one obtains an estimate of the best, weighted linear combination of the independent variables. In this case, the predicted $\hat{Y}$ variable may also be taken as an estimate of the unmeasured, theoretical Y that underlies the probit model.

The probit model has a second standard interpretation, in fact the initial sort of situation for which probit was developed. Suppose, for example, that the dependent variable is strictly dichotomous, say 0 or 1. In this case, one obtains estimates of the "weights" to attach to independent variable dimensions (plus an intercept term) which, by employing once again the assumed normality properties, allow us to predict the probability of any given observation being a 0 or 1. Finney (1971), in his classic study of (dichotomous) probit, used biological examples such as seed germination. A researcher might be interested in estimating the probability of germination of seeds, conditional upon such properties as rainfall, temperature, amount of fertilizer used, and other such independent variables. However, he can only observe whether or not the seeds actually germinated. The probit estimates and normality assumption allow for the estimation of such probabilities, conditional upon the values the case takes on the independent variables. Notice that the (erroneous) use of OLS regression may result in predictions that might be interpreted as "probabilities." However, these "probabilities" may be negative or exceed one, since the linear model is not restricted to the range of probability. (See, e.g., the OLS linear estimate of Figure 2.) Probit, on the other hand, yields probability estimates that are true probabilities, and therefore lie in the required ranges. The extension of probit to n-chotomous, ordinal, dependent variables has associated with it a parallel extension of the probabilistic predictions (e.g., a trichotomous dependent variable leads to the prediction of probabilities of being a 1, 2, or 3, if that is the coding of the variable, and of course one may obtain predictions of the probabilities of any combination as well). It is this probabilistic nature of the predictions of observed categories that led to the naming of the technique; probit is short for a "probability unit."[8]

[8] The reader should note that the assumed transformation of the linear model to the probabilistic one, viz. the cumulative normal, is only one possible "probability unit" yard stick which could be employed. E.g., Finney (1971, Section 3.8) discusses several alternatives including "logit," the most well known alternative, as well as the highly idiosyncratic assumptions that must be made about the error term to "legitimately" apply OLS regression to probabilistic questions.

To exemplify the use of probit and compare it to the estimates obtained through the use of OLS, we have chosen voting in 1972. Consider the attempt to predict the probability of a citizen voting for, say, McGovern, in that election (coded a 1) versus the probability of either abstaining or voting for Nixon (0). We might want to make the prediction of the citizen's probability of voting for McGovern, given the citizen's positions on issues. Thus, we obtain the following expression for the theoretical, linear, model underlying probit:

$$Y^* = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n \tag{7}$$

where $X_1, \ldots, X_n$ are the positions (of a given citizen) on the n issues; a, b_1, $\ldots$, b_n are the parameters which are to be estimated; and Y^* is the "theoretical, underlying" linear dependent variable. In this case, Y^* might be interpreted as measuring the "propensity" of a citizen to vote for McGovern. We observe Y (= 0 or 1), so we want to transform the linear "proness" variable to a probabilistic prediction. In the dichotomous case, it can be shown (see McKelvey and Zavoina, 1969) that the probability of voting for McGovern [or Pr (Y = 1)] and the probability of not voting for McGovern [Pr (Y = 0)], given the independent variables and estimated coefficients (denoted $\hat{a}, \hat{b}_i$), are:

$$\Pr(Y = 1/\hat{a} + \hat{b}_1X_1 + \ldots + \hat{b}_nX_n) = \Phi(\hat{Y}^*) = \Phi(\hat{a} + \hat{b}_1X_1 + \ldots + \hat{b}_nX_n) \tag{8}$$

$$\Pr(Y = 0/\hat{a} + \hat{b}_1X_1 + \ldots + \hat{b}_nX_n) = 1 - \Phi(\hat{Y}^*) = 1 - \Phi(\hat{a} + \hat{b}_1X_1 + \ldots + \hat{b}_nX_n) \tag{9}$$

where "Φ" denotes the cumulative normal distribution function. In this manner a unique probability of voting for McGovern can be determined for each citizen, with his own particular configuration of positions on the issue dimensions.

We have reported the results of such an estimation of the probability of voting for McGovern in 1972 in Table 1, using data drawn from the 1972 CPS election survey.[9] In particular, we employed the 7-point issue scales upon which the respondent is asked to locate himself.[10] The actual positions of the

[9] This study was sponsored by the National Science Foundation under Grant GS-33956, and the results were made available through the Inter-University Consortium for Political Research. The authors are grateful for the aid of the Foundation and the Consortium, but neither, of course, bears any responsibility for the analysis reported here.

[10] The formats of the 7-point issue scales are discussed at length in Page and Brody (1972) and Aldrich (1975). The latter citation also includes this example, the basis for

TABLE 1

Comparison of OLS and Probit Predictions:
Probability of Voting for McGovern, 1972

	Probit ($\hat{Y}^*$)			OLS Regression		
7-Point Issue	MLE	SE	MLE/SE	$\hat{\beta}$	SE	$F_{1\,1\,20}^{1}$
Federal Jobs	−.375	.082	−4.55*	−.087	.018	24.77*
Taxation	−.257	.066	−3.88*	−.050	.014	11.60*
Vietnam	−.593	.092	−6.46*	−.145	.020	50.71*
Marijuana	−.075	.058	−1.30	−.019	.014	1.98
Busing	−.205	.083	−2.47*	−.067	.019	12.76*
Women's Rights	−.038	.046	−0.83	−.010	.011	0.80
Rights of Accused	−.046	.068	−0.68	−.011	.015	0.50
Aid to Minorities	−.136	.072	−1.90	−.030	.017	3.13
Liberal/Conservative	−.639	.113	−5.64*	−.168	.025	43.93*
Constant	−.713			.303		
Est. R^2 =	.530			R^2 = .347		
Est. R =	.728			R = .589		
−2 x LLR =	441.64			$F_{1\,1\,20}^{9}$ = 66.06		
(chi square, 9 degrees of freedom)						
N = 1130						

*Indicates significance at .05 level (critical values, Z = 1.96, $F_{1\,1\,20}^{1} \sim 3.84$)

citizens in this analysis were determined by the employment of a scaling technique reported elsewhere.[11] In Table 1 we report the probit estimates of the linear Y* variable estimate, as well as the comparable OLS regression estimates done on the same dichotomous dependent variable.

The relevant comparisons between the two techniques are best summarized by looking at the OLS regression R^2 and the estimated R^2 of the probit estimators. In the probit case, the R^2 is a full 18% higher, indicating that approximately 53% of the variance is "explained." In the regression version of the estimates, only 35% of the variance is "explained" by the issue positions of the citizens, a reduction in explanatory power of slightly more than one-third. In this particular case, the two techniques do not differ in

the determination of this model, and how it corresponds with deductions from the spatial model of electoral competition.

[11] See McKelvey and Aldrich (1974), and Aldrich and McKelvey (1975).

tests of significance. All variables that are significant at the .05 level in OLS are so as well in the probit, and vice versa. However, in other reported comparisons of probit and OLS, there were differences in statistical significance (McKelvey and Zavoina, 1969).

The second major difference concerns individual predictions of the probability of voting for McGovern. In Figure 3, we have plotted the transformed Y* variable, which is the linear predicted "propensity to vote for McGovern" (summarized in Table 1), obtained by the transformation outlined above (i.e., solving Equation 8). The S-shape of the predicted probability of voting variable is dramatically clear. In comparison with the OLS predictions, we note that OLS predictions are strictly linear, and that they exceed both the upper and lower bounds of probability at the extremes. Thus, if our interest is in predicting the probability of voting for McGovern, we would be led to some inexplicable predictions. For example, the smallest actual predicted probability using the probit model is .00016. The regression prediction for this individual is a "probability" of voting for McGovern of −.41587. Similarly, the largest OLS "probability" is 1.62091, which has a corresponding probit predicted probability of .99999 for that case. Thus for these two individuals as well as others, the OLS prediction is nonsensical. We have plotted the predicted OLS regression line on Figure 3 over the range of actual observations. The differences in the two probabilistic predictions are quite clear, and very similar to Figure 2. OLS regression, of course, yields a strictly linear prediction, while probit leads to quite nonlinear predictions. At the two bounds of probability, the probit model curves or "flattens" to approach 0 and 1 only in the limit, while OLS exceeds the limits by large amounts. Further, within the nearly linear range of probit probabilities (e.g., between about .25 and .75), the linear increase is steeper than the comparable regression.

Our basic interest thus far has focused on "predicting" or "explaining" the dependent variable. Researchers are often concerned with the independent variables, e.g., in measuring their relative "importance" in the analysis. We might consider the probit MLE coefficient and the slope or "b" of OLS as one measure of "importance." The rank order of the size of the coefficients for the two techniques is quite similar. The only difference in the ordering is that the MLE for taxation is larger than that associated with the busing issue in the probit, while the order is reversed in the OLS estimation. However, there are greater differences in the relative sizes of the coefficients. For example, statistically significant but smaller coefficients are generally larger in comparison with the strongest independent variable (the liberal/conservative continuum) in probit than they are in the OLS case. As an illustration of this

point, the liberal/conservative coefficient is 3.4 times the size of that of taxation as estimated by OLS, while it is only 2.5 times as large in the probit estimate. This sort of result is consistent with other work. In a Monte Carlo study, Zechman (1974) has observed that OLS generally underestimates all coefficients with a trichotomous dependent variable. However, he found that the underestimation was more severe the smaller the true (and in Monte Carlo studies, the known) b and/or standardized "beta weight" in comparison with other included variables. This finding did not hold for probit estimates. Rather, they appeared unbiased.[12]

In summary, we have seen that the probit model much more adequately describes the data as seen in the much greater R^2, and it avoids the prediction of nonsensical results such as "probabilities" that exceed the true ranges of probability. We expect the same sorts of results to hold as the number of ordinal categories increases. We would also expect that the problems of underestimation and the like through the erroneous use of OLS regression would decrease as the number of categories of the observed dependent variable increases. However, a new problem is raised as we move from dichotomous to n-chotomous variables. In particular, if the dependent variable is in fact ordinal, then the numerical assignment of values is arbitrary up to the order constraint. Thus, if Y is, say trichotomous, we could assign the numbers 1, 2, 3 to the three categories in order. However, equally appropriate would be the assignment of −100000, 999, 1000. The use of n-chotomous probit is insensitive to such shifts, but OLS regression assumes that the interval properties are meaningful and could lead to dramatically different estimates in the two cases.

Discriminant Analysis

To this point, we have considered OLS regression as an estimation technique when the dependent variable is measured intervally, and probit as an estimating procedure when the dependent variable is observed only ordinally. In both cases, the value of the dependent variable for any case is assumed to be a function of a weighted linear combination of independent variables and a

[12] Zechman's work is based upon a comparison of small sample properties of probit. Recall that the properties associated with any maximum likelihood estimation method are applicable only in "large" samples. Generating small samples (N = 50) meeting the conditions of the probit model with a trichotomous, ordinal dependent variable, he observed not only that probit estimates appear to be unbiased, but that they outperform OLS regression by a wide variety of measures. This sort of result provides some comfort when the researcher is faced with small samples and an ordinal dependent variable.

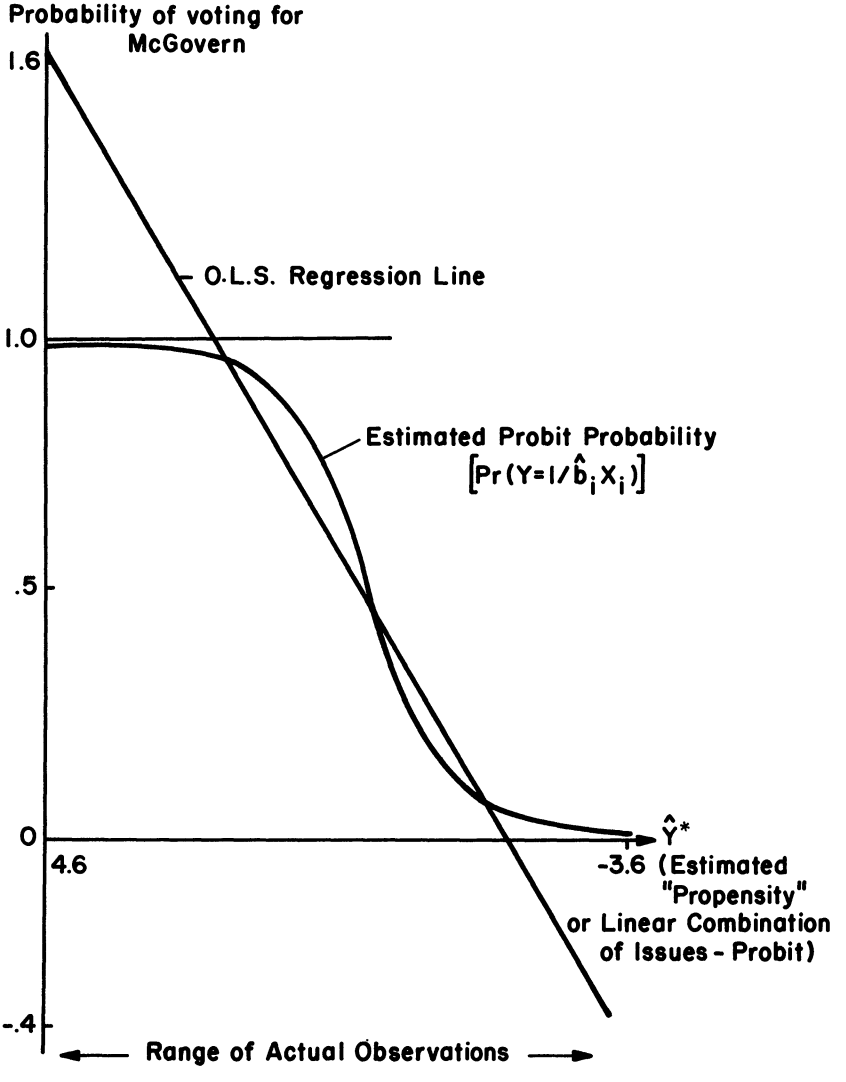

FIGURE 3
Comparison of OLS Regression and Probit Predicted Probabilities of Voting for McGovern

function of a stochastic term that is random and normally distributed. Suppose, however, that the dependent variable is only categorical. That is, we can assume that the values of the dependent variable differentiate observations by classes which are mutually exclusive but do not necessarily form any order. For example, we can classify people into racial classes. This classification scheme does not allow for the formation of ordinal, and surely not interval scales of "race." The methodological question in this case may be put: "How do we predict the category of an observation as a function of its values on independent variables?" Quite clearly, neither probit nor OLS regression can serve as an estimating model.[13] Suppose, for example, that the dependent variable is trichotomous, with categories of x, y, z. If the variable is categorically measured, then the estimating technique should yield equivalent results if the arbitrary order is y, x, z, or x, z, y, or any other permutation. Neither OLS nor probit would meet this condition. The most similar technique to those we have considered for the nominal level variable is "discriminant analysis."

We can view discriminant analysis as a problem in classification.[14] How can we "best" classify observations to the nominal categories of the dependent variable on the basis of their values on a set of independent variables? Further, for the purposes of exposition, at least, we can present the technique in two distinct parts, the estimation of the parameters with respect to the independent variables, and (as a somewhat more distinct second step than previously) the treatment of the errors of classification.

Our first problem then concerns the relationship between the independent variables, the X's, and the dependent variable Y. Y is assumed to consist of, say, m classes or categories, one of which can be written as Y^i. Our prediction of the category of Y depends upon the values the observation takes on each independent variable. These values can be shown as a vector of observations $X = (X_1, X_2, \ldots, X_n)$ which is also equivalently shown as a point in a geometrical space of n dimensions. The basic idea behind discriminant analy-

[13] The one exception is the case of a dichotomous dependent variable which can be considered ordinal, and probit appropriately applied or categorical and discriminant analysis used. We have estimated the discriminant function (to be explained below) for the McGovern voting example. This function estimates a coefficient for each issue/independent variable analogous to the probit and OLS regression estimates. Interestingly, the issue coefficients estimated by discriminant analysis are nearly identical to those obtained under the probit procedure. For example, the correlation between the two sets is .989.

[14] More extensive treatments of discriminant analysis are found in Anderson (1958) and Kort (1973).

sis then consists of dividing this n-dimensional space into m mutually exclusive and exhaustive regions, say $R_1, R_2, \ldots, R_m$, such that if an observation falls into region R_i it would be predicted as being a Y^i.

The treatment of the independent variables as defining an n-dimensional space does not differ from those techniques already discussed. What needs to be done is to consider the nature of the relationship between the "true" population classes and the independent variables. We assume, as discussed, that the space may be partitioned into separate and distinct regions—one for each category of the dependent variable. Moreover, we assume that within each population class observations are distributed normally across the independent variables, and with equal variances and covariances for each class, but allow for differing means. This normality assumption is stronger than that made earlier. Moreover, it represents a differing emphasis. In probit, for example, the normality of the stochastic term implies that, given a particular set of independent variable values and coefficients, remaining variation in Y (i.e., stochastic variance) will be normally distributed. In discriminant analysis, we must assume that the independent variables are normally distributed, given a particular value of Y. With this assumption it is possible to estimate the probability of each observation being in any one of the classes of Y. That is, we can define $P_i(x)$ as the (normality based) probability that an observation (with its particular values on the X's) would be a Y^i. Thus we could compare the probability that an observation is a Y^i versus a Y^j. In particular we would examine the $P_i(x)$ and $P_j(x)$, say by taking the ratio: $\frac{P_i(x)}{P_j(x)}$. In the dichotomous case, for example, if the ratio is greater than some constant (usually symbolized by K_{ij}), we would classify the observation as a Y^i; if it is less than K_{ij}, we would predict it as a Y^j. For this dichotomous case, this rule can be formalized as defining the regions of predicted Y classes as

$$\text{Region } R_i = \text{if } \frac{P_i(x)}{P_j(x)} > K_{ij} \tag{10}$$

$$\text{Region } R_j = \text{if } \frac{P_i(x)}{P_j(x)} < K_{ij} \tag{11}$$

We can define the "boundary" between region R_i and region R_j ("B_{ij}") as the set of points where the ratio exactly equals the constant K_{ij}. The probabilities, the $P_i(x)$'s, can be solved since we have assumed that they are all normally distributed and have the same variances and covariances. Then the only differences are in the means of the various classes of Y on the indepen-

dent variables and the actual values of any given observation on the X variables.

If we let the mean values of Y^i on the X's be symbolized by $\bar{X}_i$ (i.e., the means of the observations of category Y^i on each independent variable), and the (assumed constant) variances and covariances summarized by the matrix S, the normality and other assumptions lead to the following rather long equations for solving for the two regions:

$$R_i = x'S^{-1}\ (\bar{x}_i - \bar{x}_j) - 1/2\ (\bar{x}_i + \bar{x}_j)'S^{-1}\ (\bar{x}_i - \bar{x}_j) > \log K_{ij} \qquad (12)$$

$$R_j = x'S^{-1}\ (\bar{x}_i - \bar{x}_j) - 1/2\ (\bar{x}_i + \bar{x}_j)'S^{-1}\ (\bar{x}_i - \bar{x}_j) < \log K_{ij} \qquad (13)$$

These equations are derived by substituting the definition of a normal distribution for the $P_i(x)$'s, taking logarithms (a permissible transformation), and algebraically manipulating the various terms. The first, leftmost, term in each equation [i.e., $x'S^{-1}\ (\bar{x}_i - \bar{x}_j)$] is the only term involving the individual observations, X. This term is called the *discriminant function,* and it is a *linear* function of the observations of the dependent variable. The rest of that term consists simply of differences in means of Y^i and Y^j weighted by the variances, S. The middle term is also based on weighted combinations of means. In effect, this can be considered as a measure derived from the distance between means for categories of Y. That is, it measures the distance or separation of the means of Y^i and Y^j on the independent variables. One way of conceptualizing the purpose of discriminant analysis is to define regions of classification which maximize the variation within the predicted classes as a proportion of the total variance. This middle term provides an indication of this "discriminability" of the independent variables as predictors of Y, and statistical tests can be performed on it. The final term in these equations is the constant K_{ij}, which depends on the criterion we employ with respect to error, which we will discuss shortly.

Extending the dichotomous case to a more general n-chotomous situation is straightforward. One simply has a set of ratios of probabilities to solve simultaneously, one ratio for each pair of classes. Thus in the trichotomous case (say Y^i, Y^j and Y^k) there would be three probability ratios and inequalities to solve. To define region R_i, we would have:

$$R_i = \text{Log}\frac{P_i(x)}{P_j(x)} > \text{Log}\ K_{ij} \text{ and } \text{Log}\frac{P_i(x)}{P_k(x)} > \text{Log}\ K_{ik} \qquad (14)$$

Again, the set of points where these inequalities (and the remaining ratio) are equal form the boundaries between regions B_{ij}, B_{ik} and B_{jk}.

The undiscussed constants, the K_{ij}'s, incorporate our criterion about errors in classification and are closely related to the success of the discriminant analysis. By changing the value(s) of the constant(s), the boundaries between regions are changed, but are only changed by defining a new set of lines *parallel* to the old.

An example may help to clarify many of the points we have been making. If there are two independent variables, the space to be divided into regions is a plane, and the boundary(ies) between regions will be lines through the plane. Changing values of the constants would change the boundaries by defining new lines that would be parallel to the original regional boundaries. For example, in Figures 4A and 4B, we have drawn a plane and regional boundary lines for a di- and trichotomous dependent variable, respectively. In Figures 5A and 5B we show the effects of changing the constant term(s); the boundaries have changed, but they are parallel to the original.

As we have pointed out, any observation on x can be represented as a point in the space. Given the definitions of the regions, an observation will be classified into a population class. For example, in Figure 4B, the point w would be predicted to be in Y^1. A set of such predictions leads to a table of predicted classes of the dependent variable versus the actual observations, the table which indicates the correct and incorrect predictions.

The discriminant function for which we have developed estimates is the "best" discriminant estimate, given the assumed normality and linearity

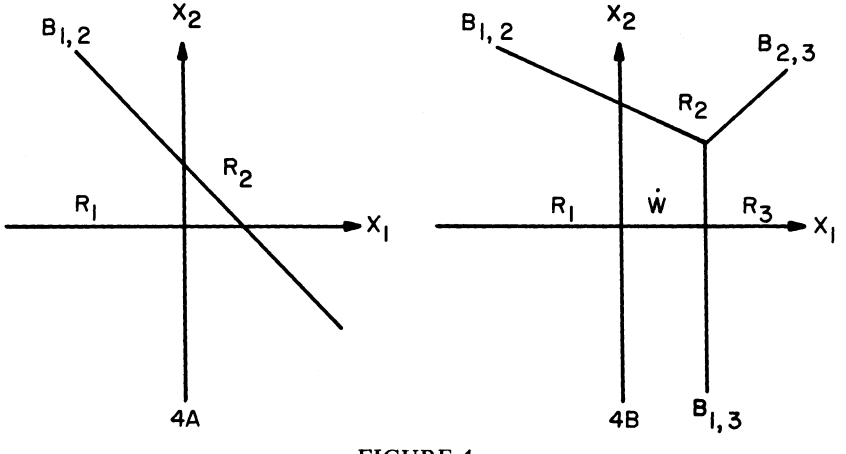

FIGURE 4
Examples of Discriminant Analysis of Di- and Trichotomous Dependent Variables

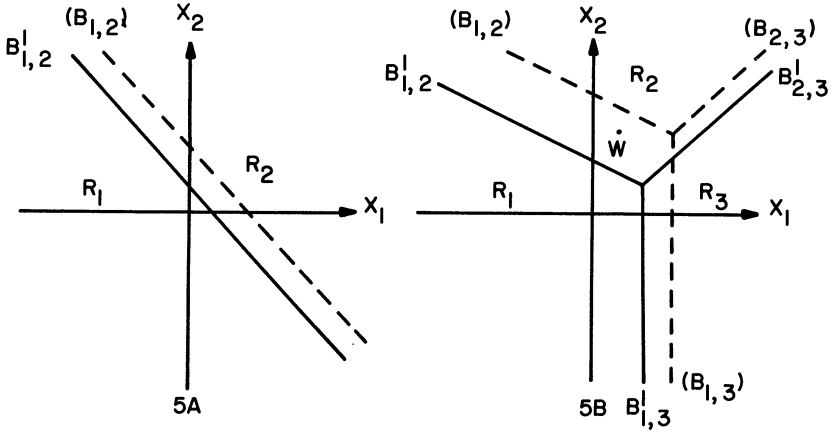

FIGURE 5
Two Examples of Discriminant Analysis with Varying Constants

constraints (i.e., the estimates of the population means and variances have a variety of desirable statistical properties). We can greatly alter the actual predictions of categories, however, by manipulating the constant terms. For example, in Figure 5B, we have altered the constant, and hence boundary lines between the regions, from those of Figure 4B. In this case, point w is now in region 2, not the initially estimated region 1, so we would estimate w as a Y^2 not a Y^1. Obviously, both the estimation of the $P_i(x)$'s and the K constants will be crucial.

A number of criteria have been proposed for the determination of these constant terms. The most obvious criterion is to let all K_{ij} be one. That is, if the ratio $\frac{P_i(x)}{P_j(x)}$ is greater than one, classify that observation in predicted class Y^i. If it is less than one, predict it as a Y^j. In other words, the criterion is simply to define the regions such that the probability of an observation being a Y^i is higher than any other class in Region R_i. Obviously, such a criterion has much intuitive appeal. However, if one approaches the problem of classification from a broader viewpoint, alternative criteria may be superior, since it can be shown that the posited rule "assumes" that there are equal probabilities of observing all categories. We are more often faced with the situation where there are very unequal probabilities of observing the various categories—race is an obvious example. Therefore we might want to incorporate additional information into the criterion. For example, we might

"weight" the boundary-defining conditions by the actually observed proportion of Y^i's and Y^j's (presumably our "best estimate" of the true proportions). Thus our constant K_{ij}'s would simply be the ratios of observed sample frequencies of Y^j to Y^i. If we let P^{Yi} and P^{Yj} be the proportions of all observed Y's which are Y^i's and Y^j's, respectively, then we can define our two criteria of determining the K_{ij} constants as:

$$K_{ij} = 1 \text{ for all i, j } \left(\text{i.e., } \frac{P_i(x)}{P_j(x)} = 1 \text{ to define } B_{ij}\right) \tag{15}$$

$$K_{ij} = \frac{P^{Yj}}{P^{Yi}} \text{ for all i,j } \left(\text{i.e., } \frac{P_i(x)}{P_j(x)} = \frac{P^{Yj}}{P^{Yi}} \text{ to define } B_{ij}\right) \tag{16}$$

As an example of the use of discriminant analysis, consider the choice among voting for Nixon, Humphrey, or Wallace in 1968. As a dependent variable, the 1968 vote is not measurable in terms of an ordinal or interval scale. Attempts to use party identification based concepts (e.g., the normal vote) which assume ordinality or more have floundered over the problem of what to do about Wallace voters.[15] One resolution is to employ discriminant analysis. As in our probit example, we will use the 7-point scales as a base to estimate the citizens' positions on the two issues of Vietnam and urban unrest.[16]

The estimations of the discriminant functions are presented in Table 2, these being the equations, as well, for the "boundary lines" to define the regions of the issue plane. We computed the constant coefficients by both assuming equal probability (i.e., $P^{Yi} = P^{Yj}$), so that K = 1 and log (K) = 0 in all pairs, and also using the proportionate vote division in the sample as an estimate of the true probabilities (in the 942 included respondents, 41.1% voted for Humphrey, 47.9% for Nixon, and 11.0% for Wallace). The two sets of regions are drawn in Figures 6 and 7, while the table of actual versus predicted vote categories is found in Table 3. Finally, we have reported a comparison table resulting from assuming that the coefficients for the two issues are equal. (The actually estimated ratios of the two coefficients for the three discriminant functions are found at the bottom of Table 2.)

First, let us consider the estimated relationships between the independent

[15] See, for example, the dialogue between Boyd (1972a, 1972b) and Kessel (1972) on this point.

[16] While these two issues were very important in that election, so certainly were many others. However, 7-point scale format data were collected for only these two in 1968. It is also convenient to be able to visualize this two-dimensional example.

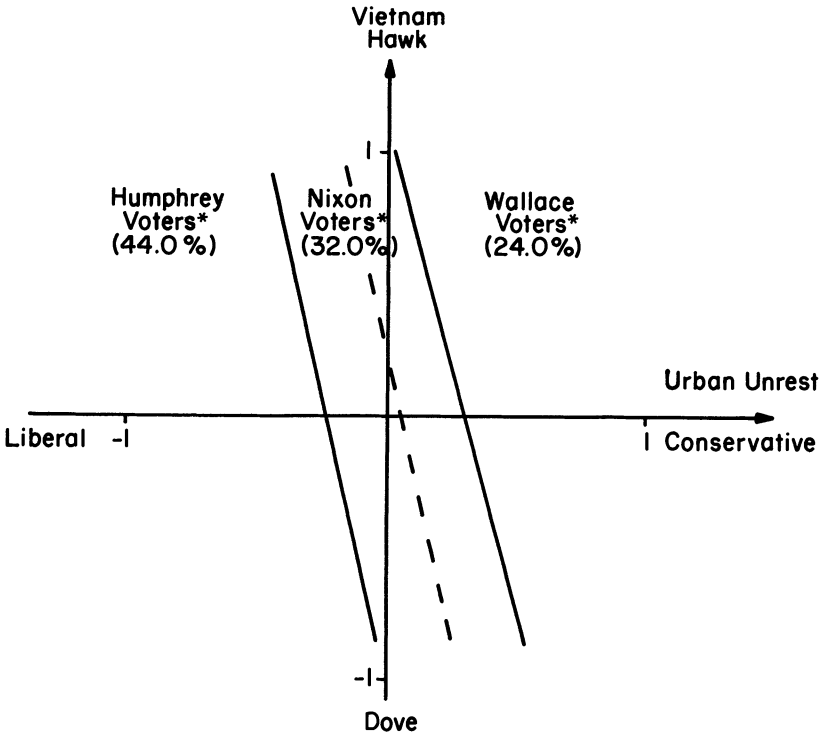

FIGURE 6
Discriminant Analysis of 1968 Voting, Based on Constants Equal to One

variables and the dependent variable, and then the error in classification. The Humphrey voters are estimated to be towards the liberal end of both issues, Wallace's region is to the right on both, and consequently Nixon voters are classified as those in between the other two. Notice, too, that the urban unrest issue is much more important than Vietnam in discriminating voters, its coefficients (the analogues to the regression $\hat{b}$'s and probit MLE coefficients) being 3.7 to 5.5 times as large as Vietnam's. Thus each candidate's region tends to cover almost the entire range of hawk to dove. Moreover, there is little difference between the discriminant functions defining the two regions, so that the Humphrey and Wallace regions in the range of actual observations do not touch. This sort of result is not typical of all applications of discriminant analysis. Indeed, the three categories of voting in 1968 appear

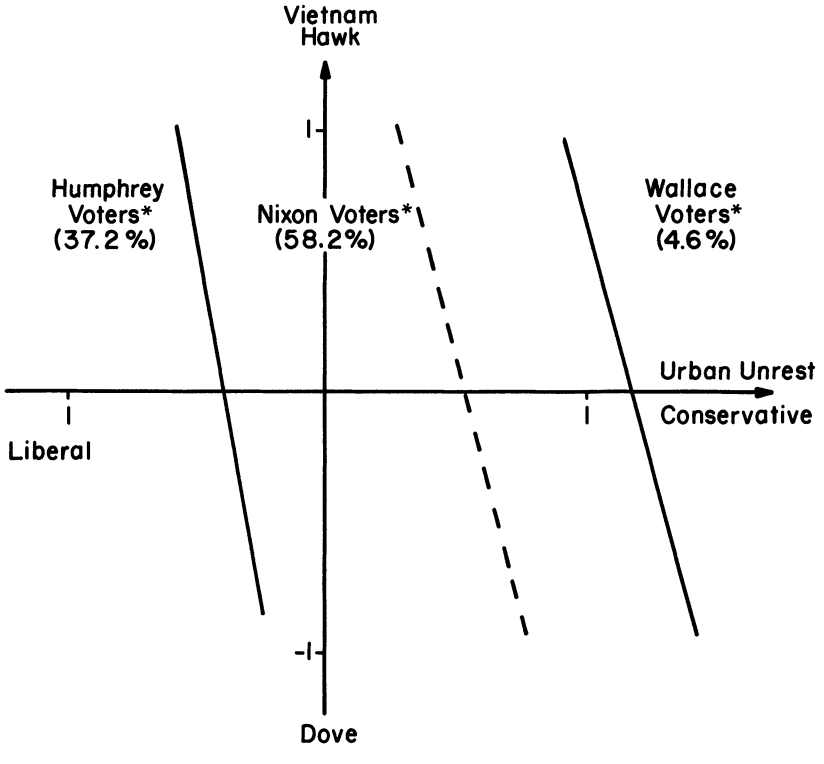

FIGURE 7
Discriminant Analysis of 1968 Voting, Based on Constants
Equal to Proportion of the Candidate Vote Actually Observed

to form an ordinal measure *with respect to the citizens' positions on these two issues.* Thus one possible use of discriminant analysis is to see if a measured dependent variable is approximately ordinally related to the independent variables of concern. This finding does *not* allow us to conclude that there is an ordinal measure of voting for the three candidates in 1968. Rather, we can only conclude on the basis of this evidence that the dependent variable is approximately ordinal over the independent variables examined. This has an equivalent interpretation: that the three discriminant functions are approaching the extreme possibility of being equal to each other, differing but by an additive constant.

The effects of changing our assumptions about the constant terms, K_{ij}, are quite clear. The Nixon region is much narrower when we assume that all K_{ij}'s are equal. Using the sample proportions as an estimate of the true probabilities increases the size of his region, so that the proportion of voters found in his region increases from 32.0% to 58.2%.

Let us turn to the question of error in classification. In the first case, where log(K) = 0 for all comparisons, we obtain 55% correct classification, with the greatest successes being in classifying Humphrey and Wallace voters (about 67% and 71% correct, respectively), our greatest error being concentrated in predicting Nixon voters. When we incorporate the "prior probabilities," our correct classifications jump by more than 7% to 62% correct. The effect of changing the boundaries is to move about 26% of Humphrey and Wallace predicted voters to the region of predicted Nixon voting. Such redefinition of boundaries greatly increases the success of classifying Nixon voters (from 41% correct to over 72% correct). This was accomplished at minimal cost in classification of Humphrey voters (an 8% reduction in correct predictions), but at a greater loss in classification of Wallace voters (down to 28% correct). However, the Wallace voters are such a relatively small group that error there can be absorbed with relatively little cost.

A comparison of these results to a comparable probit or OLS regression analysis would be useful at this point. Unfortunately, just what would constitute a comparable analysis is far from clear. First, one might run three

TABLE 2

Discriminating Voters on Issues, 1968

Discriminant Functions	Log K = 0	Log K = Log $\frac{(P^{Yj})}{(P^{Yi})}$
Z_{HN} = –.3484 –.2233 Vietnam –1.2257 Urban Unrest	(= 0)	(= + .152)
Z_{HW} = +.1379 –.6724 Vietnam –2.8962 Urban Unrest	(= 0)	(= –1.318)
Z_{NW} = +.4863 –.4492 Vietnam –1.6705 Urban Unrest	(= 0)	(= –1.471)

Relative Weightings	Vietnam/Urban Unrest
Humphrey-Nixon	0.182
Humphrey-Wallace	0.232
Nixon-Wallace	0.269

TABLE 3

Error in Classifying Voters, 1968

Predicted Vote Log (K) = Log $\frac{(P^{Yj})}{(P^{Yi})}$

		Humphrey	Nixon	Wallace			
	Humphrey	24.4%	16.5	0.3	41.1%	(307)	
Actual Vote	Nixon	12.0	34.7	1.2	47.9	(358)	
	Wallace	0.8	7.1	3.1	11.0	(82)	62.16% Correct
		37.2%	58.2	4.6	100%	(747)	
		(278)	(435)	(34)			

Predicted Vote Log (K) = 0

		Humphrey	Nixon	Wallace			
	Humphrey	27.7%	10.3	3.1	41.1%		
Actual Vote	Nixon	15.3	19.5	13.1	47.9		
	Wallace	1.1	2.1	7.8	11.0		55.02% Correct
		44.0%	32.0	24.0	100%	(747)	
		(329)	(239)	(179)			

Euclidean Equidistant (UU = Viet) Log (K) = Log $\frac{(P^{Yj})}{(P^{Yi})}$

		Humphrey	Nixon	Wallace			
	Humphrey	27.0%	11.8	2.3	41.1%		
Actual Vote	Nixon	14.6	24.6	8.7	47.9		
	Wallace	0.5	3.5	7.0	11.0		58.64% Correct
		42.2%	39.9	17.9	100%	(747)	
		(315)	(298)	(134)			

paired comparisons (i.e., Humphrey and Nixon, Humphrey and Wallace, Nixon and Wallace), thus generating six equations instead of three. Next, we must decide what to do about the dependent variable. Looking at the Humphrey-Nixon pairing, for example, what are we to do with Wallace voters? Three possibilities come to mind. We might code a "Nixon vote" variable as a one if the respondent voted for Nixon and as a zero if he voted for Humphrey or for Wallace (and perhaps for abstention as well). Alternatively, we might simply remove Wallace voters from the analysis. Finally, we might attempt to infer

what Wallace voters would have done in the absence of the Wallace candidacy on the basis of some other data (e.g., the SRC's 100-point "thermometer" evaluation of candidates). Clearly, our choice among these three possibilities (and any others that might come to mind) will greatly affect our results. The instability in such estimations is convincingly demonstrated in Table 4, which briefly summarizes probit results of these three approaches to estimating Nixon support. The MLE estimates differ considerably, in some instances being more than double those of others. As in the comparison between probit and OLS, there are some similarities (e.g., the relatively larger size of the coefficient for urban unrest), as well as some important differences (e.g., the

TABLE 4

Probit Predictions of Nixon Voting When Paired with Humphrey

	MLE	SE	MLE/SE
Wallace Voters Removed			
Urban Unrest	0.453	0.075	5.835
Vietnam	0.081	0.048	1.695
Constant	–0.114		
–2 X LLR	48.55		
(chi square, 2 degrees of freedom)		N = 860	
Wallace Voters Coded "0" (Not Vote for Nixon)			
Urban Unrest	0.235	0.070	3.367
Vietnam	0.031	0.045	0.689
Constant	–0.273		
–2 X LLR	15.14	N = 942	
Wallace Voters Vote "Preference" (Via 100° Thermometer)			
Urban Unrest	0.377	0.070	5.381
Vietnam	0.064	0.045	1.431
Constant	–0.159		
– 2 X LLR	41.296	N = 942	

Vietnam coefficient is never statistically significant at the .05 level). However, the instability of the estimates is the overriding consideration, especially when it is recalled that the Wallace voters amount to less than 10% of the sample. Similar comparisons of pairings involving Wallace are almost completely determined by our approach to the incorporation of the 40% or so of the sample who voted for the candidate not in the pair. Thus, a categorical dependent variable is wholly unsuited for OLS or probit estimation.

Conclusion

We hope we have been able to demonstrate the following points in this paper. First we tried to show that there are a variety of powerful techniques, all grounded on a solid base of statistical theory. Moreover, one or more of these techniques is suitable for whatever level of measurement is used for the dependent variable. All too often it appears that the alternatives are simply OLS regression and contingency table analysis. This implies that if one's data are measured on less than an interval scale, one has only the choice of using regression erroneously or using an appropriate but much weaker method of analysis. Alternative techniques exist. The particular examples we chose were selected for their similarities in terms of their purposes, their assumptions, and the powerful analysis they allow, while at the same time being applicable to each of the three basic measurement levels.

Second, we have argued that the "level of measurement" can be a crucial consideration, but one based upon the theory underlying the statistical procedure. As our example applications have demonstrated, employing a procedure assuming a different level of measurement can seriously affect the estimates and lead to incorrect inferences and hypothesis tests. Thus, our comparison application of OLS to an ordinal variable seriously underestimated the overall fit of the model to the data. Further, the individual coefficient estimates are known to be biased in general, and we have seen that the bias is not uniform, so that there were some changes in the order of size of coefficients and more dramatic changes in their relative magnitudes as compared to the probit estimates. There were great difficulties in attempting to specify the nominal level dependent variable example to make a comparison with probit. If our procedures are acceptable, the resultant estimates exemplify differences in hypothesis tests. The coefficient for the Vietnam issue was significant in all cases in the discriminant analysis, while it was never significantly different from zero (at the .05 level) in the probit estimates.

It is clear, then, that the levels of measurement problem is real. Yet it is

but one important factor in choosing an appropriate statistical technique. Other aspects must also be considered. For example, there are other differences in assumptions of the techniques. Probit necessarily assumes that the stochastic term is normally distributed, an assumption which may not be necessary in all instances in using OLS regression. The discriminant procedure assumes that observations on the independent variables are normally distributed within each category of the dependent variable. This assumption is much stronger than any comparable assumption made in probit or OLS. Therefore, the plausibility of these and other assumptions must be weighed along with the level of measurement. Further, the degree to which assumptions are violated must be considered. Thus, the bias in OLS when applied to an ordinal variable will be less serious the greater the number of ordinal categories (*all else remaining equal*). Finally, it must be recognized that techniques have differing properties and differing degrees of development. Regression, in particular, has been extensively analyzed and formally developed. Only it is suitable, at this time at least, for estimation of more complex relationships and multiple equation models (e.g., simultaneous equations, causal modeling, etc.). In short, the choice of an appropriate statistical procedure is complex and contingent on many criteria. The level of measurement is only one criterion; yet it is important and may well have direct consequences for the analysis.

Third, we have argued that the political researcher must very carefully examine the assumptions underlying the statistical technique, and consider their correspondence to the theoretical assumptions one has made in deriving one's hypotheses. The failure to carefully trace out this correspondence can lead to incorrect conclusions equally as well as the application of a technique to an unsuitable level of measurement. A simple illustration of this point concerns the prediction of the division of congressional seats in off-year elections. We assumed that there was relative stability in partisan preference. If this assumption were wrong (e.g., if we extended our series back through elections during the era of Republican hegemony), the model would lead to quite different estimates.

All too often, research is based on faulty priorities. We are all familiar with examples of research which appears to be based on a new statistical technique and where the substantive problem was chosen simply to show off methodological sophistication. The undue emphasis on technique must be eschewed. Yet we cannot ignore this crucial element of research. The choice of a statistical method inappropriate for the substantive and theoretical concern leads just as surely to worthless research. Yet that choice can be grounded in

the assumptions underlying the method and the decision about whether they are appropriate to the substantive problem, rather than the conventional "levels of measurement" debate.

Manuscript submitted February 5, 1975.
Final manuscript received March 7, 1975.

APPENDIX

The purpose of this Appendix is to present copies of actual computer output of the three techniques and indicate how to interpret it. It has been our experience that there can be some confusion in understanding such output the first few times a program is actually used, regardless of the comprehension of the basic statistical principles involved.

OLS regression programs are widespread, commonly used, and well documented. Most statistical packages (SPSS, OSIRIS, BMD, etc.) have an OLS program, and as well it has been our experience that many university computer facilities have their own regression packages. Therefore, they require little comment. Figure A-1 is a copy of the SPSS output which is summarized in Table 1 (in which OLS regression and probit are compared). Summary characteristics of the overall regression are printed first. Attention is usually focused on the R^2 and its root, the multiple correlation coefficient (or correlation between predicted and observed dependent variable values) as indications of the "goodness of fit." The standard error (of the estimate) and F statistic are most relevant for statistical tests of significance of the whole regression. The remaining summary statistics are useful for constructing "Analysis of Variance" (explained or "regression," unexplained or "residual," and total sum of squares and variances), such as the "ANOVA" tables found in Johnston (1972, p. 143). Following these summary calculations are statistics relevant to individual independent variables. The first column is the regression "b's" or slope coefficients (what we call "$\hat{\beta}$" 's in Table 1). The "Betas" of the second column are standardized betas, the standardization being a multiplication of the regression coefficients by the ratio of the standard deviation of the relevant independent variable to the standard deviation of the dependent variable. This standardization, therefore, puts all independent variables in comparable units (sometimes referred to as "dimensionless") to facilitate comparisons between independent variables (note that no comparable standardization exists for probit, since the standard deviation of Y is undefined). Next is the standard error of the estimated regression coefficient and the F statistic for each variable to test for significance of each

coefficient (in other programs, a t value may be produced which is simply the square root of this F value). These F's all have one degree of freedom in the numerator and degrees of freedom equal to the difference between the sample size and the number of parameters being estimated in the denominator (or 1130 −10 = 1120 in this case). Many other statistics can be obtained from regression packages. Those available under SPSS are carefully explained in the SPSS Manual (Nie, Bent, and Hull, 1970) and various updates.

Figure A-2 is a copy of the output of the probit example (also found in Table 1). The program used is that developed by McKelvey and Zavoina (to the best of our knowledge, probit is not included in any of the statistical packages). The output is well documented and rather straightforward. The "Maximum Likelihood Estimate" column is the MLE coefficient for each independent variable and the constant, comparable to the regression b-coefficients or "$\hat{\beta}$" (hence, the reference to them as "BETA(.)" on the leftmost column of the output). The standard error column is self-explanatory. The ratio of the two, in the third column of the output, is useful for tests of significance much like the individual F statistics of regression. Recall that, as in all maximum likelihood estimates, properties of the probit estimates are "asymptotic" (i.e., are applicable with large sample sizes only). The ratio MLE/SE is, in large samples, approximately a standardized normal random variable, or "Z score." Thus, this Z score can be used to test whether the coefficient is significantly different from zero, as in the case of the individual F values for OLS regression. The comparable statistic to the F of OLS regression for testing "overall significance" is −2 times the log of the likelihood ratio. This statistic is a comparison of the probability of observing this sample if the MLE estimates are correct (i.e., the estimated log of the likelihood function which is also printed) to the situation if all coefficients were zero (i.e., the null model). As stated in the output, this statistic is, in large samples, a chi-square statistic with degrees of freedom equal to the number of independent variables. Other summary indicators are found after a case-by-case residual analysis in the McKelvey-Zavoina program (and may not be found in other probit programs). These calculations are also self-explanatory and for the most part have direct analogues in OLS regression. It should be pointed out, however, that all statistics under the heading "Estimated Analysis of Variance" are just that—estimated. This is so since the Y variable is not measured intervally. The estimates are derived by arbitrarily setting the residual sum of squares so that there is an equivalent to one unit error for each case (i.e., this figure will always be equal to the sample size). Given this arbitrary setting, the other statistic estimates follow. Of course, of some interest is the percent of the bases correctly predicted, which is not estimated

FIGURE A-1
An OLS Output from SPSS

```
REGRESSION VERSIONS OF SUPPORT FUNCTION MODELS 1972                    05/16/74        PAGE   15

FILE   CSF72    (CREATION DATE = 04/29/74)

* * * * * * * * * * * * * * * * * * * * * * * * M U L T I P L E   R E G R E S S I O N * * * * * * * * * * * *    VARIABLE LIST   1
                                                                                                                 REGRESSION LIST   1
DEPENDENT VARIABLE..    VAR001     ACTUAL VOTE  MCGOVERN

VARIABLE(S) ENTERED ON STEP NUMBER  1..     VAR003     FEDERAL JOBS IDEAL
                                            VAR004     TAXATION IDEAL
                                            VAR005     VIETNAM IDEAL
                                            VAR006     MARIJUANA IDEAL
                                            VAR007     BUSING IDEAL
                                            VAR008     WOMENS RIGHTS IDEAL
                                            VAR009     RIGHTS OF ACCUSED IDEAL
                                            VAR010     AID TO MINORITIES IDEAL
                                            VAR011     LIBERAL CONSERVATIVE IDEAL


MULTIPLE R           0.58886          ANALYSIS OF VARIANCE     DF       SUM OF SQUARES          MEAN SQUARE              F
R SQUARE             0.34675          REGRESSION               9.          78.31571              8.70175           66.05674
STANDARD ERROR       0.36295          RESIDUAL              1120.         147.53916              0.13173
```

---------- VARIABLES IN THE EQUATION ----------

VARIABLE	B	BETA	STD ERROR B	F
VAR003	-0.08727	-0.14247	0.01754	24.765
VAR004	-0.04951	-0.09014	0.01454	11.600
VAR005	-0.14549	-0.20062	0.02043	50.711
VAR006	-0.01897	-0.03929	0.01346	1.984
VAR007	-0.06748	-0.10265	0.01889	12.756
VAR008	-0.00960	-0.02259	0.01074	0.800
VAR009	-0.01093	-0.02000	0.01540	0.504
VAR010	-0.02955	-0.05167	0.01671	3.127
VAR011	-0.16815	-0.21013	0.02537	43.926
(CONSTANT)	0.30271			

---------- VARIABLES NOT IN THE EQUATION ----------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F

ALL VARIABLES ARE IN THE EQUATION

SUMMARY TABLE

VARIABLE		MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
VAR003	FEDERAL JOBS IDEAL	0.39247	0.15403	0.15403	-0.39247	-0.08727	-0.14247
VAR004	TAXATION IDEAL	0.43026	0.18513	0.03110	-0.29478	-0.04951	-0.09014
VAR005	VIETNAM IDEAL	0.52882	0.27965	0.09452	-0.42883	-0.14549	-0.20062
VAR006	MARIJUANA IDEAL	0.54294	0.29479	0.01514	-0.29457	-0.01897	-0.03929
VAR007	BUSING IDEAL	0.55827	0.31166	0.01687	-0.37627	-0.06748	-0.10265
VAR008	WOMENS RIGHTS IDEAL	0.55950	0.31304	0.00138	-0.16528	-0.00960	-0.02259
VAR009	RIGHTS OF ACCUSED IDEAL	0.56271	0.31664	0.00360	-0.30653	-0.01093	-0.02000
VAR010	AID TO MINORITIES IDEAL	0.56669	0.32113	0.00449	-0.35190	-0.02955	-0.05167
VAR011	LIBERAL CONSERVATIVE IDEAL	0.58886	0.34675	0.02562	-0.47582	-0.16815	-0.21013
(CONSTANT)						0.30271	

FIGURE A-2
An Example Probit Output

```
**********************************************************************************************************************
N=CHOTOMOUS PROBIT ANALYSIS;  PROBIT MODEL ESTIMATIONS FOR EXTREMIST SUPPORT USING MCKELVEY PROBIT
                              MCGOVERN SUPPORT FUNCTION 1972 EXTREMIST SUPPORT TEST
**********************************************************************************************************************

                THE ITERATION HAS CONVERGED ON THE  4TH ITERATION.  MAXIMUM LIKELIHOOD ESTIMATES FOLLOW

---------------------------- BETAS ----------------------------     ------------------------ MU'S ------------------------

                                MAXIMUM                                                    MAXIMUM
                 REPRESENTS     LIKELIHOOD      STANDARD                                  LIKELIHOOD      STANDARD
COEFFICIENT      EFFECT OF      ESTIMATE        ERROR       MLE/SE    COEFFICIENT       ESTIMATE        ERROR       MLE/SE

 BETA( 0)        CONSTANT        -0.71277       0.06167    -11.558     MU( 1)            0.00000
 BETA( 1)        VAR #  1        -0.37502       0.08251     -4.545
 BETA( 2)        VAR #  2        -0.25658       0.06607     -3.884
 BETA( 3)        VAR #  3        -0.59260       0.09178     -6.457
 BETA( 4)        VAR #  4        -0.07534       0.05805     -1.298
 BETA( 5)        VAR #  5        -0.20547       0.08320     -2.470
 BETA( 6)        VAR #  6        -0.03796       0.04549     -0.834
 BETA( 7)        VAR #  7        -0.04552       0.06668     -0.683
 BETA( 8)        VAR #  8        -0.13613       0.07161     -1.901
 BETA( 9)        VAR #  9        -0.63921       0.11529     -5.642

LOG OF THE LIKELIHOOD FUNCTION=   -445.0191

-2.0 TIMES LOG LIKELIHOOD RATIO =    441.6399
 (THIS IS CHI SQUARED WITH   9 DEGREES OF FREEDOM)

ESTIMATED ANALYSIS OF VARIANCE

EXPLAINED SUM OF SQUARES =      1273.54577
RESIDUAL SUM OF SQUARES =       1130.00000
TOTALSUM OF SQUARES =           2403.54577
ESTIMATED R SQUARED =              0.52986

OTHER SUMMARY STATISTICS

PERCENT PREDICTED CORRECTLY =          0.82655
RANK ORDER CORRELATION= PREDICTED VERSUS ACTUAL =          0.53474
```

in the same sense but is a straightforward computation. In the dichotomous case, the predicted value is simply the category (0 or 1) which has a higher probability for that case, given the estimated coefficients and probit transformations, and is found by solving equations such as (8) and (9) in the text or their analogues in the more general, n-chotomous case.

A program for performing discriminant analysis has recently been added to SPSS and may be found in other statistical packages (though not in OSIRIS, at least not version 3). Figure A-3 is an example of the SPSS output for log K = 0 as reported in Table 2. The core of the output is listed under the heading "Discriminant Functions" and is that which is reproduced in Table 2 (note that 1 = Humphrey voter, 2 = Nixon voter, 3 = Wallace voter, "RCSELF" indicates self-placement on the Vietnam scale and "UCSELF" on the urban unrest scale–the two issue scales being modified by the scaling technique as described in the body of this paper). Preceding the discriminant functions are a variety of statistics relating to the history of the computation. The program can operate analogously to "stepwise regression" (in which each significant variable is entered in order), the basic test being whether or not it adds a significant amount to the prediction as determined by F statistics. As can be seen, both variables do add a "significant" amount. Options are also available to generate the means and variance/covariance matrix, for solving equations such as (12) and (13), as well as a correlation matrix of independent variables. Even more recent updates indicate that it is now possible to generate the regions either by using the "equiprobability" assumption (i.e., log K = 0) or by weighting the probabilities in any specified manner [e.g., log K = log $\frac{(P^{Yj})}{(P^{Yi})}$], and to output tables and/or scatter plots of predictions.

All of these programs produce a wide variety of other statistics and other sorts of information. The portions of the output we have discussed are, we believe, the most important and most commonly used results. Explanations of other portions of the output can be found in the program write-ups.

REFERENCES

Aitchison, J., and Silvey, S. D. 1957. "The Generalization of Probit Analysis to the Case of Multiple Responses," *Biometrika,* June 1957, pp. 131–140.

Aldrich, J. 1975. *Voting in Two U.S. Presidential Elections: An Analysis Based on the Spatial Model of Electoral Competition.* Unpublished dissertation, University of Rochester.

FIGURE A-3
An Example Output of Discriminant Analysis

```
VOTING BEHAVIOR AND PREFERENCE BY SPATIAL LOCATION                    05/20/74

  RCSELF      10.2813  UCSELF      80.2396

U-STATISTIC            0.75288     DEGREES OF FREEDOM    2     2  744
APPROXIMATE F         56.64926       DEGREES OF FREEDOM     4 1486.00

F MATRIX - DEGREES OF FREEDOM     2  743

GROUP VARIABLE   VOTE
                                1.00                  2.00

GROUP

      2.00                     46.339417

      3.00                    109.614517            39.957336


F LEVEL INSUFFICIENT FOR FURTHER COMPUTATION
```

********** DISCRIMINANT FUNCTIONS **********

GROUPS	1.00,	2.00,	
	VARIABLE	COEFFICIENTS (LAMBDAS)	CHOSEN COEFFICIENT (LAMBDA/MIN. LAMBDA)
	CONSTANT	-0.3484	
	RCSELF	-0.2233	-1.0000
	UCSELF	-1.2257	-5.4900

GROUPS	1.00,	3.00,	
	VARIABLE	COEFFICIENTS (LAMBDAS)	CHOSEN COEFFICIENT (LAMBDA/MIN. LAMBDA)
	CONSTANT	0.1379	
	RCSELF	-0.6724	-1.0000
	UCSELF	-2.8962	-4.3069

GROUPS	2.00,	3.00,	
	VARIABLE	COEFFICIENTS (LAMBDAS)	CHOSEN COEFFICIENT (LAMBDA/MIN. LAMBDA)
	CONSTANT	0.4863	
	RCSELF	-0.4492	-1.0000
	UCSELF	-1.6705	-3.7189

SUMMARY TABLE

STEP NUMBER	VARIABLE ENTERED	VARIABLE REMOVED	F VALUE TO ENTER OR REMOVE	NUMBER OF VARIABLES INCLUDED	U-STATISTIC
1	UCSELF		108.7951	1	0.7737
2	RCSELF		10.2813	2	0.7529

Aldrich, J., and McKelvey, R. 1975. "A Method of Scaling with Application to the 1968 and 1972 Presidential Elections," *American Political Science Review* (forthcoming).

Anderson, T. W. 1958. *An Introduction to Multivariate Statistical Analysis.* New York: John Wiley and Sons.

Boyd, R. W. 1972a. "Popular Control of Public Policy: A Normal Vote Analysis of the 1968 Election," *American Political Science Review,* June 1972, pp. 429–449.

———. 1972b. "Rejoinder," *American Political Science Review,* June 1972, pp. 468–470.

Cnudde, C. F. 1971. *Democracy in the American South.* Chicago, Illinois: Markham.

Converse, P. E. 1966. "The Concept of the Normal Vote," in Angus Campbell, Philip E. Converse, Warren E. Miller, and Donald E. Stokes, *Elections and the Political Order.* New York: Wiley and Sons.

Finney, D. J. 1971. *Probit Analysis.* New York: Cambridge University Press.

Johnston, J. 1972. *Econometric Methods,* 2nd ed. New York: McGraw-Hill.

Kessel, J. H. 1972. "Comment: The Issues in Issue Voting," *American Political Science Review,* June 1972, pp. 459–465.

Kort, F. 1973. "Regression Analysis and Discriminant Analysis: An Application of R. A. Fisher's Theorem to Data in Political Science," *American Political Science Review,* June 1973, pp. 555–559.

McKelvey, R., and Aldrich, J. 1974. "A Method of Scaling with Application to the 1968 Presidential Election," a paper presented at the Annual Meetings of the Public Choice Society, New Haven, Connecticut.

McKelvey, R., and Zavoina, W. 1969. "A Statistical Model for the Analysis of Legislative Voting Behavior," a paper presented at the Annual Meetings of the American Political Science Association, New York, New York.

Nie, N.; Bent, D.; and Hull, C. 1970. *S.P.S.S. Statistical Package for the Social Sciences.* New York: McGraw-Hill.

Page, B., and Brody, R. 1972. "Policy Voting and the Electoral Pr ocess: The Vietnam War Issue," *American Political Science Review,* September 1972, pp. 979–995.

Zechman, M. 1974. "A Comparison of Small Sample Properties of Probit and O.L.S. Estimators with a Limited Dependent Variable," unpublished paper, University of Rochester.