4-8-82

STATISTICAL ANALYSIS CENTER

81110

Aggregation Problems in the Analysis of Illinois Statewide Criminal Justice Data

November, 1980

By James R. Coldren, Jr.
Statistical Analysis Center

CRIMINAL JUSTICE INFORMATION SYSTEMS
J. David Coldren, Director

ILLINOIS LAW ENFORCEMENT COMMISSION
Daniel W. Weil, Chairman
William Holland, Acting Executive Director

Printed by authority of the State of Illinois
November 1980
Number of Copies: 125
Printing Order Number: 81-10

ACKNOWLEDGEMENTS

All Statistical Analysis Center products are the result of the efforts of many persons. This publication is no exception. In addition to the efforts of the main author in preparing this document, many SAC and CJIS staff members have devoted their time and resources to the completion of this study. I thank them one and all, for their contributions, criticisms, and most of all, for their patience.

TABLE OF CONTENTS

TABLES AND CHARTS

APPENDICES

INTRODUCTION

Criminal justice researchers and administrators often use data obtained from the archives of agencies responsible for collecting and disseminating data to make decisions. They use these available data, rather than collecting data themselves, for several reasons. Often they need the data to answer a question that must be resolved quickly. In other instances, they do not have the resources necessary to collect their own data.

Use of available data obtained from archives poses problems for researchers and administrators. The data users must accept the data as they are, even if they are unsuitable for the research or analytic task at hand.

One set of problems posed by use of archival data is caused by aggregation. Aggregation refers to the grouping of individual units of analy sis. Archival data are usually available in some aggregated form (i.e., county totals for offenses reported to police, court circuit totals for felony convictions). Aggregation problems are problems that ensue when data from an archive are available in a form of aggregation that is unsuitable for the research or analytic task at hand.

The purpose of this paper is to introduce users of Illinois statewide criminal justice data to aggregation concepts and problems. It applies accumulated knowledge in the literature dealing with aggregation problems to specific examples of problems faced by users of Illinois criminal justice data.

Aggregation problems confront researchers at different stages of research and in various ways. It is important for users of Illinois criminal justice data to understand what kinds of problems aggregation may present and what can be done about them. Such an understanding will enable them to avoid the problems or deal with them if they are unavoidable.

Aggregation problems are not a recent discovery. Robinson (1950:354) documented an aggregation problem by demonstrating that the product moment correlation (r) computed for the relationship between nativity (whether or not an individual was born in the United States) and literacy using data about individuals differed when the same statistic was computed using state percentages (percent native-born and per cent literate in each state). In another case, Blalock (1964:102-103) similarly demonstrated that correlating percent non-white and income differentials between whites and non-whites in 150 southern counties produced one result, while aggregating the counties in a number of ways produced a wide range of correlations. Robinson and Blalock (and many others) have demonstrated that the use of aggregate data can affect, perhaps even negate, research results.

Aggregation problems are described and discussed in substantive and theoretical literatures of various fields of inquiry. Aggregation issues are defined and debated. Different approaches to aggregation problems are offered. In recent writings aggregation issues and problems are evaluated and discussed in the context of general research and methodological concerns. This literature focuses on aggregation problems encountered in correlation and linear regression analysis. That focus is carried over into this paper, though the extension of aggregation problems to other kinds of analyses is touched on as well.

The first part of this paper explains what it means to aggregate data in certain ways. The second part discusses the aggregation problems posed by three major sources of statewide criminal justice data in Illinois: the Illinois Uniform Crime Reports, the Annual Report of the Administrative Office of the Illinois Courts, and the Inmate Master records of the Illinois Department of Corrections. The third part reviews some of the major issues discussed in the aggregation literature. The fourth section presents a number of sample analyses using the data sources mentioned above. The purpose of the analyses is to explore the effects of aggregation on analyses using Illinois criminal justice data, and to discuss them in the light of helpful information gleaned from the aggregation literature. The final section of the paper reviews the effect that aggregation problems posed by the major sources of Illinois statewide criminal justice data have on research, and what steps should be taken to recognize and, if possible, to avoid them.

AGGREGATION EXPLAINED

Aggregation (grouping) of data occurs in many, and often complex, ways. This section explains two common forms of data aggregation using an imaginary data set. These forms are not exhaustive of the ways in which aggregation can occur, but they provide an introduction to the aggregation problems covered in this paper.

The imaginary data set consists of 1,500 cases (a case being a person). Four variables are measured for each case: income, education, birthplace, and religion. The data are being used for a hypothetical study of the relationships between these four variables.

In one form of aggregation the values of the variables in the analysis are grouped and the arithmetic mean for each group is used as a case. Suppose in this hypothetical study that the data are aggregated by birthplace. For each birthplace coded in the data, mean income and education are calculated and birthplaces (say cities) are used as cases in the analysis. If 50 birthplaces are coded, then the number of cases in the analysis drops from 1,500 to 50. This form of aggregation has been referred to as "ecological" or "aggregate" analysis. Its main characteristic is that group (aggregate)-level data are used to measure relationships between individual units of analysis.

Another form of aggregation involves altering the categorization scheme for one or more variables in the analysis, though not reducing the number of cases. Suppose the researcher wants to analyze the relationship between income and religion. Suppose also that the income variable is coded in thousand dollar increments from $5,000 to $25,000, and that five religions are coded in the database. One approach to studying the relationship is to recategorize the income variable so that it consists of fewer categories, (i.e., $5,000-$9,999, $10,000-$14,999, $15,000-$19,999, $20,000-$25,000), and analyze the relationship using crosstabulation and other non-parametric techniques.[1] In this case, grouping occurs along a variable so that the number of categories coded for that variable changes. The level of analysis remains the same, and all 1,500 cases in the sample are used.

In each of the examples above the the 1,500 cases were grouped in a certain way to study the relationship between certain variables. It is important to understand forms of aggregation because, depending on the circumstances surrounding any research effort, a data user may be forced to confront aggregation issues. Sometimes it is necessary to choose a form of aggregation to control for the effects of a variable. In other cases, a form of aggregation can be forced on a researcher due to problems of data availability or legal restrictions on the use of individual-level data. Thus, an understanding of how and why aggregations occur is important to all researchers, especially those who have little or no control over data collection procedures.

Forced and chosen aggregation are not always mutually exclusive occurrences. A researcher can choose to aggregate in one manner to avoid problems due to forced aggregation. A researcher can be forced out of conducting individual-level analysis, but have a choice between alternative aggregation schemes. The following section describes three sources of statewide criminal justice data. As their forms of availability are explained the distinction between the different aggregation forms and options discussed above will be clarified.

AGGREGATION PROBLEMS AND ILLINOIS CRIMINAL JUSTICE DATA SOURCES

There are three major sources of statewide criminal justice data in Illinois, each of which pertains to a major sector of the criminal justice system:

1) the Illinois Uniform Crime Reports;

2) the Annual Report of the Administrative Office of the Illinois Courts; and

3) the Inmate Master records of the Illinois Department of Corrections.

Each of these is a source from which data can be obtained in a short amount of time compared to the time it would take to collect original data, and each is available to criminal justice officials, planners, and reseachers in various forms. Their forms of availability and the aggregation problems they pose are disscussed below.

Illinois Uniform Crime Reports

The Illinois Uniform Crime Reports database is a source of data about local police activities concerning offenses and arrests in Illinois. Each month approximately 1,000 police agencies in Illinois report data concerning offenses reported and arrests made in their jurisdictions to the Illinois Department of Law Enforcement (DLE). DLE manages and disseminates these data, a responsibility it took over from the Federal Bureau of Investigation in 1972 (DLE 1972:2). In 1977, the Statistical Analysis Center undertook the project of recoding the Department's computerized files containing police data into SPSS[2] format, and became a source of Illinois Uniform Crime Reports data. This paper uses offense- and arrest-related data obtained from the Statistical Analysis Center's edition of the Illinois Uniform Crime Reports (hereafter referred to as SAC-IUCR).

Various kinds of information are avaialable from the SAC-IUCR database concerning criminal offenses reported to the police. Among them are:

- number of offenses reported to the police for each offense category recognized by the Illinois Department of Law Enforcement;[3]

- number of reported offenses which police determined to have actually occurred; and

- number of reported offenses cleared by the arrest of an adult.

The SAC-IUCR data files permit analysis at five different geographic/administrative levels:

1) the police agency (municipal) level, in which each case is a police department or sheriff's office. There are approximately 1,000 police agencies;

2) the county level, in which each case is an Illinois county; data pertaining to all police jurisdictions are aggregated to form a file containing 102 cases, one case for each county;

3) the planning region level, in which each case is an Illinois Law E ,cement Commission planning region; data pertaining to all police jurisdictions are aggregated to form a file containing 20 region cases;

4) the circuit level, in which each case is an Illinois judicial circuit; data pertaining to all police departments are aggregated to form a file containing 21 circuit cases; and

5) the common characteristic group (CCG) level, in which each case is an Illinois Law Enforcement Commission Common Characteristic Group;[4] data pertaining to all police departments are aggregated to form a file containing four CCG cases.

Region, circuit, and CCG boundaries are all formed along county lines in Illinois, so aggregation at these levels is easily done.[5] The availability of SAC-IUCR data in this format enables researchers to conduct analyses of SAC-IUCR variables at five different levels of analysis, each level consisting of a different grouping of police departments.[6]

The aggregation problems posed by SAC-IUCR data are caused by limitations in the database as well as by limitations in other statewide data sources. Illinois police data are the only statewide criminal justice data in Illinois available at the municipal level. Persons wishing to study relationships between police-oriented and other variables are forced to use higher level data (i.e., county, circuit, region), or to collect original data. For those who opt to use aggregate data, the choice becomes that of aggregating at the circuit, region, or county level,[7] and the problem becomes that of choosing the best form of aggregation.

Annual Report of the Administrative Office of the Illinois Courts

The Administrative Office of the Illinois Courts serves as a repository for statewide court data. Each of the 21 circuit courts in Illinois reports data monthly to the Administrative Office concerning many aspects of court activities. The Administrative Office is responsible for maintaining and disseminating these reported data. It does this primarily through its Annual Report, from which the data used in this report were obtained.

The Annual Report data used in this paper have to do with the processing of felony cases in the Circuit Courts. Annual Report felony case data include disposition information of the following types:

- number of defendants not convicted due to

  a) reduction of charges,
  b) dismissal of charges, or
  c) acquittal;

- number of defendants convicted by

  a) guilty plea,
  b) court, or
  c) jury, and

- number of imprisonment sentences imposed on felony defen ts.

The aggregation problems posed by Annual Report data are similar to those posed by SAC-IUCR data. The Administrative Office reports district-level data for Cook County only.[8] Otherwise all data are reported in county and circuit totals. County level data can be aggregated to the circuit and region levels, but county-, circuit-, and region-level data cannot be dis-aggregated to the municipal level. Analyses using police and courts data, then, must be conducted at the county, circuit, or region level.

Inmate Records of the Illinois Department of Corrections

The Illinois Department of Corrections collects, manages, and disseminates data concerning all persons committed to state correctional facilities. For each person committed, data are available concerning the person's personal characteristics (age, race, sex), background (family, education, employment, military service, etc.), and criminal history (prior committments to the Department of Corrections, offense(s), sentence, etc.). The Department of Corrections data used in this paper are selected items from Inmate Master records at Stateville Correctional Center that have been transferred from Department of Corrections manual files to the computerized Correctional Institution Management Information System (CIMIS) files for Stateville.

Analysts relying on Inmate Record data are confronted with different aggregation problems than those posed by SAC-IUCR and Annual Report data. Since Inmate Record data pertain to inmates at various institutions, the problem of aggregating along county or other geographic or administrative lines is not a major issue.

Users of Inmate Record data are often forced to reduce the number of categories coded for certain variables in order to conduct meaningful analyses. They face the problem of choosing the best scheme from a large number of possible grouping schemes. For example, there are appproximately 1,000 job skills coded in the CIMIS database. This is too many to allow for meaningful analysis. The solution to the problem is to reclassify the skill categories into a smaller number of categories (i.e., professional, administrative, skilled, non-skilled, craftsman). Collapsing the skill variable makes analysis more manageable and meaningful by producing a smaller number of categories to be compared with the categories of other variables with small numbers of categories such as sex, marital status, and religion.

The same problem is posed by the crime, age, and education variables in CIMIS. The large number of values coded for these variables precludes meaningful analysis. In addition, two different classification schemes for the offense variable are available in the CIMIS database. Researchers comparing the offense variable with other variables have the option of using one of the two schemes for analysis.

Summary

This brief look at the three data sources covered in this paper highlights two main types of aggregation problems that face users of those data:

1) Statewide criminal justice data are available in formats that may force data users to conduct analyses using data aggregated at geographic/administrative levels. Does such aggregation affect the outcome of analyses that should be conducted at different levels? Is one aggregate level of analysis more appropriate than another?

2) Data users may also be forced to collapse the measured values of a variable. Regardless of level of analysis, is one categorization scheme better than another?

The data sources described above and used in this paper represent sources of easily accessible statewide criminal justice data. Users of those data are forced to confront certain aggregation problems. In many cases they are forced to conduct analysis at a level other than the municipal one, or they are forced to group variables to conduct research. At the same time, however, the data users have a limited choice concerning how the data may be aggregated. If data users cannot conduct research exactly the way they want to, it would be useful for them to know the effects of alternative designs, and, if possible, which of the alternative (aggregate) designs is best.

Whether or not aggregation affects analysis depends on a number of factors specific to the research problem at hand. The aggregation literature contains discussions, arguments, and explanations of aggregation problems that have arisen over the years. Consideration of the information the literature contains about aggregation problems in general will contribute to an understanding of the problems posed in the more specific case of aggregation using Illinois criminal justice data.

UNDERSTANDING AGGREGATION PROBLEMS

The accumulation of knowledge about aggregation problems has reached the point at which aggregation problems are not considered to be special problems requiring special solutions, but problems that are understandable and solvable within the bounds of basic research issues. Thus, the approach in the literature is to relate aggregation problems to concepts and problems basic to research and data analysis.

There are three keys to understanding aggregation effects. They are:

1) standardized and unstandardized measures of the relationships between variables;

2) model specification, including the concepts of unit of analysis and statistical bias;

3) grouping processes.

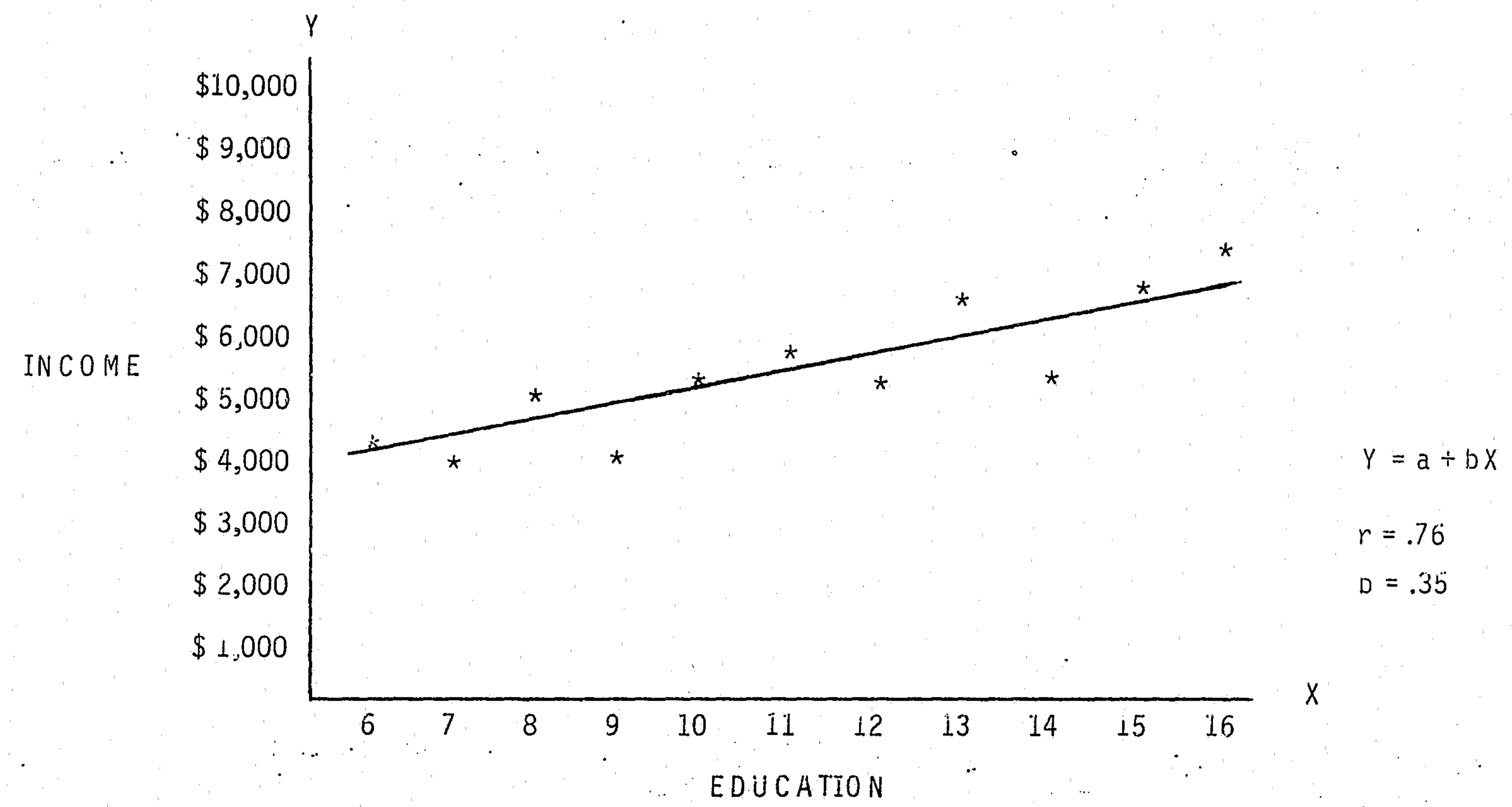
Standardized and Unstandardized Measures of the Relationships Between Variables

Researchers in the social sciences rely heavily on two statistics to measure the relationships between variables: the correlation coefficient (r) and the regression coefficient (b). Each of these statistics measures something different about the relationship between two (or more) variables. Rarely is one considered without consideration being given to the other. The researcher who understands the difference between these two measures is better able to deal with aggregation problems. The difference between correlation and regression coefficients is explained in reference to Figure A below.

The regression coefficient (b) is the slope of a line determined by the equation.

$$(1)\quad Y = a + bX.$$

FIGURE A


Y
$10,000
$ 9,000
$ 8,000
$ 7,000
$ 6,000
$ 5,000
$ 4,000
$ 3,000
$ 2,000
$ 1,000
INCOME
X
6 7 8 9 10 11 12 13 14 15 16
EDUCATION
Y = a + bX
r = .76
b = .35

It tells the researcher the magnitude of change produced in Y by a unit change in X. In the above example, a unit change (increase) in education will produce thirty-five percent of a unit change (increase) in income (b=.35). The regression coefficient also tells the researcher the direction of the relationship. If Y increases as X increases the relationship is positive. If one variable increases as the other decreases the relationship is negative. The correlation coefficient (r) is a measure of association which ranges from -1.0 to +1.0. It measures the joint variation between two or more variables, and tells the researcher how strongly they are related and what the direction of the relationship is. In the example above the correlation coefficient (r=.76) indicates that there is a fairly high degree of covariation between income and education.[9]

The correlation coefficient (r) is a standardized measure. Regardless of the unit of measurement used, it always ranges from -1.0 to +1.0. The regression coefficient (b) is not a standardized measure. It can have any value. The mathematical relationship between these two coefficients works out as follows:

$$(2) \quad r_{xy} = b_{yx} \left(\frac{sx}{sy}\right)$$

The correlation coefficient for the relationship between X and Y is obtained by multiplying the regression coefficient by the ratio of the standard deviations for each variable.[10] Likewise, in multiple regression analysis regression coefficients are standardized in the following manner:

$$(3) \quad b^{*}_{yx.z} = b^{*}_{yx.z} \left(\frac{sx}{sy}\right)$$

The standardized regression coefficient for the relationship between X and Y, with the variable Z held constant ($b^{*}_{yx.z}$), is obtained by multiplying the unstandardized regression coefficient by the ratio of standard deviations. $\left(\frac{sx}{sy}\right)$[11]

The relevance of this distinction between correlation and regression coefficients for researchers confronted with aggregation problems lies the fact that aggregation will cause a change in the ratio of standard deviations more often than in the regression coefficient. Standardized measures are more likely to be affected by aggregation than are unstandardized measures. Consider the example of the analysis of education and income above. If the relationship between income and education is studied at three different geographic levels --- county, region, and state --- it is more likely that the strength of the relationship will vary from level to level than the nature of it. Researchers facing aggregation problems should understand this characteristic of correlation and regression coefficients, and consider both in situations where aggregation occurs.[12]

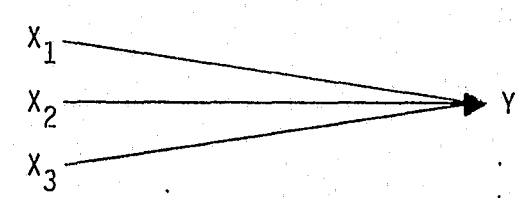
Model Specification

Model specification refers to the depiction of the interaction (or hypothesized interaction) of variables included in a research effort. Model specification can be simple or complex.

Simple level model specification entails uncomplicated statements about variable relationships such as those in Figures B and C below.

Figure B

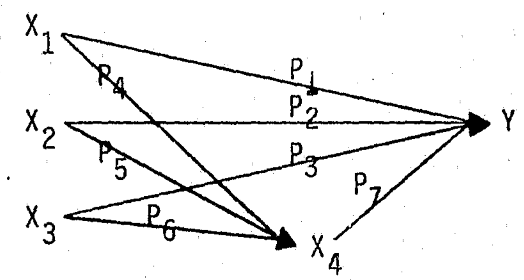
Figure C



These simple models state that variable X, or variables X_1, X_2, and X_3, have an effect on variable Y.

More complex model specification provides more information about variable relationships, as is exemplified in Figure D.

Figure D



This model depicts more complex variable relationships, and includes estimates of the strengths of the relationships (P_1 - P_7).

Properly specified models produce the best measures of relationships among variables. A model is considered properly specified when (1) the independent variables are, at most, moderately correlated with each other and uncorrelated with the error term; (2) each variable is measured with little or no error; and, (3) all relevant influences on the unit of analysis are included in the model. An improperly specified model will introduce statistical error into analysis (into measures of relationships). This type of error is called specification bias.

The unit of analysis is the subject the researcher is investigating, whose behavior he/she is describing or explaining. A unit of analysis can be an indivisible entity, or it may consist of a group of smaller units. For example, a study concerned with the behavior of police departments would have the police department as its unit of analysis, although police departments are usually composed of more than one police officer.

A change in level of analysis (i.e., from municipal to county) automatically changes the unit of analysis. If the behavior of the aggregate unit of analysis is influenced by more, or different, factors than that of the individual unit of analysis, the aggregate level model must be respecified. Failure to respecify the model at the aggregate level introduces specification bias at the aggregate level, which is called aggre-

gation bias. For example, if an analysis of police department behavior is conducted at the county level, then the new (aggregate) unit of analysis becomes the group of police departments in each county. If county-wide police department behavior differs from municipal police department behavior regarding the variables in the analysis, and if the model is not respecified, aggregation at the county level will produce biased measures.

Proper model specification at aggregate levels of analysis is posited as the key to understanding and resolving aggregation problems (Hannan 1971; Hanuscheck, Jackson, and Kain 1974; Langbein and Lichtman 1978; Erbring 1978). Model specification, however, is a problem in itself. It is difficult for the most informed researcher to correctly specify an individual-level model. Model respecification at an aggregate level is more difficult for the researcher who is unable to control the ways in which available data are aggregated. It is often difficult, if not impossible, to understand precisely how forced aggregation affects analysis. It is difficult to understand, for example, how county-wide police department behavior differs from municipal-level police department behavior, or if a difference exists at all.

Grouping Processes

The aggregation literature stresses consideration of grouping processes as a means of understanding the effect of aggregation on a model (Blalock 1964; Langbein and Lichtman 1978; Feige and Watts 1972; Shively 1969). Grouping processes refer to how groups are formed. If a researcher understands the process(es) (how) and the reason(s) (why) behind group formation, an understanding of the effect of aggregation on the model under analysis is more easily reached. Model respecification, then, is more easily done, and better measures are obtained.

It is sometimes difficult for reachers confronted with aggregate data to understand why the data are grouped the way they are. It is also difficult to understand how data are grouped, though it is possible at times to characterize aggregate data as being grouped according to one of four aggregation schemes suggested by the aggregation literature.

1) grouping by an independent variable - placing similar values of an independent variable into a certain number of groups lower than the total number of individual cases;

2) random grouping - placing cases into groups so that each case has the same, nonzero, chance of falling into any one group;

3) grouping by a dependent variable - placing similar values of the dependent variable into a certain number of groups; and

4) grouping by a variable related to the dependent and independent variables-placing similar values of a variable related to both the dependent and independent variables, and not included in the model, into a certain number of groups.

Knowledge of which of the above best characterizes the grouping processes confronted in a research effort can help a researcher understand, and perhaps avoid, bias that aggregation may introduce into analysis.

The manner in which these three factors --- measures, model specification, and grouping processes --- come into play in any research situation determines, in most cases, whether or not aggregation will introduce bias into measures of relationships between variables. Chart 1 on the following page summarizes the relationships between the three factors, which are discussed below.

Chart 1 indicates that random grouping processes do not produce aggregation bias when the individual level model is properly specified, regardless of which type of measure is used. Grouping by the dependent variable always produces bias in aggregate measures. Grouping by an independent variable almost always produces aggregation bias, except when an unstandardized measure is used. Grouping by a variable related to the dependent and independent variables almost always produces bias, though special cases in which bias is not introduced do exist. The remainder of this section explains how these factors operate to produce (or not produce) aggreagation bias. Figure E below illustrates Chart 1.

CHART 1

Factors Contributing to Bias in Aggregate Measures

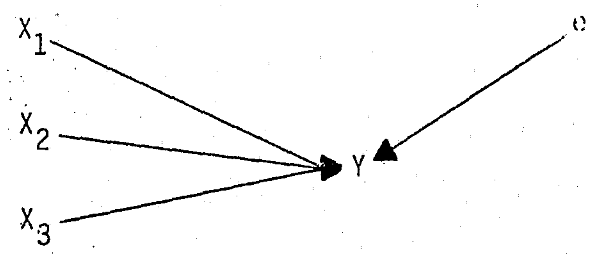
	Measure of Relationship Between Variables	GROUPING PROCESS: Independent Variable	Dependent Variable	Random Grouping	By a Variable Related to the Dependent and Independent Variables
Individual Level Model Properly Specified	Standardized Measure	Y	Y	N	MAYBE
	Unstandardized Measure	N	Y	N	Y
Individual Level Model Mis-Specified	Standardized Measure	Y	Y	Y	MAYBE
	Unstandardized Measure	Y	Y	Y	MAYBE

Y = bias will be introduced

N = bias will not be introduced

MAYBE = bias may or may not be introduced

Figure E



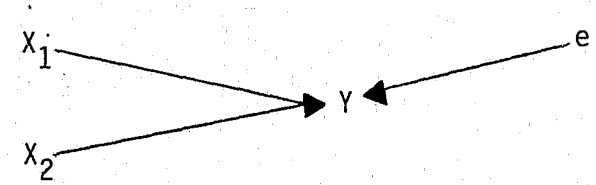
where: X_1 = resources
X_2 = crime type
X_3 = morale
Y = arrest rate
e = error

This model states that arrest rates (Y) for police departments are largely determined by the resources available to those departments (X_1), the types of crimes committed in their jurisdictions (X_2), and the morale of the officers in the departments (X_3). For the purposes of the following discussion it is assumed that this model is properly specified.

Misspecified Individual-level Model

If research is conducted to analyze the model in Figure E, and data are available (and collected) for resources (X_1), and crime type (X_2), though unavailable for police department morale (X_3), the research is conducted with a misspecified individual-level model such as that in Figure F below:

Figure F



If data are collected at the individual level (on police departments the measures in the equation.

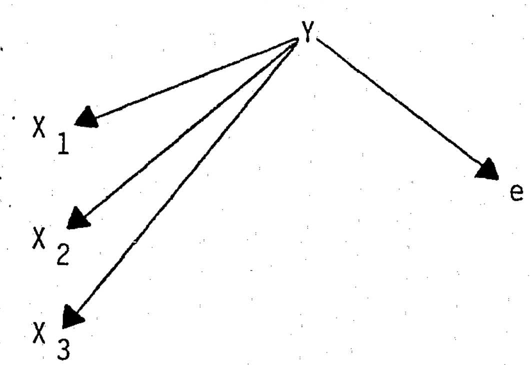
$$4)\quad Y = a+b_1X_1+b_2 X_2+e$$

will contain statistical error due to model misspecification. If aggregation occurs in any of the ways indicated in Chart 1, the specification error is reproduced at the aggregate level, whether or not aggregation bias is introduced. Thus, Chart 1 shows that bias will almost always be introduced when aggregation occurs with a misspecified individual level model.[13]

Grouping the Dependent Variable

Grouping along the dependent variable (Y) always introduces bias into aggregate measures whether or not the model is properly specified. The effect of grouping by Y, a variable related to all the X's and to e, is to alter the causal flow of the model so that Y determines, to a certain extent, the values of X_1, X_2, X_3, and e, as Figure G depicts:

Figure G



To the extent that the X's and e are related to Y, grouping by Y amounts to placing similar values of all those variables in the same groups. The result is a model in which the independent variables and the error term are all correlated; in other words, a misspecified model. The result of such an aggregation process is always to produce aggregation bias in standardized and unstandardized measures. If the individual-level model is misspecified, the bias produced at the aggregate level is compounded with specification bias:

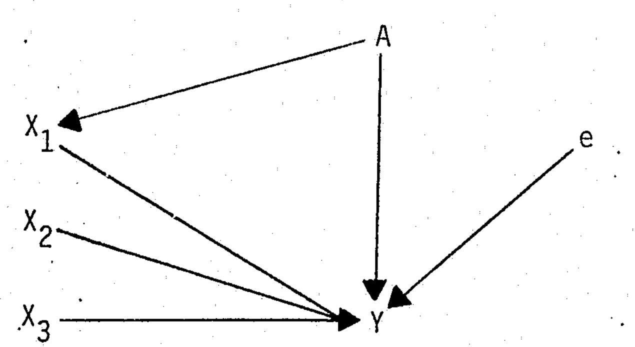
Random Grouping

If the individual-level model is properly specified and grouping occurs in a random manner, bias does not occur in standardized or unstandardized measures. Random grouping does not maximize variance along any one variable, nor does it place similar values of any variables into the same groups. Random grouping, then, does not alter the model at the aggregate level, does not affect standardized measures, and only produces bias due to misspecification at the individual level.

Grouping by a Variable Related to the Dependent and Independent Variables

In most cases of forced aggregation grouping does not occur along any one variable, or randomly. Aggregation usually occurs in complex ways, involving both the dependent and independent variables to varying degrees. When aggregation occurs along a variable related to the dependent and independent variables (say, along variable A, which is related to X_1 and Y) the model is altered in the following way:

Figure H



If the individual-level model is properly specified, whether or not bias is introduced into standardized measures depends on the extent to which both X_1 and Y are related to the grouping variable A. In rare instances they are related so that maximizing variance in X_1 and Y due to grouping by A does not change the ratio of standard deviations at the aggregate level. Usually one variable is affected more than the other and bias results in aggregate standardized measures. Grouping by A will place similar values of X_1 and Y in the same groups. The extent to which that

process amounts to grouping by an independent or dependent variable depends, again, on which variable (X_1 or Y) is more strongly related to A. Therefore, grouping by A will always introduce a certain amount of bias into unstandardized aggregate measures in the case of a properly specified individual-level model.

The effect of aggregating by A when the individual-level model is misspecified can be beneficial or harmful to aggregate measures. In some, again rare, cases the bias due to aggregation will be opposite in sign and great enough to counteract the bias in aggregate estimates due to misspecification at the individual level. For example, if X_1 is omitted at the individual level, and the effect is to bias bx_1y downward (produce a negative bias), aggregation may produce better measures than would be obtained using the misspecified individual-level model. If the bias introduced into bx_1y at the aggregate level is opposite in sign and less than twice the bias introduced into bx_1y at the individual level, then a better unstandardized measure of the relationship between X_1 and Y is obtained due to aggregation.

In most cases of forced aggregation bias is introduced into aggregate measures. In most cases of forced aggregation, also, aggregation occurs in complex ways so that an understanding of the bias introduced is not easily obtained. Consideration of the three issues discussed above --- measures, model specification, and grouping processes --- enables researchers confronted with aggregation problems to approach an understanding of them and, thus, be better able to interpret research results.

It is important to understand that only in an ideal situation can the effects of aggregation on analysis be fully understood. Just as grouping processes do not occur in truly random fashions, they do not occur in fixed processes either (i.e., along a dependent or independent variable only). The grouping processes explained in this section are meant to give data users an idea of how grouping can affect analysis, and understanding them is of little help without understanding model specification and basic statistics.

SAMPLE ANALYSES

This section presents four separate analyses using data obtained from the sources covered in this paper. The purpose of the analyses is two fold:

1) to demonstrate how aggregation of Illinois criminal justice data can affect research results; and

2) to demonstrate how the concepts stressed in the aggregation literature can be used to gain an understanding of the effects of aggregation on analyses.

Prior to presenting the analyses, a few qualifying statements need to be made.

The data used in the sample analyses consist of subsets of data obtained from the three sources covered in this paper. Only a few variables are used in the analyses, and randomly drawn samples from larger data sets are employed. The analyses, then, do not reflect the range of data available from the sources, nor do they reflect the limitations on the kinds of analyses that are possible.

No attempt is made in the examples to prove a definitive, substantive point about law enforcement or the administration of criminal justice in Illinois. The purpose of the examples is to provide an idea of what aggregation problems are and to make a statement about a methodological problem that can arise in any research field.

None of the sample analyses using police department and courts data include Chicago Police Department or Cook County Circuit Court data because the volume of offenses or cases in those jurisdictions is so much higher than those of others in Illinois that their inclusion significantly affects variable distributions. In addition, the analyses which rely on county-, circuit-, and region-level data only do not include DuPage County data because it is counted as a single circuit by the Administrative Office of the Illinois Courts and as a single region by the Illinois Law Enforcement Commisssion.

Finally, no attempt is made in this report to assess, explain, or correct inaccuracies in the data sources. It is assumed that the inaccuracies which exist are not severe enough to invalidate the points made in this paper.[14]

Three of the analyses focus on the effects of grouping by geographic area (county, judicial circuit, planning region) on correlation and simple regression analyses using IUCR and Illinois Courts data. The fourth analysis focuses on aggregation problems encountered by users of Illinois corrections data.

Example 1

This example compares the relationship between the number of criminal offenses occurring in a jurisdiction and the per capita personal income of persons living in those jurisdictions at three different levels of analysis. The offense data were obtained from the 1977 SAC-IUCR files and the income data were obtained from the 1977 files of the Regional Economics Information System. County level data are used as the individual level of analysis data in this example because income data are not available at the police department (municipal) level for all cities in Illinois. Thus, the county-level data consist of totals for offenses occurring in each county and the per capita income figure reported for each county-, and the circuit- and region-level data consist of mean values for each variable. The results of the analysis are presented in Table 1 below.

TABLE 1

Comparison of the Relationship Between Offenses Occurring and Per Capita Income at Different Levels of Analysis

Level of Analysis	Number of Cases	Correlation Coefficient	Slope (b)	$\frac{sx}{sy}$
County	100	.37*	3.31	.11
Circuit	19	.62*	6.92	.09
Region	18	.65*	7.23	.09

*=Significant at .05 level

When the relationship between the number of offenses occurring and per capita income is compared at different levels of analysis, the effect of aggregation is to increase the slope (the unstandardized regression coefficient) and, thus, the correlation coefficient. A researcher concerned about this relationship at the county level, but who uses circuit- or region-level data, may make incorrect inferences about the relationship. Aggregation along geographic lines in this case is similar to grouping along a variable related to the dependent and independent variables.

Example 2

This example compares the relationship between the number of criminal offenses occurring in a jurisdiction and the number of arrests made of adults in those jurisdictions at four different levels of analysis. The data on both variables were obtained from the 1977 SAC-IUCR files. The municipal-level data consist of totals for offenses occurring and arrests for each police department, and the county-, circuit-, and region-level data consist of mean values for each variable. Table 2 below presents the results of the analysis.

TABLE 2

Comparison of the Relationship Between Offenses Occurring and Number of Arrests of Adults at Different Levels of Analysis

Level of Analysis	Number of Cases	Correlation Coefficient	Slope (b)	$\frac{sx}{sy}$
Municipal	1053	.62*	.11	5.67
County	95	.68*	.19	3.57
Circuit	20	.57*	.12	4.67
Region	19	.65*	.14	4.67

*=Significant at .05 level

When the relationship between the number of offenses occurring and the number of arrests of adults is compared at different levels of analysis, there is no apparent aggregation effect. The county, circuit, and region slopes and correlation coefficients remain close to the measures obtained at

the municipal level, though the county-level estimates show slightly higher bias. Thus, the researcher who relies on county-, circuit-, or region-level data to estimate this relationship at the municipal level will not be confronted with aggregation bias. In this case aggregation occurs in a random manner.

Example 3

This example compares the relationship between the number of felony cases reduced to misdemeanors and two other court-oriented variables --- the number of guilty pleas entered, and the number of felony convictions --- at three different levels of analysis. The data were obtained from the 1977 Annual Report. The county-level data consist of totals for all three variables, and the circuit- and region-level data consist of mean values for each variable. Tables 3a and 3b present the results of the analyses.

TABLE 3a

Comparison of the Relationship Between Cases Reduced and Guilty Pleas Entered at Different Levels of Analysis

Level of Analysis	Number of Cases	Correlation Coefficient	Slope (b)	$\frac{sx}{sy}$
County	100	.56*	1.18	.48
Circuit	19	.53*	.98	.53
Region	18	.57*	1.34	.43

*=Significant at .05 level

TABLE 3b

Comparison of the Relationship Between Felony Cases Reduced and Felony Convictions at Different Levels of Analysis

Level of Analysis	Number of Cases	Correlation Coefficient	Slope (b)	$\frac{sx}{sy}$
County	100	.56*	1.31	.42
Circuit	19	.52*	1.10	.47
Region	18	.60*	1.59	.38

*=Significant at .05 level

In the cases of both relationships, the slopes and correlation coefficients at the circuit- and region-levels do not differ much from those obtained through county-level analysis. As in Example 2, then, a researcher who does not have access to county-level data, and who conducts analysis in a manner similar to that presented above, will not make erroneous inferences about the relationships between those court variables at the county level. Aggregation in this case also takes place in a random manner.

These three examples indicate that aggregation bias is not inevitable with Illinois criminal justice data. Two important points need to be made to put the examples in a proper perspective. First, each analysis begins with a misspecified model, so bias is present in all unstandardized measures. The second point is that most researchers will not aggregate to the circuit and region levels because the number of cases is so small at those levels, and county-level data are widely available. The main aggregation problem with police and courts-oriented analyses in Illinois lies in the lack of compatible data at the municipal level, and the subsequent forcing of county-level analysis. It is not reasonable to assume that aggregation at the county level involves random grouping processes because people do not form groups (communities, cities, police departments, etc.) randomly. It is, therefore, not reasonable to make automatic inferences from the county to the municipal level. Aggregation did not produce bias in the measures in Tables 3a and 3b, but that does not warrant the assumption that bias does not exist in the county level measures.

There is no direct way of knowing the precise effect of county-level aggregation, though steps can be taken to approach such an understanding. One way is to conduct dummy variable regression in which the effect of the grouping variable (county) is introduced as a categorical variable in a regression equation with the two variables under study.[15]

Dummy variable regression was conducted for each of the analyses presented in Tables 1-3b in the following manner:

- County-level regression analysis was conducted for each analysis using the two variables under study.

- The county variable was made into dummy variables using three categories of counties (metropolitan, non-metropolitan, and metropolitan-adjusted),[16] and regression analysis was conducted using the dummy variables.

Table 4 presents the results of these regression analyses.

TABLE 4

Comparison of Simple and Dummy Variable Regressions for Analyses in Tables 1-3b

	Relationship	R simple	R dummy	b simple	b dummy
Table 1:	Offenses and Per Capita Income	.37	.70	3.31	.62
Table 2:	Offenses and Arrests of Adults	.46	.50	.19	.23
Table 3a:	Cases Reduced and Guilty Pleas	.33	.74	1.19	.27
Table 3b:	Cases Reduced and Felony Convictions	.32	.75	1.32	.27

Table 4 indicates that, in the analyses presented in Tables 1, 3a, and 3b, knowledge of which type of county a case (county) represents improves prediction of the dependent variables substantially, and also changes the slopes. In the analysis from Table 2, dummy variable regression improves prediction very little and produces a small change in the slope. It can be assumed, then, that aggregation at the county level in Tables 1, 3a, and 3b produces bias that would not be present in individual-level analyses because the variables in the models are related to the grouping variable (county), and grouping by county systematically places values of the dependent and independent variables in the same groups.

Example 4

This example compares the relationships between four variables found in Inmate Record data using different methods of collapsing each variable. The relationships between three demographic variables --- inmate's age when committed, last school grade completed, and primary employable job skill ---and the offense variable (most serious offense on the inmate's record) are examined in a series of contingency table analyses.[17] Chart 2 below describes the different classification schemes used in this example.

The offense variable is collapsed to create two different variables --- FELONY and INDEX --- and each is categorized in two different schemes. A six- and a two-category variable are created for FELONY, and a seven- and a two-category variable are created for INDEX. The three demographic variables are categorized in nine different ways. The nine demographic grouping schemes are cross-tabulated with the four offense grouping schemes, resulting in a series of thirty-six contingency table analyses.

Aggregations of the nature employed in this example are different from those used in the three sample analyses presented above. Grouping occurs along certain chosen variables, and is controlled by the data user. Correlation and regression techniques are not at issue because most of the variables in Inmate Record data are measured by nominal or ordinal scales, and thus are not suitable for those types of analyses. Still, the aggregation problems posed by Inmate Record data require an understanding of modeling issues and the effect of aggregation on different measures.

CHART 2*

Grouping Schemes Used in Example 4

VARIABLE	GROUPING SCHEME	NUMBER OF CATEGORIES
FELONY	1) -Class X[a] -Murder -Class 1 Felony -Class 2 Felony -Class 3 Felony -Class 4 Felony	6
	2) -Class X, Murder -Class 1 through Class 4 Felony	2
INDEX	1) -Murder, Voluntary Manslaughter -Rape -Robbery, Armed Robbery -Attempted Murder, Aggravated Assault, Aggravated Battery -Burglary -Theft, Burglary from Auto -Motor Vehicle Theft	7
	2) -Murder, Voluntary Manslaughter, Rape, Robbery, Armed Robbery, Attempted Murder, Aggravated Assault, Aggravated Battery -Burglary, Theft, Burglary from Auto, Motor Vehicle Theft	2
GRADE	1) -First through Sixth -Seventh through Twelfth -First Year College through Fourth Year College	3
	2) -First through Twelfth -First Year College through Fourth Year College	2
	3) -First through Tenth -Eleventh through Fourth Year College	2
	4) -Seventh through Twelfth -First through Sixth, First Year College through Fourth Year College	2
AGE	1) -6 through 28 -29 through 56	2
	2) -6 through 19 -20 through 29 -30 through 39 -40 through 56	4
	3) -6 through 19 -20 through 34 -35 through 49 -50 through 56	4
SKILL[b]	1) -Professional/Technical, Manager/Administrator, Craftsman, Farm Worker -Machine Operator, Transport Operator -Clerical, Laborer, Service Worker, Private Household Worker	3
	2) -Professional/Technical, Manager/Administrator, Clerical -Craftsman, Machine Operator, Transport Operator -Laborer, Service Worker, Private Household Worker, Farm Worker	3

*notes to CHART2 appear on the following page

NOTES TO CHART 2

(a) A Class X felony is a special category of crime defined by Public Act 80-1099 (effective February 1, 1978). Class X felonies include aggravated kidnapping for ransom, rape, deviate sexual assault, heinous battery, armed robbery, aggravated arson, and treason.

(b) The approximately 1,000 coded skill categories recognized by the Department of Corrections were recoded to compare roughly to the occupation classification scheme used in the General Social Survey conducted by the National Opinion Research Center (See the Cumulative Codebook for the 1972-1977 General Social Survey, especially pp. 223-235.). The skill variable was first coded to form 10 occupational categories, and was then further collapsed to form the two three-category variables.

When a variable is collapsed its effect is attenuated to a certain degree because the full range of measured values is not allowed to operate in analysis. If its effect is attenuated too much, then grouping removes or diminishes its effect from the model. For example, more can be found out about the relationship between age and income through contingency table analysis if ten three-year age categories are used than if two fifteen year age categories are used.

There are a number of different measures that can be used in contingency table analysis. These measures are referred to as non-parametric statistics because the analyses relying on them do not require the strict assumptions that must be met in correlation and regression analysis.[18] In this example, three statistics are calculated for each contingency table analysis --- Chi Square, Cramer's V or Phi, and Gamma or Yule's Q --- to see if aggregation in the manners described above affects non-parametric statistics.[19] The results are presented in Table 5 below.

Table 5 indicates that grouping the demographic and offense variables in different ways does not affect the Chi Square statistic. Only one significant Chi Square statistic resulted, that for the relationship between the first Age and the first Index Crimes grouping (Age 1 and Index 1). A significant Chi Square statistic tells the researcher that the probability is high that a systematic relationship exists between the variables under study. Regardless of how the values for any of the four variables are grouped, the relationships between them are consistantly unsystematic.

Cramer's V and Gamma are measures of association (similar to the correlation coefficient) used in contingency table analysis. Cramer's V (or Phi) ranges from 0 to +1.0, and Gamma (or Yule's Q) ranges from -1.0 to +1.0. Each is interpreted like a correlation coefficient. A high value (positive or negative) indicates a strong relationship between the variables under study, and a low value indicates a weak relationship.[20] Gamma indicates the direction of a relationship while Cramer's V does not.

TABLE 5

Comparison of the Relationships Between Inmate's Most Serious Offense and Inmate's Last Grade Completed, Age, and Employable Job Skill Using Different Grouping Schemes

	Felony 1)[a]			Felony 2)			Index 1)[b]			Index 2)[c]		
	Chi Square	Cramer's V	Gamma	Chi Square	Cramer's V	Gamma	Chi Square	Cramer's V	Gamma	Chi Square	Cramer's V	Gamma
Grade 1)	NS	.16	.10	NS	.03	-.02	NS	20	-.27	NS	09	-.43
Grade 2)	NS	.17	.13	NS	.02	-.08	NS	.13	-.30	NS	.06	-.42
Grade 3)	NS	.15	-.01	NS	.07	-.15	NS	.17	-.18	NS	.08	-.24
Grade 4)	NS	.15	.11	NS	.03	-.10	NS	.15	-.21	NS	.03	-.16
Age 1)	NS	.23	.10	NS	.05	-.11	S	.30	-.26	NS	.07	-.22
Age 2)	NS	.20	.05	NS	.15	-.09	NS	.20	-.22	NS	.16	-.23
Age 3)	NS	.18	.12	NS	.06	-.01	NS	.19	-.32	NS	.18	-.44
Skill 1)	NS	.21	-.09	NS	.06	-.12	NS	.17	-.05	NS	.04	-.11
Skill 2)	NS	.17	-.07	NS	.07	-.10	NS	.14	-.02	NS	.06	-.05

NS = non significant at 05 level
S = significant at 05 level

a The Felony crime classification scheme is based on the Illinois statutory classification system, which classifies each criminal offense as a felony, misdemeanor, petty or business offense depending on the type of crime and possible severity of sentence. It is used mostly by courts and corrections agencies in Illinois (Block 1979:1-2).

b The Index crime classification scheme is based on the one used by the Federal Bureau of Investigation (Illinois Department of Law Enforcement 1979:1).

c This dichotomization of the Index crimes corresponds to the Violent versus Property crimes classification used by the Federal Bureau of Investigation (Illinois Department of Law Enforcement 1979:2).

Table 5 indicates that grouping does affect the Cramer's V (Phi) and Gamma (Yule's Q) statistics. Higher Cramer's V statistics were produced by the Felony 1 and Index 1 classification schemes, regardless of the ways in which the demographic variables were grouped, than by the dichotomized Felony 2 and Index 2 schemes. Higher Gamma statistics were produced by both Index classification schemes for the analyses involving Grade and Age than by the Felony schemes.

In the case of Cramer's V and Phi the dichotomous grouping schemes (Felony 2 and Index 2) are more appropriate for analysis. When the six- and seven-category offense grouping schemes (Felony 1 and Index 1) are used, too many cells with an insufficient number of observations result and the statistics are unreliable.[21] Collapsing the offense variable into dichotomies produced fewer cells with fewer than five observations and, thus, more reliable statistics.

In the case of Gamma and Yule's Q the Index grouping schemes produced the more reliable statistics.[22] The Gamma statistic requires that the variables under study be measured by at least an ordinal scale, a scale of qualitative differences containing more than two categories. In this case, the Index offense grouping schemes are better ordinal scales than the felony schemes. The index offenses scheme is based on a rough ordinal scale of offenses ranging from Motor Vehicle Theft to Murder and Voluntary Manslaughter. The Felony offense scheme is based on Illinois statutes, and is based on the severity of the sentence which can be imposed on convicted offenders, not on the severity of the offense. Table 5 indicates that a scale based on the severity of offense is better than one based on possible sentence severity for the purposes of contingency table analysis.

CONCLUSION AND RECOMMENDATIONS

Users of Illinois statewide criminal justice data are inevitably confronted with aggregation problems. Either they are forced to use data aggregated in ways over which they have no control, as is the case with IUCR and Illinois courts data, or they are forced to aggregate data along certain variables, as in the case with Inmate Record data. The former case presents problems of cross-level inference: whether inferences can be made from the county to the municipal (or any other) level of analysis. The latter case presents problems of grouping data in ways that are least harmful to analysis.

The most important contribution of the aggregation literature lies in its demonstration that aggregation problems can be equated with concepts and problems common to research and inference. The key to understanding aggregation problems is to determine the effect that grouping processes have on models and measures of relationships between variables. Thus, the researcher who has a grasp of such basic research issues as model specification and inferential statistics is in a good position to deal with aggregation problems. The grouping process issue is the link between general issues concerning research and inference and aggregation problems.

Most often, as in the case with IUCR and statewide court data, grouping processes and, thus, the effects of aggregation on analyses, are not easily recognized. The task for the researcher then, is to obtain as accurate an understanding as possible about grouping processes and make adjustments in analysis to correct for aggregation effects.

Researchers using Illinois statewide criminal justice data are unable to fully comprehend grouping processes for two main reasons:

1) Most statewide criminal justice research in Illinois has to be conducted at the county level, and aggregation at the county level involves complex grouping processes.

2) Statewide data sources (criminal justice related or not) do not contain enough information (variables) to correctly specify research models at individual or aggregate levels.

In spite of these limitations, it is possible to approximate grouping processes and aggregation effects, and to avoid improper inferences.

A number of considerations to be made and steps taken to confront aggregation problems are suggested below.

- Do not assume random grouping processes when confronted with data aggregated along geographic/administrative lines. Most often it will be an invalid assumption.

- Whenever possible, include the effect of the grouping variable in the aggregate-level model. Even if proper model specification is impossible, it can be determined whether or not the variables in the model are related to the grouping variable.

- If some of the variables in the model are available at the individual level, the effect of aggregation on those variables can be explored, which will shed light on the effect of aggregation on the model.

- An educated guess about how variables in a model should behave should (and usually can) be made prior to analysis. This can be done through consideration of experience with the research at hand and/or of how the variables in the model behave in similar research efforts.

- Aggregation problems posed by collapsing values of a variable(s) to conduct nonparametric analysis require the researcher to choose among a number of regrouping schemes. In these instances, care should be taken to choose a grouping scheme that will produce the most reliable statistics (i.e., leave few cells with less than five observations, and creating meaningful categories or scales).

The purpose of this paper is not to solve aggregation problems for those confronted with them. Its purpose is to alert data users to the aggregation problems posed by the data and to provide them with means of understanding and avoiding them. There are no "solutions" to the aggregation problems posed by Illinois statewide criminal justice data. There are, however, ways to understand and deal with them, as this paper has demonstrated.

APPENDIX A

Offense Categories Recognized by the Illinois Department of Law Enforcement

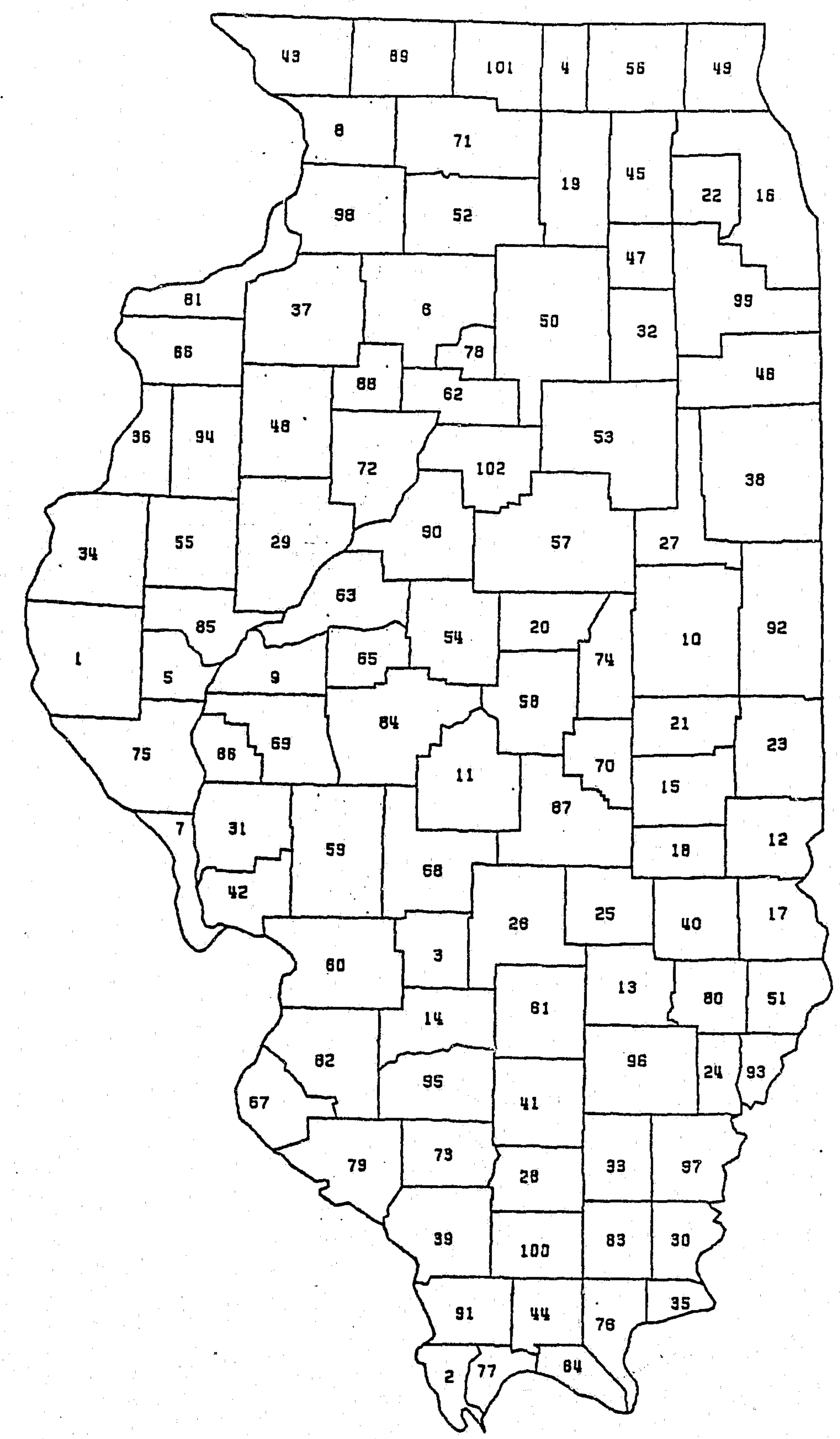
ILLINOIS UNIFORM CRIME REPORTING OFFENSE CODES

CODE 1	OFFENSE CLASSIFICATION 2	STATUTE
	HOMICIDE	
0110	Murder	38-9-1
0121	Attempt Firearm	38-8-4
0122	Attempt Knife or Cutting Instrum	38-8-4
0123	Attempt Other Dangerous Weapons	38-8-4
0124	Attempt Hands, Fists, Feet, Etc.	38-8-4
0130	Voluntary Manslaughter	38-9-2
0141	Invol Mansltr & Reckls Homi-Non-Veh	38-9-3a
0142	Invol Mansltr & Reckls Homi Veh.	38-9-3b
	FORCIBLE RAPE	
0211	Forcible Rape: Firearm	38-11-1
0212	Forcible Rape: Knife or Cutting Instrum	38-11-1
0213	Forcible Rape: Other Dang. Weapons	38-11-1
0214	Forcible Rape: Other Forcible Means	38-11-1
0220	Attempts Forcible Rape	38-8-4
	ROBBERY	
0311	Armed Firearm	38-18-2
0312	Armed Knife or Cutting Instrum	38-18-2
0313	Armed Other Dangerous Weapons	38-18-2
0320	Strong Arm No Weapon	38-18-1
0330	Attempts Armed-Firearm	38-8-4
0334	Attempts. Armed-Knife or Cut. Instrum.	38-8-4
0337	Attempts Armed-Other Dang. Weapons	38-8-4
0340	Attempts Strong Arm-No Weapon	38-8-4
	BATTERY	
0410	Aggravated Firearm	38-12-4(1)
0420	Aggravt'd Knife or Cut. Inst	38-12-4(1)
0430	Aggravt'd Other Dang Weap	38-12-4(1)
0440	Aggravt'd Hds., Fists Ft., Etc. (Inflicts Great Bodily Harm)	38-12-4(a)(c)
0445	Aggravt'd Hds Fists, Ft. Etc	38-12-4(b)(2 thru 9)
0460	Simple Battery	38-12-3
0470	Reckless Conduct	38-12-5
	ASSAULT	
0510	Aggravated Firearm	38-12-2(1)
0520	Aggravt'd Knife or Cut. Inst	38-12-2(1)
0530	Aggravt'd Other Dang Weap.	38-12-2(1)
0540	Aggravt'd Hds. Fists, Ft, Etc (Intends to Inflict Great Bodily Harm)	38-12-2(2)(10)
0545	Aggravt'd Hds, Fists Ft, Etc	38-12-2(2)(10)
0560	Simple Assault	38-12-1
	BURGLARY	
0610	Forcible Entry	38-19-1
0620	Unlawful Entry (No Force)	38-19-1
0630	Attempts Forcible Entry	38-8-4
	BURGLARY FROM MOBILE VEHICLE	
0710	Over $150	38-19-1
0720	$150 and Under	38-19-1
0750	Attempts Burglary From Mobile Veh.	38-8-4
	THEFT	
0810	Over $150	38-16-1
0820	$150 and Under	38-16-1
0850	Attempts Theft	38-8-4
	MOTOR VEHICLE THEFTS	
0910	Autos	38-16-1
0915	Trucks & Buses	38-16-1
0918	Other Vehicles	38-16-1
0920	Attempts Autos	38-8-4
0925	Attempts Trucks & Buses	38-8-4
0930	Attempts Other Vehicles	38-8-4
	ARSON	
1010	Arson-Explosive Device	38-20-1
1020	Arson-Incendiary Device	38-20-1
1030	Possession: Explosives, Incend. Dev.	38-20-2
1090	Attempts: Arson	38-8-4
	DECEPTION	
1110	Deceptive Practices	38-17-1
1120	Forgery	38-17-3
1130	Fraud	38-17-1
1140	Embezzlement	38-17-1
1150	Credit Cards	121-1/2-608 Thru 624
1160	Deceptive Altering of Coins	38-17-4
1170	Impersonating An Officer	38-17-2
1190	Attempts: Deception	38-8-4
1200	Stolen Prty., Buy., Rec., Poss	38-16-1(d)
1205	Theft By Lessee	38-16-1.1
1210	Thft. of Labor, Serv., Use of Prty.	38-16-3
1220	Theft of Lost or Mislaid Prty	38-16-2
1230	Poss. of Keys or Dev. to Coin Op. Mach	38-16-6
	CRIMINAL DAMAGE AND TRESPASS TO PROPERTY	
1310	Criminal Damage to Property	38-21-1
1320	Criminal Damage to Vehicle	38-21-1
1330	Criminal Trespass to Land	38-21-3
1340	Criminal Damage to State Sup. Prty.	38-21-4
1350	Criminal Trespass to State Sup Land	38-21-5
1360	Criminal Trespass to Vehicle	38-21-2
1370	Criminal Damage of Fire Fighting Apparatus, Hydrants or Equipment	38-21-1.1
1380	Unauth. Poss. or Storage of Weapons	38-21-6
	DEADLY WEAPONS	
1410	Unlawful Uses of Weapons	38-24-1
1420	Unlawful Sale of Firearms	38-24-3
1430	Unlawful Poss. of Firearms & Ammu.	38-24-3.1
1440	Register of Sales by Dealer	38-24-4
1450	Defacing Identi. Marks of Firearms	38-24-5
1460	Firearms & Ammu.-No. I.D. Card	38-83-2&3
1490	Attempts: Deadly Weapons	38-8-4
	SEX OFFENSES (EXCEPT FORCIBLE RAPE)	
1505	Prostitution	38-11-14
1510	Soliciting for a Prostitute	38-11-15
1515	Pandering	38-11-16
1520	Keeping a Place of Prostitution	38-11-17
1525	Patronizing a Prostitute	38-11-18
1530	Pimping	38-11-19
1535	Obscenity	38-11-20
1540	Harmful Material	38-11-21
1542	Tie-in Sales of Obscene Pub to Distrib.	38-11-22
1550	Deviate Sexual Assault	38-11-3
1555	Indecent Liberties with a Child	38-11-4
1560	Contrib. to Sexual Delinq. of a Child	38-11-5
1565	Indecent Solicitation of a Child	38-11-6
1570	Public Indecency	38-11-9
1575	Aggravated Incest	38-11-10
1580	Incest	38-11-11
1583	Rape of Mentally Deranged	38-11-1
1585	All Others	
1590	Attempts Sex Offenses	38-8-4
	GAMBLING	
1610	Bookmaking	38-28-1.1(b)
1620	Numbers-Lottery	38-28-1(a)-7
1630	Keeping a Gambling Place	38-28-3
1640	Reg. of Fed. Gambling Stamps	38-28-4
1650	Card Games: Operating	38-28-1(a)-3
1651	Card Game: Playing	38-28-1(a)-1
1660	Dice Game: Operating	38-28-1(a)-3
1661	Dice Game: Playing	38-28-1(a)-1
1670	Gambling Device	38-28-1(a)-3
1680	All Other	
	OFFENSES INVOLVING CHILDREN	
1710	Endangering Life or Health	23-2354
1720	Contrib. to Delinq. of a Minor	23-2361a
1730	Curfew	23-2371
1740	Run-Aways (Juvenile)	37-702-3a
1750	Child Abuse	23-2042
1760	Paternity	106-3/4-52
1770	Truancy	37-702-3b
1780	All Other	
	CANNABIS CONTROL ACT	
1811	Poss. of 30 Grams or Less	56-1/2-704
1812	Poss. of Over 30 Grams	56-1/2-704
1821	Manuf., Del., Posses w/Intent to Del. or Manuf. of 10 Grams or Less	56-1/2-705
1822	Manuf., Del., Posses w/Intent to Del. or Manuf. of Over 10 Grams	56-1/2-705
1830	Casual Delivery	56-1/2-706
1840	Under 18 - Delivery	56-1/2-707
1850	Production of Cannabis Plant	56-1/2-708
1860	Calculated Cannabis Conspiracy	56-1/2-709
1880	Other	
1900	Intoxicating Compounds	38-81
	CONTROLLED SUBSTANCES ACT	
2010	Manufacture-Intent	56-1/2-1401
2020	Possessing a Cont'ld Substance	56-1/2-1402
2030	Counterfeit Subs.-Manuf. or Del.	56-1/2-1403
2040	Delivery or Poss. Intent to Del	56-1/2-1404
2050	Criminal Drug Conspiracy	56-1/2-1405
2060	Licensed Operations-Regist.	56-1/2-1302
2070	Delivery to Persons Under 18	56-1/2-1407
2080	Failure to Keep Records-Oper.	56-1/2-1306
2090	Other	
	HYPODERMIC SYRINGES & NEEDLES ACT	
2110	Possession or Sale	38-22-50 & 51
2120	Failure to Keep Records	38-22-52
	LIQUOR CONTROL ACT VIOLATIONS	
2210	Sales to Minors, Drunkards, Etc.	43-131
2220	Illegal Possession by Minor	43-134a
2230	Illegal Consumption by Minors	43-134a
2240	Misrepres. of Age by Minor	43-134a
2250	Other	
2300	Solicit of Alcoholic Beverages	38-26-1-2
	MOTOR VEHICLE OFFENSES	
2410	Driving Under the Infl.-Alcohol	95-1/2-11-501
2420	Driving Under the Infl.-Drugs	95-1/2-11-501
2430	Trans. of Alcoholic Liquor	95-1/2-11-502
2440	Reckless Driving	95-1/2-11-503
2445	Hit and Run	95-1/2-11-403
	MOTOR VEHICLE OFFENSES (Continued)	
2450	Drag Racing	95-1/2-11-504
2455	No Registration	95-1/2-3-701
2460	Revoked, Cancelled Registration	95-1/2-3-702
2465	Improper Use of Registration	95-1/2-3-703
2470	No Drivers License	95-1/2-6-101
2480	Suspend Revoked Drivers License	95-1/2-6-303
2490	Unlawful Use of Drivers License	95-1/2-6-301
2495	Flee or Attempt to Elude Police Off	95-1/2-11-204
2500	Criminal Abortion	81-17
	DISORDERLY CONDUCT	
2805	Vagrancy	24-11-5-4
2807	Drunkenness (Local Laws)	
2810	Prowler	38-26-1(a)-(1)
2820	Telephone Threat	38-26-1(a)-(2)
2830	Obscene Phone Calls	38-26-1(a)-(2)
2840	False Fire Alarm	38-26-1(a)-(3)
2850	Bomb Threat	38-26-1(a)-(4)
2860	False Police Report	38-26-1(a)-(5)
2870	Peeping Tom	38-26-1(a)-(6)
2880	Confinement to Prevent a Crime	38-200-1
2890	Other - Not Drunkenness	38-26-1(a)-(1)
2900	Air Rifles	38-82
3000	Fireworks	127-1/2-128
3100	Mob Action & Related Offenses	38-25
3200	Armed Violence	38-33A-2
3300	Public Demonstration	38-85
3400	Looting	38-42
3500	Prty Forfeit. (Pub. Nuis.)	38-37
3610	Interfer. with a Pub. Instit	38-21.2-2
	INTERFER WITH PUB OFFICERS	
3710	Resis. or Obstruct a Peace Off	38-31-1
3720	Refusing to Aid an Officer	38-31-8
3730	Obstructing Justice	38-31-4
3740	Concealing or Aiding a Fugitive	38-31-5
3750	Escape	38-31-6
3760	Other	
3800	Interfer. w/Judical Proced	38-32
3810	Contempt of Court	38-1-3
3910	Bribery	38-29 & 29A & 33
3960	Intimidation	38-12-6
3970	Extortion	Title 18 U S Code 875-876
4000	Violation of Civil Rights	38-13
4100	Criminal Defamation	38-27
	KIDNAPING	
4210	Kidnaping	38-10-1
4220	Aggravated Kidnaping	38-10-2
4230	Unlawful Restraint	38-10-3
4310	Possession of Burglary Tools	38-19-2
4410	Draft Cards Destruction or Mutil.	38-90-11
4510	Probation - Violation	38-1005-6-4
4625	Parole - Revocation	38-1003-3-9
4710	Suspicion	38-107-2
4720	AWOL & Desertion	ART. 85-U C M J.
4730	Illegal Entry - Alien	Title 8 U S Code 1325
5000	Other Criminal Offenses	
5060	Other Traffic Offenses	
5080	Other Departmental Services	

APPENDIX B

Maps Detailing Illinois County, Court Circuit, and Planning Region Boundaries

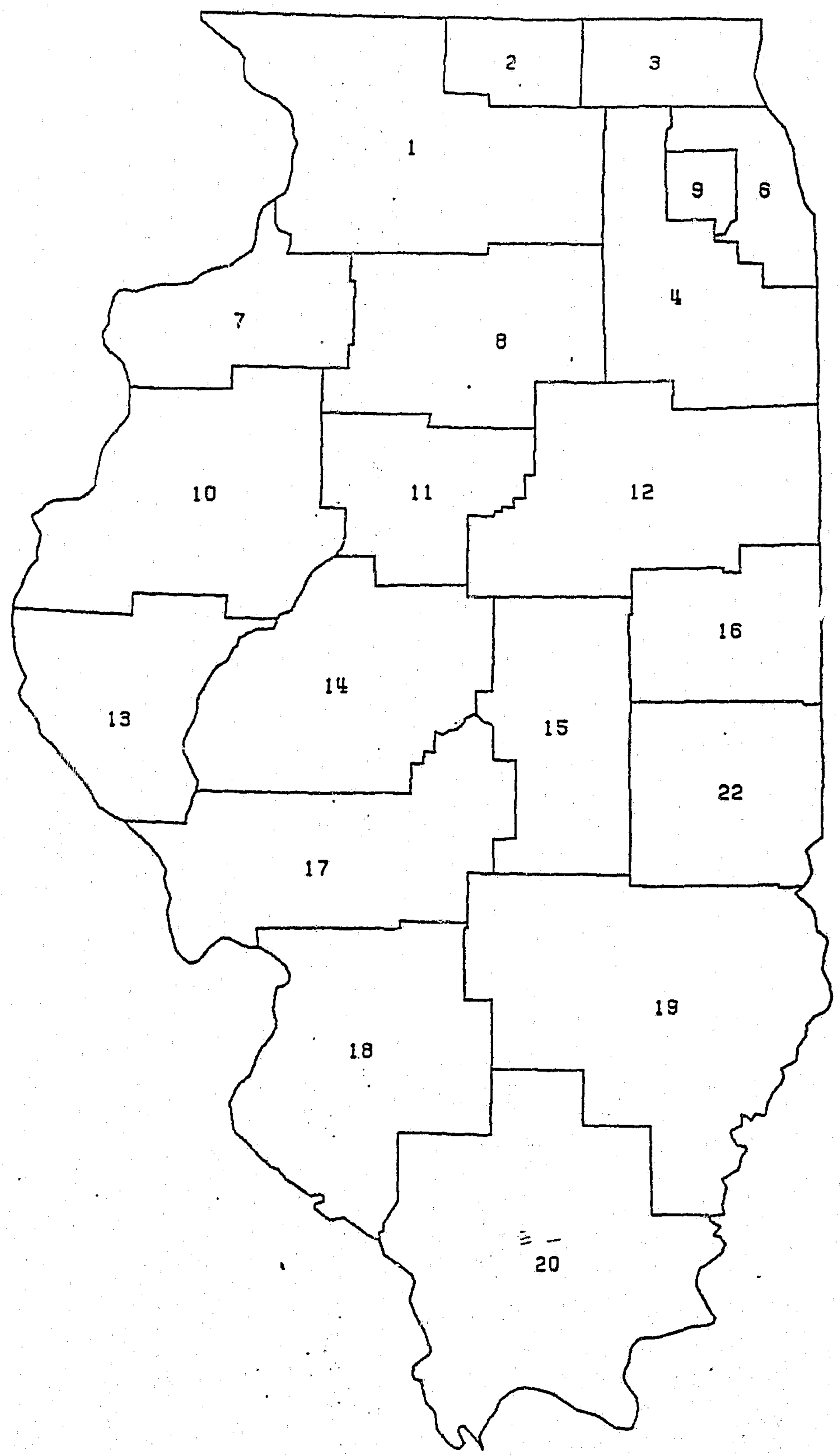
ILLINOIS COUNTY REFERENCE MAP



1 = ADAMS
2 = ALEXANDER
3 = BOND
4 = BOONE
5 = BROWN
6 = BUREAU
7 = CALHOUN
8 = CARROLL
9 = CASS
10 = CHAMPAIGN
11 = CHRISTIAN
12 = CLARK
13 = CLAY
14 = CLINTON
15 = COLES
16 = COOK
17 = CRAWFORD
18 = CUMBERLAND
19 = DE KALB
20 = DEWITT
21 = DOUGLAS
22 = DUPAGE
23 = EDGAR
24 = EDWARDS
25 = EFFINGHAM
26 = FAYETTE
27 = FORD
28 = FRANKLIN
29 = FULTON
30 = GALLATIN
31 = GREENE
32 = GRUNDY
33 = HAMILTON
34 = HANCOCK
35 = HARDIN
36 = HENDERSON
37 = HENRY
38 = IROQUOIS
39 = JACKSON
40 = JASPER
41 = JEFFERSON
42 = JERSEY
43 = JO DAVIESS
44 = JOHNSON
45 = KANE
46 = KANKAKEE
47 = KENDALL
48 = KNOX
49 = LAKE
50 = LA SALLE
51 = LAWRENCE
52 = LEE
53 = LIVINGSTON
54 = LOGAN
55 = MC DONOUGH
56 = MC HENRY
57 = MC LEAN
58 = MACON
59 = MACOUPIN
60 = MADISON
61 = MARION
62 = MARSHALL
63 = MASON
64 = MASSAC
65 = MENARD
66 = MERCER
67 = MONROE
68 = MONTGOMERY
69 = MORGAN
70 = MOULTRIE
71 = OGLE
72 = PEORIA
73 = PERRY
74 = PIATT
75 = PIKE
76 = POPE
77 = PULASKI
78 = PUTNAM
79 = RANDOLPH
80 = RICHLAND
81 = ROCK ISLAND
82 = SAINT CLAIR
83 = SALINE
84 = SANGAMON
85 = SCHUYLER
86 = SCOTT
87 = SHELBY
88 = STARK
89 = STEPHENSON
90 = TAZEWELL
91 = UNION
92 = VERMILION
93 = WABASH
94 = WARREN
95 = WASHINGTON
96 = WAYNE
97 = WHITE
98 = WHITESIDE
99 = WILL
100 = WILLIAMSON
101 = WINNEBAGO
102 = WOODFORD

ILEC/CJIS—STATISTICAL ANALYSIS CENTER GRAPH

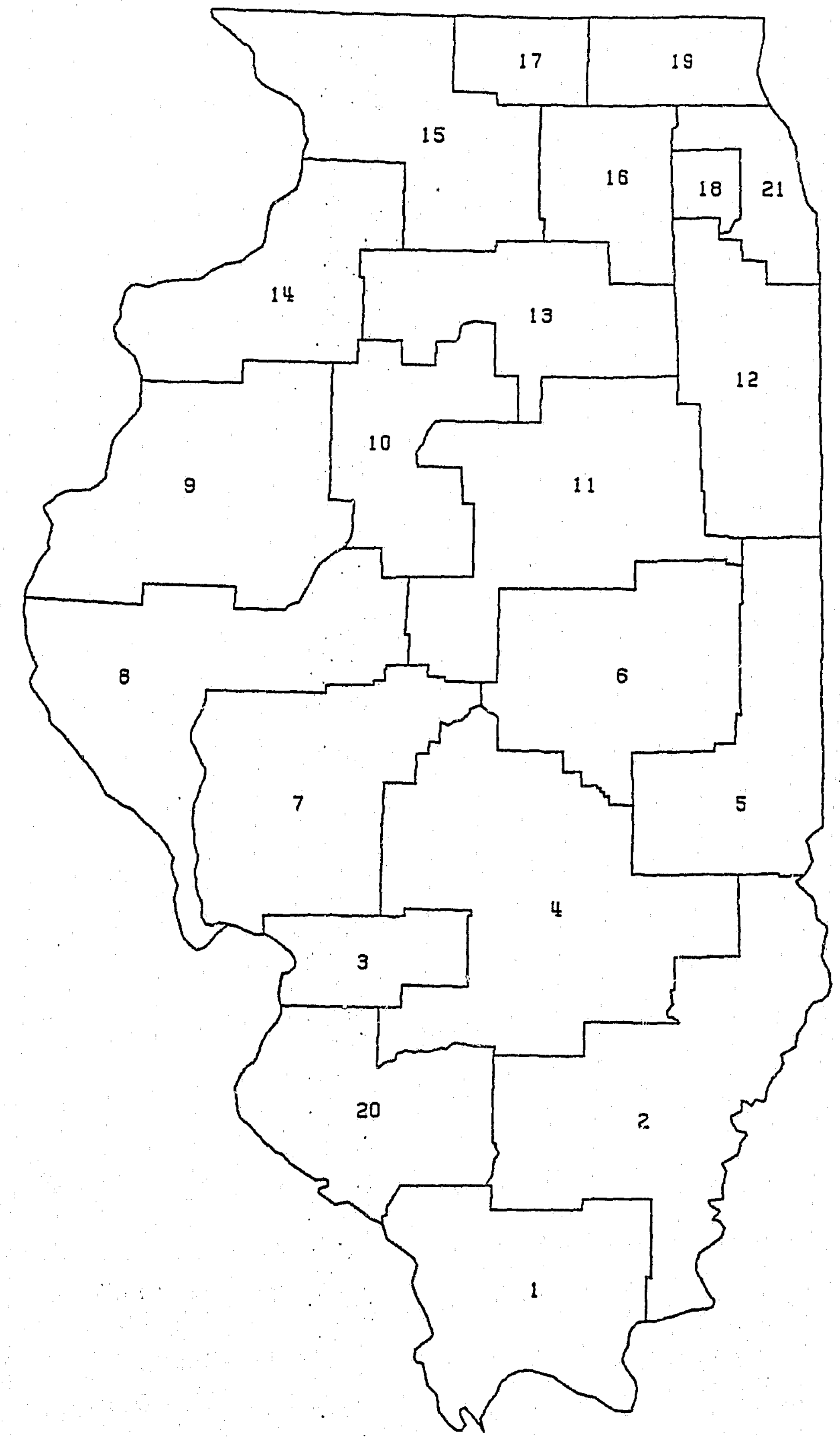
ILEC PLANNING REGION REFERENCE MAP



ILEC REGIONS ARE REFERENCED BY NUMBER. THE NUMBER OF EACH REGION APPEARS WITHIN ITS BOUNDARIES. THERE ARE NO REGIONS NUMBERED 5 OR 21.

ILEC/CJIS--STATISTICAL ANALYSIS CENTER GRAPH

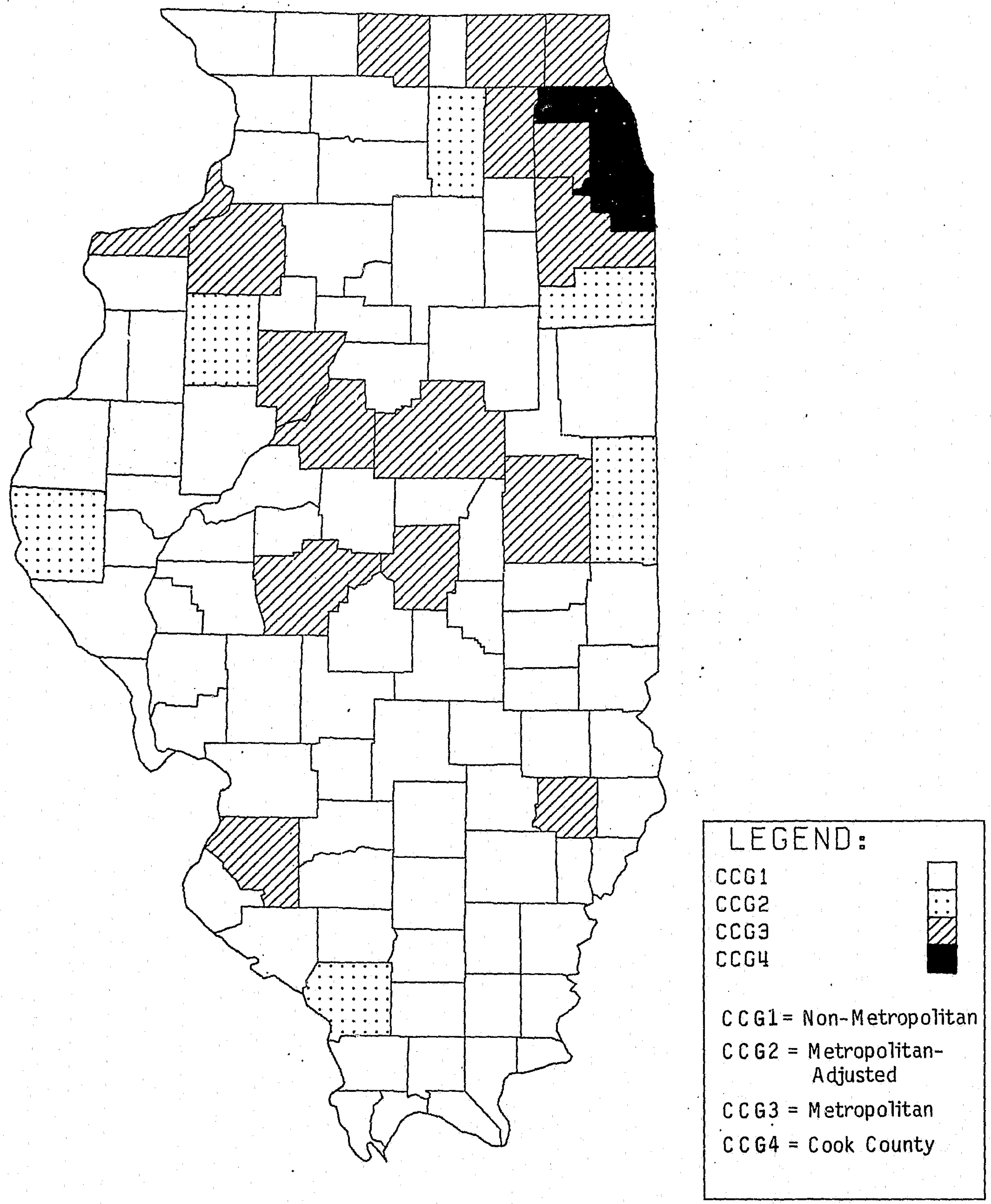
ILLINOIS CIRCUIT COURT REFERENCE MAP



ILLINOIS CIRCUIT COURTS ARE REFERENCED BY NUMBER. THE NUMBER OF EACH CIRCUIT APPEARS WITHIN ITS BOUNDARIES.

ILEC/CJIS--STATISTICAL ANALYSIS CENTER GRAPH

ILLINOIS COMMON CHARACTERISTIC GROUPINGS



ILEC/CJIS--STATISTICAL ANALYSIS CENTER GRAPH

BIBLIOGRAPHY AND REFERENCES

Alexander, K.L., and L.J. Griffin

1976 "School district effects on academic achievement: a reconsideration." American Sociological Review. 41: 144-152.

1976a "Reply to Bidwell and Kisarda." American Sociological Review. 41: 755-762.

Alker, H.R., Jr.

1969 "A typology of ecological fallacies." pp. 69-86 in M. Dogan and S. Rokkan (eds.), Quantitative Ecological Analysis in the Social Sciences. Cambridge: MIT.

Bidwell, C.E., and J. D. Kasarda

1976 "Reply to Hannan, Freeman, and Meyer, and Alexander and Griffin." American Sociological Review. 41: 152-160.

1975 "School district organization and student achievement." American Sociological Review. 40: 55-70.

Blalock, H.

1964a Causal Inference in Non-Experimental Research. Chapel Hill: University of North Carolina Press, 95-114.

1964b Social Statistics. San Fransisco: Mc Graw-Hill Book Company, 361-426.

Blau, P.M.

1960 "Structural effects." American Sociological Review. 25: 178-193.

Davis, J. A., J. L. Spaeth, and C. Huson

1961 "A technique of analyzing effects of group composition." American Sociological Review. 26: 215-225.

Duncan, O.D., and B. Davis

1953 "An alternative to ecological correlation." American Sociological Review. 18: 665-666.

Erbring, L.

1978 "Individuals writ large: on the virtues of 'ecological' analysis." Paper presented at the Annual Meetings of the American Sociological Association.

Farkas, G.

1974 "Specification, residuals and contextual effects." Sociological Methods and Research. 2: 333-363.

Feige, E.L., and H. Watts

1972 "An investigation of the consequences of partial aggregation of micro-economic data." Econometrica. 40: 343-360.

Goodman, L.

1959 "Some alternatives to ecological correlation." American Journal of Sociology. 64: 610-625.

Grunfeld, Y., and Z. Grilches

1960 "Is aggregaton necessarily bad?" Review of Economics and Statistics. 42: 1-13.

Hammond, J.L., Jr.

1973 "Two sources of error in ecological correlations." American Sociological Review. 38: 764-777.

Hannan, M.T.

1970 "Problems of aggregation and disaggregation in sociological research." Institute for Research in Science Working Paper in Methodology, N.C. Chapel Hill: University of North Carolina.

1971 Aggregation and Disaggregation in Sociology. Lexington, Mass.: D.C. Heath and Co., 91-109.

Hannan, M., and L Burstein

1974 "Estimation from grouped observations." American Sociological Review. 39: 374-392.

Hannan, M.T., J.H. Freeman, and J.W. Mayer

1976 "Specification of models for organizational effectiveness: a comment of Bidwell and Kasarda." American Sociological Review. 41: 136-143.

Hanuschek, E.A., J.E. Jackson, and J.F. Kain

1974 "Model specification, use of aggregate data, and the ecological correlation fallacy." Political Methodology. 1: 87-106.

Hauser, R.M.

1974 "Contextual analysis revisited." Sociological Methods and Research. 2: 365-375.

Illinois Department of Law Enforcement

1972 Crime in Illinois.

Illinois Law Enforcement Commission

1979 Fiscal Year 1979 State Plan.

Langbein, L.I., and A.J. Lichtman

1978 Ecological Inference. Beverly Hills: Sage Publications, Inc.

Lazarsfeld, P. and H. Menzel

1961 "On the relation between individual and collective properties." pp. 422-440 in A. Etzioni (ed.), Complex Organizations: a Sociological Reader. New York: Holt, Rinehart, and Winston.

Lichtman, A.J.

1974 "Correlation, regression and the ecological fallacy: a critique." Journal of Interdisciplinary History. 4: 417-433.

Meckstroth, T.W.

1974 "Some problems in cross-level inference." American Journal of Political Science. 18: 45-66.

Orcutt, G.H., Watts, and L.B. Edwards

1968 "Data aggregation and information loss." American Economics Review. 58: 773-787.

Peston, M.H.

1959 "A view of the aggregation problem." Review of Economics Studies. 27: 58-64.

Robinson, W.S.

1950 "Ecological correlations and the behavior of individuals." American Sociological Review. 15: 351-357.

Shively, W.P.

1974 "Utilizing external evidence in cross-level inference." Political Methodology. 1: 61-74.

Tannenbaum, A.S., and J. Bachman

1964 "Structual versus individual effects." American Journal of Sociology. 69: 540-554.

Valkonen, T.

1969 "Secondary analysis of survey data with ecological variables." Social Science Information 8: 33-36.

NOTES

[1]Crosstabulation and non-parametric techniques refer to statistical analysis techniques that are used mainly for analysis of variables measured with nominal and ordinal scales.

[2]SPSS is the Statistical Package for the Social Sciences, a system of computerized programs for data analysis. See Nie, et. al., SPSS, McGraw-Hill Book Co., 1970,1975.

[3]See Appendix A for a list of these offenses.

[4]CCG's are analytical units formed by classifying all Illinois counties (except Cook County) as Metropolitan, Non-Metropolitan, or Non-Metropolitan with a city over 25,000 population. Cook County is counted as a single CCG. See Illinois Law Enforcement Commission, Fiscal Year 1979 State Plan, pp. i,I-3, for a more detailed description of CCG's.

[5]See Appendix C for maps detailing boundaries for Illinois counties, circuits, regions, and CCG's.

[6]It is important to understand that, regardless of the geographic level at which they are used, SAC-IUCR offense data are aggregated in some manner because they are not incident-level data. Each case in the SAC-IUCR offense files represents a grouping of crime incidents (offenses). For example, SAC-IUCR offense data are available in two main formats: county-yearly and agency monthly. In the county-yearly files each case (crime category) contains the total number of reported offenses for each county for one year. In the agency-monthly files each case contains the total number of reported offenses for each police agency for one month. The SAC-IUCR files contain incident-level data for property crimes, arrests, and homicides. See the SAC publication, Illinois Uniform Crime Reports User's Guide and Codebooks, for a detailed explanation of the different formats that SAC-IUCR data are available in.

[7]Aggregation at the CCG level is not included as an option because four cases is too small a sample for most analyses.

[8]See the SAC report, A Guide to the Sources of Data on Criminal Cases Processed in the Cook County Circuit Court for a detailed explanation of Cook County Circuit Court organization and data collection.

[9]See Blalock (1960), Chapters 17 and 18, for a detailed explanation of correlation and regression coefficients.

[10] Standard deviations are measures of dispersion which indicate the amount of spread a variable has around its arithmetic mean. See Blalock (1960:80-82,100) for an explanation of standard deviations.

[11] Standard regression coefficients are used in multiple regression analysis to measure (in standardized units) the effect an independent variable has on the dependent variable when the effects of the other variables in the equation are held constant. See Blalock (1960:450-453) for a discussion of standardized regression coefficients.

[12] See Langbein and Lichtman (1978:33-38) for a discussion of aggregation analysis using standardized and unstandardized measures.

[13]In rare cases, when aggregation occurs along a variable related only to the independent variables in a model, aggregation can produce better estimates than would be obtained at the individual level with a misspecified model. This is called aggregation gain. It occurs when aggregation produces a better specified model. See Langbein and Lichtman (1978:28-31) for a discussion of aggregation gain.

[14] A complete analysis of the inaccuracies in the major Illinois criminal justice data sources has not been published. The Statistical Analysis Center publication, Illinois Uniform Crime Reports User's Guide and Codebooks, however, provides valuable information concerning inaccuracies in IUCR data.

[15] If it is not possible to conduct dummy variable regression with the available data, a more indirect approach is possible. This involves finding out how variables behaved in analyses similar to the one at hand.

[16] This three-category scheme is adopted from the Common Characteristic Grouping (CCG) variable coded in the SAC-IUCR data files. It represents a crude scale of rural to urban county types.

[17] Contingency table analysis is another term for crosstabulation analysis. It is a statistical technique for analyzing the relationships between variables measured with ordinal or nominal scales. See Blalock (1960), Chapter 15 for a discussion of contingency table analysis.

[18] See footnote 1, page 4 of this paper.

[19] In all contingency table analyses involving 2 X 2 tables, the Phi coefficient is reported in place of Cramer's V, and Yule's Q is reported in the place of Gamma. Phi and Yule's Q are special statistics used only for analyses involving 2 X 2 tables. They are interpreted in the same manner as their counterparts, however.

[20] See Blalock (1960:297-299,421-424) for a more detailed discussion of Gamma and Cramer's V. See also Nie, et. al. (1970,1975:224,228)for discussions of both statistics.

[21] It is generally accepted that the Chi Square statistics produced from contingency tables containing more than a few cells with fewer than 5 observations are unreliable Blalock (1960:285-286). Cramer's V and Phi are calculated using the Chi Square statistic.

[22] Gamma and Yule's Q are not based on the Chi Square statistic.

<u>SAC Publications</u>

DATA SOURCES ON THE INCIDENCE OF ARSON IN ILLINOIS, by Chip Coldren (November, 1977)

CRIME RATES WORKBOOK, by Ruth A. Perrin (December, 1977)

DATA ON EXTORTION IN ILLINOIS, by Ruth A. Perrin and James R. Coldren Jr. (April, 1978)

REPORT ON TECHNICAL ASSISTANCE PROJECT: DOMESTIC VIOLENCE SURVEY, by L. Edward Day (September, 1978)

A METHODOLOGICAL REVIEW OF "THE IMPACT OF MANPOWER SERVICES ON ILLINOIS OFFENDERS," by George W. Knox, by Carolyn R. Block (October, 1978)

A GUIDE TO THE SOURCES OF DATA ON CRIMINAL CASES PROCESSED IN THE COOK COUNTY CIRCUIT COURT, by Karen P. Smith and James Zuehl (December, 1978; revised June, 1979)

ILLINOIS UNIFORM CRIME REPORTS USER"S GUIDE AND CODEBOOKS, by Linda Kok (February, 1979)

ILLINOIS UNIFORM CRIME REPORTS USER"S GUIDE UPDATE FOR 1977 DATA, by Linda Kok (February, 1980)

ILLINOIS VICTIM SURVEY DATA: A GUIDE TO THEIR USE, by Ruth A. Perrin (October, 1977; revised March, 1979)

COMPARING ILLINOIS POLICE DATA TO COURTS AND CORRECTIONS DATA: THE PROBLEMS AND A SUGGESTED SOLUTION, by Carolyn R. Block (March, 1979)

A LOOK IN THE BLACK BOX: THE TRANSFORMATION OF ROBBERY INCIDENTS INTO OFFICIAL ROBBERY STATISTICS, by Richard Block and Carolyn R. Block (June, 1979)

ROBBERY AND BURGLARY 1972-1977 TRENDS FOR ILLINOIS CITIES, COUNTIES AND ILEC PLANNING REGIONS, by L. Edward Day (June, 1979)

GUIDE TO ILLINOIS PAROLE AND RELEASE DATA, by Carolyn R. Block (July, 1979; revised November, 1980)

DESCRIPTIVE TIME SERIES ANALYSIS FOR CRIMINAL JUSTICE DECISION MAKERS: LOCAL ILLINOIS ROBBERY AND BURGALRY, by Carolyn R. Block (November, 1979)

INMATE PROFILE ANALYSIS, by Edward F. Maier and Stephen F. Tapke (November, 1979)

DATA SOURCES ON PROBATION, CONDITIONAL DISCHARGE, SUPERVISION AND PERIODIC IMPRISONMENT IN ILLINOIS, by James R. Coldren Jr. (December, 1979)

The COMPILER, editor: Eric Newcomer (newsletter published bi-monthly)

ANNUAL PRIVACY AND SECURITY AUDIT REPORT, period ending June 30, 1979, by Illinois Criminal Justice Information Council

PATTERNS OF CHANGE IN CHICAGO HOMICIDE: THE TWENTIES, THE SIXTIES AND THE SEVENTIES, by Carolyn R. Block and Richard Block (April, 1980)

DECISIONS AND DATA: THE TRANSFORMATION OF ROBBERY INCIDENTS INTO OFFICIAL ROBBERY STATISTICS, by Richard L. Block and Carolyn R. Block (July, 1980)

AGGREGATION PROBLEMS IN THE ANALYSIS OF ILLINOIS STATEWIDE CRIMINAL JUSTICE DATA, by James R. Coldren, Jr. (November, 1980)

GUIDE TO ILLINOIS FIREARM DATA, by Carolyn R. Block (November, 1980)

END