Regression Analysis:

How This Technique Is And Can Continue to Be Used Effectively in Criminal Justice Research

---

A Thesis

Presented to

the Chancellor's Scholars Council of

The University of North Carolina at Pembroke

By

Danielle Carden

April 28, 1998

Faculty Advisor's Approval ______________________

Date 4-30-98

A

Table of Contents

Multiple Regression Analysis:
How This Technique Is and Can Continue To Be Used Effectively in Criminal Justice Research

Thesis: Multiple regression analysis, a very effective and powerful analytic tool, is one of the most popular statistical techniques used in the behavioral or social sciences, particularly in the Criminal Justice field. This technique is and can continue to be used effectively if the researcher follows the specific assumptions that this method dictates.

I. Introduction to Statistics

A. The definition of statistics is presented.
B. Descriptive statistics is defined.
C. Inferential statistics is defined.

II. Introduction to Multiple Regression Analysis

A. The definition and equation for multiple regression analysis are presented.
B. Why regression analysis is used in the behavioral sciences.
C. How multiple regression developed.
D. The pros and cons of multiple regression analysis.
  1. The pros and cons of multiple regression analysis as presented by Cohen and Cohen.
  2. The pros and cons of multiple regression analysis are presented by Aiken and West.
E. Why multiple regression is a powerful and versatile technique used in the behavioral sciences.
  1. Multiple regression is versatile because the regression equation can automatically handle the researcher's data and analyze it in the hopes of finding the answer to their research hypothesis.
  2. Multiple regression can handle many types of data.
    a. Nominal data is presented in a multiple regression analysis setting.
    b. Ordinal data is presented in a multiple regression analysis setting.
    c. Interval data is presented in a multiple regression analysis setting.
    d. Ratio data is presented in a multiple regression analysis setting.
  3. Three reasons from James Stevens about using multiple regression analysis.
  4. The best reason for using multiple regression analysis from James Jaccard, et al.
F. The design structure for using multiple regression analysis is established by Richard J. Shavelson.

III. Assumptions Needed For Using Multiple Regression Analysis As Presented By Richard J. Shavelson.

A. One assumption to using multiple regression analysis is independence.
B. The second assumption to using multiple regression analysis is normality.
C. The third assumption to using multiple regression analysis is homoscedasticity.
D. The fourth assumption to using multiple regression analysis is linearity.
E. Additional assumptions needed for multiple regression analysis as defined by William D. Berry and Stanley Felman.

IV. Problems With Multiple Regression Analysis That Should Be Avoided

A. One problem that should be avoided with multiple regression analysis is multicollinearity.
   1. An explanation of multicollinearity is provided.
   2. Symptoms of multicollinearity are described.
   3. What the researcher should do if multicollinearity does exist in his or her research.

B. Another problem that exists with multiple regression analysis is specification error.
C. A third problem with multiple regression analysis is measurement error.

V. The Effectiveness of Multiple Regression Analysis in Criminal Justice Research

A. James P. McGregor argues that the regression model is overused by social scientists.
B. Introduction to specific situations where multiple regression analysis has been used effectively in criminal justice research.

VI. Multiple Regression in Evaluating Opinions on the Death Penalty

A. Multiple regression can determine the characteristics of those in favor of the death penalty.
B. Multiple regression can determine the characteristics of those who disapprove of the death penalty.
C. How the death penalty opinion study was executed by Marla Sandys and Edmund F. McGarrell.

VII. Multiple Regression and How It Can Be Used to Evaluate the Effectiveness of Legislative Statutes

A. Introduction to Bill C-127, a Canadian rape reform bill that was studied by Rita Gunn and Rick Linden through using multiple regression analysis.
B. Logistic regression analysis of Bill C-127 is described.
C. An analysis of Gunn and Linden's findings are presented.

VIII. Using Multiple Regression to Evaluate Deterrence Theory Models

A. The definition of deterrence theory models is presented.

B. The research design model question is approached by Andy Anderson, et al. in determining how to use multiple regression analysis to evaluate deterrence theory models.
C. The research model and findings of Anderson, et al.'s study on deterrence theory models is explained.

IX. Evaluating Multiple Regression in Predicting Criminal Recidivism

A. An introduction to the study, "Predicting Criminal Recidivism: A Comparison of Neural Network Models with Statistical Methods" by Jonathan Caulkins, et al. is provided.
B. Neural Network models are explained.
C. The findings from Caulkins, et al. are provided.
D. A similar study executed by Brent B. Benda, et al. undergoes analysis.
E. The findings from Benda, et al. are provided and explained.

X. Multiple Regression Analysis is Evaluated Through Using A Statistical Program, Microcase

A. A description of the Microcase statistical program is given.
B. The three Microcase examples/tables are presented.
C. The data presented in Table 1--The Percentage of White, Homeless and/or Unemployed Single men Who Commit Burglary is explained.
D. The data presented in Table 2--The Number of Larceny and Auto Thefts That Occur in a Metropolitan Area As They Effect the Population of 1990 is explained.
E. The data presented in Table 3--The Number of Larceny and Auto Thefts That Occur in a Rural Area As They Effect the Population of 1990 is explained.
F. The data in all three tables is compared and contrasted as the effectiveness of multiple regression analysis .

XI. Conclusion

A. A summation of multiple regression analysis is presented.
B. Why multiple regression is one of the best techniques to use in Criminal Justice research.
C. How multiple regression analysis can continue to be used effectively in Criminal Justice research.

Regression Analysis:
How This Technique Is and Can Continue to Be Used Effectively in Criminal Justice Research

Introduction to Statistics

Statistics is the collection and analysis of information, called data, which numerically answers questions or hypotheses that a researcher has developed about a particular topic that concerns him or her (Weinberg 1). Statistics developed from two areas. The first of these areas came from nations such as England where the government wished to collect data on its population. The origin of the word "statistics" came from this idea. As Frederick D. Herzon and Michael D. Hooper state in their book, *Introduction to Statistics for the Social Sciences,* "The word 'statistics' originally came into the language as meaning the collection, compilation, and analysis of facts about the state" (2). This type of statistics came to be known as descriptive statistics, or those that "...*describe* the data that have been (or will be) collected" (Weinberg 2). The second area statistics developed from was gambling. Herzon and Hooper state, "The second major source of statistics lay in the desire of certain people who gambled to predict the outcome of games of chance. This interest led to the development of the field of probability. The development of probability theory allowed statisticians to make inferences about the outcomes of situations in which a random mechanism was present" (2). This type of statistics is known as inferential statistics, or statistics that allow the researcher to generalize information from the sample studied, to the entire population (Weinberg 2).

Introduction to Multiple Regression Analysis

Many statistical procedures exist in order to analyze information. Multiple regression analysis, a very effective and powerful analytic tool, is one of the most popular statistical techniques used in the behavioral or social sciences, particularly in the Criminal Justice field. This technique is and can continue to be used effectively if the researcher follows the specific assumptions that this method dictates.

Regression analysis can be defined as "...a highly general and therefore very flexible data-analytic system that may be used whenever a quantitative [numeric] variable (the dependent variable) is to be studied as a function of, or in relationship to, any factors of interest (expressed as independent variables)" (Cohen and Cohen 3). This procedure is calculated through using the equation:

$$Y = a_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots b_k X_k + e$$

(Lewis-Beck 48-49). In simpler terms, regression analysis is a mathematical procedure that allows one to study many independent variables, X, as they affect one dependent variable, Y.

Regression analysis, a very complex subject, has two main forms. It can involve two variables and be called bivariate analysis; however, it may involve many variables and be considered a multivariate analysis. The researcher must determine which type of regression analysis--bivariate or multivariate they plan to use for their research. The type of analysis a researcher uses depends greatly upon what types of factors the researcher is trying to find. Once the researcher has decided upon bivariate or multivariate regression analysis, more specific types of regression analysis such as straight line regression or curvilinear regression, etc. must be decided upon to determine the causal relationship between the variables.

In the behavioral sciences, simple experiments which only have one independent or causal variable (X) producing an effect on the dependent variable (Y) can be used. However, many times several causal variables may have an effect on the dependent variable. Cohen and Cohen explain this:

> Consider a representative set of dependent variables: epinephrine secreted, distant word associations, verbal learning, school achievement, psychosis, anxiety, aggression, attitude toward busing, income, social mobility, birth rate, kinship system. A few moments' reflection about the causal nexus in which each of these is embedded suggests a multiplicity of factors, and possibly further multiplicity in how any given factor is represented....But MRC [Multiple Regression Correlation] makes possible the use of paradigms of the form Y= f(C, D, E, etc.), which are far more efficient than the strategy of studying multiple factors at one time. Moreover, causal analysis, utilizing interlocking regression equations as formal models, achieves an even greater degree of appositeness to complex theories (Cohen 7).

Once the researcher has defined his or her variables for study, two options are available. The researcher can do a simple experiment using each variable or the researcher can use one simple experiment using multiple regression analysis which allows the researcher to look at all of the independent variables' effects upon a dependent variable at one time, thus saving him or her time. One large experiment usually produces the same results as many simple experiments, therefore, many researchers choose multiple regression to execute their experiments.

Multiple regression analysis or multiple regression correlation developed because complex research hypotheses could not be tested using other research methods. In their book, *Multiple Regression: Testing and Interpreting Interactions,* Leona S. Aiken and Stephen G. West describe the background to multiple regression analysis. This procedure developed due to "...interest in complex hypotheses that are not adequately represented by simple, additive linear regression equations is common in many disciplines" (Aiken and West 2).

The procedure of multiple regression analysis did not develop overnight, however. In fact, it was in development for many years before its initial utilization in the behavioral sciences. As Aiken and West state:

> The general methods of structuring complex regression equations to test such hypotheses explicitly were first proposed over two decades ago in the social sciences. Cohen (1968) proposed multiple regression (MR) analysis as a general data analytic strategy. According to this strategy, any combination of categorical and continuous variables can be analyzed within a multiple regression (MR) framework simply through the appropriate dummy coding of the categorical variables. Interactions can be represented as product terms, and curvilinear relationships can be represented through higher order terms in the regression equation (3).

They also mention that early attempts to create multiple regression were made by many other sociologists such as Allision, Blalock, and Southwood (Aiken and West 3-4).

Jacob and Patricia Cohen, authors of *Applied Multiple Regression/Correlation Analysis,* state "Historically, MRC arose in the biological and behavioral sciences around the turn of the century in the study of the natural covariation of observed characteristics of sample subjects" (Cohen and Cohen 4). However, Cohen and Cohen also state that the recent popularity of multiple regression has not been looked upon as a favorable research tool and that:

> ,,,MRC [Multiple Regression Correlation], because of its association with nonexperimental , observational, survey-type research, came to be looked upon as less scientifically respectable than AV/ACV [Analysis of Variance/Analysis of Covariance], which was associated with experiments. The recent development

> of causal analysis, formal systems of inference based on nonexperimental data, with its heavy dependence on regression analysis, has tended to offset this onus (5).

Aiken and West tend to agree with Cohen and Cohen about this issue. They believe that multiple regression correlation has not been used as frequently as it may be used thus leading them to comment, “Despite the availability of general procedures for testing interactions and curvilinear effects within a MR [multiple regression] framework, the actual practice of researchers in many areas of social science, business, and education indicates these strategies have only rarely been followed” (4). Aiken and West have theorized “We believe this underutilization of MR approaches stems in large part from several impediments that arise when researchers actually attempt to utilize the general procedures that have been outlined and to interpret their results” (5), as to why multiple regression has not been used as frequently as other research strategies. They continue to believe these impediments to be “...procedures for displaying and probing significant interactions have not been readily available. That is, once an interaction was found to be significant, exactly what should on do next?” (5) and “...another impediment to the use of MR with interaction terms: the lack of invariance of the MR results even under simple linear transformations of the data” (5).

When interpreted, Aiken and West mean that once an experiment is determined to have more than one causal variable, researchers do not know how to effectively analyze and conclude their resulting data. The researchers seem to have not found any inequality between the impact of causal variables. Therefore, it seems Aiken and West conclude that many other research procedures exist that will find the same information that multiple regression would also find, so to other researchers it makes sense to choose those other research strategies, whether simpler than multiple regression or not. However, it is also known, that in Criminal Justice research, multiple regression analysis is overused. In fact, it is used for types of research that it should not be used for.

Despite these statements from Aiken and West, multiple regression analysis is a very powerful analytic tool, described at length by Cohen and Cohen. They state:

> The MRC system has other properties that make of it a powerful analytic tool: it yields measures of the magnitude of the 'whole' relationship of a factor to the dependent variable, as well as of its partial (unique, net) relationship, that is, its relationship over and above that of other research factors (proportions of variance and coefficients of correlation and regression) (3-4).

Also, multiple regression has the necessary statistical tools to test research hypotheses, estimate data, and analyze the data or "In short, ...it is a versatile, all-purpose system for analyzing the data of the behavioral, social, and biological sciences and technology" (Cohen and Cohen 6).

Multiple regression is versatile because the regression equation can automatically handle the researcher's data and analyze it in the hopes of finding the answer to their research hypothesis (Cohen 6). As Cohen and Cohen state, "...in the MRC system, the latter has 'built-in' effect-size measures that are unit-free and easily understood and communicated" (7). Also, the most important feature of multiple regression is the ability to handle many types of variance, "MRC's capability for assessing unique variance, and the closely related measures of *partial* correlation and regression coefficients it provides, is perhaps its most important feature, particularly for observational (nonexperimental) studies (Cohen and Cohen 9). MRC can also be cosidered a flexible tool because it can use data in many forms. Cohen and Cohen state, "The capacity of MRC to use information in almost any form, and to mix forms as necessary, is an important part of its adaptive flexibility. Were it finicky about the type of input information it could use, it could hardly function as a *general* data-analytic system" (11). This is particularly helpful to a researcher in the behavioral sciences. Oftentimes, the data one is looking at is something that can not be easily observed, such as the number of particular crimes committed in a specific population area. However, if the number of crimes can be observed, it is not always easy to find a reason or cause for the number of crimes that have been committed. Therefore, a researcher often chooses multiple regression analysis to help him or her to determine the specific reasons such as the criminal's education, economic background, religion, or other influences that may have affected his committing the particular crime.

The types of data, nominal, ordinal, interval, and ratio that Multiple Regression Correlation can use. Nominal information is qualitative or non-numerical information, that simply gives information. Multiple regression is typically not used with this type of data; however, Cohen and Cohen describe a way in which multiple regression can be used with nominal variables. They state:

> Traditional MRC analysis was generally restricted t o quantitative scales with (more or less) equal intervals. But much information in the behavioral sciences is not quantitative at all, but qualitative or categorical, or measured on "nominal" scales.....the qualitative information which constitutes nominal

> scales may be expressed quantitatively, and used as independent variables in MRC (10).

Another type of data that a researcher can use multiple regression analysis with is ordinal data. Ordinal data is nominal data, but it has rank order to it. Cohen and Cohen state, "Whether simple rank order values are used, or they are expressed as percentiles, deciles, or quartiles, these properties of ordinal scales are the same" (10). A third type of data that a researcher can use multiple regression with is interval data. Interval data, like ordinal data, are ranked, but they have a true zero point to start from. Ratio scales are similar to interval scales, therefore the last two levels of data previously discussed are usually combined to be called interval-ratio. Multiple regression is typically used for data on the interval-ratio level rather than the nominal or ordinal level. Cohen and Cohen rationalize this by saying, "Most psychological measures and sociological indices are at this level, for example, the scores of tests of intelligence, special abilities, achievement, personality, temperament, vocational interest, and social attitude" (10).

Aiken and West discuss the reasons a researcher can and should choose multiple regression analysis when analyzing their research data. The main reason that one should choose multiple regression analysis is its reliability and statistical power. "Measurement error in individual predictors produces a dramatic reduction in the reliability of the higher order terms constructed from them. In turn, reduced reliability of higher order terms increases their standard errors and consequently reduced the power of their statistical tests" (Aiken and West 139). Therefore, it makes sense to utilize multiple regression analysis to analyze the research data. For instance, the multiple regression equation has an error constant, *e*, in the equation to keep the possibility of error by the researcher, or from the existing data, in his or her data analyzation as small as possible. The error constant is different for each equation that the researcher attempts; however, for that equation it is constant.

James Stevens in his book, Applied Multivariate Statistics For The Social Sciences, also details reasons for using multiple regression or multiple criterion measures. "There are three good reasons why the use of multiple criterion measures in a study...is very sensible" (Stevens 2). His reasons are that variables, specifically in this case treatments, will have different effects on different subjects or people. Also, the research hypothesis can be explained more specifically if the researcher studies more variables and their effects on the subject rather than simply studying only one variable. Finally, Stevens discusses cost as a major issue when obtaining data, and he, like many researchers, believes that looking at various dependent variables on a subject is less costly than other research techniques and enables them to gather more information to support or not support their research hypothesis (2).

Perhaps the best explanation for why multiple regression is such a popular technique is given by James Jaccard, et al., in the book *Interaction Effects in Multiple Regression.* In this book, Jaccard and others describe the six different relationships that can be described through multiple regression analysis. These are the direct causal relationship, the indirect causal relationship, the spurious relationship, the bidirectional or reciprocal causal relationship, the unanalyzed causal relationship, and the moderated casual relationship. The explanations of these are as follows:

> 1. A *direct* causal relationship is one in which a variable, X, is a direct cause of another variable, Y. (i.e., it is the immediate determinant of Y within the context of the theoretical system)
>
> 2. An *indirect* causal relationship is one in which X exerts a causal impact on Y, but only through its impact on a third variable, Z.
>
> 3. A *spurious* relationship is one in which X and Y are related, but only because of a common cause, Z. There is no formal causal link between X and Y.
>
> 4. A *bidirectional of reciprocal* causal relationship is one in which X has a causes an influence on Y, which, in turn, has a causal impact on X.
>
> 5. An *unanalyzed* relationship is one in which X and Y are related, but the source of the relationship is unspecified.
>
> 6. ...a *moderated* causal relationship is one in which the relationship between X and Y are related, but the source of the relationship between X and Y is moderated by a third variable, Z. In other words, the nature of the relationship between X and Y varies, depending on the value of Z (Jaccard 7).

For graphical representation of these relationships see Appendix A

Richard J. Shavelson in his book *Statistical Reasoning for the Behavioral Sciences, Second Edition,* defines the purpose of multiple regression analysis, but then goes on the state the specific design structure for the analysis. If the researcher does not design the multiple regression analysis to his or her specific data, problems may arise. Shavelson argues, " The design issue is one of matching the research design to the substantive problem without distorting the problem. This means that the behavioral scientist must have a firm grasp of a variety of research designs, along with the logical and statistical rationale underlying them, and build the design most appropriate for the phenomenon being studied" (1). Shavelson lays out three specific design requirements for multiple regression analysis. These are:

> 1. There is, one dependent (or criterion) variable and two or more independent (or predictor) variables.
>
> 2. All variables are continuous.
>
> 3. The minimal sample size needed to provide adequate estimates of the regression coefficients is something like 50 cases, and a general rule of thumb is that there should be at least about 10 times as many cases (subjects) as independent variables.

Assumptions Needed For Using Multiple Regression Analysis

Shavelson also laid out four assumptions that one must have in order to use multiple regression analysis to analyze date. These include independence, normality, homoscedasticity, and linearity:

> In order to use MRA to test hypotheses statistically, we make the following assumptions:
>
> 1. *Independence.* The scores for any particular subject are independent of the scores of all other subjects.
>
> 2. *Normality.* In the population, the scores on the dependent variable are normally distributed for each of the possible combinations of the levels of the X variables.
>
> 3. *Homoscedasticity.* In the population, the variances of the dependent variable for each of the possible combinations of the levels of the X variables are equal.
>
> 4. *Linearity.* In the population, the relation between the dependent variable and an independent variable is linear when all other independent variables are held constant (593).

In addition to these four assumptions, William D. Berry and Stanley Felman, authors of *Multiple Regression in Practice,* mention seven other assumptions one must have to use multiple regression analysis. These are:

1. All variables must be measured at the *interval level* and *without error.*

2. For each set of values for the k independent variables, ($X_1, X_2...X_k$), e=0 (i.e, the mean value of the error term is 0).

3. For each set of values for the k independent variables, VAR = ^{2} (i.e., the variance of the error term is constant).

4. For any two sets of values for the k independent variables, COV = 0 (i.e., the error terms are uncorrelated; thus there is no auto correlation).

5. For each X_i, COV= 0 (i.e., each independent variable is uncorrelated with the error term.)

6. There is no *perfect collinearity* --no independent variable is perfectly linearly related to one or more of the other independent variables in the model.

7. For each set of values for the k independent variable, E_j is normally distributed (Berry 10-11).

Problems With Multiple Regression Analysis That Should Be Avoided

If one follows Shavelson's design for multiple regression analysis, he or she should not run into any data analytic problems. However, several potential problems with multiple regression analysis do exist. These problems are known to researchers as multicollinearity, specification error, and measurement error.

Multicollinearity exists when an independent variable is perfectly correlated with one or more of the other independent variables in the research study. Michael Lewis-Beck in the book, *Applied Regression: An Introduction* describes the problem and ways it can be eliminated. "For multiple regression to produce the 'best linear unbiased estimates,' it must meet the bivariate regression assumptions, plus one additional assumption: the absence of *perfect multicollinearity*. That is, none of the independent variables is perfectly correlated with another independent variable or linear combination of other independent variables" (Lewis-Beck 58).

A list of symptoms of multicollinearity in the researcher's data collection include:

1. One rather sure symptom of high multicollinearity is a substantial R for the equation, but statistically insignificant coefficients.

2. A second, weaker, signal is regression coefficients which change greatly in value when independent variables are dropped or added to the equation.

3. A third, still less sure, set of symptoms involves suspicion about the magnitudes of the coefficients.

4. A fourth alert is a coefficient with the 'wrong' sign. Obviously, this last symptom is feeble, for knowledge of the 'right' sign is often lacking (Lewis-Beck 59-60).

If symptoms of multicollinearity are apparent, the researcher must look to his actual data analysis, usually done with a computer program, to determine if indeed multicollinearity does exist. This can be done, as Lewis-Beck describes, by looking at the "...bivariate correlations among the independent variables, looking for coefficients of about .8, or larger....if none is found, one goes on to conclude that multicollinearity is not a problem" (60). If multicollinearity is found, the researcher must determine if he or she should throw out from the data one or more of the independent variables to see if those variables are causing a multicollinearity problem or if some other influence is affecting the research.

Specification error occurs when the researcher chooses the wrong type of regression model to use to interpret his or her data findings. This type of error can occur in two forms with the first of these being that the regression model has the proper variables specified as independent or dependent variables but the form of the relationship is wrong. As Berry explains, "The regression model assumes that the relationships between the independent variables and the dependent variable are both linear and additive [as defined by the multiple regression equation]. If these assumptions are violated, the least squares estimation will be biased" (18). The second type of specification error appears when the researcher uses the wrong independent variables in the analysis of data. Berry explains this by saying, "Either one or more variables that should have been in the model are omitted, or one or more variables that should not have been included are, or both" (18). The problem of specification error can not be easily detected once the analysis of data is complete, so the best option in avoiding this problem is for the researcher to make sure that the variables he or she is using are the correct ones to prove the research hypothesis because as Berry states, "...there is no clear-cut way of knowing, after the fact, that a substantively important variable has been excluded from the analysis" (25).

Another problem that may occur in regression analysis that is perhaps larger than multicollinearity or specification error is that of measurement error. As Berry states, "Measurement error is one of the most important methodological problems facing the social sciences, and it can have a major impact on the estimation of regression coefficients in otherwise well-specified models.... No measure is ever perfect; the question is how much errors present and what impact it will have on the analysis of data" (26). As specification error had two types, measurement error also has two types. These are non random measurement error and random measurement error. Non random measurement of error occurs when the researcher measures an additional variable along with the variable they are actually interested in studying. Random measurement error into measures for many reasons such as error by the researcher in creating his or her research categories, guessing by the researcher, or coding mistakes (Berry 26-27). The best way to avoid random measurement error is to make an effort to reduce it from the beginning of the research project. Also, the researcher can design data collection and data coding to reduce the chance of measurement error occurring at any point throughout the research process (Berry 33).

The Effectiveness of Multiple Regression Analysis in Criminal Justice Research

James P. McGregor, author of "Procrustus and the Regression Model: On the Misuse of the Regression Model," has argued that the regression model is overused by social scientists but it is not worthless. He states:

> Is the regression model then worthless to social scientists? Absolutely not. While routine application of regression methods in the context of discovery may be dangerous, use of regression in the context of confimation is quite useful.. If one's theory explicitly permits the use of the linear, additive model, then it is entirely appropriate to apply it, provided that the data assumptions of the model are met. More importantly, bivariate regression can be used to discover constants and confirm laws" (803).

Also, "...in the great majority of the cases where the linear, additive regression model is not suitable, we need to strive harder to develop other forms of relationships among...phenomena" (McGregor 804). The following sections will provide the support needed for McGregor's statements.

Multiple Regression in Evaluating Opinions on the Death Penalty

Multiple regression analysis, as mentioned, allows the researcher to analyze multiple variables and their effects upon one particular cause. In criminal justice, a researcher can analyze concrete data as to who commits particular crimes such as rape, murder, burglary, arson, etc., i.e. how many white, unemployed, and homeless men commit burglaries. Multiple regression can also measure abstract data such as how people feel on a certain topic, i.e. how people view applying capital punishment or how people feel about sex offenders, who move into their neighborhoods, registering with their local law enforcement agencies. Multiple regression allows the researcher to find specific or concrete information, rather than information that only provides general information.

Sandys and McGarrell in their article "Attitudes Toward Capital Punishment: Preference for the Penalty or Mere Acceptance?", used multiple regression to define the particular type of person who favors capital punishment as well as those who disapprove of capital punishment. These authors based their research strategy on the statement "Recent national polls find that approximately 70% of respondents indicate they are in favor of the death penalty for first-degree murderers when asked a general favorableness-type question such as: Do you generally favor or oppose the death penalty for persons convicted of first-degree murder?" (191). Although it was known that 70 percent of people were in favor of the death penalty, the research wished to find out if specific information such as race, gender, political party, etc. influenced one's opinion on capital punishment. In itself, the percentage provided information for the researchers; however, it was not as concrete as it could have been, thus the researchers, Sandys and McGarrell, wished to analyze the data further.

To begin their study, Sandys and McGarrell, began as most researchers in any field would, by doing a literature review. In this review, they found abundant amounts of prior research on the subject. In one study, "American Death Penalty Opinion, 1936-1986: A Critical Examination of the Gallop Polls," by the researcher Robert M. Bohm, as quoted by Sandys and McGarrell, "Ten variables [including race, gender, income/socioeconomic status, political party, and region] were examined for possible association with support for capital punishment" (193). Sandys and McGarrell continued their research by looking at a study, "Death Penalty Opinion in the Post-Furman Years," by James Alan Fox, Michael L. Radelet, and Julie L. Bonsteel. Sandys and McGarrell quote, "Fox et al. (1990-1991) examined a number of these same variables in a multivariate analysis using data collected in the General Social Survey for the period 1972-1988" (193). From looking at Sandys and McGarrell's literature review before looking at their actual research, it is obvious why multiple regression analysis is used more often than any other statistical technique in criminal justice research. Simply put, through using multiple regression analysis more specific observations about who favors capital punishment become apparent which provide more information than a simple percentage, thus leading the researcher to make more accurate generalizations as to how the rest of the public views capital punishment.

Sandys and McGarrell's in their research had hoped to "analyze the role of demographic and political variables as well as the goals of punishment variables in explaining such support" and to "...extend the prior research not only by examining the role of these variables in determining support for capital punishment but also by considering these issues in relation to support for alternatives to capital punishment and support for capital punishment for specific subgroups (juveniles, the mentally retarded)" (194). To conduct their study, Sandys and McGarrell chose the population of Indiana as their sample population, then devised a telephone survey to narrow down the sample population. Upon completion of the telephone survey, Sandys and McGarrell found that "90% of the respondents were White and a comparable percentage had completed at least a high school education. A majority of the respondents also were married. The modal income of the respondents was between $20,001 and $30,000....In terms of political orientation, self-described conservatives outnumbered self-described liberals almost two to one" (196). Once the telephone survey was completed, a questionnaire was distributed to these respondents. Through this questionnaire and the multiple regression analysis that followed, Sandys and McGarrell found several bivariate and multivariate relationships in the answers of the 76 percent of those surveyed who were in favor of capital punishment. Although much more time consuming than conducting a simple poll of Indiana residents and thus finding a percentage, multiple regression enables the researcher to find much more accurate and varying opinions on capital punishment than a percentage allows one to know.

Multiple Regression and How it Can Be Used to Evaluate the Effectiveness of Legislative Statutes

Multiple regression analysis can also be effective to analyze the impact that legislative statutes have on punishing criminals. In one study, "The Impact of Law Reform on the Processing of Sexual Assault Cases", executed by Gunn and Linden, logistic regression analysis was utilized to analyze the impact of Bill C-127, a bill theat reformed the former sexual assault statutes previously on the books in Winnipeg, Manitoba. Gunn and Linden describe that Bill C-127, introduced into the Parliament of Canada on January 4, 1983, changed the previous sexual assault laws by:

> --the reclassification of the crime of "rape" to three levels of sexual assault based on aggravating factors;
>
> --disqualification of evidence concerning the complainant's background that is not pertinent to the case (rape shield law);
>
> --removal of spousal immunity;
>
> --making sexual assault gender neutral; and
>
> --changing the rules of evidence concerning consent, corroboration, and recent complaint (156).

Gunn and Linden state the legislation was enacted as "Proponents hoped the new legislation would encourage more victims to report sexual assaults, ensure that victims were treated more fairly, and increase the conviction rate for sexual assaults, thereby enhancing the deterrent effect of the law" (156).

For their study, Gunn and Linden chose logistic regression as the most applicable statistical technique to use to evaluate the implications of Bill C-127. They state:

> Because the law was clearly intended to change the way in which sexual assault cases were treated by the courts, we used logistic regression to examine the impact of several independent variables on charges and convictions before and after the new legislation. Logistic regression allowed us to determine if the rape reform legislation led to changes in the way in which these variables affect change and conviction (158).

Using data taken from police records for the two years prior to Bill C-127 and the two years after its enactment, Gunn and Linden undertook a comparison of data to evaluate any changes that the enacted legislation may have influenced. From the police reports, Gunn and Linden selected a random sample of all sexual assault cases *reported* in Winnipeg. Gun and Linden then narrowed their sample. They state:

> From the random sample, cases which were not relevant to the evaluation were selected out. The 1981-1982 sample included the crimes of rape, attempted rape, indecent assault, sexual intercourse with a female between 14 and 16 [years of age], sexual intercourse with a female under 14 [years of age], gross indecency and buggery. To permit a comparison with the post-reform sexual assault data, we excluded incest cases except when combined with any of the above offenses. The 1984-1985 sample was drawn from all cases that involved a complaint classified as sexual assault, sexual assault with a weapon, threats or causing bodily harm and aggravated sexual assault. As in the pre-reform sample, incest was included only when combined with one of the three levels of sexual assault (159).

After implementing logistic regression analysis, Gunn and Linden reported their findings, "Analysis of the filtering of cases through the Winnipeg justice system showed that the number of cases reported to the police increased by 66% following the implementation of C-127. The conviction rate increased from 12.7% of the pre-reform cases to 29.1% of those post-reform cases that resulted in conviction" (160-61).

Not only did Gunn and Linden look at the effects Bill C-127 had on conviction rates, but they also utilized logistic regression analysis to determine if, in the two samples, "...extralegal variables involving the character of the victim made a difference in the founding, charging, prosecution and convicting of sexual assault cases" (162). Through this analysis, Gunn and Linden found that the variables OVKNOTH (if the complainant or victim had knowledge of the accused but the accused was not a parent), NOOFVICT (if there was more than one victim), BADCHAR (if there was evidence alleging that the complainant was of "bad character" or "promiscuous"), and INJURIES (if there were injuries suffered by the complainant) influenced the victim into reporting the assault (162).

Gunn and Linden's results showed overall that the reformation of the rape law has not caused a higher conviction rate as well as finding that character variables continue to have an impact on sexual assault charges. However, their study did show and effective way to utilize multiple regression analysis in criminal justice research, particularly how the passing of a particular piece of legislation that influences how crimes are reported.

Using Multiple Regression to Evaluate Deterrence Theory Models

Multiple regression has also been used effectively when evaluating models of deterrence theory. Deterrence theory, as explained by Anderson, et al. in the article "Models of Deterrence Theory," is "Deterrence theory in its classical form holds that crime is deterred by the threat of punishment...Threat of punishment involves (1) severity of punishment and (2) the probability of punishment" (236). In their study of deterrence theory models, Anderson et al. wished to explain why one commits a crime, such as the "...gain [from the crime], punishment [if caught in or after the commission of the crime], risk [of getting caught after the crime has been committed]... (236). Anderson et al. also wished to analyze "...the roles of the attending social and economic costs and controls (respect of community, family, friends, future job chances, etc)?" (236).

To execute their study, Anderson et al. faced one problem--What research design model to use. They state:

> Potential gain, risk, and punishment are likely to be sufficiently interrelated that it would be difficult to disentangle their separate effects. Using multiple regression models is a solution only if there are enough points in the right regions of the space and if one knows the correct form of the model.
>
> The ideal solution would be a large experiment that manipulates and fully crosses gain, risk, and punishment; randomly assigns people to these conditions; and measures the resulting criminal behavior. Of course, this cannot be done. The question becomes what compromises can be made that will still permit the modeling of the design calculus (237).

To rid themselves of these problems, Anderson et al. chose to use Rossi vignette analysis which uses "...(1) the weights (regression coefficients or simple transformations of regression coefficients) given particular pieces of information and (2) the model combining the weighted pieces of information" (239). In simpler terms, Anderson et al., describes Ross vignette analysis as, "Ross vignette methodology permits us to assess the importance of each set of information using standard forms of the linear [regression] model. The unique and independent contributions of the separate pieces of information can be obtained" (240).

Anderson et al., in their deterrence model analysis, looked at the variables of risk, gain, jail and fine, and type of crime through comparisons of four groups (race X sex) (242-43). In their analysis, Anderson et al. first used "...single-equation models that regress subjective probability on the 5 deterrence variables, Gain, Risk, Fine, and Jail and the dummy variables coding Crime Type" (243) To do this, Anderson et al. set up the equation:

$$Y= B_o + B_1G + B_2R + B_3J + B_4F + \quad BDi + E$$

"were G=Gain in dollars, R= Risk as 0-100 scale, J= Jail in years, F=Fine in Dollars, and the D_i are 0-1 dummies coding crime categories 2 through 7 on Table 1*, with crime 1, armed robbery, being the category excluded to avoid singularity, and E is a stochastic error term with the usual OLS [ordinary least squares regression] assumptions. Y is the subjective probability on a 0-100 scale" (244).

***see Appendix B for Table 1**

The results found by Anderson, et al. were Risk most influenced whether one committed a particular type of crime (247). Anderson, et al. then took their findings and converted them to logarithms in order to use OLS regression analysis (248-49). Through doing this, Anderson, et al. only found a slight difference from their original data, but "Because of the sample problems, we are reluctant to generalize these results" (254).

Anderson, et al. in addition to the previously mentioned data analysis, then effectively utilized multiple regression analysis to add additional variables to their study: education, age, and yearly personal income (254). Therefore, though Anderson et al. did not begin their data analysis specifically with multiple regression analysis due to the complications it would cause, they were able to finish their data analysis through using multiple regression alone, thus showing its effectiveness for use in criminal justice research.

Evaluating Multiple Regression in Predicting Criminal Recidivism

In the study, "Predicting Criminal Recidivism: A Comparison of Neural Network Models with Statistical Methods," written by Caulkins, et al., the researchers compared multiple regression analysis alongside other methods to determine how effective these methods were in determining the characteristics of which type of criminal is likely to return to prison upon release--criminal recidivism. As Caulkins et al. explain, "Predicting which offenders pose a high risk of criminal recidivism has long been an important concern in criminology" (227).

Currently many scales exist to tabulate which criminals have the highest recidivism rate. As Caulkins et al. explains, "...prediction scales [have] turned to statistical methods for combining predictor variables, including multiple regression methods, predictive attribute analysis and association analysis methods and multidimensional contingency table analysis" (227).

To understand the Caulkins, et al. study and how effective it shows multiple regression to be, one must understand what a neural mode is. Caulkins, et al. states, "Neural network models have some unique properties and advantages, but to a large extent neural networks can be viewed simply as non-linear multiple regression models that use a new class of nonlinear forms" (228). Caulkins et al. then explain why they undertook the comparison of neural models with other current statistical methods. They state:

> Neural network models, thus, provide substantial flexibility in which much of the model structure itself is estimated empirically from patterns recognized in the data. This capability may be useful in situations where there are: (a) inadequate theories for full model specification, (b) rich collections of independent variables with complex interactions, (c) subtle nonlinearities, or (d) distinct submodels of unique behaviors. Predicting criminal recidivism...seemed promising for neural network models because there is little consensus on the appropriate theoretical bases for predicting recidivism...(229).

Caulkins et al., when comparing neural models to other statistical techniques, utilized data from a study done by Gottfredson and Gottfredson in 1979. Gottfredson and Gottfredson had observed the recidivism rates of 3,508 criminals over a two year period after their release from federal prison. Offenders were released in 1970 and 1972. From these offenders, Gottfredson and Gottfredson studied twenty-nine predictor variables which fell into several broad categories including offenses committed by the criminal, his or her past criminal record, their social history, and if the criminal had made any lifetime adjustments because of his or her prison term (230). The results of the Caulkins et al. study were, "The neural network model performed nearly identically to the regression model on the four performance measures fro the three different variable subsets of the Gottfredson and Gottfredson data" (252), therefore as concluded by Caulkins, et al., "...network models do not offer an advantage over multiple regression analysis in predicting recidivism in the Gottfredson and Gottfredson data set" (234). Because of this it seems safe to conclude that in evaluating recidivism rates or evaluating any topic with multiple variables, assuming the appropriate data are available, multiple regression analysis is a safe statistical procedure to utilize in one's data evaluation.

Benda, et al. authors of "Recidivism Among Boot Camp Graduates: A Comparison of Drug Offenders to Other Offenders", also used types of regression analysis, particularly bivariate regression, ordinary least squares regression, and logistic regression to determine which offenders return to boot camp. Boot camp is "...programs [that] provide very controlled and disciplined military environments designed to instill and reinforce inner control over undesirable impulses [such as criminal activity]..." (242). In their study, Benda et al. compared the recidivism rate of drug offenders to that of other offenders. As reasons for this study Benda et al. states:

> Despite the high proportion of drug offenders in boot camps and the theoretical grounding for these programs, there is a paucity of research on this type of intermediate intervention. There are very few studies of criminal recidivism among participants of boot camp programs, especially among drug offenders. Yet, recidivism is the "bottom line" measure of program effectiveness, and certainly information is needed about how drug offenders are responding to boot camp in comparison to other offenders (243).

The main purpose of the Benda et al. study was: "...(a) to investigate several demographic predictors of criminal recidivism among the graduates of the only boot camp in Arkansas, and (b) to determine how predictive of recidivism being a drug offender is in comparison to other well-documented predictors" (243). Also:

> ...despite programmatic assumptions undergirding boot camps that they instill control over natural urges and provide social learning processes that positively change drug using behavior, these programs presently have shown, at best, only modest impact on recidivism. The paucity of boot camp recidivism studies indicates the need for further study of what factors predict continued unlawful behavior among participants in this type of intervention, especially among those involved in drug use (Benda et al. 244).

For their study Benda et al. studied 792 males who had graduated from the boot camp during a three year period. The characteristics or variables that may have influenced the recidivism rate studied by Benda et al. included the subjects age, their education, length of their prison sentence, their number of previous violations, race, marital status, and if they had recidivisized after a six-month follow-up period and a twelve-month follow-up period (244-45). The particular boot camp studied had, as major components: military training, education, substance abuse treatment, work, and mental health treatment (246).

For their particular study, Benda et al. used two separate types of multiple regression. As stated:

> To determine which of the data...were predictors of the two measures of recidivism for both follow-up periods, two statistical procedures with different underlying assumptions were used. Ordinary least squares (OLD) regression procedures were employed to examine predictors of return survival....Logistic regression was performed to identify which variables predicted membership in the dichotomous outcome of return versus no return (246).

Benda et al., in their research and subsequent data analysis resorted to using these two types of multiple regression analysis to get accurate research results due to the potential problems caused from using each type of multiple regression analysis alone. They state:

> For example, logistic regression, similar to OLS regression, controls for the influence of other predictors in examining the effect of a particular factor. Unlike OLD regression, however, logistic regression treats the dependent variable or outcome as a probability and is appropriate when the outcome is dichotomous. OLD regression is not appropriate for dichotomous outcome variables, especially when there are not comparable numbers of cases in each category, because of the violations of several assumptions. This is why the present study was designed with the two distinct outcome measures of return survival (the number of months in the community) and of return versus no return (dichotomy) (247).

Benda et al. first analysis of the boot camp data was a bivariate analysis, which was done to rule out any potential problems with multicollinearity. This bivariate analysis found that all age variables were highly related to one another, thus only age at first arrest was included in the two types of multiple regression analysis that were to follow. Benda et al., when analyzing their data with ordinary least squares regression, found that for those who recidivate, the type of offense, particularly if one was a drug offender, had a tremendous effect on the likelihood of a criminal committing another crime in his or her lifetime. Not only did the type of offense one had committed before boot camp effect his or her likelihood to recidivate, but also the number and type of offense one committed while in the boot camp program effected the likelihood of one to recidivate. Finally, another variable that played predominantly in whether one would recidivate was the person's race.

Logistic regression was to measure what variables influenced one to return or not return to prison. This found that the type of offense committed and race of the offender were the only predictors. Also, if one had committed a drug offense, he or she was more likely to return to boot camp. To sum up their study, Benda et al. state:

> The results of the present study are of particular interest for two reasons. First, this study finds that whether or not there are drug offenses is an important factor among what are typically the best predictors of criminal recidivism. Indeed, age, race, marital status, education, reading scores, age at first arrest, and past record in crime are among the very strongest predictors of criminal activity at various stages in the life span. The failure of these factors to predict recidivism in the present study is likely attributable to the homogeneous population of the offenders entering the boot camp in Arkansas (249).

Multiple Regression Analysis is Evaluated Through using A Statistical Program, Microcase

In order to determine how multiple regression can continue to be used effectively, three multiple regression analyses were done through the pre-programmed computer statistical software package, *Criminology: An Introduction Using Microcase, Third Edition* by Rodney Stark. These three examples were created from the FIFTYC component, a file described as a "...data file [that] is based on the 50 states and includes 110 variables" (3). Using ordinary least squares regression, these variables included number of larcenies, per capita income, and whether a person attends church or not, among others.

The three examples are represented in the following tables: Table 1--The Percentage of White, Homeless, and/or Unemployed Single men Who Commit Burglary; Table 2-- The Number of Larceny and Auto Thefts That Occur in a Metropolitan Area As They Effect the Population of 1990; and Table 3-- The Number of Larceny and Auto Thefts That Occur in a Rural Area As They Effect the Population of 1990.

Table 1

The Percentage of White, Homeless, and/or Unemployed Single Men Who Commit Burglary

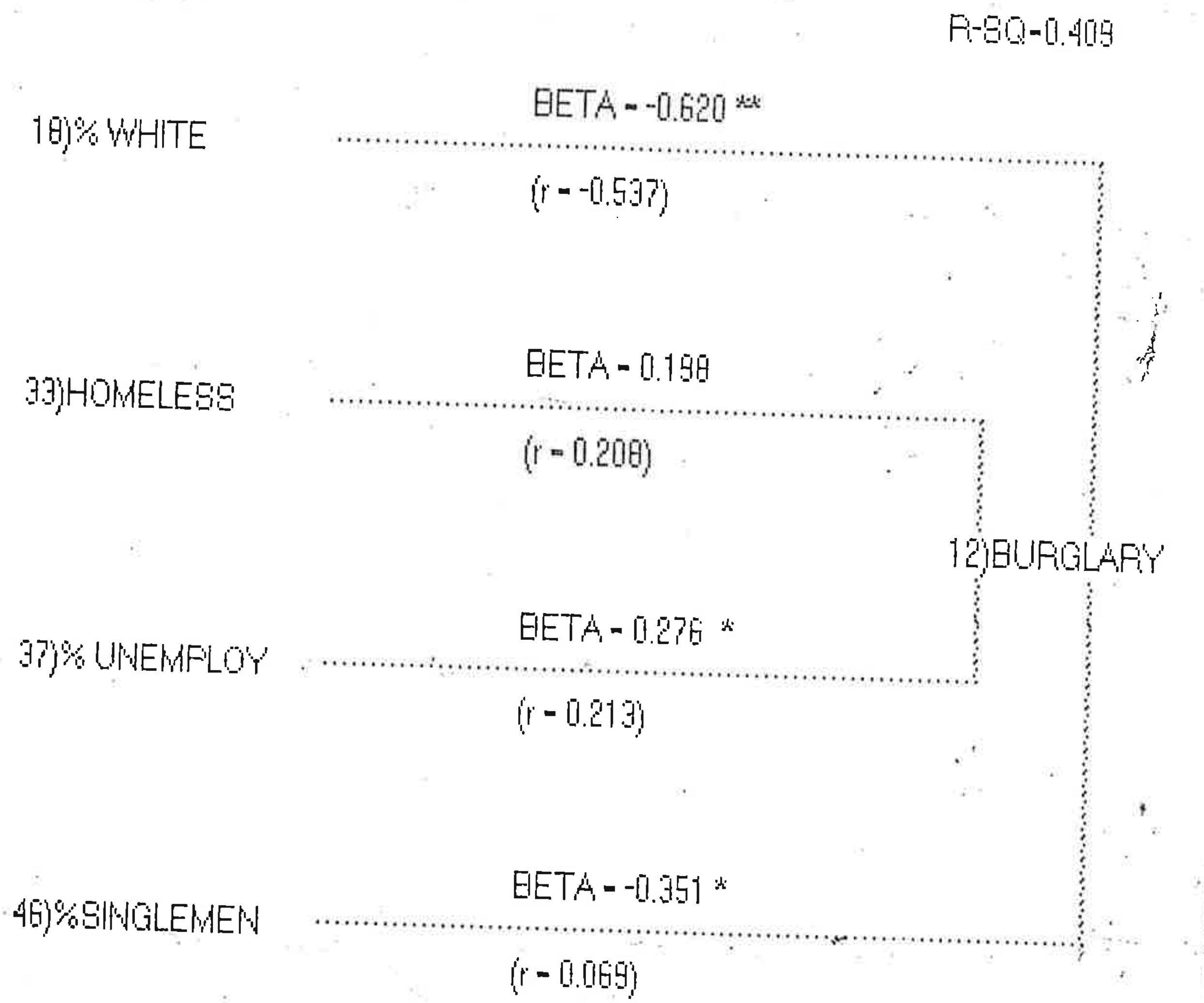


Table 2

The Number of Larceny and Auto Thefts That Occur in a Metropolitan Area As They Effect the Population of 1990

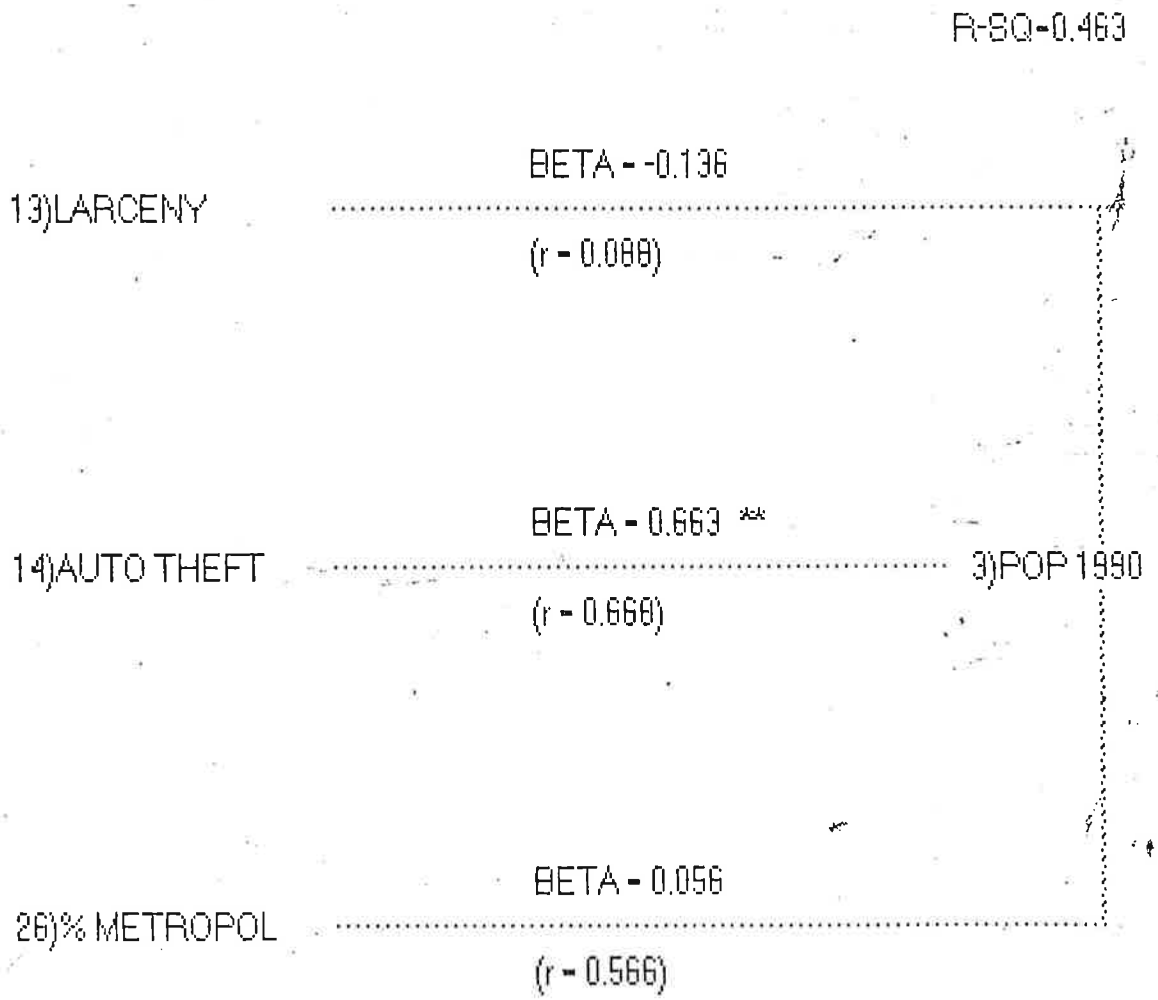
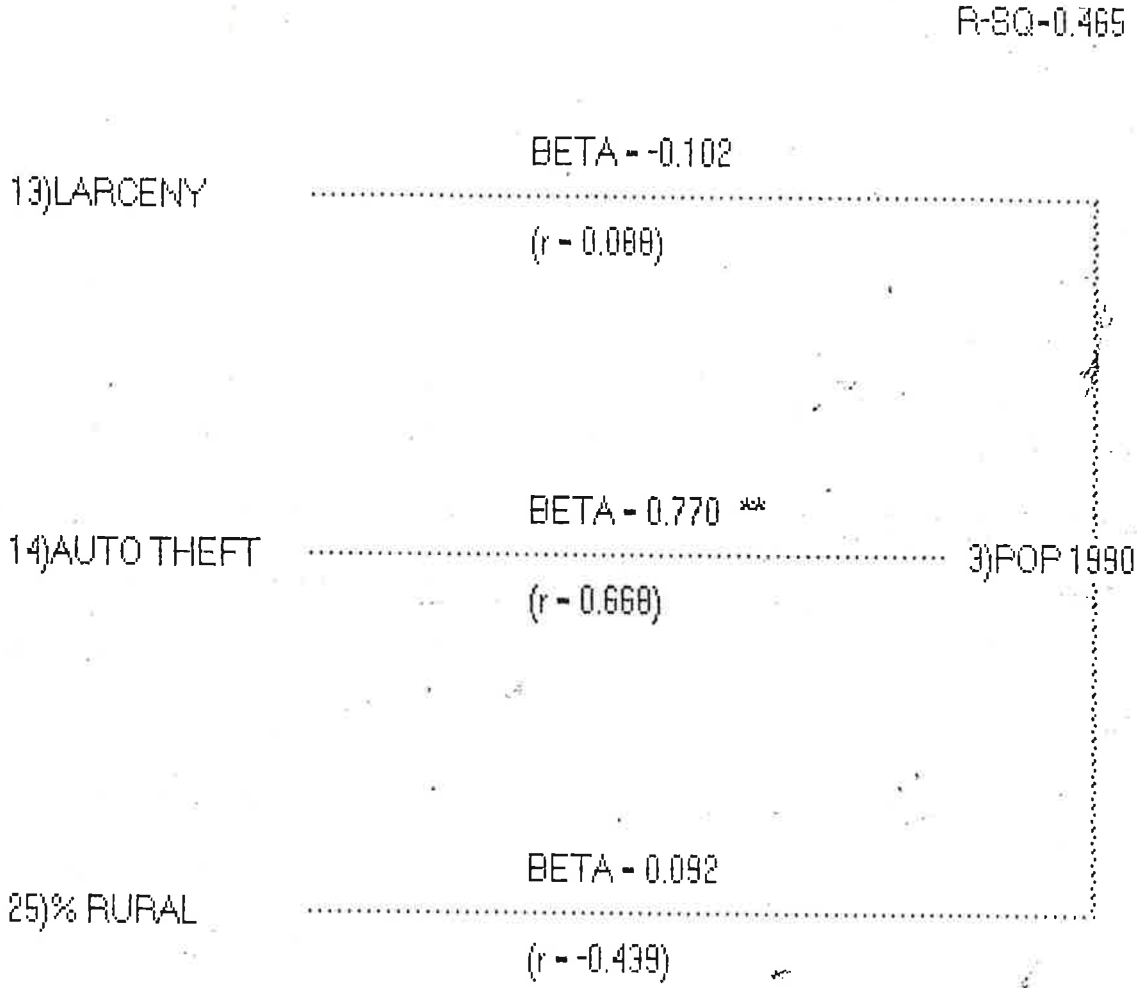


Table 3

The Number of Larceny and Auto Thefts That Occur in a Metropolitan Area As They Effect the Population of 1990



Before any explanation for the tables is given on an individual basis, several graphical notations must be explained. For instance, on each table three common notations exist. These include Beta=____, R^2, and R. Beta=____ occurs over the graphic lines for each variable described and represents the standardized partial slopes and, in the multiple regression equation are represented as the B_k term in each B_kX_k portion of the equation. Joseph F. Healey, in the text *Statistics: A Tool For Social Research Fourth Edition*, describes what a Beta weight actually describes. Healey states, "The beta-weights will show the amount of change in the standardized scores of the independent [variable] while controlling for the effects of other independents" (445). R and R^2 respectively represent the multiple correlation coefficient and the coefficient of multiple determination. Healey describes these two coefficients as, "The value of the latter statistic represents the proportion of the variance in Y that is explained by all the independent variable combined" (446).

For Table 1, the dependent variables for %White, Homeless, %Unemployed, and %Single men were chosen in order to evaluate which one of these variables, if any, had the most effect over someone committing burglary. From the analysis, it appears from the R values of each variable that the variables of % White, Homeless and %Unemployed effect whether one commits a burglary more so than the variable of %Single men. This comes from taking the absolute value of the R value. The closer the R value is to one, the more it effects the dependent variable. However, this does not tell us much. For this data to be effective, the researcher needs to go further. This means the researcher should evaluate all races, those with jobs and homes, as well as those who are married in order to determine if indeed being homeless and unemployed does greatly effect the number of burglaries committed in the fifty states. Luckily, for the researcher, all of these variables can be thoroughly evaluated using multiple regression analysis.

For Table 2, the number of larcenies and auto thefts that occur in metropolitan areas were evaluated as to how they effect the United States population of 1990. The variables of auto theft and larceny were evaluated as two different variables because they are two different crimes committed. *The Roxbury Criminal Justice Dictionary* by Dean Champion, describes auto theft as "Any stealing of a motorized vehicle" (11) and larceny as "Unlawful taking, carrying, leading or riding away of property from the possession of another; included shoplifting, pocket-picking, thefts from motor vehicles, and thefts of motor vehicle parts or accessories" (71). In this study, it was found that auto theft (R= 0.668) had the highest effect on the population of 1990. However, being in a metropolitan area (R= 0.566) also effected the amount of auto thefts and their effects on the population of 1990. When the two variables, auto thefts and metropolitan area, are compared to the larceny variable (R=0.088) it is found that larceny barely has any effects on the population at all.

In Table 3, larceny and auto thefts and their effects on the 1990 population were also examined, however these variables were examined in a rural setting. It was found that larceny and auto thefts effect the population the same as they did in a metropolitan area. However, it was found that if these crimes are committed in a rural area, (R= 0.439), they effect the population slightly less than those in a metropolitan area.

The data presented in Tables 2 and 3 tell us one apparent thing--auto theft is a larger crime. Or is it really? Even though it appears through multiple regression analysis that auto theft effects the population more than larcenies, the data is skewed because auto theft is one of the most, if not the most reported, uniform crime in the country. The Uniform Crime Report, "[An] Annual publication by Federal Bureau of Investigation that describes crime from all reporting law enforcement agencies in the United States" (Champion 126). Most auto thefts are reported to police so that a person's insurance company will reimburse them for the car's value. Most larcenies, unless they cause major damage to one's property, are not reported. This means that larcenies may effect the population of 1990 more than is admitted through our data. Although these problems with our data occurred, the researcher should not abandon multiple regression analysis. Multiple regression analysis did show us that auto thefts and larcenies effect the population to varying degrees and leaves the researcher room to proceed with more research in an attempt to prove his or her research hypothesis.

Conclusion

Multiple regression analysis, although seemingly complex, as most statistical procedures are, is actually one of the simplest procedures to execute. For instance, if one wishes to study the effects that ten variables have on one variable, only one equation is needed. Also, through the utilization of multiple regression, the researcher can decipher which one of the ten variables, if any, effect the one variable the most--if this is the research hypothesis the researcher is looking for. It is possible that he or she may be looking for which one of the ten variables as the least effect on the one, dependent variable. By using multiple regression analysis, the researcher is able to determine almost anything that he or she may be looking for concerning the variables and the research hypothesis design. Also, the researcher may be able to find new information that he or she may not have been looking for, and this information may or may not be relevant to his or her research hypothesis. However, these characteristics are what make it a useful research tool to use, in any discipline, but particularly those in the social science discipline.

To be effective, the researcher must always remember to follow the guidelines set by other researchers. These guidelines include the design structure for using multiple regression analysis and the assumptions of independence, normality, homoscedasticity, and linearity proscribed by Shavelson. One must also take into account the assumptions provided by William D. Berry and Stanley Felman.

Another requirement to using multiple regression analysis one must always consider is the potential problems with multiple regression analysis that may arise. These are multicolinearity, specification error, and measurement error. However, if one follows the proscribed design structure and necessary assumptions, these problems need not arise. If these problems do arise, multiple regression analysis has enables the researcher to find them at the onset of the research, thus enabling the researcher to take the appropriate steps necessary to rid the equation of problems. This may include throwing out the variables causing the problems completely or by simply restructuring the variable itself to be rid of the aforementioned problem or problems.

Although many, such as James P. McGregor, feel that multiple regression analysis is an overused technique, this is not so. In many cases such as Sandys and McGarrell's death penalty opinion study multiple regression is the most favorable technique to use. This was also apparent in other studies such as Gunn and Linden's Bill C-127 rape reform law study, Anderson et al.'s deterrence theory model study, and Caulkins, et al. And Benda et al's studies on criminal recidivism. In using the Microcase pre-programmed data, the studies executed enabled one to, as previously mentioned, to find which independent variables effected the dependent variables the most or least. The Microcase studies also enabled one to find new research avenues that need further exploration, which indeed is not a terrible thing--in fact it tends to be a very good happening to occur to the researcher.

In conclusion, it has been stated previously that multiple regression analysis, a very effective and powerful analytic tool is one of the most popular statistical techniques used in the behavioral or social sciences, particularly in the Criminal Justice field. This technique, as has been shown, is and can continue to be used effectively if the researcher follows the specific assumptions that this method dictates. It has also been shown that even if multiple regression does not specifically show the researcher the results he or she wants to prove or disprove, it can provide another avenue for the researcher to explore in an effort to find out exactly he or she wished to prove or disprove in the original research hypothesis.

Until recently, I would say that the population of 1990 was a good dependent variable. It is not a good independent variable. It is possible that the population of

1990 can be a dependent variable; however, it is a better independent variable. As an independent variable, the researcher would be able to determine how the difference in population effects the crime rate. As an dependent variable, it is very hard to tell how certain crimes or the crime rate effect the population. For this project, I was unaware of this finding until the projects completion. However, this does not mean that my research is completly wrong. The application of multiple regression analysis was perfectly accurate for this thesis. The findings; however, may not be accurate. As the computer disk which held my data, MICROCASE, has fell victim to a computer error, the findings can not be re-ran. I apologize for any inconviences this may cause any future researchers of multiple regression analysis.

Works Cited

Aiken, Leona S. and Stephen G. West. *Multiple Regression: Testing and Interpreting Interactions.* Newbury Park, CA: SAGE Publications, 1991.

Anderson, Andy B., et. al. "Models of Deterrence Theory." *Social Science Research* 12 (September 1983): 236-62.

Benda, Brent B., et al. "Recidivism Among Boot Camp Graduates: A Comparison of Drug Offenders to Other Offenders." *Journal of Criminal Justice* 24.3 (1996): 241-253.

Berry, William D. and Stanley Feldman. *Multiple Regression in Practice.* Beverly Hills, CA: SAGE Publications, 1985.

Caulkins, Jonathan, et al. "Predicting Criminal Recidivism: A Comparison of Neural Network Models with Statistical Methods." *Journal of Criminal Justice* 24.3 (1996): 227-40.

Champion, Dean J. *The Roxbury Dictionary of Criminal Justice, Key Terms and Major Court Cases.* Los Angeles, California: Roxbury Publishing Company, 1997.

Cohen, Jacob and Patricia Cohen. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences Second Edition.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1983.

Gunn, Rita and Rick linden. "The Impact of Law Reform on the Processing of Sexual Assault Cases." *The Canadian Review of Sociology and Anthropology* 34 (May 1997): 155-74.

Healey, Joseph F. *Statistics: A Tool For Social Research Fourth Edition.* Belmont, California: Wadsworth Publishing Company, 1996.

Jaccard, James, et.al. *Interaction Effects in Multiple Regression.* Newbury Park, CA: SAGE Publications, Inc., 1990.

Lewis-Beck, Michael S. *Applied Regression: An Introduction.* Beverly Hills, CA: SAGE Publications, 1980.

McGregor, James P. "Procrustus and the Regression Model: On the Misuse of the Regression Model." *PS: Political Science and Politics* 26 (December 1993): 801-04.

Sandys, Marla and Edmund F. McGarrell. "Attitudes Toward Capital Punishment:

Preference For The Penalty or Mere Acceptance?." *Journal of Research In Crime And Delinquency.* 32.2 (May 1995): 191-213.

Shavelson, Richard J. *Statistical Reasoning for the Behavioral Sciences, Second Edition.* Boston, MA: Allyn and Bacon, Inc., 1988.

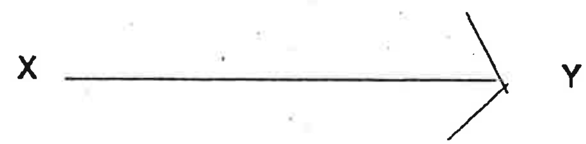
Stark, Rodney. *Criminology: An Introduction Using Microcase Third Edition.* Bellevue, Washington: Microcase Corporation, 1995.

Stevens, James. *Applied Multivariate Statistics for the Social Sciences.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers, 1986.

Weinberg, Sharon and Kenneth Goldberg. *Statistics for the Behavioral Sciences.* Cambridge: Cambridge University Press, 1990.

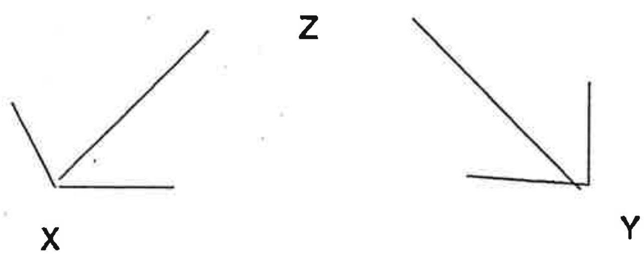
Appendix A
Six Types of Causal Relationships*

*Adapted from James Jaccard, et. al. p. 9. (see references)

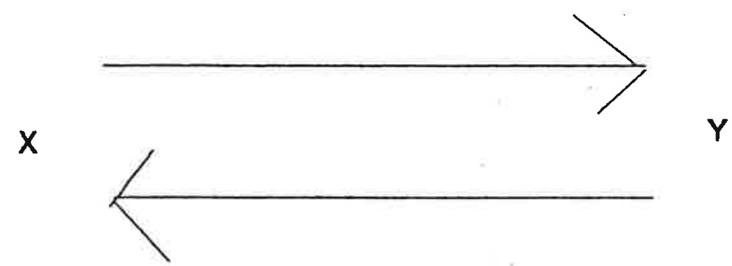


Direct Relationship

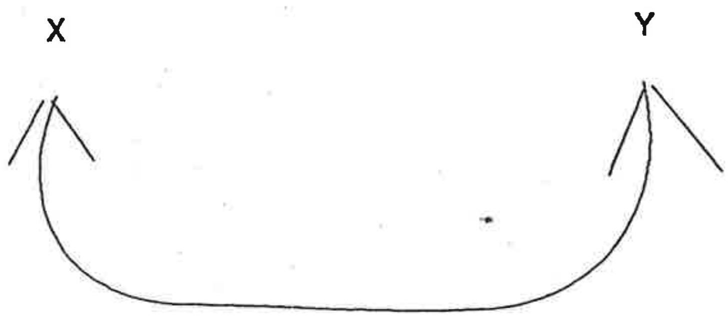
Indirect Relationship



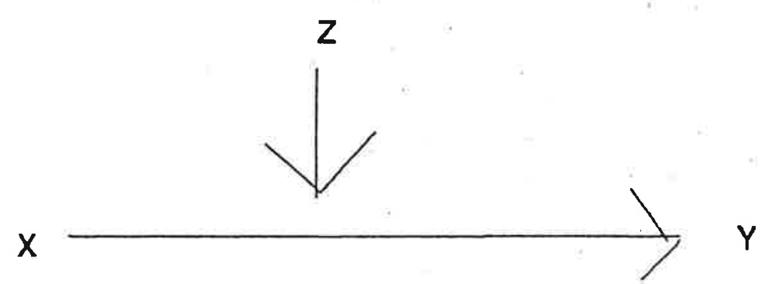
Spurious Relationship



Bidirectional or Reciprocal Relationship



Unanalyzed Relationship



Moderated Relationship

Appendix B

Table 1 from Anderson, et al.

TABLE 1
Vignette Variables and Values

	Type of crime[a]	Risk	Jail	Fine	Gain
1	Armed robbery	0	0	0	5
2	Purse snatching	5	probation[b]	100	25
3	Tax evasion	10	.5	200	50
4	Stealing from a store	20	1	400	100
5	Selling heroin	30	2	600	250
6	Selling marijuana	40	4	800	500
7	Embezzlement	50	6	1000	750
		60	8	2000	1000
		70	10	4000	2000
		80	15	6000	4000
		90	20	8000	6000
		95	25	10,000	8000
					10,000
					20,000
					30,000
					40,000
					50,000
					75,000
					100,000

[a] Three noneconomic crimes also were used but these results will not be considered here.
[b] For the data examined in this paper, probation was arbitrarily coded as .25 year.