Discriminant analysis is a multivariate statistical technique that researchers have used successfully in several recent policing studies. This technique allows researchers to divide the sample into meaningful groups that more adequately represent real-life situations and to analyze simultaneously multiple variables that have the potential of explaining group placement. This paper accomplishes four objectives: (1) it describes the general benefits and appropriateness of using discriminant analysis in policing research; (2) it provides an overview of navigating the printed discriminant analysis report; (3) it provides a research case example to demonstrate how discriminant analysis was a tenable procedure when examining police learning strategies, and (4) it compares the value of regression/logistic regression analysis and discriminant analysis in policing research.

The policing profession is a human enterprise; consequently, it is complicated and has many variables. From the earliest research in policing, researchers have largely accepted the assumptions of positivism that a single, objective reality exists, which consists of inter-related variables, and the goal of science is to better understand this reality. Thus, in policing research, much like other social science research, rationalistic designs have attempted to identify, isolate, and measure variables to assist in uncovering these laws and explain reality.

Questions such as description, reliability, causal explanation, prediction, and control are important to the positivist paradigm of research (Carr & Kemmis, 1986). A number of statistical techniques have simplified this rather complex process. Statistical techniques such as analysis of variance (ANOVA) usually examine only one or two variables at a time. Increased access to computers and statistical software has allowed researchers to include more variables in the analysis, and the use of techniques such as regression and factor analysis has increased (Conti, 1993). The criminal justice abstract database has created an index of criminal justice studies conducted between 1990 and 2000 according to their use of these statistical techniques: ANOVA, 15; factor analysis, 48; regression analysis, 73; and logistic regression, 86. While the numbers for policing research conducted between 1990 and 2000 are relatively smaller, they show a similar pattern in use of statistical techniques: ANOVA, 13; factor analysis, 26; regression analysis, 19; logistic regression, 16.

Michael L. Birzer is associate professor and director of the School of Community Affairs at Wichita State University. Delores E. Craig-Moreland is an associate professor at Wichita State University.
Despite the prevalence of these statistical techniques in the criminal justice and policing research literature, some concern exists that these procedures may not adequately address the research needs for the field (Miller & Tewksbury, 2001). This concern centers on the argument that traditional statistical techniques do very little to offer a complete explanation of the phenomenon under study or, as DiCristina asserts, “quantitative research may have an explanation, but this is not the same as being justified” (1997, p. 191). Traditionally, policing researchers have tended to isolate variables, run statistical analyses, and generalize the results to the population under study. Approaches such as these may be problematic for researchers attempting to understand such a complex area as policing.

One response to the need to better understand complex phenomena found within the policing field has been the growth of naturalistic designs that rely heavily on qualitative data (Pogrebin, 2003). Many naturalistic designs combine both qualitative and quantitative data-gathering strategies. Furthermore, triangulation designs that allow researchers to use multiple methods to measure the phenomenon under study are increasingly receiving attention in the literature (Berg, 2001; Creswell, 1998). The use of triangulation designs, when appropriate, is also an effective method for assessing validity (Bayens & Roberson, 2000). What has prompted researchers to move toward the inclusion of more qualitative data in research designs? A major factor in this shift may be that social scientists investigate a situation because they already know something about it and have a desire to understand it better.

Traditional statistical procedures rely on computer software to make crucial data decisions. In policing research this may not be the most appropriate method when the study objective is to understand or explain a phenomenon. However, one statistical procedure, discriminant analysis, allows researchers to retain this decision-making authority. Discriminant analysis allows researchers to make meaningful decisions about the data and to impose sense upon data. Moreover, as a multivariate procedure, it enables simultaneous consideration of several interacting variables that make up the complicated exploratory set.

Although discriminant analysis can bridge the gap between traditional research design and the needs qualitative data address, it does not have a long history of use in criminal justice research. From 1990 through 2000, 18 studies in the criminal justice abstracts (Silver-Platter) database used discriminant analysis; 11 studies used discriminant analysis during this same timeframe in policing research.

Discriminant Analysis

Discriminant analysis is “a powerful multivariate statistical procedure which allows the investigation of the differences between two or more groups in relationship to several
variables simultaneously” (Klecka, 1980, p. 7.) With discriminant analysis, as with other multivariate techniques, the emphasis is on analyzing the variables together rather than singly. In this way, researchers can consider the interaction of multiple variables. Discriminant analysis is useful when known and distinct groups exist (Greenberg, 1979; Marriott, 1974). To use discriminant analysis at least two groups must exist, which differ in several variables. Furthermore, the variables must be measurable at the ratio or interval levels. Discriminant analysis enables researchers to identify such relationships between qualitative criterion variables and quantitative predictor variables (Kachigan, 1986).

In the social sciences, the use of discriminant analysis allows researchers to identify underlying dimensions that help understand differentiation of selected groups. Because policing often entails groups of people and comprises many variables that may distinguish groups, discriminant analysis is an ideal research technique for police researchers. Discriminant analysis classifies existing groups and allows researchers to investigate differences between these groups based on a set of variables. An important feature of discriminant analysis pertains to the individual case that is the basic unit of analysis (Huberty & Barton, 1988). Thus, discriminant analysis is a technique researchers can use to focus on the grouping of people and to analyze the interrelationships of multiple variables to explain a person’s placement in a specific group. This process contrasts with univariate analyses in which researchers examine individual variables separately, dissociating them from the total person who is a synergistic composition of these variables.

Researchers can use the results of discriminant analysis for several purposes, including the prediction of group membership and description of the ways groups differ (Dattalo, 1994). For researchers to accomplish these two purposes requires that the discriminate analysis produce three pieces of crucial information. The first is the discriminant function, a formula that contains the variables and their discriminant coefficients, which can be used to place people in groups. The second is the structure matrix, which names the discriminant function so qualitative terms exist to explain any interaction among variables when distinguishing among groups. The third is the classification table, which indicates the accuracy of the discriminant function in correctly placing people in a given group (Klecka, 1980).

When researchers know that distinct groups exist, discriminant analysis is very useful. The number and type of groups can differ widely and depend largely on the study conducted. For example, the criminal justice studies that have been conducted by using discriminant analysis have been varied: factors that discriminate psychopathology and competence to stand trial (Rosenfeld & Wall, 1998); factors that discriminate a person who feels safe walking at night from the person who does not (Ganjavi, LeBrasseur, & Whissel, 2000); factors that discriminate between gang or non-gang homicide (Bailey &
Unnithan, 1994); factors that discriminate between good performance and bad performance among special event security officers (Leeds & Rains, 1995); factors that discriminate dispositions of felony court cases (Eisenstein & Jacob, 1977); understanding how offenders identify and discriminate between the concepts of free will and determinism in terms of their own self-concepts (Holbert & Unnithan, 1990). Policing studies that have used discriminant analysis include factors that discriminate learning strategies police use (Birzer & Nolan, 2002), factors that discriminate job similarities and differences among university police departments (Mullins, 1986), and factors that discriminate various police and prosecutorial responses to sexual assaults against women (Kerstetter, 1990).

**Group Designation**

One important step in designing the research that will use discriminant analysis is the delineation of groups. This important step allows the researcher to impose sense on the data and to maintain control of the research. In real life, most police practitioners know that certain groups exist; often the most relevant research question relates to what factors can account for a person's placement in that group.

The attributes researchers use to distinguish among the groups are called discriminating variables, or predictors. In general, there is no limit on the number of discriminating variables as long as the total number of cases exceeds the number of variables by more than two (Klecka, 1980). These variables must be measured at the interval or ratio level of measurement to calculate the means and variances. However, some researchers believe there are some exceptions to this general rule. Huberty (1994) points out that in cases where the predictor, or discriminating variable, is categorical in nature this may not necessarily affect the analysis adversely. For example, Bailey and Unnithan (1994) conducted research on gang homicides in California and used discriminant analysis. Their study measured many of the predictor variables categorically rather than continuously, and they reported rather robust results. Similarly, Holbert and Unnithan (1990) investigated the criminal self-perception of adult offenders by using the free will/determinism dichotomy. Their study measured the independent variable of self-concept as a dichotomous variable, but this violation did not adversely affect the analysis in light of the fact that the study examined only a few of the many possible variables in defining an offender's self-perception of criminality.

Likewise, there is no limit on the number of groups one can include in an analysis, and the chances for the correct classification of cases in the proper group increase with large groups (Spearing & Woehlke, 1989). The number of groups most likely depends on the research problem or the theoretical specification. A further consideration is the number
of groups that actually exist and what groups the research seeks to discriminate among. Most of the research in policing and criminal justice conducted with discriminant analysis reveals that the most common grouping pattern is a dichotomous division. For example, Toseland (1982) divided respondents into two groups: those fearful of crime and those that were not; Bailey and Unnithan (1994) divided homicides into gang related and non-gang related; and Kerstetter (1990) divided prosecutorial decisions into two groups—(1) to file felony charges and (2) not to file felony charges.

Criteria for Selection
Traditionally, researchers test null hypotheses, which state no significant relationship between independent and dependent variables: the groups do not differ. Hypotheses in studies that use discriminant analysis may not follow the traditional pattern. Instead these studies use the format of stating that it is possible to discriminate between the groups by using the discriminating variables. This formula simply substitutes the specific variables and groups from the study for the general terms (Hill, 1992). However, specific research questions are answered by specific statistical and post hoc tests, and discriminate analysis has its own. It is also possible to use research questions instead of hypotheses in studies that use discriminant analysis. Here the question merely asks whether it is possible to discriminate among the groups by using the discriminating variables.

Regardless of whether a study uses research questions or hypotheses, it must state the criteria for accepting the outcome of the analysis. Two criteria are appropriate for judging the acceptance of the discriminant analysis as useful. One criterion is that the discriminant function the analysis produces is describable by using the structure coefficients of the analysis; researchers often use a coefficient value of .3 or greater as the criterion for determining whether the analysis will use variables from the structure matrix. The other criterion is that the discriminant functions correctly classify a certain percentage of the cases in a sample. If a discriminant function can be described in a meaningful way, and it correctly classifies cases into the proper group, then it is judged useful.

Navigating the Printed Report
The SPSS (Statistical Package for the Social Sciences) output that reports discriminant analysis resembles that of a regression analysis. At first glance a discriminant analysis report looks very confusing; however, focusing on certain parts makes it very understandable. These parts include the within-groups correlation matrix, the canonical discriminant functions, the structure matrix, the unstandardized canonical discriminant function coefficients, and the classification results.
Pooled Within-Groups Matrix
Because discriminant analysis is a multivariate technique, it examines many variables simultaneously. The effects of a variable may not be discernable if it shares variance with other variables. The printed report reveals the pooled within-groups matrix, which helps identify variables that share covariance. This matrix shows the strength of the relationship between corresponding pairs of variables for the cases within each of the groups identified for the analysis. The researcher should look for those variables with a high-shared variance, and, if present, remove one of the sets of variables from the analysis. In short, shared variance between variables is not meaningful for purposes of discriminant analysis and should be discarded.

Canonical Discriminant Function
The canonical discriminant function portion of the report contains several pieces of information. Eigenvalues appear in this table. Large eigenvalues are associated with useful functions within the discriminant analysis (Norusis, 1988). Chi-square information indicates the likelihood that the groups of the means are the same (Conti, 1993). What is most important is the canonical correlation, which tells how useful the discriminant function is in explaining group differences. Squaring the canonical correlation provides the proportion of variation in the discriminant function explained by the groups (Klecka, 1980). Thus, if the groups do not differ in the variables analyzed, then all correlation will be low, because we cannot create correlation when none already exists.

The Structure Matrix
The structure matrix is one of the most important parts of the report because it depicts correlation coefficients that indicate how closely a variable and the discriminant function are related. High coefficients are of interest to the researcher. A high coefficient indicates that the information in the function is very similar to the variable. This is useful because the researcher will use high coefficients and other variables with high coefficients to name the discriminant function (Klecka, 1980). Likewise, low coefficients would not be useful as they indicate that the overall function and the variable have very little in common.

Unstandardized Canonical Discriminant Function
The unstandardized canonical function coefficients section of the report contains the information for composing the discriminant function. It indicates the variables and coefficients that should be included in each function. It is important to point out that each discriminant analysis produces one less function than the total number of groups. For example, if a study investigates stress levels of police officers, and the instrument the
study uses places officers into one of three stress indicator groups, the canonical discriminant function reports only two functions.

The listed variables with their coefficients and signs are put together in a mathematical statement to express each function. Discriminant scores are then computed by taking the original value for a case and multiplying it by the coefficient for that variable; the products are then added along with the constant term. “The constant term is an adjustment for the means, so that the mean discriminant score will be zero over all the cases” (Klecka, 1980, p. 24). This discriminant score appears like a mathematical formula and is used for classification of each case into a group.

**Classification Table**

The classification table indicates the accuracy of the discriminant function in correctly grouping the cases used in calculating the discriminant analysis (Conti, 1993; Conti & Kolody, 1998). “Classification is the process by which a decision is made that a specific case belongs to or most closely resembles one particular group” (Klecka, 1980, p. 42). The classification table displays the predicted group membership based on the discriminant score for each group and shows the accuracy of placing the group members in their original groups.

As a direct measure of the predictive accuracy, this percentage is the most intuitive measure of discrimination. One should, however, judge the magnitude of this percentage in relation to the expected percentage of correct classifications if assignments were made randomly (Klecka, 1980). That is, if the analysis involved two groups, there is a 50% likelihood that random assignment could classify the cases correctly. Thus, the researcher must interpret the accuracy of the classification results in relationship to what one can expect from random assignment.

The measure of predictive accuracy is the most intuitive measure of discrimination. The researcher should judge the magnitude of this percentage in relation to the expected percentage correctly classified as if assignments were made randomly (Klecka, 1980). For example, if the analysis involved three groups, the likelihood that the cases could be correctly classified by random assignment is 33.3%. The researcher must interpret the classification results while remaining cognizant of what a random assignment would be.

**Police Learning Strategy Research**

One of the authors recently used discriminant analysis in research that investigated learning strategies police officers use. We discuss this research here to exemplify how
discriminant analysis was a successful multivariate technique when a study identifies multiple groups.

Learning strategies are techniques an individual uses to approach specific learning situations (Conti & Kolody, 1995) and accomplish a learning task (Fellenz & Conti, 1989). Learning is a continuous and complex process for police officers and may take place in many situations. Police officers are typically involved in tasks such as identifying problems and potential solutions, assessing community needs, evaluating options, and implementing the most appropriate option. Many of these tasks they learn informally and in the field when they must make daily and sometimes immediate decisions tailored to specific situations.

The author administered the Assessing the Learning Strategies of Adults (ATLAS) instrument, which is based on the research of Conti and Kolody (1998), to a sample of police officers who were employed by a large Midwestern urban police department. These police officers were assigned to either community oriented policing duties or traditional beat patrol assignments. The ATLAS places participants into one of three learning strategies groups: Problem Solvers, Engagers, or Navigators.

The set of discriminant variables the researcher used to explain placement in these groups consisted of age, gender, prior education, job assignment, and years of experience. The study used the following discriminant function to classify the cases and serve as guide for describing future placement of respondents into these groups:

\[
D = 1.375 \times \text{gender} + 0.204 \times \text{age} - 0.179 \times \text{years of experience} - 1.099 \times \text{job assignment/community policing or traditional patrol} - 0.170 \times \text{educational level of participant} - 5.171 \times \text{constant}.
\]

Three variables had sufficient coefficients (cutoff point at .03) to include in the interpretation of the meaning of the discriminant function: gender (.31), job assignment (-.40), and age (.36). Since the coefficients for all three variables were similar in value, they carried equal weight in naming the discriminant function. Two of the five variables (experience and education) had low coefficients; therefore, they were not used in the interpretation of the meaning of the discriminant function. Based on the strength of the variables (gender, job assignment, and age), the researcher named a discriminant function. Examining these variables simultaneously allowed the researcher to glean a better understanding of the officers in the three learning strategy preference groups.

The research also reinforced the argument for multivariate designs. The researcher determined from other analyses (i.e., chi-square and analysis of variance) that individual variables by themselves were not significant in describing learning strategy group placement. The study needed a multivariate procedure that could add meaning and help
describe but not necessarily predict the learning strategies police officers used. In this sense, a multivariate procedure such as discriminant analysis proved useful in further understanding this complex phenomenon.

The discriminant function allowed the researcher to describe more fully what takes place among people within the learning strategy groups. The discriminant analysis revealed that the variables of age, gender, and job assignment were significant in describing the process. Moreover, discriminant analysis allowed the researcher to examine multiple variables simultaneously to see whether they could describe learning strategy group placement. It is important to note that in the discriminant function, it is not any one variable independently that explains group placement, but it is a constellation of variables.

**Discriminant Analysis & Multiple Regression Analysis**

Some researchers who rely heavily on multiple regression analysis may ask the question, why not use regression? This valid question deserves consideration. Recall that discriminant analysis has several tools for the interpretation of data (Klecka, 1975). The most powerful of these is the structure matrix, which allows the researcher to name the discriminant function. Thus, while regression analysis allows for the identification of variables and the strength of these variables in contributing to the prediction or description of the criterion variable, discriminant analysis provides the researcher with the tools to describe the differences between the groups (Klecka, 1975).

While multiple regression analysis allows the researcher to say that variables X and Z, for example, contribute to recidivism, discriminant analysis allows the researcher to distinguish between those inmates with a high risk for recidivism and those inmates with a low risk (assuming that the sample comprised these two groups). Thus, the regression formula would allow the researcher to predict recidivism while discriminant analysis would allow the researcher to distinguish characteristics of group members. Note how regression analysis focuses on the criterion variable while discriminant analysis focuses on the people in the groups.

Regression analysis and discriminant analysis do have some features in common because they are part of a general class of multivariate statistics. Indeed, when one calculates multiple regression analysis with a dichotomous criterion variable, the two types of analysis are very similar (Greenberg, 1979). The purpose of discriminant analysis is to provide “a powerful technique for examining differences between two or more groups of objects with respect to several variables simultaneously” (Klecka, 1975, p. 435). On the other hand, multiple regression analysis focuses primarily on predicting the criterion variable. These are quite different purposes. Moreover, discriminant analysis then includes the tools to make meaning of these differences.
Logistic Regression

Some researchers have begun to use logistic regression analysis, a relatively new technique that examines the relationship among variables. Researchers can use both discriminant analysis and logistic regression for prediction or explanation (description). In logistic regression, similar to bivariate or multiple regression, there is only one dependent variable; however, the dependent variable is dichotomous in nature (Huck, 2000; Walsh & Ollenburger, 2001). This condition is equivalent to having just two groups in discriminant analysis, which, in some research studies, may indeed be a limitation. In discriminant analysis you can have more than two groups.

Klecka (1980) points out that discriminant analysis does not describe the groups and discriminating variables as independent and dependent; in other words, the analysis does not suggest a cause and effect relationship. Logistic regression does designate groups in this manner. Thus, logistic regression suggests causation while discriminant analysis shows the nature of differentiation. For example, in the police learning strategies study discussed previously, the researcher grouped participants by use of learning strategies and used demographic variables to see whether it was possible to discriminate among the learning strategy groups. If the study had used logistic regression analysis, the conclusion may have been that these demographic variables caused the learning strategy preference. Of course, this assumption would be spurious; however, it would be reasonable to conclude that a relationship exists between these factors and learning strategy group preference.

As previously pointed out, discriminant analysis produces the structure matrix, which allows the researcher to name the discriminant function and, thus, explain or describe the process that separates the groups. In the social sciences this explanation may be in large part what we are interested in. In essence, we want to explain human behavior. Logistic regression produces an odds ratio, which allows the researcher either to predict or to describe the likelihood that a person is one group or the other. While this outcome is desirable in some studies, it may be, in some studies, more desirable to be able to describe the process that separates the groups and, therefore, causes the person to be in one group. For example, it may be more interesting to discern factors that describe how a community policing officer learns than merely to predict the officer's learning strategy preference.

With the use of discriminant analysis, the researcher tries to interpret and give meaning to the differences that are present inherently in the data. Discriminant analysis helps organize and make these differences clear. With logistic regression, a main objective in research seems to be creating models. Thus, the interest lies in trying to find a good set of independent variables that can help predict group membership on the dependent variable. For this modeling process to be successful, the researcher must possess a thorough...
understanding of his or her discipline to decide which of a vast array of variables to measure (Huck, 2000).

In logistic regression modeling, the researcher may very well run different models with different variables and then select the final one. Hence, a different combination of variables may give different results. This result may in one sense limit the objectivity of the study when the final model selected is merely a function of the a priori. The situation is quite different when using discriminate analysis. With discriminate analysis, any subjectivity that goes into naming the discriminate function comes after the process is run. That is, two people may look at the structure matrix and name it differently because of the way they look at the same combination of variables, or different researchers may use a different cutoff point for the coefficients in the matrix (for example, .4 instead of .3) to use in the naming process. However, these decisions do not affect the statistical outcome of the process. When choosing between logistic regression and discriminate analysis, police researchers should consider (1) the purpose of the analysis (classification or description); (2) the characteristics of the sample; and (3) the tenability of assumptions.

The work of Crawford and Burns (2002) illustrates the opportunities available and tradeoffs in use of discriminate analysis compared with logistic regression. Their study of resisting arrest and use of force against police featured analysis of data by means of a series of three logistic regressions employing 15 independent variables.

The research involved three categories of arrest conditions: compliance, resisted by passive means, and resisted by violent means. The first logistic regression compared those who complied with arrest and the group formed by combining the passive resisters with the violent resisters. Male officers were less likely to encounter resistance, suspects most likely to resist arrest were angry and aggressive and impaired by drugs, and the situation most likely to produce resistance was domestic violence. The strongest predictor of resisting arrest was an angry/aggressive suspect. This variable is a problem because the information for the variable was taken from police reports about a suspect’s initial response to arrest. Virtually every arrest where the suspect responded with violence was reported as angry/aggressive. In other words, this independent variable is essentially identical to the dependent variable category of violent resisting arrest.

The second logistic regression involved a comparison of those who offered passive resistance to arrest with all other arrestees. The results showed that significant predictors included suspect features of being angry/aggressive and impaired by drugs. No officer or situational variables were significant.

The third logistic regression compared violent resisting arrest with all others. The significant predictors included officers with more years on the police force, suspects
reported as angry/aggressive and impaired by drugs, and witnesses present at the arrest doubling the likelihood of violent resisting of arrest.

The biggest bonus in use of discriminant analysis for this research would be that it allows a comparison of compliant arrests, passive resistance arrests, and violent resistance arrests. This research does not help us to understand what distinguishes these two very different degrees of resistance. Discriminant analysis permits more than two groups in the dependent variable.

At the same time that discriminant analysis offers this simultaneous consideration of the three arrest responses, it would limit use of many of the variables in this study or require a different level of measurement of the variable. The three police officer variables included one continuous variable of length of time on the force and two dichotomous variables of gender and whether the office was white. No means of converting the two dichotomous variables is possible. The eight suspect variables were all dichotomous, but three of the variables would readily convert to continuous variables with probable improvement in precision in the research. One could convert “Suspect young” to “Suspect age”; “Suspect impaired by alcohol” to “Blood alcohol level”; “Suspect impaired by drugs” to parts per million of the drug onboard in the drug test results, “PPM of drugs onboard.” The situation variables were four dichotomous measures that included two readily convertible to continuous measures. “Arrest at night” converts readily to “Time of arrest,” and “Witnesses present” converts to “Number of witnesses/onlookers.”

Summary and Conclusion

Discriminant analysis is a multivariate procedure that has much potential in police research. Unlike univariate techniques, it can allow the simultaneous analysis of many variables in the complex phenomenon of policing; in allowing this complexity, it more closely reflects real life than the univariate process of isolating variables for analysis. In allowing the researcher to impose sense on the data by forming groups that are known to exist and that are meaningful, it borrows from the qualitative approach. These two strengths make discriminant analysis an ideal technique to use when analyzing social science data.

To be recognized as legitimate, theoretical claims must be in the language of members of the intellectual community who are the direct audience for the work—readers, students, colleagues (Alford, 1998). Discriminant analysis not only employs the academic rigors and language of research but also provides a means to understand what the data actually mean when applied in real-life situations. The strength of discriminant analysis is the structure matrix, which allows the researcher to name and, therefore, better under-
stand the process that distinguishes the groups. This outcome, in many policing studies, is the heart of what we are trying to do.

Both discriminant analysis and regression analysis allow the researcher to explore the interaction of several variables together. The researcher must choose which one makes the most sense to him or her. More important, the researcher should choose the technique that most appropriately gleans the answers to the research questions that he or she has framed. Univariate procedures cannot address the complicated human interactions policing involves. A true understanding of the interactions occurring requires a multivariate approach so researchers can see how their variables interact. Although discriminant analysis has had limited use in policing research, the results of recent studies that used the technique offer much promise. Discriminant analysis has the potential to be a viable procedure for both descriptive and predictive purposes.

References


