FINAL REPORT

LEAA/NILCJ GRANT #79-NI-AX-0057

PROJECT TITLE: Confirmatory and Exploratory Analysis of the Spatio-Temporal Properties of Crime Data

PRINCIPAL INVESTIGATORS: Reginald G. Golledge
Professor of Geography
University of California,
Santa Barbara

Lawrence J. Hubert
Professor of Education
University of California,
Santa Barbara

FINAL REPORT: EXECUTIVE SUMMARY

Confirmatory and Exploratory Analysis of the Spatio and Temporal Properties of Crime Data

LEAA/NILCJ GRANT #79-NI-AX-0057

by

Reginald G. Golledge
Professor of Geography
University of California, Santa Barbara

and

Lawrence J. Hubert
Professor of Education,
University of California, Santa Barbara

January 1981

Executive Summary

The purpose of this research was to develop procedures for analyzing selected spatial and temporal characteristics of crime data. Three particular problem areas were to be investigated: (a) selected aspects of the aggregation process; (b) confirmatory and exploratory methods for examining crime data summarized in proximity matrices; and (c) an examination of spatial autocorrelation procedures. In each case an attempt was made to examine the problem in terms of a randomization approach rather than by conventional statistical procedures. The randomization procedure was based on the work of Hubert (1978a,b, 1979a,b), Hubert and Baker (1978a,b,c) Hubert and Levin (1976b), and Mantel (1967). In the course of the general set of investigations, spillover effects occured in that our examination of methodologies such as hierarchical clustering and multidimensional scaling, and the use of the combinatorial strategies associated with the quadratic assignment procedure, stimulated research on diverse topics such as developing a model for roll-call cohesion methods (a combinatorial strategy) and examining the potential use of non-Euclidean geometries in the study of cognitive maps (a variation of the standard Euclidean measures used in calculating proximities and an extension of the "perception of crime" research in paper #4). While these latter two papers are somewhat peripheral to the main theme of the research, they do represent areas of considerable interest in disciplines such as geography, cognitive science, psychology, and political science.

The major problems investigated during the course of the project and outlined in the final report can be summarized as follows:

1. Aggregation in Data Tables: Implications for Evaluating Criminal Justice Statistics

a) SUMMARY

Data collected on a set of objects (e.g., cities) over a set of attributes (e.g., time points) can be subjected to a variety of aggregation schemes. For example, if a hypothesized pattern over the attributes is to be confirmed (e.g., a temporal increase in homicide rate), the data could be first aggregated over cities and then compared to the hypothesized pattern. Alternatively, the correspondence for each city could be separately assessed and the individual city indices aggregated. The stage at which aggregation takes place affects the size of the final measure of confirmation as well as its significance, but unfortunately, in opposite ways. Preliminary data aggregation typically leads to larger summary statistics and larger significance levels. The conflicting notions of size and significance are first formalized in detail when the basic data are single numerical values obtained for each object-attribute pair. Extensions are presented to multi-group concordance, hierarchical aggregation schemes, and to object data defined by pairwise proximities between the attributes.

b) DISCUSSION*

In the very first example illustrating two aggregation schemes, a rank transformation was used within each row and on the rank sums for each column. This process ensures that each object or row contributes "equally" and thus, some degree of natural comparability exists between the summary statistics obtained for the two aggregation schemes. To develop more explicit relationships in terms of formulas, however, the transformations used in most of the paper were based on obtaining z-scores. In the jargon of statistics, observations within rows are aligned for location and scale. This convention allowed precise connections to be developed between the two aggregation schemes both in terms of summary indices (e.g., for r_A and r_B) and Z-statistics (e.g., for Z_A and Z_B) (for definition of all these summary indexes see pages 9-12 of the first paper presented in the main section of this report).

Matrix extensions offer a great deal of flexibility in defining different relationships among the attributes, but unfortunately, the problem of defining a transformation on the aggregate data matrix also makes it very difficult in general to develop precise formulas for connecting the two aggregation schemes. As an example of this problem, suppose we are given the basic object by attribute data table and define an n x n matrix for each object (e.g., city) as follows: the entry in the u^{th} row and v^{th} column is +1 if $x_{iu} > x_{iv}$; -1 if $x_{iu} < x_{iv}$, and 0 if $x_{iu} = x_{iv}$. If we treat the n(n-1) entries in each such matrix as a sample, normalize to z-scores in the usual way, and carry out

the same redefinition for a criterion set of values $y_1, y_2, \ldots, y_n$, the average correlation r_B is actually the average Kendall Tau_b statistic of each row against the criterion (cf. Hays, 1960). However, to obtain an analogue of r_A, a similar transformation to signs must be performed on the aggregated scores from the standardized object by attribute table. This discontinuity in strategy prevents any simple way of defining a relationship between the summary measures for the two aggregation schemes. We would still expect data aggregation to give a larger descriptive measure and a greater significance level, but it is not clear how these expectations could be formalized as a parallel to our previous equations (1) and (2) (given on pages 10 and 11 of the first paper in the main section of the report).

As another point of clarification, we note that the normalization within rows of an object by attribute data table may not be the only natural transformation to carry out. Instead, suppose z-scores are obtained within columns and for each object i we define an n x n matrix having an entry in the u^{th} row and v^{th} column of $\frac{1}{I} z_{iu} z_{iv}$. Here, z_{iu} and z_{iv} are z-scores for attributes u and v, respectively. If we aggregate over the I matrices (treating objects as if they were "subjects"), the correlation matrix among attributes is generated. Pattern comparisons are important here in the context of what is called a multi-trait multi-method matrix; consequently, some of the same aggregation principles discussed previously appear important to distinguish in these applications as well (see Campbell and Fiske, 1959; Hubert and Baker, 1978).

The methods of data aggregation discussed in the paper represent both ongoing procedures used in geography for aggregating data and alternatives to those standard procedures. For example, Harries (1973) used preliminary aggregation procedures when he averaged violent crime rates for 189 SMSA's for the five year period 1965-69 and calculated simple correlations between violent crime and population over the SMSA's. Alternative procedures for determining correlations between violent crime and population are given in our discussion of the statistics r_B and $r_{s(ave)}$. Pyle (1974) used correlational aggregation when he examined actual crime rates per 1000 persons for nine crime types, plus armed robberies per 1000 commercial structures and rates of residential burglary per 1000 dwelling units. Pyle calculated all pairwise correlations for the entire study area (Summit County) and for a subset of the study area (Akron) and then attempted to illustrate differences between the correlations. This is similar to the procedures used in discussing multiple group concordance in this paper.

Harries (1974) also used preliminary data aggregation when he correlated city size with crime rates averaged over index crimes for a five year time period - a procedure that could be extended by using the matrix comparison procedures developed in the latter section of this paper. Other examples of preliminary and correlational data aggregation procedures can be found in the growing literature on the use of canonical correlation in geography (Monmonier and Finn, 1973; Clark, 1975). However, the exact procedures detailed in this paper focusing on rank orders and Z-statistics for both preliminary and correlational

aggregation procedures, to our knowledge, have not appeared in the geographical literature.

As one final observation, it should be noted that the two-group discussion developed in the paper was concerned with the concordance between two classes even though the various summary indices were subject to modification by the degree of internal concordance. In other words, we were not explicitly interested in assessing large within group homogeneity per se. Given the original I x I intercorrelation matrix, however, and a hypothesized split of the I objects into T groups (e.g., into two disjoint subsets), we may also wish to test whether there is more concordance within the groups than expected under some chance model. This topic has been discussed in detail elsewhere for the null conjecture that the given partition was chosen at random from all possible partitions with the same number of classes and objects in each. Thus, we would hope to reject the randomness assumption if the within group concordance was substantially greater than the between group concordance, i.e., the a priori partition is reflected in the size of the correlations in the original I x I matrix. For a complete discussion, the reader is referred to Hubert and Levin (1976).

* References that are cited are given in the appropriate paper presented in the main section of this report.

2. Assessing Homogeneity in Cross-Classified Proximity Data

a) SUMMARY

Given an arbitrary proximity matrix that is cross-classified according to two dimensions, a nonparametric strategy generalizing Friedman's (randomized blocks) analysis-of-variance method, is suggested for testing the saliences of the dimensions. Straightforward extensions of the approach can be given for more than two dimensions and/or when only the ordering of the proximity values is of interest.

b) DISCUSSION

The major contribution of this paper is in the use of arbitrary proximity measures and the development of a strategy for blocking on the levels of one (or more) a priori dimension(s) when evaluating the differences over a second. The strategy being proposed is really very general even though the illustration we have used in explaining the method contained the three explicit classification dimensions of space, time, and crime type. For instance, since any two of the dimensions could in fact have been considered the major classification facets of interest, proximity measures could have been obtained between profiles over the m cities and our interests directed toward the two dimensions of crime type and time. The basic inference principles would remain the same and the analyses would be carried out as before. Hopefully, our discussion will allow researchers to assess dimensional salience in data sets that are

not easily studied by more standard analysis-of-variance schemes because of an unusual proximity measure. Moreover, the possibility of relying on only nonmetric comparisons among proximities should provide a nice tie-in to the current emphasis in nonmetric clustering and scaling in the social and behavioral sciences.

3. Unidimensional Seriation: Implications for Evaluating Criminal Justice Data

a) SUMMARY

The problem of validating a given unidimensional scale (i.e., an ordering of a set of objects along a single dimension) is discussed in terms of a few simple properties of the data used to obtain the scale. Based on a set of asymmetric proximity values as raw data, a distinction between analyzing absolute value information or sign information is presented that leads to a formal test of whether a given scale is being reliably represented. In short, a scale is generated from absolute value information but validated through sign information. A numerical example which deals with the perception of homicide rate over 15 of the largest SMSA's is included as an illustration of the general methodological discussion.

b) DISCUSSION & EXTENSION*

Given the basic context of unidimensional seriation as developed in the earlier sections of this paper, a variety of

additional topics could be pursued. We mention only a few in passing to give some indication of the current research efforts in this direction. For example, the type of inference strategy that was proposed for evaluating the pattern of signs can be extended to compare two arbitrary skew-symmetric matrices (see Hubert and Schultz, 1976). Thus, it is possible to evaluate the consistency between two skew-symmetric interaction matrices where the latter may be based on migration data at two time points or from two different demographic subgroups. Secondly, from a combinatorial optimization point of view, several very elegant theoretical paradigms have been introduced recently for characterizing a discrepancy between a given seriation and the original asymmetric data, e.g., see Bowman and Colantoni (1973) and Merchant and Rao (1976). Along these same combinatorial optimization lines, a general strategy has been suggested (Hubert, 1980) for locating and seriating only a part of a proximity matrix that appears to be most consistent with the basic underlying spatial model. This latter technique can assist in identifying subsets of an object set that can be seriated well and those subsets that are not represented satisfactorily along a continuum.

As one example of particular importance we note that the topic of criminal mobility could define one of the more interesting applications for unidimensional scaling in the criminal justice area. For example, based on movement data from place of residence to place of the committed crime, we may wish to rate a set of geographical areas in terms of criminal attractivity, with the possible goal of comparing these rates

over different crime types, age groups, and so on. Typically, the basic data are flow statistics for a set of n localities defined by the number of people, m_{ij}, who travel from region i to j. Our aim is to model these data in terms of the distances among the localities and their assumed placement along an attractivity continuum.

Following Tobler's (1979) lead, the simplest model we consider is defined in terms of the skew-symmetric matrix $q^*_{ij} = m_{ij} - m_{ji}$, which specifies the degree to which j attracts more from i than it exports. We assume that these statistics conform to a model defined as

$$q^*_{ij} \approx \frac{A_j - A_i}{d_{ij}} \qquad (1)$$

where d_{ij} is the distance between locations i and j and $A_1, A_2, \ldots, A_n$ define attractivities along a single dimension. Obviously, since distances are typically known, our analysis task is to estimate the n attractivities, which in turn scales the n localities according to attractivity along a continuum.

Tobler (1979) discusses in detail two major approaches to the estimation of attractivities. The first is called the potential method in which the A 's are given implicitly by the matrix equation

$$\begin{bmatrix} \sum_{i\neq 1} \frac{1}{d_{i1}} & -\frac{1}{d_{12}} & -\frac{1}{d_{13}} & \cdots & -\frac{1}{d_{1n}} \\ -\frac{1}{d_{21}} & \sum_{i\neq 2} \frac{1}{d_{i2}} & -\frac{1}{d_{23}} & \cdots & -\frac{1}{d_{2n}} \\ \vdots & & & & \\ -\frac{1}{d_{n1}} & -\frac{1}{d_{n2}} & -\frac{1}{d_{n3}} & \cdots & \sum_{i\neq n} \frac{1}{d_{in}} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} \quad \begin{bmatrix} \sum_i q^*_{i1} \\ \sum_i q^*_{i2} \\ \vdots \\ \sum_i q^*_{in} \end{bmatrix}$$

Since this system is not of full rank, Tobler suggest letting $A_1 = 0$ and deleting the first row and column of the coefficient matrix. The system can then be solved by inverting the reduced coefficient matrix and using it as a premultiplier on the right-hand vector. The second procedure may generate different estimates since it is based on minimizing the least-squares criterion:

$$\sum_{i,j} \left(q^*_{ij} - \frac{A_j - A_i}{d_{ij}} \right)^2 \qquad (2)$$

Again, a system of equations similar to that given above is generated that is not of full rank. Based on the coefficient matrix having diagonal entries

$$\sum_{\substack{i=1 \\ i\neq k}}^{n} \left(\frac{1}{d^2_{ki}} + \frac{1}{d^2_{ik}} \right)$$

and off-diagonal entries

$$-\left(\frac{1}{d^2_{ik}} + \frac{1}{d^2_{ki}} \right),$$

and the right-hand vector defined by an arbitrary entry

$$\sum_{\substack{i=1 \\ i\neq k}}^{n} \left(\frac{q^*_{ik}}{d_{ik}} - \frac{q^*_{ki}}{d_{ki}} \right),$$

the deletion option for $A_1 \equiv 0$ can then be used to obtain a closed-form solution through matrix inversion.

Although Tobler's discussion is very elegant, seemingly minor modifications in the way the model is stated will eventually lead to several useful simplifications. In particular, since we assume q^*_{ij} should be "close" to $(A_j - A_i)/d_{ij}$ and the d_{ij}'s are known, it should also be true that $d_{ij}\ q^*_{ij}$ is "close" to $(A_j - A_i)$. Thus, the original gradient notion characterized as a division by d_{ij} is redefined by a distance weighting of the observed skew-symmetric proximities. Continuing in this way it should also be true that $(d_{ij}\ q^*_{ij})^2$ is "close" to $(A_j - A_i)^2$. Our problem is now reduced to fitting the entries in a symmetric matrix $\{(d_{ij}q^*_{ij})^2\}$ by a squared distance matrix $\{(A_j - A_i)^2\}$. This latter task can be approached by the type of eigenvector analysis introduced in the main paper. Or, if we assume $|d_{ij}q^*_{ij}| \simeq |A_j - A_i|$, the strategies developed by Defays (1978) and De Leeuw and Heiser (1977) could be followed.

By taking explicit advantage of a model equivocation, we can concentrate on the skew-symmetric proximities $t_{ij} \equiv d_{ij}q^*_{ij}$, which are supposedly defined by the simple differences between attractivities. For example, a least-squares loss-function would minimize

$$\sum_{i,j} (t_{ij} - (A_j - A_i))^2 ,$$

which is equal to a criterion weighted by the squared distances d^2_{ij}:

$$\sum_{i,j} d^2_{ij} \left(q^*_{ij} - \left(\frac{(A_j - A_i)}{d_{ij}}\right)\right)^2 .$$

The least-squares measure used by Tobler in (2) is similar in general form to this latter expression but is unweighted.

The redefinition of the estimation problem to use t_{ij} may seem trivial but it leads immediately to several convenient results. For example, the least-squares estimate of A_i subject to the constraint that $\sum_{i=1} A_i = 0$ is given by $t_{.i}/n$. This same estimate is also obtained by the potential method, and consequently, both methods lead to the same solution in this context. From a slightly more general perspective, suppose we fit a matrix of the form $\{C_j - C_i\}$ to our arbitrary skew-symmetric matrix $\{t_{ij}\}$ by maximizing the correlation between the corresponding entries. Again, the solution is obtained when C_i is defined as $t_{.i}/n$, and when used to define the matrix $\{C_j - C_i\}$, these values induce a correlation of

$$\sqrt{\frac{2}{n}\ \frac{\sum_j \left[\sum_i t_{ij}\right]^2}{\sum_j \sum_i t^2_{ij}}}$$

All of these last results are very close to some work by Noether (1960) on paired comparison scaling.

It should be apparent that many different approaches could be developed for estimating the attractivities $A_1, \ldots, A_n$ from a skew-symmetric matrix t_{ij} or q^*_{ij} merely by varying the explicit form of the model used and the loss function. This arbitrariness is troublesome since minor variations can dramatically affect the final estimation process (see Noether, 1960). To provide some hedge, and as we have suggested before, it may be appropriate to rely only on the absolute-value data in the estimation stage and use the sign data as a strategy for validating the order of the estimates along the continuum. Sign information has the nice property of being independent of the form of the gradient model

being used as long as the signs are governed by differences in attractivities.

* References that are cited are given in the appropriate paper presented in the main section of this report.

4. Proximity Matrix Reorganization and Hierarchical Clustering

a) SUMMARY

Connections between hierarchical clustering and the seriation of objects along a continuum that depend on the patterning of entries in a proximity matrix are pointed out. Based on the similarity between the central notion of an ultrametric in hierarchical clustering and what is called an anti-Robinson property in seriation, it is suggested that both data analysis procedures are compatible. In fact, preliminary seriation of a proximity matrix may help verify the adequacy of the results obtained from a hierarchical clustering or suggest alternatives that may be better. A numerical example using data from the criminal justice area is included.

b) DISCUSSION*

The idea of using a seriation of the object set prior to looking for a specific clustering reappears continually in the

literature although in many disguised forms. The key to recognizing this general paradigm is by the presence of some object ordering before a final clustering is given. Obvious examples would include Hartigan's (1975) leader algorithms, Matula's sequential graph coloring schemes (Matula, Marble, and Isaacson, 1972), Fisher's (1958) single variable clustering, and Szczotka's (1972) notion of an admissible partition. Implicitly or explicitly all of these methods rely on an object ordering, typically as an initial organizing step prior to a final clustering based on partitioning the reordered matrix in some particular way, e.g., seriating the "break points" that define the possible subsets of a partition.

We do not wish to advocate the superiority of seriation over a particular method of HC or conversely. Instead, our aim is to point out their complementary nature and how clustering and seriation could be used together to justify a specific analysis. Looking at one's data in ways that could suggest alternative interpretations may seem to be a very obvious tactic. Unfortunately, it is easily forgotten when a scheme is available that promises to give a single best answer and without the ambiguity that is usually attached to a more intuitive data analysis strategy.

* References that are cited are given in the appropriate paper presented in the main section of this report.

5. Generalized Procedures for Evaluating Spatial Autocorrelation

a) SUMMARY

Several generalizations of the usual spatial autocorrelation indices are developed based on the notion of matrix comparisons. These extensions are immediate from a related literature in biometrics and psychology, and in fact, the spatial autocorrelation inference task can be considered a special case of a much more general inference paradigm. The connections between matrix comparison and spatial autocorrelation are sketched including a way to define spatial autocorrelation statistics that only depend on the order of the entries in a proximity matrix.

b) DISCUSSION*

Autocorrelation measures, as traditionally used by geographers, describe the pattern of an observed variate over a map system and imply something about the predictablity of the map or the structure. By generalizing spatial autocorrelation measures using the randomization model as a base, a number of advantages accrue over the classical models based on specific distributional assumptions for the data. In general, the model proposed in this paper generates an immediately assessible inference paradigm for situations that would be very difficult to handle in a classical framework. In fact, it may well be that standard tests of significance are at times inappropriate for classical SA measures. By using a randomization model and the

complete enumeration process, significance levels appropriate for SA measures can be obtained.

In geography it is common to distinguish between two different approaches to spatial autocorrelation. One is tied to expressing spatial autocorrelation in a lagged form and depends on calculating and expressing covariances between different data values at different lagged distance or directional lengths. The second approach examines spatial autocorrelation in terms of the influence each observation is assumed to have on other observations. Our emphasis is more in line with the second approach rather than the first. Regardless of which approach is used, similar types of problems face the individual attempting to assess spatial autocorrelation and similar problems face researchers attempting to use and extend the procedure. Gattrell (1979) states the first of these as the need to specify alternate forms of the "distance" concept that provides the base for the calculation of spatial autocorrelation effects. At least on the variate side of the problem we have shown how spatial autocorrelation can be generalized to such alternate measures, including Mahalanobis distances, correlation coefficients, and any other arbitrarily defined indices of proximity. More general distance measures can be handled directly by defining the weight matrix appropriately, e.g., Mahalanobis distances in some generalized multidimensional space.

The extension of SA measures to data structures that are nominal or ordinal produces indices comparable with standard SA measures without the need to adhere to the stricter parametric assumptions necessary to generate an inference model for those

standard measures. Basing the spatial autocorrelation index on randomization overcomes one of the more critical problems currently being faced by geographers. For example, Haining (1980) argues that with respect to the Cliff, et al. study of measles data for Southwest England (1973), the absence of information on the joint sampling distribution of the average correlations, together with the small sample sizes involved and the generally insignificant values assumed by the correlations, tend to cast doubt on their interpretation of the measles epidemic as having a central place type diffusion structure. Developing an autocorrelation index using the randomization model clearly overcomes the first of these deficiencies, although it does not solve the second problem - that of defining a satisfactory model base for the interpretation of results.

The use of spatial autocorrelation in geography to compare observed and theoretical or expected map patterns has in the past been limited by the problems involved in measuring the degree of departure from randomness. As Dacey (1968b) and Cliff and Ord (1973) have found out, rejecting a hypothesis of randomness based on Poisson models cannot be taken as indicative of apparent contagion. While Besag (1972, 1974) has examined this problem in more detail, the inference problems raised in the geographic studies can be approached using randomization procedures to construct a reference distribution against which to measure the magnitude of deviations. Thus by using such an index and reaching a stage where a hypothesis of randomness is rejected, the researcher may feel more at ease assuming that the patterns examined are being produced by similar processes.

In summary, the approach we have described for measuring SA has at least four major advantages. The first relates to the generality of the paradigm and the fact that many different measures, even those tailor-made for specific substantive problems, can all be placed under one common framework and tested for significance using the same type of randomization argument. It is somewhat inappropriate, however, to view the general notion of matrix comparison as a competitor to the traditional way of handling SA tasks since special cases of matrix comparison have been used for some time. Nevertheless, there is an obvious inherent value in offering alternatives that may be more suited for particular research applications than the I and c indices. An obvious example would be in our ability to deal with more than one variable at a time in assessing SA through a multivariate measure of distance, defining the entries in the matrix $\underset{\sim}{C}$. Second, the randomization strategy itself can be approached through Monte Carlo sampling, bypassing the optimistic use of asymptotic distributional results of possibly unknown accuracy. These latter large sample size results are very spotty and do not cover all the SA statistics that could be defined in our framework. Third, by placing SA into a larger matrix comparison structure, an obvious pedagogical advantage is achieved. This is analogous to the perspective provided by understanding the general linear model even though the special cases of anlysis of variance and regression may continue to be the most popular alternatives as implemented by routines that are specialized from the more general structure. Fourth, once a comprehensive framework is understood, further work on the framework itself

immediately suggests many associated results that are pertinent to a class of measures. Thus, once the commonality of analysis tasks are recognized, there is an obvious broader purpose taken on by the research enterprise.

* References that are cited are given in the appropriate paper presented in the main section of this report.

6. Inference Models for Roll-Call Cohesion Measures

a) SUMMARY

An inference model for the percentage voting agreement measure is introduced that takes into account the composition of the parent group in evaluating the cohesion of a subgroup. The combinatorial strategy extends to a number of more general indices related to the original agreement measure.

b) DISCUSSION*

The problem of defining inference models for measures of roll-call cohesion has been recognized in the literature. For instance, as one way of developing a more reasonable inference structure, Born and Nevison (1975) introduced a probability measure based on the cumulative distribution of votes in the more inclusive body. Although this is a step in the appropriate direction, there are at least three limitations on the Born and Nevison approach. First of all, since the probability measure

requires a rather sophisticated understanding of probability theory and deviates markedly from the justification behind the indices typically used in the literature, acceptance of the new statistic may be very slow in coming, particularly since extensive tables and/or the use of specifically designed computer programs are required for its calculation. Secondly, the probability measure is essentially limited to votes that have 2-values (Aye and Nay) and voting options with k alternatives are not easily incorporated within the paradigm. Finally, the Born-Nevison measure is really a significance level, and therefore, it is heavily dependent on the size of the subgroup being considered. For example, two subgroups with the same values on a more traditional index of cohesion could also have very different probability values depending on the sizes of the two subgroups.

It would seem more appropriate to consider a statistic that has properties similar to the Pearson correlation coefficient; i.e., an index that would give an indication of the absolute level of cohesion irrespective of sample size but have an associated significance statement that was dependent on the number of observations, or in our case, on the number of voters in the subgroup as well as in the more inclusive body. Using these three concerns as our motivation, it is relatively straightforward to carry out the original Born-Nevison goal of providing a suitable inference model for a measure of cohesion through the well-known percentage voting agreement measure.

* References that are cited are given in the appropriate paper presented in the main section of this report.

7. Some Comments on Non-Euclidean Mental Maps

a) SUMMARY*

The widespread acceptance of Euclidean geometry as the most appropriate for representing space, predisposes a certain type of perspective on the world. Within the Euclidean framework, space is conceived as being isotropic - that is, the same geometric relations hold in all parts of the space. A second important concept is that of parallelism - that is, parallels do not converge. Accepting these concepts readily allows us to implement perhaps the best known and most widely used formula in the discipline of geography, that of measuring inter-point distances in N-dimensional Euclidean spaces. This formula is:

$$D_{ij} = \left[\sum_{k} \left(|x_{i,k} - x_{j,k}| \right)^{r} \right]^{\frac{1}{r}} \qquad (1)$$

where: D_{ij} is the distance between two arbitrarily defined points i and j;

$x_{i,k}$ is the coordinate for point i on the k^{th} dimension;

r is the exponent to which displacement in any dimension are taken in the particular distance formulation (i.e., the Minkowskian Metric).

While this general Minkowskian formula is well known, almost invariably the Euclidean (r = 2) is preferred to other

Minkowskian metrics such as r = 1 (the Manhattan or city block metric) or r = ∞ (the dominance metric or SUP-metric) (see Krause, 1975). In this latter metric, the distance between any pair of points is defined as the longest side of the right-angled triangle constructed in the space about the points (i.e., $d_{ij} = \max_{k}(x_{ik}, x_{jk})$).

In this paper we examine some characteristics of spatial cognition that indicate that the use of any of these Minkowskian metrics may not be appropriate for representing cognitive information, and we present a survey of elliptical and hyperbolic spaces (Riemann manifolds) that could potentially be used to represent such data.

b) DISCUSSION

If we examine configurations of points that have been generated using interpoint distance estimation or a set of points located in a space which is defined solely in terms of inter-point distances, then we are confronted with a space which implies neither a specific dimensionality nor is there implicitly embedded within it a coordinate system. If we can further find key nodes in this particular space, we might imagine that as distance increased from each key node, the probability of a fold, crack, tear, hole, or other warpage of the space would increase considerably. Assume further that the key nodes are not uniformly spaced. If one were then to construct a set of Thiessen polygons for this set of non-uniformly spaced nodes, there would be considerable variations in the distance of the edges of the polygons from the key nodes. One might further

expect that, in those areas of each polygon which are most distant from the key node, information about the area may be least and the probability of distortion or warpage might increase except along the dominant gradient or link path between adjacent nodes. Using conventional ideas from probabilistic market area analysis and the recent suggestions of Tobler (1976) concerning non-constant warpages and the development of a probability surface for distortions in "mental maps", one can visualize a mental map of the previous set of locations with the major distortions or warpings occurring away from the major nodes and the primary paths that connect them, and increasing in probability of occurrence in the more inaccessable or distant paths of each Thiessen polygon. The result would be to produce a map which associates points or key locations with error probability surfaces. The probability of warping would, therefore, be non-constant as direction changes from key nodes, or as distance increases from the key node to different edge segments of the Thiessen polygon.

The building of such a probability surface is simple in Euclidean space, but the specification of the surface parameters is somewhat more complicated even in a simple Riemann space of constant curvature. Detailed examination of over 200 individual configurations of 48 locations for the city of Columbus has shown that location errors differs considerably across the individual maps and that there is definitely both a directional and distance component to the distribution of errors (e.g., see Rivizzigno, 1976; Spector, 1978; Golledge and Spector, 1978; Gale, 1980). What is more, this error surface undulates depending on the

activity pattern of the individual, for as information about different segments of the environment increases, the probability of maintaining a constant location error diminishes. At various time periods, therefore, the manifold in which the points are located can be warped differentially. If one were to obtain a cross-section through time of a series of these manifolds, one should be able to recreate a history of the main repetitive components of an individual's spatial behavior for that time period.

Obviously, the first thing to be done is to attempt to define the appropriate parameters which describe the Riemann space in which a number of subjective configurations exist. Once this has proved to be a feasible operation, then expanding the work to cover manifolds produced at different stages of the environmental learning process and recreating histories of spatial behaviors associated with each manifold would seem to be an intriguing direction for constructive use of current work on mental maps.

An alternate way of envisioning the warped manifolds suggested in the previous paragraph is to introduce the concept of a mean information field with holes. Imagine a standard grid with a series of familiarity measures allocated to each grid cell. Imagine also that there are two "holes" in the mean information field where zero information is recorded, and two major peaks (one towards the S.W. corner of the field and another towards the N.W. corner). If one contoured the mean information field, the holes would stand out in the two dimensional Euclidean representation of it. However, if we collapsed the field so as

to eliminate the holes, the configuration that resulted would more readily be described in non-Euclidean terms. Such a warped field more closely approximates the sketch maps drawn by individuals with incomplete information about test environments; the consequent shortening of distances across places with low information levels and the exaggeration of distances where information is consistently high mirrors many of the types of distortions recovered from individual configurations of urban areas in other published research (Golledge and Rayner, 1975; Golledge, Rivizzigno, and Spector, 1976; Rivizzigno, 1976; Spector, 1978).

As a further step in an attempt to define the types of metrics most suitable for the representation of subjective configurations of places, current work at U.C. Santa Barbara is aimed at defining configurations of places in Riemann spaces such that an index of fit between subjective and objective configurations mapped onto the same space can be obtained. In general, it would appear that questions related to the suitability of representing mental maps in metric spaces need to be answered before much confidence can be placed in widespread use of such maps in conventional geographic work.

* References that are cited are given in the appropriate paper presented in the main section of this report.

Possible Future Research

Our experiences with proximity matrices of varying types leads us to suggest that considerable future research is warranted on developing methods to handle related proximity matrices (square and rectangular), asymmetric and/or skew-symmetric data matrices and incomplete data matrices. Each of these is a common occurrence when dealing with criminal justice data, and each presents a set of different methodological problems which need to be solved. For example, a general method for determining which of a set of outcomes from alternative analyses of the same data set best represent the structure of the original data matrix is needed. An example might be where two clustering algorithms are used on the same data set or where multidimensional scaling output is obtained in several dimensions. In several of our papers we suggested extensions of our methods to cover asymmetric proximity values - a topic that is currently being researched in many social and behavioral sciences. When dealing with large data sets consisting of subjective evaluations (preferences, choices, perceptions), it is frequently impractical or impossible to collect evaluations on all object combinations, and thus researchers are forced to use incomplete experimental designs in the data collection phase. The effect that differing levels of incompleteness have on aggregation processes, matrix matching procedures, homogeneity measures, and correlational type measures needs to be further explored.

Since subjective data is primarily non-metric, attention could also be directed towards examining non-metric equivalents of a variety of general matrix measures; possible alternatives for doing these are discussed in several of the papers presented in the main body of the report.

TABLE OF CONTENTS

PART III

Appendices

INTRODUCTION

INTRODUCTION

(Extracts from the original proposal)

Over the course of the present century the predominant philosophies of data analysis have taken a somewhat circular course. Early methodologists spent a great deal of time and effort attempting to understand empirically obtained information from a point of view that was more or less intuitive. Data were plotted, reorganized, and often simply inspected to determine their structure and inherent pattern. For example, in his recent address to the American Statistical Association, Box (1976) provided an interesting historical account of how many of Fisher's contributions to statistical theory arose from just such a pragmatic approach to data analysis. As the field of statistics matured, however, substantive researchers relied more heavily on hypothesis testing, with the result that data analysis was in part reduced to examining only the outcomes of significance tests. The widespread availability of the digital computer has further accelerated this trend, and in fact, today, data is often collected in a computer compatible medium that is appropriate for direct submission to some standard statistical package. As a consequence of automation, the researcher may have only a limited understanding of the raw data producing the significance tests and almost none of the practical intuition that at one time was considered absolutely essential in explaining any observed pattern within a given data set.

To be of maximum utility to researchers working in the crime and criminal justice areas, the methodological studies supported

by NIJ would ideally involve data analysis strategies that have a simple conceptual basis, maintain a close contact with the data, be easy to use, and be capable of providing meaningful ways of evaluating the variables or entities of interest to the law enforcement community. As will be indicated, one particular nonparametric statistical orientation has been developed within the past few years that appears to have these characteristics and that at the same time holds the promise of bringing data analysis back to its intuitive beginnings. After the introduction of several general problem areas, we will briefly review this class of nonparametric strategies and indicate how they could be used in attacking significant areas of concern in the fields of crime and criminal justice that involve the patterning of data, such as spatial and temporal variations in criminal activity and its environmental correlates. Since much of the necessary methodology is available in a more or less theoretical form in the psychology and geography literature, the major contribution of our research effort will be in terms of applications that have importance for the research program of NIJ.

The statistical techniques of interest have their origins in operations research, combinatorial mathematics, and graph theory, and for our purposes can be discussed very generally under the label of Generalized Concordance. Although the background necessary for this methodology is not new and much of its basic formulation dates back at least to the early 1950's, only the recent access to digital computers can allow its routine use in the analysis of data. Special cases of this orientation have been applied most widely in disciplines such as electrical

engineering (Younger, 1963), economics (Simpson and Tsukui, 1965), and management science (Blin and Whinston, 1975); and in fact, only within the past five years has it been recognized that these tools are appropriate for the analysis of behavioral science data as well. With this motivation in mind, the major goal of this research is to first introduce selected significant topics in criminal justice research (e.g., those concerned with data aggregation, spatio-temporal variations, changes in scale, and spatial autocorrelation) and then indicate in a heuristic manner what implications a theory of Generalized Concordance would have for these areas of analysis. The research program for which funding is requested would illustrate and develop exportable analytical procedures for examining these problems in the context of crime and criminal justice data sets.

Problem Areas

Although many different data analysis tasks could be approached with the class of strategies we envision, for the purpose of discussion, three areas will be identified and treated as illustrations of the type of problems we would wish to pursue. Since our intent is to offer demonstrations of a class of analysis schemes in the criminal justice area, our efforts are partially dependent on previously collected data. Consequently, the following discussion will offer some insights into our general thinking but it is also possible to modify our efforts if alternative and particularly important data sets are made available, e.g., those developed under that auspices of NIJ but

not generally referenced in the public domain.

Aggregation/Disaggregation

The aggregation problem is probably one of the most serious unsolved difficulties faced by data analysts in both the social and physical sciences. In its most general form it can be defined as the loss of information that occurs when data collected at one level are summarized or aggregated into larger units, or decomposed and disaggregated into smaller units. The effects of data aggregation in statistical analysis has recently been discussed in the geography literature (Clark and Avery, 1976; Smit, 1978; and Clark and Avery, 1978). They suggest that one of the most serious aggregation problems in the social sciences occurs in "aggregation bias" in correlation and regression analysis. This type of bias is said to be manifest as the inflation of macro-level coefficient estimates above the corresponding values of the coefficients estimated from micro-level data. Blalock (1971) has argued that it is incorrect to assume that relationships existing at one level of analysis will necessarily demonstrate the same strength at another level. Furthermore, estimates derived from aggregate data are valid only for the particular system of observational units employed. The consequences of using potentially biased estimates in correlation and regression procedures as substitutes for the true micro-level estimates are most serious when conducting causal inference analyses on statistical output. Against this background, one major problem of interest for us in the context of aggregation/disaggregation is to examine the structure of data in

collection units at various scales and at various time periods and determine at what levels of aggregation or disaggregation the patterns evidenced in those data sets breakdown or become different. From the past work of the authors, it appears that elements of such difference can be identified and studied using a strategy called "Generalized Concordance", and specifically, with those techniques that evaluate the degree of concordance within and between the subsets that define a partition of a set of proximity matrices. For example, each proximity matrix could represent the intercorrelations among a set of crime-related variables for a particular geographic entity; the partition would then represent a first level of possible aggregation. In short, the intent will be to determine if the information contained in data sets for various spatial units and/or for various years is consistent or inconsistent, and to assess the interaction between the spatial and temporal dimensions of data by analyzing changes in data patterns as expressed in various proximity values. The methods to be developed and/or applied will be appropriate for fully metric data sets (such as crime occurrence statistics) or for data sets that at best contain only ordinal information in the available proximities.

As indicated above, the generalized concordance procedures to be used in this research agenda appear appropriate in evaluating the structure of a given data set and whether this structure is repeatable as levels of aggregation or changes in scale occur. As part of this project we will take selected crime and criminal justice statistics that have been subject to various aggregation procedures and assess whether the aggregation process

destroys the inherent structure of the data existing at previous levels and superimposes a new and perhaps less meaningful structure on the data itself.

Although a variety of methods exist within the field for examining problems of aggregation (Duncan and Davis, 1953; Goodman, 1953, 1959; Blalock, 1964, 1971; Cliff and Ord, 1973; Curry, 1966; Hannan, 1971a, 1971b, 1972), most of these procedures focus on the grouping aspects of aggregation and discuss ways in which such procedures may effect causal relationships. The techniques most frequently discussed include random grouping; grouping to maximize the variation in the independent variable; grouping to maximize the variation in the dependent variable; and grouping on the basis of spatial proximity. The approach suggested in this proposal is different since we are concerned with the structure of data at various scales, between various areas, and over various temporal dimensions. Thus, in contributing to a deeper understanding of the aggregation problem itself, we will naturally complement the the existing work on spatial autocorrelation. Finally, we note the difference between the procedures suggested here that use proximity measures (but not necessarily direct spatial measures) to test for similarity of structure in data matrices and the work of Clark and Avery (1976) who used spatial measures of proximity in their discussion of the aggregation problem in spatial data.

Spatial and Temporal Structures

A second problem area is to develop and test methods for assessing concordance in complementary data sets at different

scales, for different time periods and in different spatial contexts. Data that are assembled by geographical units over successive time periods can be analyzed by searching for patterns that occur and re-occur at various spatial scales and with varying temporal frequencies, by first generating similarity or proximity measures of criminal activity and environmental and socio-economic variables, and secondly, searching for similarity of structure or pattern. The methods proposed can be used on metric and non-metric data, and are substantially different from other multivariate methods (such as principal components, factor analysis, discriminant analysis, and canonical correlation) currently used on crime statistics (see Carter, 1974; Nettler, 1974; and Pyle, 1974).

Spatial Autocorrelation and Generalized Concordance

Current treatments of spatial and temporal aspects of crime statistics have failed to handle the autocorrelation problem that is inherent in existing data sets. The Generalized Concordance procedures developed in this research will allow us to evaluate the extent of this problem in data sets and to illustrate procedures for handling this problem that are substantial generalizations of existing methods for estimating spatial autocorrelation effects (e.g., Cliff and Ord, 1973). Ideally, we would be able to disentangle temporal effects from data categorizations, such as those based on spatial properties, and vice versa, and provide answers to questions such as, "Is there more consistency in criminal activity of type Y between years than there is within areas of scale X?".

The current methodological literature in geography typically views that evaluation of spatial autocorrelation as distinct from many of the problems encountered in analyzing data defined in terms of proximities. In general, the hypothesis testing strategy used in the assessment of spatial autocorrelation is reserved for the relationship between geographical contiguity and a variable available in each of the given regions, e.g., variables such as unemployment, proportion of elderly, wealth, productivity, frequency of crime type, and so on. When interpreted appropriately, however, the exact same statistical principles provide a very powerful class of data analysis strategies for confirming the presence of structure within any matrix that contains numerical information among a set of geographical entities. The possibility of carrying out such an extension has been mentioned in the geographical literature (Hubert, 1978a), but the area has not been pursued to the depth that would identify the range of potential applications.

The link between the ideas of spatial autocorrelation and concordance can be extended to allow the researcher to move beyond static or cross-sectional data analysis into a dynamic or temporal analysis. This leads to the development and use of multitrait-multimethod analytical techniques (Hubert and Baker, 1978a) that attempt to find how information is "nested" in various data sets - e.g., if the temporal dimension is more significant than the spatial or if the local dimension is more significant than the regional. For example, a typical data structure that would involve multitrait-multimethod techniques is seen below (Fig. 1). In this problem, spatio-temporal

FIGURE 1

		Geog. Area A					Geog. Area B					Geog. Area C				
		Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5
Geog. Area A	Yr 1															
	Yr 2															
	Yr 3															
	Yr 4															
	Yr 5															
Geog. Area B	Yr 1															
	Yr 2															
	Yr 3															
	Yr 4															
	Yr 5															
Geog. Area C	Yr 1															
	Yr 2															
	Yr 3															
	Yr 4															
	Yr 5															

confounding in the data can be unpacked and the significant data structure exposed (Hubert and Baker, 1978a).

Methodology

One of the major practical difficulties faced by all behavioral scientists in analyzing data concerns the choice of formal techniques that are intended to be of aid in developing reasonable substantive interpretations. Much of the time the final selection of a statistical tool is guided either by tradition in the researcher's field, or at the other extreme, because one particular procedure happens to be in vogue. In either case, the chosen methodology may not be the most appropriate way to answer the specific questions posed. The difficulty of choosing a statistical tool becomes even more acute when a research problem cannot be easily rephrased within an omnipotent general linear model since there are few alternative paradigms that are broad enough to formulate sufficiently powerful analyses. Consequently, because of the general inflexibility of statistical schemes that do not rely on rather strong parametric assumptions, novice researchers tend to limit the questions they ask to those that fit neatly within the analysis of variance context and its derivations, or alternatively, embrace some other familiar strategy that may not be suitable for the particular application at hand.

The brief overview to follow sketches one special case of a more comprehensive strategy called Generalized Concordance. This special case will be referred to as the quadratic assignment paradigm (QA) and incorporates a variety of disparate data structures that may be reflected in a proximity matrix defined on

the objects from some set S. The term "proximity" is used generically and merely denotes a numerical measure of relationship between pairs of objects from S. Obviously, there will be many instances in which the tasks faced by an applied scientist in analyzing a proximity matrix do not fall into one of the categories that can be handled by the type of strategy illustrated below; nevertheless, the quadratic assignment approach still appears flexible enough to give a general point of departure for many of the problems an individual faces in choosing an appropriate methodology, and, more importantly, is broad enough to provide a general organizing principle for an extensive theoretical analysis of structure within a proximity matrix. In addition, QA forms the basis for the more general strategy of Generalized Concordance discussed in Hubert (1979c).

Some Overview Details of the Quadratic Assignment Approach

The general quadratic assignment approach to data analysis can be formulated in a relatively easy manner. It is assumed that the collected data are on n objects from a set S that, for convenience, is denoted by $\{O_1, O_2, \ldots, O_n\}$. The term "object" is intended to be extremely general and could refer to individuals, areas, societies, crime types, and so on. Furthermore, it is assumed that the data on these n objects can be reduced to a single numerical proximity value defined for each ordered object pair. For instance, if the objects are individuals, the numerical value could be an index obtained by measuring similarity over a set of behavioral symptoms, or possibly the objects could be cities and the numerical values could be

measures of similarity between profiles of crime types. To formalize this concept in more detail, a data matrix $\underset{\sim}{Q}$ is defined to be an nxn matrix, where both the u^{th} row and the u^{th} column refer to object O_u, and the entry in row u and column v is denoted by $q(O_u, O_v)$. Typically, it is assumed that $q(O_u, O_u) = 0$ for $1 \leq u \leq n$.

In addition to the data matrix Q, a structure matrix C is specified that represents the type of hypothesis the researcher wishes to evaluate against his data or, alternatively, the type of structure he may wish to identify in his data. These two aims represent what can be referred to as the confirmatory and exploratory data analysis problems, respectively. The rows and columns of $\underset{\sim}{C}$ are labeled by the integers 1, 2 ..., n, and $c(r, s)$ is the entry in row r and column s of C; typically, $c(r, r) = 0$ for $1 \leq r \leq n$.

To connect these two matrices $\underset{\sim}{Q}$ and $\underset{\sim}{C}$ in a more formal way, suppose ρ is an arbitrary permutation of the integers 1, 2, ..., n and ρ_I the identity permutation that maps each integer back to itself. If we define

$$\Gamma(\rho) = \sum_{r,s} q\left(O_{\rho(r)}, O_{\rho(s)}\right) c(r,s),$$

then $\Gamma_{(\rho_I)}$ is merely the sum of products of the corresponding elements between $\underset{\sim}{Q}$ and $\underset{\sim}{C}$. More generally, $\Gamma(\rho_I)$ is the sum of the products of the corresponding elements between $\underset{\sim}{C}$ and $\underset{\sim}{Q}_\rho = \left\{q(O_{\rho(r)}, O_{\rho(s)})\right\}$, where the u^{th} row and column of $\underset{\sim}{Q}_\rho$ is the row and column previously labeled $\rho(u)$ in $\underset{\sim}{Q}$.

Given this notation, the exploratory and confirmatory data analysis problems can be rephrased. In particular, the

confirmatory problem is based on $\Gamma(\rho_I)$ and the researcher wishes to determine whether the given $\underset{\sim}{Q}$ matrix mirrors the structure in $\underset{\sim}{C}$. Since $\Gamma(\rho_I)$ is treated as an unnormalized correlation between $\underset{\sim}{C}$ and $\underset{\sim}{Q}$, the additional problem of evaluating the relative size of $\Gamma(\rho_I)$ by some significance test has to be addressed. As mentioned below, this assessment is approached by considering all possible indices $\Gamma(\rho_I)$ that could be obtained by varying the permutation ρ. Alternatively, the confirmatory task merely wishes to identify those permutations that optimize $\Gamma(\rho_I)$.

Introductory Comments on Confirmatory and Explanatory Inference

In the literature on data analysis over the last twenty years, a distinction between exploratory and confirmatory procedures has become very popular (see Kaiser, 1970; Hildebrand, Laing and Rosenthal, 1977; Tukey, 1962). An exploratory strategy typically involves the use of an analysis technique on a given data set with the aim of identifying interesting relationships, patterns, and the like (Brown, Odland and Golledge, 1970; Lee and Egan, 1972; Phillips, 1973; Harries, 1974; Pyle, 1974). Alternatively, a confirmatory approach requires the test of an *a priori* conjecture that is generated from a source distinct from the data to be used for the purpose of validation. This latter test in our context will be correlational, and thus the term "confirmation" is given a limited meaning here that does not imply the absolute correctness of a hypothesis. Since a correlational analysis can never exclude all competing explanations, we will argue when it is justified that the pattern of data is not unrelated to the conjectured pattern.

It may be obvious that confirmatory analyses would be desirable adjuncts to many of the current exploratory methods used in the study of proximity matrices (such as clustering and multidimensional scaling), but very few techniques have been proposed that could help carry out such a program with any degree of rigor. Thus, users of the newer data reduction procedures lack confirmatory techniques even of a correlational nature and must rely on intuitive arguments based on whatever additional information is available for the objects being studied. Although this practice is commendable given the current state-of-the-art, it is now possible to proceed one step further using the correlational methods presented in this proposal and incorporate the same information relevant to a post-hoc explanation more directly in a confirmatory manner.

EXAMPLE 1:

The field of criminal justice provides a very interesting application that can be used to introduce some of the necessary concepts of the confirmatory QA approach. For example, Pyle (1974) and Harries (1974) in their studies of the connections between crime and socio-spatial characteristics of urban areas suggested that sets of supplementary variables can help explain similarities of crime patterns among subareas of different environments. Using a given variable (such as income level or the predominant ethnicities in the subareas) it is possible to divide the subareas into sets, where the objects within each set contain the same value of the variable being considered. Here

the inference problem of interest is whether the data produced by the experiment reflects the partitioning of the areas according to the given variable. As always, we can assume that some measure of proximity is available between areas based, for example, on a summary measure of crime similarity.

To illustrate the conceptualizaiton with a more detailed discussion, suppose we are given a set of 100 cities and have obtained some type of proximity between each pair based on similarity of committed crimes. These measures are contained in the matrix $\underset{\sim}{Q}$. Furthermore, assume the 100 cities can be partitioned into K subsets based on the type of law enforcement program in operation or any other group of variables that would serve to characterize criminal justice systems. It could be conjectured a priori that cities within a subset might be very similar in the pattern of crime, whereas cities in different subsets would demonstrate a reduced level of correspondence. As a mechanism to embody this conjecture within QA, the second nxn matrix $\underset{\sim}{C}$ is used to specify the hypothesized organization of the empirical proximity matrix. For example, since the partition under consideration groups the 100 cities into K subsets, the 100 x 100 structure matrix is divided into K submatrices each of a size that corresponds to the number of cities in the subset. From our conjecture, cities within a subset are likely to be similar, and thus, all elements, $c(O_u, O_v)$, in the on-diagonal submatrices are set equal to unity, except that by convention the cell elements actually on the main diagonal are set to zero. Furthermore, it is hypothesized that a correspondence between cities in different subsets is unlikely, and consequently, the

cells of the off-diagonal sub-matrices are set to zero. In other words, distinct city pairs that should be similar are assigned a 1 and all other pairs are assigned a 0 which implies that $\Gamma(\rho_I)$ is merely the sum of all within subset proximities. Obviously, other homogeneity indices could be defined by varying the structure matrix $\underset{\sim}{C}$.

Given a data matrix $\underset{\sim}{Q}$ and a structure matrix $\underset{\sim}{C}$ as presented in this example, the confirmatory problem is one of comparing $\underset{\sim}{Q}$ and $\underset{\sim}{C}$ and assessing whether the pattern represented by $\underset{\sim}{C}$ is also present in $\underset{\sim}{Q}$ (or conversely). In particular, if the structure defined by the matrix $\underset{\sim}{C}$ is not reflected in $\underset{\sim}{Q}$ then the value of $\Gamma(\rho_I)$ should not be unusually large (or small) compared to the distribution we would expect if all labelings of the rows and corresponding columns of Q were equally likely. The index $\Gamma(\rho_I)$ is evaluated for each permutation ρ and the frequency table constructed for all n! (possibly nondistinct) values of Γ, generating what is typically called a permutation distribution. The statistic $\Gamma(\rho_I)$ is then compared to this distribution and if $\Gamma(\rho_I)$ is at a suitably extreme percentage point, the hypothesis of an equally likely a priori labeling is rejected in favor of the structure defined by $\underset{\sim}{C}$. Typically, the actual permutation distribution is too computationally laborious to obtain each time a new data matrix is obtained; however, the mean and variance parameters needed for approximate tests can be obtained by formula, or alternatively, approximate permutation tests can be constructed from random samples of the complete permutation distribution. This latter option is illustrated in many of the published papers listed in a later section.

Exploratory Analyses and the Quadratic Assignment Approach

Instead of attempting to confirm whether a given partition (such as that used in assessing similarity in crime type by area) can be used to explain the patterning of the proximities, suppose our concern is to locate "good" partitions in an exploratory or post hoc fashion. More specifically, assume our interest is in finding possibly "good" partitions that have the same general structure as the conjectured partition in our example. As one possible approach, suppose the form of the structure matrix is fixed as is, i.e., representing subsets of size $n_1, \ldots, n_K$, but we attempt to rearrange the rows and simultaneously the columns of the proximity matrix in such a way that the patterning of entries in the rearranged $\underset{\sim}{Q}$ matrix is similar to the fixed $\underset{\sim}{C}$ matrix. Once such a reorganization is effected, the four objects that are now in the first n_1 rows (and, also, in the first n_1 columns) would represent one class, the second n_2 rows and columns would define a second class, and so on.

The rather loosely defined goal of reorganizing the $\underset{\sim}{Q}$ matrix until it "fits" the form specified by the $\underset{\sim}{C}$ matrix can be made more precise by defining two major subtasks. First of all, the correspondence between the $\underset{\sim}{C}$ matrix and a reorganized $\underset{\sim}{Q}$ matrix must be measured in some way to determine if an adequate "fit" has been achieved. Although a number of measures of correspondence are available, the simple sum of the products of corresponding elements in the two matrices used in the confirmatory context appears to be one of the most natural indices to consider. As a second subtask, it is necessary to

define some procedure that can be used to rearrange the $\underset{\sim}{Q}$ matrix. For example, the ordering of the corresponding rows and columns of the $\underset{\sim}{Q}$ matrix could be modified by interchanging the positions of those two objects that will increase (or decrease) the index Γ the most. The pairwise interchange process is repeated until no pairwise interchange can increase (or decrease) the value of Γ; that is, until a local optimum has been achieved.

The basic features of an exploratory Quadratic Assignment approach to data analysis are evident in the example presented above. The $\underset{\sim}{Q}$ matrix contains a measure of relationship (or proximity) between the pairs of objects under study and serves as the empirical information to be analyzed. The $\underset{\sim}{C}$ matrix, on the other hand, specifies the structure that the researcher assumes the observed proximity measures to have if only the $\underset{\sim}{Q}$ matrix could be reorganized appropriately. Finally, a combinatorial optimization heuristic is used to reorder the rows and corresponding columns of the $\underset{\sim}{Q}$ matrix until a high degree of correspondence with the structure matrix is obtained. Since the end result of the exploratory mode of the QA approach is simply a final ordering of the objects, it is up to the researcher to develop a substantive interpretation for the obtained reorganization of the proximity matrix.

PART I

Methodologies for Examining Spatial and Temporal Patterns of Crime Data

78710

AGGREGATION IN DATA TABLES : IMPLICATIONS FOR EVALUATING CRIMINAL JUSTICE STATISTICS*

by

L.J. Hubert, R.G. Golledge,
C.M. Costanzo, and T. Kenney**

* This project is funded by LEAA/NILCJ, Grant #79-NI-AX-0057 and NSF SOC-77-28227.

** L.J. Hubert and R.G. Golledge are Professors of Education and Geography (respectively) at the University of California at Santa Barbara.

C.M. Costanzo and T. Kenney are graduate students in Geography and Education (respectively) at the University of California at Santa Barbara.

Aggregation in Data Tables: Implications for Evaluating Criminal Justice Statistics

Introduction

One common form of data encountered in the criminal justice area can be represented schematically as an object-by-attribute matrix or table:

Attribute (e.g., crime type)

Object (e.g., SMSA/city/neighborhood)	1	2	3	...	n
1	x_{11}	x_{12}	x_{13}	...	x_{1n}
2	x_{21}	x_{22}	x_{23}	...	x_{2n}
3	x_{31}	x_{32}	x_{33}	...	x_{3n}
⋮	⋮	⋮	⋮		⋮
I	x_{I1}	x_{I2}	x_{I3}	...	x_{In}

Here, for example, the term "object" may refer to a city, prison, or an individual and the term "attribute" to a time point, attitudinal variable, or experimental condition. The symbol x_{ij} denotes

a numerical value that attribute j attaches to object i where $1 \leq i \leq I$ and $1 \leq j \leq n$. If the objects were cities and the attributes were time points, x_{ij} may give the rate of homicide for city i at time j.

When faced with data organized in such a matrix, a reasonably sophisticated researcher may determine whether there are differences among the n attribute levels (i.e., among the n columns) by using a repeated measures analysis of variance procedure (see Morrison, 1976, pp. 141-153). Although our presentation deals at least initially with data in exactly this same form, our concerns are more detailed than performing an omnibus statistical test. Specifically, the discussion below develops some basic principles for comparing the data in an object by attribute data matrix to a pattern conjectured to hold within each of the rows. Usually, we wish to confirm that a conjectured pattern is present, and consequently, that some *a priori* notion of structure is supported.

We start with a simple example involving real data to illustrate a counterintuitive result due to aggregating over the rows of an object by attribute matrix. This first example serves the immediate purpose of motivating a formal investigation of several analysis problems relating to aggregation that can be phrased for the type of data matrix represented above.

Example

The data given in Table 1 provide the rates of homicide (per 100,000) over a ten year period (1968 to 1977) for the fifteen largest Standard Metropolitan Statistical Areas (SMSA's) whose geographical regions have remained unchanged throughout this period (see Appendix I). By inspecting the entries for each SMSA, it would appear that a general increase is present in the homicide rate from 1968 to 1977, except possibly for the last two or three years. In fact, since the trend is strong even in the presence of the obvious inconsistencies to a perfect pattern, we may be content with a simple visual inspection and interpretation. In more ambiguous contexts, however, it would be of value to have a formal inference strategy for confirming whether an *a priori* conjecture of an increase in rate is reasonable, or more generally, for confirming any conjecture that is based on a source independent of the data itself.

Table 1 here

Given the general problem of pattern confirmation, two more or less obvious evaluation methods could be followed:

Preliminary data aggregation: If the data of Table 1 were aggregated over rows to produce a single homicide value for each year, these summary values could then be compared, say, to an expected pattern of increase. Intuitively, if a general trend exists within each row, it would also be apparent and possibly enhanced in the aggregation. As an illustration of the mechanics of such a strategy, suppose the rates

Table 1(a)

Homicide Rates (per 100,000) for the Fifteen Largest SMSA's where Geographical Regions have Remained Unchanged Throughout the Period

1968-1977

	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977
Chicago, Ill.	10.7	11.6	12.9	13.1	11.5	14.1	15.9	13.9	13.4	13.7
Los Angeles - Long Beach	8.6	9.7	9.4	10.7	12.8	12.4	12.9	14.3	13.8	16.0
Philadelphia	6.7	7.5	9.3	10.9	10.7	11.5	11.9	12.0	10.0	8.9
San Francisco - Oakland	7.7	9.5	8.3	9.4	8.6	10.9	11.6	12.4	12.2	11.9
Boston, Mass.	4.4	4.2	4.4	4.8	4.6	5.7	5.6	5.3	3.8	3.6
Pittsburgh, Pa.	2.9	3.2	4.4	4.1	3.9	4.2	5.3	5.6	4.8	4.8
Baltimore, Md.	13.6	13.4	13.2	17.3	17.6	15.4	16.8	14.8	11.0	10.2
Cleveland, Ohio	9.6	13.8	14.5	14.8	16.1	15.1	17.2	17.0	15.1	14.2
Anaheim-Santa Ana-Garden Grove	2.3	2.5	2.7	2.5	4.7	3.0	3.7	3.8	4.7	4.0
San Diego, Ca.	3.8	4.2	4.1	5.1	3.8	6.0	7.4	6.7	6.3	6.8
Miami, Florida	12.5	12.8	15.6	17.1	14.3	15.7	17.4	18.3	13.6	15.6
Milwaukee, Wis.	3.9	3.3	3.8	4.3	4.3	5.0	5.1	5.2	4.8	4.5
Seattle-Everett	4.6	5.4	4.4	4.5	4.5	4.6	6.4	5.7	4.4	4.3
Cincinnati	5.0	7.3	6.4	8.4	7.5	7.0	7.8	6.4	6.6	7.5
Buffalo, N.Y.	4.1	4.4	5.7	6.4	6.6	5.8	6.0	6.1	5.0	6.0

(Source: Federal Bureau of Investigation, Crime in the United States, Uniform Crime Reports, Washington, D.C., Govt. Printing Office, 1968-1977).

Table 1(b)

Rank Orders of

Homicide Rates (per 100,000) for the Fifteen Largest SMSA's where Geographical Regions have Remained Unchanged Throughout the Period 1968-1977

	Year									
	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977
Chicago,Ill.	1	3	4	5	2	9	10	8	6	7
Los Angeles - Long Beach	1	3	2	4	6	5	7	9	8	10
Philadelphia	1	2	4	7	6	8	9	10	5	3
San Francisco - Oakland	1	5	2	4	3	6	7	10	9	8
Boston,Mass.	5	3	4	7	6	10	9	8	2	1
Pittsburgh,Pa.	1	2	6	4	3	5	9	10	8	7
Baltimore,Md.	5	4	3	9	10	7	8	6	2	1
Cleveland,Ohio	1	2	4	5	8	7	10	9	6	3
Anaheim-Santa Ana-Garden Grove	1	3	4	2	10	5	6	7	9	8
San Diego,Ca.	2	4	3	5	1	6	10	8	7	9
Miami, Fla.	1	2	6	8	4	7	9	10	3	5
Milwaukee,Wis.	3	1	2	5	4	8	9	10	7	6
Seattle-Everett	7	8	3	5	4	6	10	9	2	1
Cincinnati	1	6	3	10	8	5	9	2	4	7
Buffalo, N.Y.	1	2	4	9	10	5	7	8	3	6

a) Column Sums	32	50	54	89	85	99	129	124	81	82
b) Col-Sums Ranks	1	2	3	7	6	8	10	9	4	5
c) Perfect Temporal Ranking	1	2	3	4	5	6	7	8	9	10

Correlation (b) and (c) = .552 = r_s (Spearman rho)

Z-statistic = 1.65 = $\sqrt{n-1}\ r_s$

within each row are first ranked from 1 to n according to their size - ties are broken using the conservative procedure of assigning ranks contrary to the expected temporal increase. The rank sums within each column are then ranked from 1 to n according to their size and correlated with the integer pattern 1, 2, ..., n representing a perfect temporal increase. This ordering process on column sums produces what may be called a "consensus ranking" (e.g., see Kendall, 1970, pp. 101-102); we are essentially using Spearman's rank order correlation r_s to measure the degree of correspondence to our conjecture. (It should be noted that we rely on the notion of a perfect temporal increase only as an illustration and as a conjecture that someone may wish to test. We are making no statement about the truth of this conjecture and, in fact, the data themselves may suggest that a non-linear relationship is more appropriate.)

Based on well-known formulas, the significance of the rank order correlation r_s can be assessed in the usual way. For example, under the independence hypothesis that all n! permutations of the first n integers are equally likely to be the consensus ranking, the expectation of r_s is 0 and its variance is 1/(n-1). Thus, a Z-statistic would be defined very simply as $\sqrt{n-1}\ r_s$. In the example, the column sums are 32,50,54,89,85,99,129,124, 81,82, producing the consensus ranking of 1,2,3,7,6,8,10,9,4,5 and a correlation with the perfect pattern of .552. Assuming the adequacy of a normal approximation, the associated Z-statistic of 1.65 would be declared significant, but just barely. Although there is an apparent decrease in the column sums for 1975 to 1977, the rather strong upward trend in the earlier years still produces a rather substantial correlation to the pattern of a strict temporal increase. Obviously, if we had an a priori reason to conjecture

the curvilinear trend that seems to be reflected in the column sums, this pattern could have been used to define the criterion instead.

Correlational aggregation: Instead of aggregating over cities to obtain a single value for each year, suppose the rows of Table 1 are considered separately. In particular, the entries (i.e., ranks) within each row are first compared (i.e., correlated) with the expected pattern of increase; the aggregation is then carried out over the I correlations to provide a single average measure of correspondence. It is apparent that aggregation now forms the last step rather than the first.

As a simple numerical example using the data of Table 1, 15 correlations would be generated: .71,.96,.47,.87,-.08,.81,-.25, .51,.77,.78,.42,.76,-.30,.17,.36, giving an average value, $r_{s(ave)}$, of .465. If it is assumed that all permutations of the integers within each row are equally likely, the expectation of $r_{s(ave)}$ is 0 and its variance is $1/I(n-1)$. Thus, the Z-statistic, $\sqrt{I(n-1)}\, r_{s(ave)}$, would be 5.40, giving a much smaller (i.e., better) significance level than the Z-statistic generated under preliminary data aggregation. Counter-intuitively, however, $r_{s(ave)}$ is smaller than r_s even though the former is more "significant".

The stage at which aggregation takes place appears to affect dramatically the significance of the final summary statistic as well as its size, and unfortunately, in opposite ways. Extrapolating from the simple example of Table 1, preliminary data aggregation will lead to a larger summary correlational measure. However, this larger correlation will generally be less significant when compared to the

average of the separate correlations for each of the rows.

We now formalize these general conclusions in greater detail and extend their range of validity beyond the context of simple rank orders and the use of Spearman's correlation. Our extensions later will include the possibility of comparing matrices rather than sequences and will address the problem of multi-group concordance as well. In this latter instance, the pattern that is used as a target (e.g., a pattern analogous to the conjecture of homicide increase) is only implicitly given by a second object by attribute data table (or set of matrices) collected on a second group of objects.

As one final motivating comment, we note that the practical implications of a more formal analysis are important for the way in which crime data are reported and kept. The aggregation level at which the strongest correlational pattern can be found may be at odds with the level of aggregation that would allow the pattern to be most easily detected statistically. Thus, by pragmatically picking an aggregation level that seems to "smooth out" the data the best, we also may limit the types of relationships that could later be determined as statistically relevant, particularly if researchers have only the secondary summaries at hand. These same relationships are well-known in multi-factor analysis of variance. Once data are collapsed (aggregated) over the levels of a factor, the associated sum of squares becomes part of a usually larger error term needed in tests of significance; consequently, these latter tests tend to be less sensitive.

Some Formal Details

The two schemes for aggregation presented in the last section can

be charaterized more formally. For notation we assume the object by attribute data table of the form $\{x_{ij}\}$ given in the introduction and let the predicted criterion (or pattern) over the attributes be defined by the numerical sequence $y_1, y_2, \ldots y_n$. Also, for algebraic convenience no secondary transformations to ranks are used; instead, to produce comparability across rows and to allow each object to have an equal weight, the entries within each row are standardized to z-scores, i.e., $\sum_k x_{ik} = 0$ and $\sum_k x_{ik}^2 = n$ for $1 \leq i \leq I$. This canonical form is assumed thoughout the paper.

Preliminary data aggregation: Based on the n column sums, $\sum_i x_{ij}$ for $1 \leq j \leq n$, let r_A be the correlation between these sums and the criterion $y_1, y_2, \ldots, y_n$. Under the standard permutation model of independence (see Bradley, 1968, pp. 73-76), all n! orderings of the column sums are equally likely. Consequently, this assumption produces a "null" model in which r_A has expectation 0 and variance $1/(n-1)$.

Correlational aggregation: If we denote the correlation between the n elements in the ith row of the object by attribute data table and the criterion by r_i, then the average correlation over the I rows is merely $r_B = \frac{1}{I}\sum_i r_i$. Based on the standard Friedman randomness model (see Bradley, 1968, pp. 123-129), all orderings of the entries within each row are equally likely. This assumption generates $(I!)^n$ equally likely realizations of the whole table. As for r_A, the expectation of r_B is 0; however, the variance of r_B is now $\frac{1}{I(n-1)}$.

Using these two aggregation schemes, the values of r_A and r_B and the associated Z-statistics, $Z_A = \sqrt{n-1}\, r_A$ and $Z_B = \sqrt{I(n-1)}\, r_B$ can be compared through simple algebra. For example, as shown in

Appendix II,

$$r_A = r_B/\theta, \qquad (1)$$

where $\theta = \sqrt{\frac{1 + (I-1)r_{ave}}{I}}$ and r_{ave} is the average of all $\binom{I}{2}$ correlations between pairs of rows. Since r_{ave} is less than or equal to 1, θ must also be less than or equal to 1. Moreover, θ can be 0 if and only if $r_{ave} = \frac{-1}{I-1}$, which implies a lack of variance in the column sums (i.e., the sums are all zero) and an undefined correlation r_A. Thus, without loss of any essential generality, it is assumed that θ is positive and less than or equal to 1, i.e., $0 < \theta \leq 1$.

The implications of (1) are somewhat surprising but are consistent with what is generally known about aggregation phenomena. First of all, $|r_A|$ is always greater than or equal to $|r_B|$ with equality only when $r_{ave} = 1$, i.e., when all rows are identical in their standardized scores. Consequently, any non-trivial variability across rows will result in $|r_A|$ being greater than $|r_B|$. Secondly, as $I \to \infty$,

$$\theta \simeq \sqrt{r_{ave}}$$

and

$$r_A \simeq r_B / \sqrt{r_{ave}}.$$

This relationship is analogous to a correction for attenuation in the psychometrics literature (see Lord & Novick, 1968, pp. 69-74). Here, r_A can be loosely interpreted as a correlation between a "true score" (defined from column data) and an "infallible variate" (defined by the criterion); r_B is the correlation between an imperfect measure of the "true score" (defined by an average over the separate row data) and an infallible variate (again defined by the criterion).

In terms of Z-statistics, Appendix II proves the relation given in (2):

$$Z_B/Z_A = \sqrt{1 + (I-1)r_{ave}} \qquad (2)$$

Thus, if $r_{ave} \geq 0$,

$$|Z_B| \geq |Z_A| .$$

In words, if there is some positive degree of correspondence among the rows, i.e., $r_{ave} > 0$, the Z-statistic for r_A will be less extreme than the Z-statistic for r_B. Equality exists only if $r_{ave} = 0$; at the other extreme, if $r_{ave} = 1$, $Z_B = \sqrt{I}\, Z_A$.

In summary, the greater the internal correspondence as measured by r_{ave}, the greater the discrepancy between the Z-statistics and the closer r_A and r_B become. This can be seen in the simple equality

$$\frac{Z_B}{Z_A} = \sqrt{I}\,\frac{r_B}{r_A} ,$$

suggesting that as r_B and r_A get closer together, Z_B and Z_A get more

discrepant. We also note that since $-1 \leq r_A \leq +1$,

$$-\sqrt{\frac{1 + (I-1)r_{ave}}{I}} \leq r_B \leq \sqrt{\frac{1 + (I-1)r_{ave}}{I}} .$$

This last expression provide simple bounds on the average correlation r_B in terms of the degree of internal concordance. The less the internal concordance or correspondence, the tighter the bounds.

Returning to the simple data of Table 1, the various relationships developed above can be verified numerically. Table 2 is an analogue of Table 1 and provides the z-scores necessary for the calculations given below. First of all, since $r_{ave} = .479$, $\theta = .717$. Thus, $r_A = .727 = r_B/\theta = .521/.717$. Similarly, $Z_B/Z_A = 2.776$, which is equal to $\sqrt{1+(I-1)r_{ave}}$, or alternatively, to $\sqrt{I}\,\frac{r_B}{r_A}$. Finally, if we assume r_{ave} is given as .479, the possible algebraic bounds on r_B are from -.717 to +.717. Thus, the actual size of .521 for r_B is rather substantial given the degree of observed concordance among the SMSA's.

Table 2 here

Multiple-Group Concordance

A fixed set of scores $y_1, y_2, \ldots, y_n$ was assumed to define the hypothesized pattern or criterion in the last section. Suppose now that no such static conjecture is available and instead we are given a split of the I objects into a first set of I_1 and a second set of I_2, where $I_1 + I_2 = I$ (e.g., a split into Eastern and Western cities).

Table 2

z-scores of Homicide Rates (per 100,000) Standardized within Rows

	Year									
	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977
Chicago	-1.66	-1.03	-0.13	0.01	-1.10	0.71	1.97	0.57	0.22	0.43
Los Angeles	-1.53	-1.04	-1.17	-0.60	0.33	0.15	0.37	0.99	0.77	1.74
Philadelphia	-1.87	-1.41	-0.37	0.55	0.44	0.90	1.13	1.19	0.03	-0.60
San Francisco	-1.53	-0.45	-1.17	-0.51	-0.99	0.39	0.81	1.29	1.17	0.99
Boston	-0.35	-0.65	-0.35	0.24	-0.06	1.56	1.41	0.97	-1.24	-1.53
Pittsburgh	-1.75	-1.38	0.10	-0.27	-0.52	-0.15	1.21	1.58	0.59	0.59
Baltimore	-0.30	-0.39	-0.47	1.24	1.36	0.45	1.03	0.20	-1.39	-1.72
Cleveland	-2.54	-0.46	-0.12	0.03	0.67	0.18	1.22	1.12	0.18	-0.27
Santa Ana	-1.26	-1.03	-0.80	-1.03	1.52	-0.45	0.36	0.47	1.52	0.71
San Diego	-1.24	-0.93	-1.01	-0.24	-1.24	0.44	1.51	0.98	0.67	1.05
Miami	-1.49	-1.33	0.17	0.97	-0.53	0.22	1.13	1.61	-0.90	0.17
Milwaukee	-0.88	-1.89	-1.05	-0.20	-0.20	0.98	1.15	1.32	0.64	0.14
Seattle	-0.42	0.78	-0.72	-0.57	-0.57	-0.42	2.27	1.22	-0.72	-0.86
Cincinnati	-2.22	0.35	-0.66	1.57	0.57	0.01	0.90	-0.66	-0.43	0.57
Buffalo	-1.90	-1.52	0.11	0.99	1.25	0.24	0.49	0.62	-0.77	0.49
Column Sums	-20.94	-12.38	-7.64	+2.16	+0.63	+5.23	+16.96	+13.47	+0.24	+1.90
Criterion: z-scores	-1.57	-1.22	-0.87	-0.52	-0.17	0.17	0.52	0.87	1.22	1.57

$r_A = .727$

$r_B = .521$

$$\frac{Z_B}{Z_A} = \frac{\sqrt{I(n-1)}\ r_B}{\sqrt{n-1}\ r_A} = \frac{\sqrt{15(9)}\ .521}{\sqrt{(9)}\ .727} = 2.776$$

Although each group could be discussed separately, our more immediate concern is with measuring the concordance between the two given groups. Intuitively, the criterion scores $y_1, y_2, \ldots, y_n$ are replaced by a set of rows from one of the two groups. The first group then serves as a target for the second and conversely (see Schucany and Frawley, 1973).

Assuming the common standardization to mean 0 and variance 1 for each row, the two aggregation schemes discussed previously have natural analogues in the two-group context. First of all, the data in each group could be aggregated and the two sequences of scores correlated to obtain r_A^*. Alternatively, each row in group 1 could be correlated with each row in group 2 and the $I_1 I_2$ correlations averaged to give r_B^*. Obvious extension of the proofs given in Appendix II for r_A and r_B would generate rather simple relationships among r_A^* and r_B^*. For example, $r_A^* = r_B^*/\theta_1\theta_2$, where $\theta_1 = \sqrt{\frac{1 + (I_1-1)r_{ave1}}{I_1}}$, $\theta_2 = \sqrt{\frac{1 + (I_2-1)r_{ave2}}{I_2}}$, and r_{ave1}, and r_{ave2} are the average intercorrelations within each of the two groups. Thus, $|r_A^*| \geq |r_B^*|$.

A loose correspondence to reliability theory is again possible. If $I_1 \to \infty$ and $I_2 \to \infty$, $r_A^* \simeq r_B^*/\sqrt{r_{ave1}\ r_{ave2}}$, implying that r_A^* is an analogue of a disattenuated correlation. In other words, to generate r_A^*, we merely correct r_B^* for the lack of perfect concordance within each of the two groups separately.

Using these same ideas, we find $\frac{Z_B^*}{Z_A^*} = \sqrt{1 + (I_1-1)r_{ave1}}\sqrt{1+(I_2-1)r_{ave2}}$.

Thus, if $r_{ave1} \geq 0$ and $r_{ave2} \geq 0$, $|Z_B^*| \geq |Z_A^*|$. Finally, r_B^* can be bounded by $\pm\sqrt{\frac{1 + (I_1-1)r_{ave1}}{I_1}}\ \sqrt{\frac{1 + (I_2-1)r_{ave2}}{I_2}}$.

Returning to our numerical example, suppose we split our 15 SMSA's into two groups - (Chicago, Philadelphia, Boston, Pittsburgh, Baltimore, Cleveland, Miami, Milwaukee, Cincinnati, Buffalo) and (Los Angeles-Long Beach, San Francisco-Oakland, Anaheim-Santa Ana-Garden Grove, San Diego, Seattle-Everett) on the basis of an east-west dichotomy. Here, $I_1 = 10$, $I_2 = 5$, $r_{ave1} = .558$ and $r_{ave2} = .519$. Thus, $\theta_1 = .776$ and $\theta_2 = .784$. The relationships among the correlations r_A^* and r_B^* and the Z-scores Z_A^* and Z_B^* are now immediate. For example, $r_A^* = .657$ $= r_B^*/\theta_1\theta_2$, $\frac{Z_B^*}{Z_A^*} = \frac{8.484}{1.971} = \sqrt{1 + (I_1-1)r_{ave1}}\ \sqrt{1 + (I_2-1)r_{ave2}} = 4.304$. Finally, r_B^* is bounded by $\pm\ \theta_1\theta_2$ or $\pm\ .608$. Thus, the actual value of .399 for r_B^* is substantial given the algebraic upper bound. As an interesting substantive observation, r_{ave1} and r_{ave2} are both greater than r_{ave} found in the previous section, and moreover, the values of r_A^* and r_B^* are both smaller than their counterparts r_A and r_B. Descriptively, there appears to be a greater correspondence between all the SMSA profiles and a strict temporal increase than there is between the SMSA profiles grouped according to an east-west dichotomy.

The two-groups results given above can themselves be extended to T groups in an obvious way. We give some of the necessary formulas but omit any numerical example since it would parallel the two-group split very closely. Suppose now there are T groups with $I_1, I_2, \ldots, I_T$ objects in each. Moreover, let r_A^Δ denote the average of the intercorrelations of the type r_A^* among the T sequences obtained by aggregating within each of the groups separately; let r_B^Δ denote a similar average using correlations of the form r_B^*. More specifically, if $r^*_{A_{uv}}$ denotes the correlation (of the form r_A^*) for groups u and v

and $r^*_{B_{uv}}$ denotes the correlation (of the form r_B^*) for groups u and v, then $r_A^\Delta = \frac{1}{\binom{T}{2}} \sum_{u<v} r^*_{A_{uv}}$; $r_B^\Delta = \frac{1}{\binom{T}{2}} \sum_{u<v} r^*_{B_{uv}}$; and

$r_A^\Delta = \frac{1}{\binom{T}{2}} \sum_{u<v} \frac{r^*_{B_{uv}}}{\theta_u\theta_v}$. No immediate simplification is possible, however, since the terms $r^*_{B_{uv}}$ and $\theta_u\theta_v$ depend on the same subscripts. If we assume that all component correlations $r^*_{B_{uv}}$ are non-negative (as they typically would be), then it is easy to show that $r_A \geq r_B^\Delta \geq 0$. Finally, a related expression holds for the associated Z-statistics:

$$\frac{Z_B^\Delta}{Z_A^\Delta} = \sqrt{\frac{\binom{T}{2}}{\sum_{u<v} \frac{1}{I_u I_v}}}\ \frac{\sum_{u<v} r^*_{B_{uv}}}{\sum_{u<v} \frac{r^*_{B_{uv}}}{\theta_u\theta_v}}.$$

Although we have implicitly assumed that our initially given objects were elemental in some given sense, it should be clear that the T group analysis merely sets up another object by attribute data matrix that could be analyzed as such. In this case, each group would correspond to a single object and the aggregate sums once standardized define the observations within a row.

We might also note in passing another alternative for measuring the concordance among T groups. Schucany and Frawley (1973) consider an index based on the products of the T aggregated sums rather than averaging over the pairwise statistics appropriate in two-group splits. An analysis similar to that given above could be formulated for these alternatives but since the extension is peripheral, it will not be pursued.

Alternative Null Models

In testing average correlations, such as r_B and r_B^*, the emphasis throughout has been on a null model that assumes all permutations of the observations within rows of an object by attribute data matrix are equally likely (referred to as the Friedman assumption). Other models are possible, however, based on weaker conditions. For example, suppose we wish to test the significance of r_B for a perfect temporal trend. For each of the I rows of the object by attribute matrix we first obtain the correlation to the set of criterion scores $y_1, y_2, \ldots, y_n$ and to its exact opposite $y_n, y_{n-1}, \ldots, y_1$. This process generates two correlations of equal absolute value but of opposite sign, i.e., if r_i denotes the correlation of row i to $y_1, y_2, \ldots, y_n$, then $-r_i$ is the correlation to $y_n, y_{n-1}, \ldots, y_1$. In short, two dependent samples are obtained consisting of I pairs of observations. If the conjecture is correct that the pattern $y_1, y_2, \ldots, y_n$ is reflected across the rows, then the first member of each pair should be positive and the second negative. Thus, the pattern can be confirmed as a special case of the common two-dependent sample framework.

To be more specific, we choose as our test statistic $T = \frac{1}{2I} \sum_i (r_{i1} - r_{i2}) = \frac{1}{I} \sum_i r_i = r_B$, where r_{i1} is the correlation between the ith row and $y_1, \ldots, y_n$ and r_{i2} is the correlation between the ith row and $y_n, \ldots, y_1$. If our conjecture is approximately correct, the values of T should be relatively large and positive. (To those readers familiar with non parametric statistics, it should be apparent that our problem could also be phrased in terms of a single sample of correlations $r_1, \ldots, r_I$; for example, see Bradley, 1968, pp. 76-79).

Assuming a "null" hypothesis that all 2^I assignments of signs to the absolute differences $|r_{i1} - r_{i2}|$ in the statistic T are equally likely, i.e., that there is no correspondence between the rows of the object by attribute matrix and the criterion or its inverse, the expectation and variance of T can be found easily (see Hubert and Schultz, 1976). The approximate significance test is based on $Z_* = \frac{T - E(T)}{\sqrt{V(T)}} = \frac{Ir_B}{\sqrt{\sum_i r_i^2}}$, where $E(T) = 0$, $V(T) = \frac{1}{I^2} \sum_i r_i^2$. In terms of the previous example, $Z_* = 3.178$ which is substantially smaller than $Z_B = 6.052$. In fact, as long as $\frac{\sum_i r_i^2}{I} > \frac{1}{n-1}$, this later Z-statistic, Z_*, will be less than that obtained under the previous Friedman model. However, since Z_* is based on a weaker condition, it still may be preferable. Obviously, this newer model is irrelevant when preliminary data aggregation is considered since only the two values of $\pm r_A$ would be available. Finally, an extension of these ideas to the multi-group case is immediate. For instance, for two groups $E(r_B^*) = 0$; $V(r_B^*) = \frac{1}{(I_1 I_2)^2} \sum_{\substack{i \in \text{group 1} \\ j \in \text{group 2}}} r_{ij}^2$, where r_{ij} is the correlation between the ith row in group 1 and the jth row in group 2. Thus, for our numerical example, the associated Z-statistic would be 5.740 which is again substantially less than the Z-statistic, Z_B^*, of 8.484.

Hierarchical Clustering (Aggregation)

The algebraic relations between the r_A^* and r_B^* correlations developed in the context of two group concordance have rather significant implications for how data should be reduced and interpreted. In particular, these distributions are relevant to how the objects are classified and clustered. In general, given an object by attribute data matrix and the associated $\binom{I}{2}$ correlations between the I rows, we may wish to cluster or partition the objects (e.g., cities) in such a way that similar cities are placed together in a single class and dissimilar cities are kept apart. We have assumed up to this point that any such grouping of the cities would be given a priori, whereas now, our interest shifts to identifying such groupings post hoc. As a convenience, we emphasize only hierarchical strategies, i.e., clustering schemes that define complete sequences of partitions of the object set. Many of the comments we make, however, can be extended rather easily to alternative methods.

Most hierarchical clustering procedures start with a trivial partition of the object set in which each object forms a separate class by itself. As the clustering progresses, pairs of classes that are the "closest" are successively united until a second trivial partition is reached with all I objects placed together. Thus, if there are K classes at a given stage, that particular pair is united to form a new class if it is the most similar among all possible $\binom{K}{2}$ pairs that are candidates for consolidation. Consequently, if we start with the I x I intercorrelation matrix among the I objects, we need a procedure for continually respecifying the similarity between a new class formed

at a particular state and all remaining classes. Once such a definition is available, the partition sequence is constructed more or less automatically.

Some notation may help clarify the intuitive description of the clustering process given above. Suppose the K classes at a given stage are labelled by $C_1, C_2, \ldots, C_K$, and the measure of similarity or proximity (as yet undefined) for the pair C_u and C_v chosen from the $\binom{K}{2}$ possible matchings is denoted by s_{uv}. Initially when K=I, s_{uv} is provided by the correlation between rows u and v. At a later stage we can assume without loss of generality that the pair C_1 and C_2 is the most similar, and consequently, the subsets C_1 and C_2 are united to form $C_1 \cup C_2$. The proximities between $C_1 \cup C_2$ and $C_3, \ldots, C_K$ have to be redefined, which implies that different procedures for calculating the new proximities will lead to different hierarchical sequences and different clustering methods.

The differences between r_A^* and r_B^* characterize two different processes when used to define the similarity between $C_1 \cup C_2$ and C_k. For correlations of the form r_A^*, we can aggregate over the rows encompassing the sets C_1 and C_2 and over the rows encompassing C_k. Alternatively, if we treat $C_1 \cup C_2$ as one class and C_k as a second, the average correlation r_B^* specifies a second possible measure of similarity. Obviously, the clustering result may vary as a consequence of these two options since r_A^* and r_B^* differ as a function of the internal concordances within $C_1 \cup C_2$ and C_k. There is no "right" way to proceed, however, and in fact, both procedures lead to rather simple formulas for redefining the similarity between $C_1 \cup C_2$ and C_k.

For example, let c_{ij} be the sum of all intercorrelations from C_i to C_j, $\bar{c}_{ij}$ the corresponding average, and I_i and I_j the number of objects in C_i and C_j, respectively. Then, for the r_A^* option, the similarity between any two distinct groups C_k and $C_{k'}$ is $r^*_{A_{kk'}} = \frac{c_{kk'}}{\sqrt{c_{kk}c_{k'k'}}} = \frac{\bar{c}_{kk'}}{\sqrt{\bar{c}_{kk}\bar{c}_{k'k'}}}$.

Thus, it is trivial to show that the similarity between C_k and C_1UC_2 is $\frac{c_{1k} + c_{2k}}{\sqrt{c_{kk}(c_{11} + c_{22} + 2c_{12})}}$. Computationally, at the stage of K classes, $C_1,\ldots,C_K$, we need the K x K matrix that has entries c_{uv} in the uth row and the vth column. At the new stage involving K-1 classes $C_1UC_2, C_3,\ldots,C_K$, only the collapsed K-1 x K-1 matrix given below has to be retained:

	C_1UC_2	...	C_k	...	C_K
C_1UC_2	$c_{11}+c_{22}+2c_{12}$	...	$c_{1k}+c_{2k}$	...	$c_{1K}+c_{2K}$
.	.		.		.
.	.		.		.
.	.		.		.
C_k	$c_{1k}+c_{2k}$	...	c_{kk}	...	c_{kK}
.	.		.		.
.	.		.		.
.	.		.		.
C_K	$c_{1K}+c_{2K}$	...	c_{kK}	...	c_{KK}

Since the original I x I matrix does not have to be stored and referred to as the clustering procedes, the procedure based on r_A^* is combinatorial in the sense of Lance and Williams (1967).

Analogous results hold for the r_B^* option if we define the similarity between any two distinct groups C_k and $C_{k'}$ as $r^*_{B_{kk'}} = \bar{c}_{kk'}$. It is easy to show that the similarity between C_1UC_2 and C_k is $\frac{c_{1k} + c_{2k}}{(I_1+I_2)I_k}$.

At the stage of K classes, only a K x K matrix need be retained with class sizes on the main diagonal and c_{ij}'s in the off-diagonal positions. Thus, this second procedure is also combinatorial.

Using Table 2 for a numerical example, the two clustering alternatives would lead to partition sequences that are identical except for a minor difference early in the hierarchy that affects only the order in which these small groups were eventually merged together. In each case, the level at which the three groups were defined produced the same decomposition: one object set containing Cincinnati alone; a second object set containing Seattle-Everett, Baltimore,and Boston; and a third containing the rest. Looking at the rank correlations to a perfect temporal increase for an interpretation, the three object set is defined by negative correlations, Cincinnati has the lowest positive value, and the large object set contains those SMSA's with the largest positive correlations.

Comparisons of Matrices

The previous discussion has been phrased for sequences of n numerical variables defined for each member of a set of I objects. The same correlational relationships also hold when matrices are available containing pairwise relationships among the n attributes. For example, consider the data in

Table 3. This matrix is defined for the Chicago SMSA (others have been compiled for all other cities used in the sample), and contains the correlations among the seven index crimes based on data over the ten years from 1968 to 1977.

Table 3 here

Given some similar results in the geographical literature (e.g., Harries, 1974, p. 39), we could expect the correlations among the 7 crimes for each city to demonstrate the well-known subdivision into crimes against persons and crimes against property. Or in other words, the correlations within the two sets, (murder, rape, robbery, assault) and (burglary, larceny, auto theft), should be larger than the correlations across sets. To formalize this conjecture in a comparison paradigm, a target or criterion could be set up as a matrix of the following form:

Crime	Murder	Rape	Robbery	Assault	Burglary	Larceny	Auto theft
Murder	X	1	1	1	0	0	0
Rape	1	X	1	1	0	0	0
Robbery	1	1	X	1	0	0	0
Assault	1	1	1	X	0	0	0
Burglary	0	0	0	0	X	1	1
Larceny	0	0	0	0	1	X	1
Auto theft	0	0	0	0	1	1	X

Crime

Rather substantial correlations would be expected between each of the 5 matrices and this given pattern if the person versus property split were represented in the correlations between crime types.

TABLE 3

CORRELATIONS (7 CRIMES): SMSA #1

Chicago

1.000	0.753	0.519	0.714	0.852	0.763	0.772
0.753	1.000	0.893	0.906	0.690	0.439	0.282
0.519	0.893	1.000	0.781	0.308	0.021	0.039
0.714	0.906	0.781	1.000	0.670	0.520	0.240
0.852	0.690	0.308	0.670	1.000	0.938	0.622
0.763	0.439	0.021	0.520	0.938	1.000	0.684
0.772	0.282	0.039	0.240	0.622	0.684	1.000

In general, we treat the $\binom{7}{2}$ entries above the main diagonal in each city matrix as a sequence and the $\binom{7}{2}$ such entries in the hypothesized matrix as a pattern. If the standard normalization to z-scores is performed on each such sequence, a value of .167 would be obtained for r_A and a smaller value of .096 for r_B. Exactly the same reasons for this difference in correlations exist as before since we have merely reinterpreted each matrix as a sequences of values. In both cases, the correlations are positive (as expected) but they are also very small. This last fact suggests that a split into property and person crime is not very salient when the profiles are correlated over time, particularly when the size of these correlations are compared to the strong split observed for profiles correlated over localities, e.g., see Harries (1974, pp. 41-43).

When devising descriptive statistics it may be appropriate to destroy the matrix character of the entries and operate as if we merely had sequences. In terms of Z-statistics, however, the two aggregation schemes would need different variance terms that respect the internal structure of the matrices for each city or for their aggregate sum. Such variance terms are available in the literature (see Hubert and Schultz, 1976) and would generally lead to the same type of relationship among Z-statistics as obtained in the sequence context. Since the variance terms are much more complicated and do not lead to any simple algebraic results, this extension is not discussed in any formal way.

Discussion

In the very first example illustrating the two aggregation schemes, a rank transformation was used within each row and on the rank sums for each column. This process ensures that each object or row contributes "equally" and thus, some degree of natural comparability exists between the summary statistics obtained for the two aggregation schemes. To develop more explicit relationships in terms of formulas, however, the transformations used in most of the paper were based on obtaining z-scores. In the jargon of statistics, observations within rows are aligned for location and scale. This convention allowed precise connections to be developed between the two aggregation schemes both in terms of summary indices (e.g., for r_A and r_B) and Z-statistics (e.g., for Z_A and Z_B).

Matrix extensions offer a great deal of flexibility in defining different relationships among the attributes, but unfortunately, the problem of defining a transformation on the aggregate data matrix also makes it very difficult in general to develop precise formulas for connecting the two aggregation schemes. As an example of this problem, suppose we are given the basic object by attribute data table and define an $n \times n$ matrix for each object (e.g., city) as follows: the entry in the uth row and the vth column is +1 if $x_{iu} > x_{iv}$; -1 if $x_{iu} < x_{iv}$; and 0 if $x_{iu} = x_{iv}$. If we treat the $n(n-1)$ entries in each such matrix as a sample, normalize to z-scores in the usual way, and carry out the same redefinition for a criterion set of values $y_1, y_2, \ldots, y_n$, the average correlation r_B is actually the average Kendall Tau_b statistic of each row against the criterion (cf. Hays, 1960). However, to obtain an analogue of r_A, a similar transformation to signs must be performed on the

aggregated scores from the standardized object by attribute table. This discontinuity in strategy prevents any simple way of defining a relationship between the summary measures for the two aggregation schemes. We would still expect data aggregation to give a larger descriptive measure and a greater significance level, but it is not clear how these expectations could be formalized as a parallel to our previous equations (1) and (2).

As another point of clarification we note that the normalization within rows of an object by attribute data table may not be the only natural transformation to carry out. Instead, suppose z-scores are obtained within columns and for each object i we define an n x n matrix having an entry in the uth row and vth column of $\frac{1}{I} z_{iu} z_{iv}$. Here, z_{iu} and z_{iv} are z-scores for attributes u and v, respectively. If we aggregate over the I matrices (treating objects as if they were "subjects"), the correlation matrix among attributes is generated. Pattern comparisons are important here in the context of what is called a multi-trait multi-method matrix; consequently, some of the same aggregation principles discussed previously appear important to distinguish in these applications as well (see Campbell and Fiske, 1959; Hubert and Baker, 1978). Since a separate paper is planned on this topic, it will not be pursued any further now. We merely comment that different normalizations of an object by attribute table may be appropriate for different purposes.

The methods of data aggregation discussed in the paper represent both ongoing procedures used in geography for aggregating data and alternatives to those standard procedures. For example, Harries (1973) used preliminary aggregation procedures when he averaged violent crime rates for 189 SMSA's for the five year period 1965-69 and calculated simple correlations between

violent crime and population over the SMSA's. Alternative procedures for determining correlations between violent crime and population are given in our discussion of the statistics r_B and $r_{s(ave)}$. Pyle (1974) used correlational aggregation when he examined actual crime rates per 1000 persons for nine crime types, plus armed robberies per 1000 commercial structures and rates of residential burglary per 1000 dwelling units. Pyle calculated all pairwise correlations for the entire study area (Summit County) and for a subset of the area (Akron) and then attempted to illustrate differences between the correlations. This is similar to the procedures used in discussing multiple group concordance in this paper.

Harries (1974) also used preliminary data aggregation when he correlated city size with crime rates averaged over index crimes for a five year time period - a procedure that could be extended by using the matrix comparison procedures developed in the latter section of this paper. Other examples of preliminary and correlational data aggregation procedures can be found in the growing literature on the use of canonical correlation in geography (Monmonier & Finn, 1973; Clark, 1975). However, the exact procedures detailed in this paper focusing on rank orders and Z-statistics for both preliminary and correlational aggregation procedures, to our knowledge, have not appeared in the geographical literature.

As one final observation, it should be noted that the two group discussion was concerned with the concordance between two classes even though the various summary indices were subject to modification by the degree of internal concordance. In other words, we were not explicitly interested in assessing large within group homogeneity per se. Given

the original I x I intercorrelation matrix, however, and a hypothesized split of the I objects into T groups (e.g., into two disjoint subsets), we may also wish to test whether there is more concordance within the groups than expected under some chance model. This topic has been discussed in detail elsewhere for the null conjecture that the given partition was chosen at random from all possible partitions with the same number of classes and objects in each. Thus, we would hope to reject the randomness assumption if the within group concordance were substantially greater than the between group concordance, i.e., the a priori partition is reflected in the size of the correlations in the original I x I matrix. For a more complete discussion, the reader is referred to Hubert and Levin (1976).

REFERENCES

Bradley, J.V., Distribution-Free Statistical Tests, Englewood Cliffs, New Jersey: Prentice-Hall, 1968.

Campbell, D.T. and Fiske, D.W., Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix, Psychological Bulletin, 1959, 56, 81-105.

Clark, D., Understanding Canonical Correlation Analysis, CATMOG #3, Norwich, UK: Geo Abstracts, Ltd., 1975.

Harries, K.D., Spatial Aspects of Violence and Metropolitan Population, The Professional Geographer, 1973, 25 (1), 1-6.

Harries, K.D., The Geography of Crime and Justice, New York: McGraw-Hill, 1974.

Hays, W.L., A Note on Average Tau as a Measure of Concordance, Journal of the American Statistical Association, 1960, 55, 331-341.

Hubert, L.J. and Baker, F.B., Analyzing the Multitrait-Multimethod Matrix, Multivariate Behavioral Research, 1978, 13, 163-179.

Hubert, L.J. and Levin, J.R., Evaluating Object Set Partitions: Free Sort Analysis and Some Generalizations, Journal of Verbal Learning and Verbal Behavior, 1976, 15, 459-470.

Hubert, L.J. and Schultz, J.V., Quadratic Assignment as a General Data Analysis Strategy, British Journal of Mathematical and Statistical Psychology, 1976, 29, 190-241.

Kendall, M.G., Rank Correlation Methods (4th Edition), London: Griffin, 1970.

Lance, G.N. and Williams, W.T., A General Theory of Classificatory Sorting Strategies I: Hierarchical Systems, Computer Journal, 1967, 94, 373-380.

Lord, F.M. and Novick, M.R., Statistical Theories of Mental Test Scores, Reading, Mass.: Addison-Wesley, 1968.

Monmonier, M.S. and Finn, F.E., Improving the Interpretation of Geographical Canonical Correlation Models, The Professional Geographer, 1973, 25, 140-142.

Morrison, D.F., Multivariate Statistical Methods (2nd Edition), New York: McGraw-Hill, 1976.

Schucany, W.R. and Frawley, W.H., A Rank Test for Two Group Concordance, Psychometrika, 1973, 38, 249-258.

Appendix I: Definitions

1. Standard Metropolitan Statistical Area (SMSA):

Most standard metropolitan statistical areas include at least one city of 50,000 inhabitants or more. This is not an absolute requirement as it was for the standard metropolitan area. Instead, two cities with contiguous boundaries and with a combined population of at least 50,000 (often referred to as twin cities) may serve the purpose if they constitute, for general economic and social purposes, a single community, and if the smaller of the two has a population of at least 15,000. There is also the provision that where, in two or more adjacent counties, each has a city of 50,000 inhabitants or more (or twin cities such as those just described) and if the cities lie within 20 miles of each other (city limits to city limits), they will be included in a single Standard Metropolitan Statistical Area unless there is definite evidence that the two urban areas are not economically and socially integrated.

2. r_s = Spearman's rank order correlation coefficient. Defined by cumulative ranks within each column and correlating the rank order of the column sums with an expected temporal rank order.

3. $r_{s(ave)}$ = rank order correlation coefficient obtained by correlating the ranks in each row with an expected temporal rank order and averaging the vector of individual correlation coefficients.

4. r_A = Pearson correlation coefficient defined as in (2) above but substituting z-scores for ranks.

5. r_B = Pearson correlation coefficient obtained as in (3) above but substituting z-scores for ranks.

6. θ = statistic defined to illustrate algebraic relations between r_A & r_B and Z_A & Z_B. For large data sets, $\theta \to \sqrt{r_{ave}}$.

7. r_A^* = a "correlation" coefficient representing a statistical relation between <u>two groups</u> defined on a single data set in which data are first transformed to z-scores within rows and then summed over columns for each group. The resulting column sums are correlated.

8. r_B^* = a "correlation" coefficient representing a statistical relation between <u>two groups</u> defined on a single data set in which data are first transformed to z-scores within rows. All between row correlations across groups are calculated and averaged.

9. r_A^Δ = a "correlation" coefficient representing a statistical relation among <u>T-groups</u> defined on a single data set. Correlations of the form r_A^* are averaged over all $\binom{T}{2}$ pairs of groups.

10. r_B^Δ = a "correlation" coefficient with form similar to r_B^* (above) but averaged over all $\binom{T}{2}$ pairs of groups.

<u>Appendix II</u>

(i) Show: $r_A = r_B/\theta$, where

$$\theta = \sqrt{\frac{1 + (I-1)r_{ave}}{I}} .$$

We assume the following restrictions on x_{ik} (as always) and, without loss of generality, similar constraints for y_k:

$$\sum_k x_{ik} = \sum_k y_k = 0 ;$$

$$\sum_k x_{ik}^2 = \sum_k y_k^2 = n .$$

By definition,

$$\frac{n\sum_k c_k y_k - \sum_k c_k \sum_k y_k}{\sqrt{\left[n\sum_k c_k^2 - \left(\sum_k c_k\right)^2\right]\left[n\sum_k y_k^2 - \left(\sum_k y_k\right)^2\right]}}$$

where $c_k = \sum_i x_{ik}$. Using the restrictions on x_{ik} and y_k given above,

r_A =

$$\frac{\sum_k c_k y_k}{\sqrt{n}\sqrt{\sum_k c_k^2}} = \frac{\sum_k c_k y_k}{\sqrt{n}\sqrt{\sum\sum_{ki} x_{ik}^2 + \sum_k \sum_{i \neq i'} x_{ik}x_{i'k}}} =$$

$$\frac{\sum_k c_k y_k}{\sqrt{n}\sqrt{nI + nI(I-1)r_{ave}}} = \frac{\sum_k c_k y_k}{nI\sqrt{\frac{1 + (I-1)r_{ave}}{I}}} =$$

$$r_B \Big/ \sqrt{\frac{1 + (I-1)r_{ave}}{I}}$$

(ii) Show: $\frac{Z_B}{Z_A} = \sqrt{1 + (I-1)\, r_{ave}} = \sqrt{I}\,\frac{r_B}{r_A}$.

We know $Z_B = \sqrt{I\,(n-1)}\; r_B$ and $Z_A = \sqrt{n-1}\; r_A$.

Thus, $Z_B/Z_A = \dfrac{\sqrt{I\,(n-1)}\; r_B}{\sqrt{n-1}\; r_A} = \sqrt{I}\,\dfrac{r_B}{r_A}$.

Using the relationship $r_A = r_B/\theta$, $Z_B/Z_A = \sqrt{I}\,\theta\dfrac{r_A}{r_A} = \sqrt{1 + (I-1)r_{ave}}$.

ASSESSING HOMOGENEITY IN CROSS-CLASSIFIED PROXIMITY DATA

By

L.J. Hubert, R.G. Golledge

C.M. Costanzo, G.D. Richardson

The University of California, Santa Barbara

Partially supported by LEAA/NILCJ 79-NI-AX-0057 and NSF SOC-77-28227.

Assessing Homogeneity in Cross-Classified Proximity Data

It is common to see data classified according to various a priori dimensions. For example, a set of latency measures may be cross-classified according to subject and task, a given crime rate may be coded by city, year, and type of offense, and school achievement scores may be presented in terms of sex and SES level. Since classification facets supposedly provide information useful in explaining the variability inherent in the measure under study, one of the first problems faced by any applied statistician interested in analyzing such data is to develop suitable methods for evaluating the effects of a priori classifications. Multivariate analysis-of-variance and its specialization to the study of profiles [Morrison, 1976] are obvious strategies to consider. In fact, they may even be too obvious since their general applicability could limit our concern with exploring alternative inference paradigms that might be more appropriate in specific instances.

Standard analysis-of-variance methods and their correlates are all rather specialized and keyed to particular ways of interpreting data and to certain measures of proximity (e.g., covariances) among the objects under study. More pointedly, classical statistical inference may not complement in any natural way many of the newer data reduction strategies of scaling and cluster analysis that are becoming very popular in the social and behavioral sciences. These latter methods have proved highly successful in describing structure within an object set and have been implemented with a variety of arbitrary measures of proximity having few if any well-developed parallels within the traditional confines of statistics. Because of these substantive developments, it would be of obvious value to have confirmatory inference strategies for testing the saliency of a priori dimensional constraints based on an unspecified proximity function.

In recent years various analysis-of-variance analogues have been proposed for arbitrary proximity measures when the object set is subject to a single classification dimension [Hubert & Levin, 1976; Mielke, Berry, & Johnson, 1976]. Since little work has addressed the more difficult task of two or more classification dimensions, we use this lack as a basic motivation for the paper. The problem emphasized is one of analyzing an arbitrary proximity matrix that contains cross-classified information on the similarities within a set of N objects. For convenience, the initial discussion assumes that the N objects are actually N space-time units, i.e., m cities observed over n years so mn = N. The proximity information per se will be given by a Euclidean distance measure between the profiles defined over the seven index crimes for each pair of space-time units. Based on this structure and data, the aim is to evaluate temporal and spatial homogeneity; that is, to assess whether any evidence exists for an increased similarity among profiles associated with the same levels in the temporal or spatial variables. It is important to remember, however, that the space-time interpretation is used here solely for convenience of exposition. Any cross-classification having a similar form and/or any measure of proximity is subject to the same analysis schemes to be developed in the following pages.

Background

The most convenient way to introduce our approach to the assessment of profile homogeneity in cross-classifications is to introduce an example that can be used throughout the discussion. Table 1 presents a 30 by 30 symmetric matrix containing measures of proximity between the profiles for 6 cities (Chicago, Los Angeles, Philadelphia, San Francisco, Boston, and Pittsburgh) at 5 time points (1969, 1971, 1973, 1975, 1977). Each entry in the matrix corresponds to a Euclidean distance measure between the profiles for two city-

time units over the seven index crimes (murder and non-negligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny/theft, motor vehicle theft). The entries across crimes were first made comparable by a transformation to Z-scores, i.e., normalizing the rates within each particular crime type to mean 0 and variance 1. The final proximity or distance between profiles was then obtained by taking the square root of the sum of the squared differences over the seven index crimes. Obviously, other measures could be used as well, e.g., Mahalanobis distances, correlations, and so on. In all cases, however, the analysis procedure to be presented remains the same.

Table 1 here

Considering the complete 30 by 30 matrix of proximities in Table 1, the assessment of spatial homogeneity should depend on the level of elevation for the entries in the 6 "same-city" blocks (each of size 5 X 5) on the main diagonal. Similarly, if we reorganize Table 1 appropriately, temporal homogeneity should be reflected in the level of elevation for each of the 5 "same-time" blocks (each of size 6 X 6) on the main diagonal. In either case a natural evaluation strategy would first define an index for the degree of elevation in the on-diagonal blocks, and then specify a procedure for assessing the relative size of the observed index compared to some chance model.

Numerous procedures have been proposed for analyzing data of the form represented in Table 1 when the proximities are actually correlations. In this case the tables are usually referred to as multitrait-multimethod matrices [Schmitt, Coyle, & Saari, 1977]. Although the literature in this area is very extensive, the work of Hubert and Baker [1978, 1979] is the most relevant to our discussion since both of these last papers develop indices of elevation for the entries within the on-diagonal blocks as well as associated significance tests. The significance tests are based on generating a reference

distribution under the assumption that all possible assignments of the space-time variables are equally-likely (a random assignment of the N objects to the m cities and n times). If the observed index is extreme with respect to this reference model, a conjecture of randomness is rejected, implying a statistically significant degree of spatial (or temporal) homogeneity. In Hubert and Baker [1978], indices based on average within-group proximities are used; nonmetric comparisons are developed in the second paper by comparing the size of the proximities within the on-diagonal blocks to those outside. For further details, the reader is referred to the two original sources.

The approach to evaluating space or time coherence just sketched has two unfortunate drawbacks. First of all, the null model assumes that all assignments of the N objects to the space-time classifications are carried out at random. Thus, an index that supposedly measures spatial homogeneity must be evaluated against a reference distribution contaminated with temporal homogeneity. We, in fact, compare the observed index for a given classification dimension (a partition of the N object set) with all possible partitions having the same number of classes and number of objects in each. No attempt is made to control for the second classification facet. For example, if there are strong spatial effects and moderate temporal effects, it may be impossible to detect the latter since the reference distribution used for evaluating temporal homogeneity does not remove the effects of spatial homogeneity. Secondly, the extensions of these notions to nonmetric comparisons among the proximities in the original N X N matrix requires a substantial amount of computation; even simple moment formulas for the indices are very cumbersome to derive. Although the sole reliance on the ordering of the proximities in a matrix may be desirable in defining a measure of elevation for the on-diagonal blocks, the computational burden is so great that the practical usefulness of these extensions is limited. As before, the developments to

date in the non-metric context confound spatial and temporal homogeneity in the construction of the appropriate reference distribution.

Given a cross-classified matrix, such as Table 1, there is a very simple device for solving both of the difficulties of confounding and nonmetric computation at the same time. The strategy can be viewed as a direct extension of well-known principles in multi-factor analysis-of-variance since we will block on the level of the spatial classification when assessing temporal similarity and conversely. The next section introduces the paradigm in greater detail and illustrates the increased sensitivities of the method.

Formal Details

To provide some simple notation, suppose that N profiles are defined over a set of T variables. As always, the N profiles are cross-classified by m levels of factor A (A_1, A_2, ..., A_m) and n levels of factor B (B_1, B_2, ..., B_n) so that N = mn. In terms of the previous example from Table 1, factor A corresponds to city, factor B to time, m = 6, n = 5, and N = 30. The proximity function is assumed to be some distance measure between the profiles corresponding to different city-time pairs, where each of the T variables is typically standardized to mean zero and variance one. Based on this notation, the complete proximity information can be represented in the schematic form given in Table 2. The main diagonal is assumed irrelevant and will always consist of zeros.

Table 2 here

Since the roles of factors A and B are interchangable, our discussion can be limited to the assessment of profile homogeneity within the levels of factor A, e.g.,to spatial homogeneity if the levels of A correspond to m cities. One obvious index is the sum of proximities within the levels of A represented by the proximities within the m triangles indicated in Table 2. If we rewrite

Table 2 as Table 3, the appropriate proximities are now represented by the three indicated diagonal strips. In either case, we denote the observed sum by r_{obs}.

Table 3 here

One possible null model for evaluating r_{obs} that is not affected by the homogeneity within levels of the nuisance factor B is best explained using the form of Table 3. Within each level of factor B the m! reorderings of the levels of factor A are assumed equally likely. Since there are n levels of B, this implies $(m!)^n$ equally likely realizations of the complete matrix. Each such realization leads to a value for r, and when tabled, these values form a reference distribution to evaluate the size of the observed index, r_{obs}. In the usual interpretation, the hypothesis of randomness (blocking on factor B) would be rejected if r_{obs} were sufficiently small; the proportion of realizations giving values of r as small or smaller than r_{obs} would be reported directly as the significance level.

As an alternative explanation for the null model, we note that r_{obs} is an index calculated for a partition of the N objects into m groups of n objects each. The reference distribution is a tabling of the index r over all partitions of exactly this same form that are subject to the blocking on time. Each such partition has the property that the n objects in each group are representative of all n levels of factor B. Or in other words, the partitions so constructed are intended as possible comparisons for the original decomposition into cities in which the levels of the second temporal factor are balanced.

Besides controlling for the effect of factor B in assessing homogeneity within the m levels of A, blocking also decomposes the inference problem in a very convenient way. In Table 3, the cities within levels of factor B are permuted to obtain the reference distribution but never across the levels of factor B. If we look at an arbitrary section of Table 3 for the u and v levels

of factor B, and permute the rows and columns separately and at random, part of the variability in the overall index r can be identified. In fact, if r_{uv} represents this contribution, then

$$r = \sum_{u>v} r_{uv} .$$

Moreover, since the separate indices of the form r_{uv} are independent in pairs, the first two moments of r can be obtained very simply:

$$E(r) = \sum_{u>v} E(r_{uv});$$

$$V(r) = \sum_{u>v} V(r_{uv}).$$

The complete enumeration of r requires an evaluation over all $(m!)^n$ possible realizations of Table 3, but the formulas given above show that moments can be obtained for each separate index r_{uv} and then merely summed to obtain those for r.

		B_v			
		A_1	A_2	...	A_m
	A_1	d_{11}	d_{12}	...	d_{1m}
	A_2	d_{21}	d_{22}	...	d_{2m}
B_u	.	.	.	.	.
	.	.	.	.	.
	.	.	.	.	.
	A_m	d_{m1}	d_{m2}	...	d_{mm}

Then from the literature [Puri & Sen, 1971],

$$E(r_{uv}) = 1/m \sum_i \sum_j d_{ij}$$

$$V(r_{uv}) = 1/(m(m-1)) \{(1/m) Q_1 - (Q_2 + Q_3) + mQ_4\}$$

where

$$Q_1 = (\sum_i \sum_j d_{ij})^2 \ ;$$

$$Q_2 = \sum_j (\sum_i d_{ij})^2 \ ;$$

$$Q_3 = \sum_i (\sum_j d_{ij})^2 \ ;$$

$$Q_4 = \sum_i \sum_j d_{ij}^2 \ .$$

In terms of our Table 1 data reorganized in the form of Table 3 for evaluating spatial homogeneity, we obtain $\Gamma_{obs} = 101.75 = \Gamma_{21} + \Gamma_{31} + \ldots + \Gamma_{54} = 5.64 + 11.03 + 7.78 + 16.85 + 13.32 + 6.74 + 14.87 + 12.45 + 6.91 + 6.16$. Thus, since the corresponding expectations would be 17.987, 20.490, 19.382, 23.797, 22.155, 19.873, 22.985, 21.990, 20.047, 20.637, $E(\Gamma) = 209.340$.

Analogously, $V(\Gamma) = V(\Gamma_{21}) + \ldots + V(\Gamma_{54})$. Therefore, $V(\Gamma) = 108.174$ and the associated Z-statistic for Γ_{obs} is

$$Z = \frac{\Gamma_{obs} - E(\Gamma)}{\sqrt{V(\Gamma)}} = \frac{101.75 - 209.34}{\sqrt{108.174}} = -10.34.$$

A simple numerical example

The data given in Table 1 are too extensive to illustrate the specifics of a complete enumeration. Consequently, a much smaller artificial example is presented in this section before we return to the more realistic illustration later on.

Suppose we have three cities and two time points and are provided with the following proximity information:

	City 1		City 2		City 3	
	Time 1	Time 2	Time 1	Time 2	Time 1	Time 2
City 1						
Time 1	0					
Time 2	1	0				
City 2						
Time 1	2	4	0			
Time 2	3	2	2	0		
City 3						
Time 1	2	3	2	2	0	
Time 2	7	3	4	1	1	0

Assuming our concern is with temporal homogeneity, we would sum the entries in the diagonal strips of this matrix (see Table 3) and obtain $\Gamma_{obs} = 12$, i.e., based on the three city to city matrices of size 2 X 2 from the matrix, we obtain $\Gamma_{21} = 4$, $\Gamma_{31} = 5$, $\Gamma_{32} = 3$, and $\Gamma_{obs} = 4 + 5 + 3 = 12$. Blocking on the spatial variable (and considering the three city to city blocks), there are $(2!)^3 = 8$ equally likely realizations of the matrix that are used for generating the reference distribution, i.e., the two time points within city 1 can be reordered, the two within city 2, and finally, the two within city 3:

City 1	City 2	City 3
T_1T_2	T_1T_2	T_1T_2
T_2T_1	T_1T_2	T_1T_2
T_1T_2	T_2T_1	T_1T_2
T_1T_2	T_1T_2	T_2T_1
T_1T_2	T_2T_1	T_2T_1
T_2T_1	T_2T_1	T_1T_2
T_2T_1	T_1T_2	T_2T_1
T_2T_1	T_2T_1	T_2T_1

From our previous formulas

$$E(r_{21}) = 11/2 ;$$

$$E(r_{31}) = 15/2 ;$$

$$E(r_{32}) = 9/2 ;$$

$$E(r) = 11/2 + 15/2 + 9/2 = 35/2,$$

and

$$V(r_{21}) = 9/4 ;$$

$$V(r_{31}) = 25/4 ;$$

$$V(r_{32}) = 9/4 ;$$

$$V(r) = 9/4 + 25/4 + 9/4 = 43/4.$$

All of these numerical relationships can be verified from the complete enumeration.

If our interest is in spatial homogeneity, a similar procedure could be followed. The on-diagonal triangles in the matrix give $r_{obs} = 1 + 2 + 1 = 4$. Blocking on the time variable there are $(3!)^2 = 36$ equally likely realizations of the matrix. The associated index values could be tabled in the same manner as for the temporal variable.

Clarifications and Extensions

Complete Enumeration Versus Approximate Tests

If the size of the cross-classification matrix is small enough, the exact distribution of r_{obs} can be obtained by a complete listing under the hypothesis of randomness. Typically, however, the problem is so large that approximations of some sort are necessary. For example, a simple Cantelli bound assures us that the significance level for any form of reference distribution whatsoever will be no larger than $1/(Z^2 + 1)$, where Z is the standardized value of r_{obs} based on the exact moments. The adequacy of an assumed normal approximation is

still unknown; consequently, it is probably best to rely either on this crude bound whenever possible, or better yet, construct an approximation to the complete enumeration by sampling. Here, M partitions of the required form would be selected at random and with replacement and the r indices obtained for each. The significance level reported is the number of indices as small or smaller than r_{obs}. In fact, since r_{obs} can itself be treated as a random draw under the null hypothesis, the sample itself is assumed to be of size M + 1 and includes the actual value of r_{obs}. For a more extensive discussion of these Monte Carlo testing strategies, the reader is referred to Edgington [1969].

For the data in Table 1 and based on an M of 999, the following significance levels would be obtained for the assessment of spatial and temporal homogeneity.

	Spatial	Temporal
r_{obs} :	101.75	271.03
E (r) :	209.34	285.063
V (r) :	108.174	4.837
Z :	-10.34	-6.38
Monte Carlo significance level	.001	.001
Cantelli significance	.01	.03

As is apparent, the Cantelli values are very conservative but would suffice for rejection of the randomness conjecture at the traditional .05 significance level. In fact, since tables of the 999 values of the statistic in both cases display patterns that are very close to normal, an optimist might even compare the observed Z-statistics to the standard normal percentage points.

The Z-statistics given above are obviously very large in absolute value and indicate that both the spatial and temporal classifications could help

explain some of the differences in the proximities. It is interesting to note that different conclusions would result without blocking if the Hubert-Baker null model discussed in the introduction were followed. In particular, the spatial statistic would be given a very large Z-value of -11.13 and the temporal statistic would produce a positive Z-value of .72. Since the latter is obviously non-significant (lower tail), temporal differences are overwhelmed by spatial differences in the construction of the reference distribution. Unless the spatial distinction is introduced as an explicit blocking variable, the temporal factor is not identified as being salient.

Connections to Randomized-Block Designs

To illustrate how the blocking strategy is really a generalization of what is already done routinely in randomized-block analyses-of-variance, suppose we are interested in evaluating spatial homogeneity and use the symbol O_{ij} to denote city j and time i. In terms of a simple cross-classified table, these symbols can be written as:

Time	City: A_1	A_2	...	A_m
B_1	O_{11}	O_{12}	...	O_{1m}
B_2	O_{21}	O_{22}	...	O_{2m}
.	.	.		.
.	.	.		.
.	.	.		.
B_n	O_{n1}	O_{n2}	...	O_{nm}

Except for the use of non-numerical objects, this table resembles the form of a two-way analysis-of-variance layout with one observation per cell. In

fact, if the objects O_{ij} were actually numerical values and our proximities in Table 1 were actually squared distances, then our previous assessment procedure reduces to a test of the A factor based on Friedman's null model. More specifically, we block on the B factor and consider all permutations of the observations with each row equally likely. The measure r_{obs} is obtained over all $\binom{n}{2}$ ordered row pairs as

$$r_{obs} = \sum_{u>v} \left[\sum_j (O_{uj} - O_{vj})^2 \right],$$

which is numerically equal to n times the within column sum of squares. Since the total sum of squares is constant, r_{obs} varies as a direct function of the sum of squares attributable to factor A.

In the more general context discussed earlier in which an arbitrary measure of proximity identifies the relationship between objects, exactly the same Friedman model is being used. Since the index r_{obs} can be written as

$$\sum_{u>v} r_{uv},$$

the possibly more complex term r_{uv} takes the place of $\sum_j (O_{uj} - O_{vj})^2$. We also note that a complete enumeration for, say, the A spatial factor, would actually require $(m!)^{n-1}$ realizations of the cross-classified table rather than $(m!)^n$. As in the standard Friedman context, the first row of objects can be considered constant. For instance, in our simple example of a complete enumeration, $(2!)^3 = 8$ partitions were listed, but only $(2!)^2 = 4$ were necessary since each index value in the listing is repeated (a multiple of) 2! = 2 times.

Nonmetric Extensions

The analysis procedure we have described for a cross-classified proximity matrix is based on measures of the form r_{uv}, and more specifically, on the submatrices from which they are derived. Consequently, the assessment of spatial homogeneity in our example depends on the component indices r_{21}, r_{31}, ..., r_{54} since r_{obs} is defined by their sum. To take one simple illustration, the following 6 X 6 submatrix is used to generate r_{21}:

1969

1971	Chicago	L.A.	Phil.	S.F.	Boston	Pitts.
Chicago	.59	3.03	3.32	3.22	3.97	3.84
Los Angeles	3.54	1.02	5.36	1.69	5.25	5.48
Philadelphia	1.46	3.62	1.67	3.81	2.46	2.34
San Francisco	3.06	1.96	4.75	.90	4.54	4.77
Boston	3.21	4.07	2.55	3.82	1.02	2.08
Pittsburgh	3.12	4.48	.96	4.87	1.65	.44

It should be apparent that any transformation could be carried out on these submatrices and the analysis developed in exactly the same manner. As possibly the most relevant definition, each entry could be replaced by an integer that specifies how many proximities are strictly smaller within that entry's given row and column. Based on the 6 X 6 matrix given above, this would give

1969

1971						
0	3	6	4	8	7
7	0	9	2	8	10
1	7	2	8	5	4
4	2	8	0	7	9
4	9	4	7	0	2
6	9	1	10	3	0

Obviously, transformations of this nonmetric form would lead to an inference strategy that depends only on the order of the proximities in a cross-classified matrix and not on their actual numerical values. Most importantly, such redefinitions are trivial to accomodate and no major changes are required in the associated inference strategy.

Extensions to More than Two Dimensions

The cross-classified matrices we have considered up to this point have only included two dimensions. Extensions to more than two are immediate, however, and offer no major difficulties. For example, suppose we have three factors A, B, and C with m, n, and t levels, respectively. If an evaluation of A is desired, then factors B and C are simply combined to produce an aggregate factor with nt levels. Otherwise, the evaluation process remains unchanged. It should be apparent that any number of dimensions and levels could be handled by the simple expedient of constructing aggregate factors from all dimensions except the particular one under test.

Discussion

The major contribution of this paper is in the use of arbitrary proximity measures and the development of a strategy for blocking on the levels of one (or more) a priori dimension(s) when evaluating the differences over a second. The strategy being proposed is really very general even though the illustration we have used in explaining the method contained the three explicit classification dimensions of space, time, and crime type. For instance, since any two of the dimensions could in fact have been considered the major classification facets of interest, proximity measures could have been obtained between profiles over the m cities and our interests directed toward the two dimensions

of crime type and time. The basic inference principles would remain the same and the analyses would be carried out as before. Hopefully, our discussion will allow researchers to assess dimensional salience in data sets that are not easily studied by more standard analysis-of-variance schemes because of an unusual proximity measure. Moreover, the possibility of relying on only nonmetric comparisons among proximities should provide a nice tie-in to the current emphasis in nonmetric clustering and scaling in the social and behavioral sciences.

REFERENCES

Edgington,E.S. Statistical Inference: The Distribution-Free Approach. New York: McGraw-Hill, 1969.

Hubert, L.J. & Baker, F.B. Analyzing the multitrait-multimethod matrix. Multivariate Behavioral Research, 1978, 13, 163-179.

Hubert, L.J. & Baker, F.B. A note on analyzing the multitrait-multimethod matrix: An application of a generalized proximity function comparison. British Journal of Mathematical and Statistical Psychology, 1979, 32, 179-184.

Hubert, L.J. & Levin, J.R. Evaluating object set partitions: Free-sort analysis and some generalizations. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 459-470.

Mielke, P.W., Berry, K.J., & Johnson, E.S. Multi-response permutation procedures for a priori classifications. Communications in Statistics -- Theory and Methods, 1976, 5, 1409-1424.

Morrison, D.G. Multivariate Statistical Methods (2nd Edition). New York: McGraw-Hill, 1976.

Puri, M.L. & Sen, P.K. Nonparametric Methods in Multivariate Analysis. New York: Wiley, 1971.

Schmitt, N., Coyle, B.W., & Saari, B.B. A review and critique of analyses of multitrait-multimethod matrices. Multivariate Behavioral Research, 1977, 12, 447-478.

Table 1

30x30 Symmetric Matrix Containing Measures of Proximity Between the Profiles for 6 Cities at 5 Time Points

	Chicago					Los Angeles					Philadelphia					San Francisco					Boston					Pittsburgh				
	1969	1971	1973	1975	1977	1969	1971	1973	1975	1977	1969	1971	1973	1975	1977	1969	1971	1973	1975	1977	1969	1971	1973	1975	1977	1969	1971	1973	1975	1977
Chicago																														
1969	0																													
1971	.59	0																												
1973	2.08	1.83	0																											
1975	2.81	2.60	.93	0																										
1977	2.53	2.48	1.39	1.12	0																									
Los Angeles																														
1969	3.07	3.03	2.83	2.99	3.17	0																								
1971	3.54	3.39	3.02	3.15	3.53	1.02	0																							
1973	4.02	3.79	2.78	2.62	3.28	1.77	1.36	0																						
1975	4.47	4.18	2.98	2.77	3.59	2.62	2.04	.94	0																					
1977	5.07	4.76	3.66	3.59	4.38	3.08	2.53	1.59	1.19	0																				
Philadelphia																														
1969	2.86	3.32	4.34	4.66	3.87	4.56	5.36	5.90	6.54	7.22	0																			
1971	1.46	1.82	2.98	3.42	2.74	3.62	4.23	4.76	5.30	6.00	1.67	0																		
1973	1.50	1.77	2.39	2.71	1.96	3.16	3.81	4.15	4.70	5.40	2.06	.89	0																	
1975	1.63	1.68	1.74	1.93	1.36	2.86	3.41	3.55	4.03	4.80	2.82	1.59	.87	0																
1977	2.15	2.45	2.64	2.72	2.05	3.07	3.84	4.05	4.66	5.42	2.12	1.66	1.04	1.16	0															
San Francisco																														
1969	3.39	3.22	3.16	3.21	3.44	1.96	1.69	2.44	3.04	3.71	5.00	3.81	3.52	3.13	3.63	0														
1971	3.06	2.88	2.69	2.70	3.03	1.96	1.77	2.29	2.80	3.65	4.75	3.53	3.20	2.68	3.23	.90	0													
1973	3.88	3.71	2.52	1.91	2.47	2.41	2.42	1.71	2.10	3.05	5.31	4.28	3.58	2.84	3.27	2.50	2.05	0												
1975	4.54	4.30	2.89	2.21	2.99	3.09	2.93	1.86	1.88	2.76	6.12	5.05	4.34	3.54	4.05	3.17	2.74	.93	0											
1977	4.52	4.31	2.92	2.30	3.04	2.82	2.68	1.55	1.69	2.51	6.10	5.06	4.33	3.58	4.03	3.07	2.71	.91	.45	0										
Boston																														
1969	3.52	3.97	4.77	4.97	4.15	4.57	5.25	5.94	6.63	7.39	1.94	2.46	2.74	3.37	2.74	4.58	4.54	5.30	6.19	6.13	0									
1971	3.21	3.60	4.28	4.47	3.74	4.07	4.60	5.33	5.98	6.78	2.55	2.39	2.60	3.06	2.70	3.82	3.80	4.72	5.61	5.56	1.02	0								
1973	3.20	3.48	3.77	3.87	3.26	3.61	3.99	4.63	5.25	6.04	3.40	2.81	2.73	2.93	2.81	3.15	3.24	4.04	4.89	4.82	2.09	1.23	0							
1975	4.69	4.80	4.63	4.60	4.33	4.22	4.16	4.72	5.19	5.90	5.54	4.72	4.53	4.46	4.63	3.28	3.64	4.40	5.06	4.95	4.15	3.24	2.16	0						
1977	3.84	4.10	3.99	3.98	3.57	3.43	3.71	4.28	4.89	5.67	4.28	3.74	3.50	3.52	3.35	3.17	3.27	3.03	4.63	4.48	3.01	2.23	1.32	1.74	0					
Pittsburgh																														
1969	3.34	3.84	4.74	4.99	4.26	4.70	5.48	6.08	6.77	7.51	1.21	2.34	2.67	3.29	2.45	5.00	4.77	5.41	6.27	6.23	1.38	2.08	3.03	5.14	3.79	0				
1971	3.12	3.61	4.53	4.79	4.09	4.48	5.29	5.87	6.55	7.27	.96	2.12	2.43	3.05	2.19	4.87	4.61	5.22	6.06	6.02	1.65	2.27	3.16	5.28	3.91	.44	0			
1973	3.14	3.63	4.44	4.67	3.93	4.59	5.39	5.90	6.56	7.30	1.03	2.15	2.38	2.97	2.09	4.96	4.67	5.17	5.99	5.97	1.68	2.29	3.12	5.25	3.87	.53	.42	0		
1975	2.66	3.11	3.66	3.82	3.09	3.96	4.73	5.12	5.75	6.54	1.40	1.81	1.76	2.22	1.29	4.37	4.02	4.34	5.15	5.13	1.90	2.17	2.72	4.78	3.37	1.28	1.11	.91	0	
1977	2.96	3.42	4.03	4.19	3.47	4.14	4.96	5.38	6.05	6.79	1.27	2.08	2.07	2.57	1.56	4.62	4.31	4.63	5.44	5.40	1.91	2.36	2.99	5.07	3.62	1.06	.84	.66	.50	0

Table 2

Schematic Form for the Cross-classified Proximity Matrix

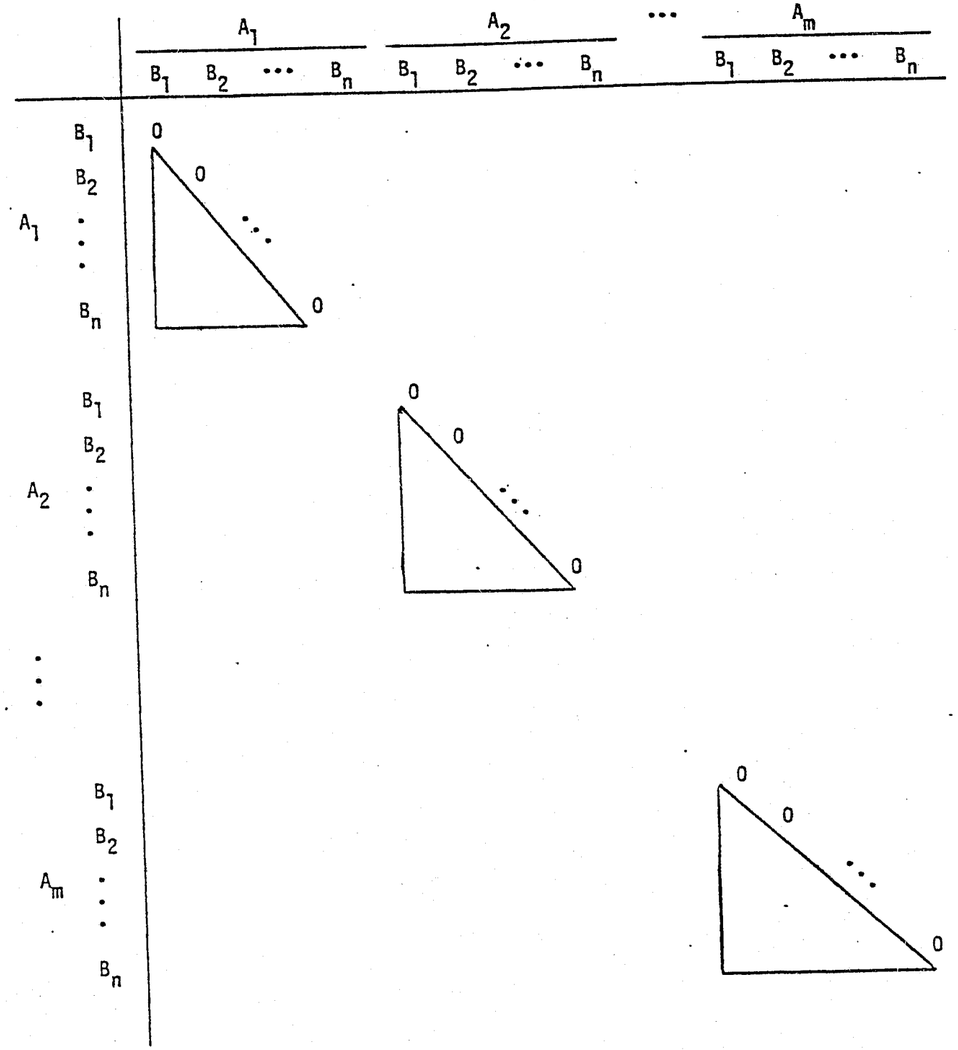
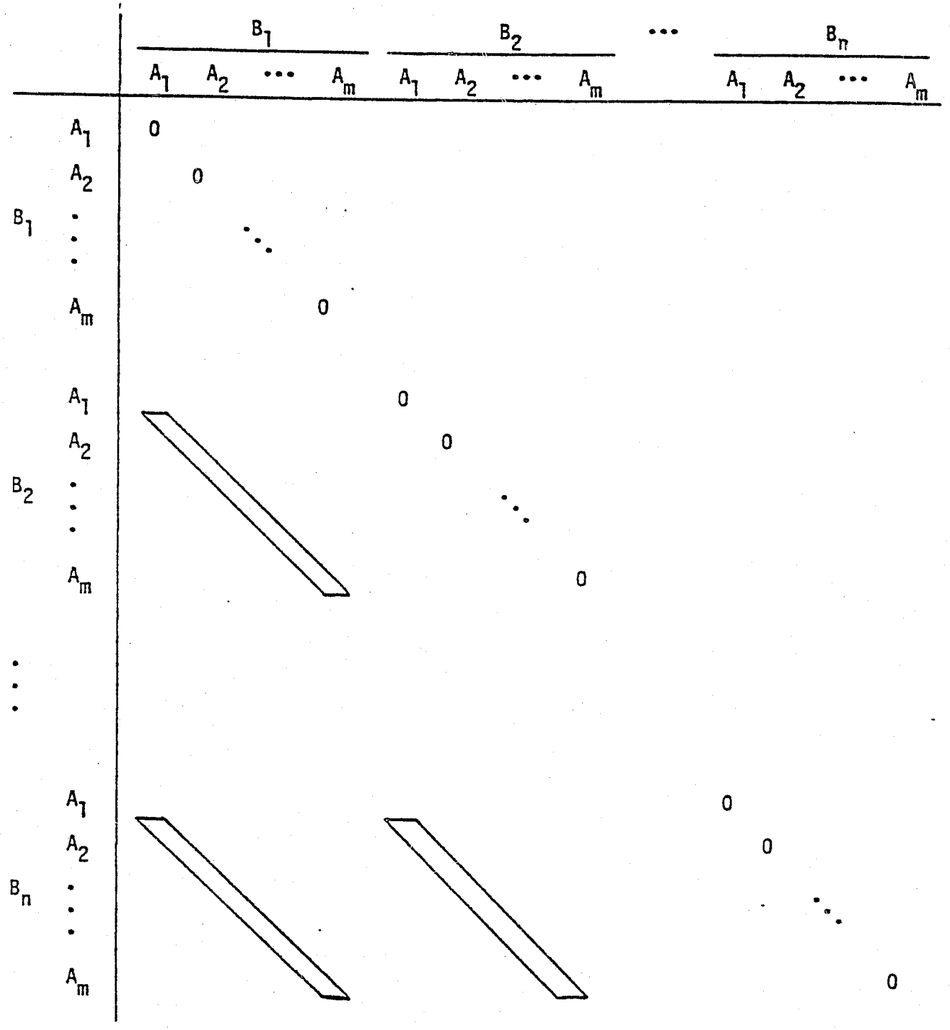


Table 3

Reordered Form of the Matrix in Table 2



78712

Unidimensional Seriation:
Implications for Evaluating Criminal Justice Data*

L.J. Hubert, R.G. Golledge
T. Kenney & D. Richardson

The University of California, Santa Barbara

* Partial support for this research was provided by LEAA/NILCJ Grant #79-NI-AX-0057 and NSF-GSOC-77-28227.

Introduction

The term "unidimensional seriation" refers to the task of ordering a set of objects along a continuum based on some measure of proximity or similarity defined between all members of the set. Typically, such orderings are constructed by a computational procedure that optimizes the correspondence between the given proximities and the distances between the objects generated by a specific placement. Since the optimization strategy, the induced distances, the index of correspondence, and the original proximity measures are as yet unspecified, many different variations on the seriation task can be (and have been) proposed. To limit the scope of our discussion, however, we emphasize only a few general tactics that seem particularly relevant to the study of criminal justice data. Our goal in the process is to develop several methodological points that are important to keep in mind when applying these methods.

Although a specific numerical illustration is given in a later section to suggest how parts of our discussion are relevant to the perception of levels of criminal activity over a set of SMSA's, it may help the reader now if a few other possible applications are mentioned. For example, if the objects in the set S are attitudinal statements regarding the possible treatment of suspects, the proximity measure could denote the proportion of individuals who endorse one statement over a second, or possibly, the proportion who believe one statement represents a more lenient position than a second. In either instance, the concern is with ordering (or scaling) the statements along a lenient-strict dimension based on the observed proportions: in the former case of asking for differential endorsement, it may also be desirable

to place individuals at their "ideal" points along the same continuum as a procedure for evaluating variation in attitude. As a second possible application, we may be interested in the degree to which various groups or types of criminals view different crimes, and specifically, with whether their ordering of crimes according to, say, seriousness reflects a perception of anti-social behavior contrary to the legal code or usual community norms. Each individual, for example, could first provide a rank order of the seriousness of a set of criminal situations; the degree of differential evaluation would then be defined by the proportion of the group who assess one behavior as more serious than a second. Again, single dimension scalings could be constructed and compared to various conjectures as to how the objects being scaled should be ordered and spaced. As a final introductory application, the objects could represent geographical locations and the proximities could indicate the degree of spatial interaction in terms of the movement of criminals from the location of residence to the location of the committed crime. Here, the placement of the spatial locations along the single dimension would represent differential degrees of attractivity for criminal behavior.

The topic of unidimensional seriation is very broad indeed; consequently, our presentation can in no way be seen as complete although we do spend a substantial amount of time reviewing relevant background material that is not explicitly represented in the later numerical example. It is hoped that our generality will convince the reader that many alternatives could be reasonably followed when attempting to validate a given seriation even when we are restricted to using internal evidence or information from the available proximities. Throughout the presentation an effort is made to rely on models

and procedures with minimal assumptions; in fact, much of the discussion is concerned with simple schemes for testing whether a specific reordering of the proximity information displays a conjectured pattern, or alternatively, whether any reordering whatsoever will display it. This concern with simplicity is carried over into a numerical example that deals with the perception of homicide level in 15 major SMSA's. We will use this illustration later as a concrete referant for the general orientation to be developed below.

Cross-Validation

The problem of cross-validating a given statistical result reoccurs continually in the applied methodological literature, particularly for the newer multivariate procedures that are becoming commonplace in the behavioral and social sciences, e.g., regression, discriminant analysis, canonical correlation, and so on. All of these techniques have the property of optimizing a model with respect to a single sample; consequently, since it is always hoped that our statistical models have some greater generality, it is important to assess the loss in "fit" that would result when an estimated model is applied in a new context. In the simple regression framework, for instance, the resulting prediction equation does the "best" it can with the available data but there is a great likelihood of a lower correlation between the predicted and actual scores in a different sample. In the current jargon, there should be "shrinkage" in the size of the correlation when the regression equation is cross-validated against a new data set.

Estimates of shrinkage are obviously important in any model that requires optimization with respect to fallible information, but unfortunately, the collection of new data is usually very expensive. Because of these costs, various procedures have been suggested over the years to obtain reasonable measures of shrinkage without replicating a complete study. In regression,

for instance, the most obvious scheme requires a reasonably large data base and a split of the available data into two parts - the first set is used to find the equation and the second operates as a replication. More recently, "sample reuse" procedures have been suggested that successively eliminate single observations (or groups) and then recalculate the desired equation on what remains. Since the resulting equation at each stage can be used to predict various characteristics of the eliminated observations, an estimate of shrinkage can be defined by aggregating various functions of these predictions over all possible observations (or groups) that are subject to elimination. As one important application, this sample reuse strategy is part of a well-known BMDP program for estimating misclassification probabilities in discriminant analysis.

From a more general perspective, the methods of sample reuse all try to evaluate the adequacy of a given result based on internal evidence within a single data set. Our task in the seriation context is similar since it would be very helpful if we could rely on internal information to validate the basic unidimensional model being assumed. Although we may have to be more subtle when dealing with the large class of schemes that can be proposed for the ambiguous task of unidimensional seriation, it still appears possible to define several general principles of assessment that can help in attacking the problem of scale cross-validation. At the very least, we should be able to emphasize different aspects of a proximity measure collected on a set of objects, or alternatively, use different optimization criteria that should lead to the same (or different) scales if the basic unidimensional conjecture is approximately correct. Unfortunately, what we can propose is not a routinized algorithm that will provide a definitive and final conclusion in all cases. As a more modest objective, an orientation toward unidimensional models is pointed out that should sensitize the applied researcher to the problems of cross-validation and suggest several heuristic guidelines that may help the process along.

Simple unidimensional models

To provide a general introductory context for discussing what is meant by internal cross-validation and how it can be carried out informally in a familiar behavioral research paradigm, suppose we have a set of five attitudinal statements regarding the rights of suspects. It is conjectured that these statements fall along a graded continuum from left to right, with the left representing a greater protection of individual liberties and, for the lack of a better label, the right representing a greater protection of society's concerns. For each pair of statements, a proportion, p_{ij}, of individuals is available representing the number who have assessed statement i as being more protective of individual rights than statement j; we assume $p_{ij} + p_{ji} = 1$. Thus, the complete set of proportions can be organized into a matrix $\underset{\sim}{P}$ of the form

Statement	1	2	3	4	5
1	X	p_{12}	p_{13}	p_{14}	p_{15}
2	p_{21}	X	p_{23}	p_{24}	p_{25}
3	p_{31}	p_{32}	X	p_{34}	p_{35}
4	p_{41}	p_{42}	p_{43}	X	p_{45}
5	p_{51}	p_{52}	p_{53}	p_{54}	X

If the spatial model is appropriate and the five statements are correctly ordered according to their numerical indices, i.e., statement 1 is most supportive of individual rights and statement 5 least supportive, then the manifest proportions should reflect this ordering up to a reasonable level

of random variability or error. In particular, the propositions should become more extreme as the distance between the statements increases and whenever statement i is placed to the left of j, p_{ij} should be larger than p_{ji}. Furthermore, within each row of P, the entries to the right of the main diagonal should all be greater than or equal to .50 and increase (or at least never decrease) moving away from the diagonal. Conversely, the entries to the left of the main diagonal should all be less than or equal to .50 and decrease (or at least never increase) moving to the left. As a simple example, the matrix given below displays a perfect pattern:

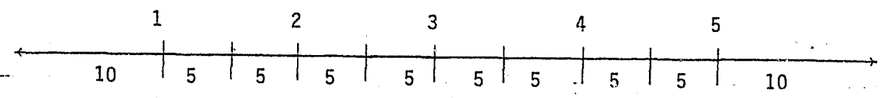
Statement	1	2	3	4	5
1	X	.55	.60	.65	.70
2	.45	X	.55	.60	.65
3	.40	.45	X	.55	.60
4	.35	.40	.45	X	.55
5	.30	.35	.40	.45	X

Since simple spatial models of the type just proposed place very severe constraints on the manifest data, it is conceivable that these constraints could also help in verifying the reasonableness or unreasonableness of the model itself. For instance, if a matrix of proportions is reordered as well as possible to obtain an appropriate gradient on only the above diagonal entries, the basic spatial model that assumes the proportions should become less extreme as distances decrease should also produce other patterns in the reordered matrix. Specifically, the set of entries either above or below the main diagonal should all be greater than or equal to .50 and

moving to the left away from the main diagonal within a row, the entries should display a gradient opposite that obtained moving toward the right. In other words, if we reorder the matrix to obtain an explicit gradient above the main diagonal, the degree to which these other two properties also hold is an informal indication of the correctness of the basic model. For a more general discussion of these spatial assumptions, the reader should consult Coombs (1956; 1964).

To illustrate the important point that not all matrices of proportions must necessarily demonstrate the characteristics we would be looking for, it is easy to construct an alternative spatial scheme that generates proportions having an above-diagonal gradient but not the other two properties. In particular, suppose the proportions follows a Coombsian model in which the statements are ordered from 1 to 5 along a continuum but the 60 individuals who generate the proportions are also placed along the scale according to the following distribution (indicated below the line):

Statement



Number of subjects in various intervals

Each subject chooses or endorses statement i over statement j if his/her (ideal) point is closer to i than j. Summing over all 60 individuals, the following matrix of proportions would be obtained:

Statement	1	2	3	4	5
1	X	.25	.33	.42	.50
2	.75	X	.42	.50	.58
3	.67	.58	X	.58	.67
4	.58	.50	.42	X	.75
5	.50	.42	.33	.25	X

Even though the necessary above-diagonal gradient is achieved, the gradient below the diagonal is the exact opposite of what was expected before; moreover, not all the above- (or below-) diagonal entries are greater than or equal to .50. The above-diagonal gradient is perfect but the failure of the other two conditions suggests correctly that all subjects do not respond to the statements in exactly the same way. In fact, if we initially believe that our proportions follow a Coombsian model of this type, an informal evaluation of the conjecture could be developed around the presence of a below-diagonal gradient that is the same as the one forced on the above-diagonal elements. For a more complete discussion of the patterns in a Coombsian model of this latter form, the reader is referred to Greenberg (1965).

Extensions to general asymmetric proximities

Although the illustration just used deals with a rather specific context, the procedures generalize easily to arbitrary asymmetric measures of proximity. For example, suppose our objects are spatial locations and q_{ij} represents the degree of interaction or flow from i to j, e.g., a function of the number of people who move from i to j, and possibly, the interpoint distance

between i and j. In general, if the basic asymmetric measure q_{ij} is defined appropriately in terms of the unidimensional spatial model being assumed, the skew-symmetric measure $q^*_{ij} \equiv q_{ij} - q_{ji}$ can play the role of the previous proportions. Again, the conditions we would look for are algebraic, and real data may only approximately represent these perfect patterns even when our conjecture of a particular underlying spatial model is reasonably correct. Error-free algebraic conditions, however, still provide a perspective on what to look for in assessing whether a conjectured model is contradicted by the data.

Given the usual spatial model without individual ideal points, it is apparent that the proportions p_{ij}, or the possibly more general skew-symmetric measures q^*_{ij}, contain two types of information. The signs of $p_{ij} - p_{ji}$ or of q^*_{ij} indicate the ordering of object i relative to j from left to right; the magnitudes $|p_{ij} - p_{ji}|$ or $|q^*_{ij}|$ indicate the degree of spatial separation between i and j. If we relied on the sign information only and reordered the matrix to achieve all +'s, say, above the main diagonal, we would then expect the entries to display particular patterns when reordered in the same way, e.g., the values of $|p_{ij} - p_{ji}|$ or $|q^*_{ij}|$ within each row should never decrease moving away from the main diagonal in either direction. Conversely, if an appropriate pattern can be achieved in, say, the absolute values, then based on the sign information, all the +'s should be either above or below the main diagonal (due to the symmetric nature of the matrix of absolute values, the position of the +'s in this last case is not specified and the direction of the ordering is arbitrary). In summary, an effective assessment strategy would force one pattern using the signs or absolute values and then proceed to evaluate the degree to which the second pattern is also achieved. Compared to many of the data-reduction strategies in the behavioral sciences, this procedure is

allowed to fail and carries with it the possibility of at least an informal internal cross-validation.

The type of informal assessment procedure we suggest can be described rather succinctly. It is conjectured that a particular model holds and will generate an (approximately) perfect matrix pattern characterized by a set of necessary algebraic conditions. If one such condition is selected and used as an optimization criterion, the degree to which the other properties are also satisfied is an informal indication of the validity of the basic model. Ideally, the necessary conditions are selected in such a manner that optimizing with respect to one does not automatically optimize a second unless the model is reasonably correct. At least intuitively, for example, the gradient conditions using the original proportions within rows above and below the main diagonal, or the sign and absolute-value information for $p_{ij} - p_{ji}$ or q^*_{ij}, satisfy such a general independence condition. From a more operational perspective, we could first use the absolute-value data to obtain one seriation by any method that uses a symmetric matrix as input; the sign information would then provide an appropriate source of information for evaluating the resulting scale. This general tactic will be the one used later.

Some Computational Details on Informal Cross-Validation

The obvious computational problem posed by the scheme of informal cross-validation is to reorder the matrix of proximities to display one of the patterns expected under the assumed model. For example, if the sign and magnitude information are considered separately in a matrix that originally contains q^*_{ij} in the i^{th} row and j^{th} column, two different optimization tasks can be defined - using signs, the matrix could be reordered to force all

+'s above the main diagonal, or in terms of absolute magnitudes, the matrix could be reordered to produce the desired gradient within each row. In either case, a host of computational strategies is available, ranging from integer programming and branch-and-bound to eigenvector analyses and heuristic methods. For reviews the reader is referred to Hubert (1976), Baker and Hubert (1977), and Hubert and Schultz (1976).

Since data sets that would be of interest in the criminal justice area may be very large, we present only one well-known alternative for scaling symmetric matrices and use it in our later example. We assume that s_{ij} denotes a general symmetric measure of proximity (e.g., the absolute values $|p_{ij} - p_{ji}|$ or $|q^*_{ij}|$) and $A_1, A_2, \ldots, A_n$ denote positions for the n objects along a continuum. Obviously, our concern is with estimating the A_i's.

As one approach, suppose f is some function and we wish to estimate $A_1, A_2, \ldots, A_n$ such that

$$\sum_{i,j} f(s_{ij})(A_j - A_i)^2$$

is maximized subject to the normalization constraint $\sum_{i=1}^{n} A_i^2 = 1$. The solution is the largest eigenvector (normalized to unit length) of the matrix

$$\begin{bmatrix}
\sum_j f(s_{1j}) & -f(s_{12}) & -f(s_{13}) & \cdots & -f(s_{1n}) \\
-f(s_{12}) & \sum_j f(s_{2j}) & -f(s_{23}) & \cdots & -f(s_{2n}) \\
-f(s_{13}) & -f(s_{23}) & \sum_j f(s_{3j}) & \cdots & -f(s_{3n}) \\
\vdots & \vdots & \vdots & & \vdots \\
-f(s_{1n}) & -f(s_{2n}) & -f(s_{3n}) & \cdots & \sum_j f(s_{nj})
\end{bmatrix}$$

and the largest eigenvalue is the maximum (see Hall, 1970; Guttman, 1968; Morse, 1972). The sum of the estimates can be shown equal to 0. Obviously, the estimates provide a seriation of the n objects along the single dimension in addition to spacing information. Although we are now removed from an explicit consideration of algebraic conditions, it is hoped that the seriation induced by the eigenvector analysis would still lead to an appropriate pattern on the signs. This latter information was not used in obtaining the estimates for $A_1, A_2, \ldots, A_n$, and consequently, it provides a source of data that could be used in verifying the adequacy of the eigenvalue seriation.

Formal Cross-Validation

In our previous discussion of informal cross-validation, it was hoped that optimizing one property would also lead to the optimization of a second. Although the degree to which this goal is achieved can be evaluated more or less intuitively, it may be straightforward to proceed one step further and carry out a formal test. For example, suppose $T = \{t_{ij}\}$ denotes an arbitrary skew-symmetric matrix-- t_{ij} may represent q^*_{ij} or be defined as $p_{ij} - p_{ji}$ if we use proportions. When T is reordered on the basis of absolute-value information only, it is possible to assess statistically whether a preponderance of +'s or -'s appear above the main diagonal. More explicitly, the number of +'s can be compared to what is expected under the null conjecture that the ordering of T produced by optimizing the first property was actually chosen at random from all possible orderings. In the discussion that follows we emphasize only a test on signs even though the same principles could be used more generally for other consistency measures once some seriation is generated by optimizing a second property. The test on signs is particularly simple to implement and has some very nice relationships back to the literature on nonparametric statistics.

To be more explicit about the formal inference problem on signs, suppose we define, based on the matrix T, a "sign" matrix $A = \{a_{ij}\}$:

$$a_{ij} = \begin{cases} +1 & \text{if } t_{ij} > 0; \\ 0 & \text{if } t_{ij} = 0; \\ -1 & \text{if } t_{ij} < 0. \end{cases}$$

Furthermore, let S denote the sum of the above-diagonal elements in A when A is rearranged to conform to the best reordering of T constructed from absolute-value information. If we assume, under a null model, that the sign data bears no relationship to the absolute-value data, then S should not be unusually extreme compared to its distribution when A is reordered at random. Conversely, large positive or negative values of S would imply a consistency in the sign and absolute-value data that in turn would lend some formal credence to the reasonableness of the underlying spatial model.

From published Monte Carlo studies (see Hubert and Schultz, 1975) and discussions of similar statistics that appear in the literature (e.g., Ager and Brent, 1978), it appears that under the conjecture of randomness and for reasonably large n, the S statistic can be considered normal with mean zero and a variance defined by

$$V(S) = \frac{1}{6}\{\sum_{i,j} a_{ij}^2 + 2\sum_i [\sum_j a_{ij}]^2\}.$$

[A formal proof of asymptotic normality could follow Kendall's discussion (Kendall, 1970, p. 72-74) assuming $\sum_{i,j,k} a_{ij}\, a_{ik}$ is of order n^3].

As a final descriptive measure of the degree to which the matrix A is reordered appropriately, a measure suggested by Ager and Brent (1978) can be adopted:

$$D = \frac{|S|}{\sum_{i<j} a_{ij}^2}.$$

Here, $|S|$ denotes the absolute difference between the number of +'s and -'s above the diagonal. The term $\sum_{i<j} a_{ij}^2$ is the number of above-diagonal non-zero entries, and thus, defines the maximum value of the numerator.

Several special cases of these formulas deserve particular mention. If there are no off-diagonal zeros in A and no intransitivities (i.e., an intransitivity is a triple $\{i,j,k\}$ for which $a_{ij} = +1$, $a_{jk} = +1$ and $a_{ik} = -1$), then our inference problem is equivalent to the comparison of two united rankings based on Kendall's tau statistic. The numerator of tau is S and the variance term reduces to $n(n-1)(2n+5)/18$, which is the standard expression used for a significance test (Kendall, 1970). The general variance term given above is also appropriate when there are ties in one of the rankings and this ranking defines the matrix A, or even when intransitivities exist in A. A particularly simple formula results when there are T intransitive triples and no off-diagonal zeros: $\frac{1}{18}(n(n-1)(2n+5)-48T)$. Obviously, for $T = 0$ this latter formula reduces to the standard expression for untied rankings. Also, when T is 0, the well-known Goodman-Kruskal Gamma statistic is equivalent to our measure D up to an absolute value.

Example

As an illustration of the ideas we have just presented, Table 1 gives a matrix of proportions among 15 of the larger SMSA's. The entries in Table 1 were obtained from a group of 101 undergraduate and graduate students at the University of California, Santa Barbara by first asking each person to rank the 15 SMSA's in terms of perceived homicide rate per 100,000, and then evaluating the number of times a particular SMSA was ranked lower than a second. For instance, since the Boston (row) - Buffalo (column) proportion is .32, 32 percent of the students ranked Boston lower than Buffalo and

64 percent ranked Boston higher than Buffalo. The order of the SMSA's in the rows and columns was obtained from the eigenvalue analysis presented earlier using the identity function for $f(\cdot)$ and the absolute difference matrix generated from Table 1. These estimated locations $A_1, A_2, \ldots, A_{15}$ are also given in the table.

As is apparent from Table 1, the gradient conditions on the absolute differences are almost perfect. Similarly, the sign matrix clearly represents an abundance of +'s above the main diagonal (in fact, the pattern in this case is exact). Although somewhat obvious here, the index value D of 1.00 generates a highly significant Z-value of 5.20, based on a variance of 408.33 for S. In summary, a strong unidimensional scale underlies the Table 1 data; the pattern is very clear when the SMSA's are ordered according to the estimates $A_1, A_2, \ldots, A_{15}$ generated from the eigenvector analysis.

Table 1 here

Since the scale just demonstrated is subjective, it is of some substantive interest to compare this aggregate perception with reality. Based on the homicide rate per 100,000 in 1977, the 15 SMSA's should have been ordered as follows (rates are given in brackets):

> Los Angeles-Long Beach (16.0), Miami (15.6), Cleveland (14.2), Chicago (13.7), San Francisco-Oakland (11.9), Baltimore (10.2), Philadelphia (8.9), Cincinnati (7.5), San Diego (6.8), Buffalo (6.0), Pittsburgh (4.8), Milwaukee (4.5), Seattle-Everett (4.3), Anaheim-Santa Ana-Garden Grove (4.0), Boston (3.6).

Obviously, some SMSA's are perceived as having a much higher homicide rate than they actually do (e.g., Boston and Pittsburg), and some are underrated (e.g., Miami). Overall, however, there is a fair degree of consistency in the objective and subjective orderings. The Spearman rank order correlation between the two rankings for the 15 SMSA's is .507, which is significant

TABLE 1

Matrix of Proportions Among 15 SMSA's Based on 101 Subjects:
The Rows and Columns are Ordered According to the Eigenvector Estimates of Location

	Eigenvector (A_1's)	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	(m)	(n)	(o)
(a): Seattle-Everett	-.26	X	.63	.63	.76	.69	.78	.83	.85	.87	.92	.93	.89	.92	.92	.97
(b): San Diego	-.18	.37	X	.51	.63	.62	.67	.76	.83	.84	.87	.86	.86	.86	.93	.96
(c): Anaheim-Santa Ana-Garden Grove	-.17	.37	.49	X	.58	.59	.68	.74	.81	.83	.80	.84	.80	.88	.90	.97
(d): Milwaukee	-.14	.24	.37	.42	X	.54	.58	.66	.72	.77	.76	.84	.84	.82	.88	.97
(e): Miami	-.10	.31	.38	.41	.46	X	.54	.73	.63	.71	.67	.73	.72	.74	.83	.94
(f): Buffalo	-.08	.22	.33	.32	.42	.46	X	.55	.55	.64	.68	.70	.72	.75	.80	.92
(g): Baltimore	-.07	.17	.24	.26	.34	.27	.45	X	.51	.62	.59	.65	.72	.70	.79	.93
(h): Cincinnati	-.05	.15	.17	.19	.28	.37	.45	.49	X	.62	.63	.72	.68	.68	.81	.88
(i): Cleveland	-.02	.13	.16	.17	.23	.29	.36	.38	.38	X	.52	.54	.56	.54	.68	.86
(j): Boston	-.02	.08	.13	.20	.24	.33	.32	.41	.37	.48	X	.53	.59	.61	.72	.87
(k): Pittsburgh	+.01	.07	.14	.16	.16	.27	.30	.35	.28	.46	.47	X	.55	.62	.66	.82
(l): Philadelphia	+.02	.11	.14	.20	.16	.28	.28	.28	.32	.44	.41	.45	X	.55	.66	.78
(m): San Francisco-Oakland	+.03	.08	.14	.12	.18	.26	.25	.30	.32	.46	.39	.38	.45	X	.61	.77
(n): Los Angeles-Long Beach	+.11	.08	.07	.10	.12	.17	.20	.21	.19	.32	.28	.34	.34	.39	X	.70
(o): Chicago	+.90	.03	.04	.03	.03	.06	.08	.07	.12	.14	.13	.18	.22	.23	.30	X

(one-tailed) at the usual .05 level (e.g., a Z-statistic of 1.90 is generated).

Discussion and Extensions

Given the basic context of unidimensional seriation as developed in the earlier sections, a variety of additional topics could be pursued. We mention only a few in passing to give some indication of the current research efforts in this direction. For example, the type of inference strategy that was proposed for evaluating the pattern of signs can be extended to compare two arbitrary skew-symmetric matrices (see Hubert and Schultz, 1976). Thus, it is possible to evaluate the consistency between two skew-symmetric interaction matrices where the latter may be based on migration data at two time points or from two different demographic subgroups. Secondly, from a combinatorial optimization point of view, several very elegant theoretical paradigms have been introduced recently for characterizing a discrepancy between a given seriation and the original asymmetric data, e.g., see Bowman and Colantoni (1973) and Merchant and Rao (1976). Along these same combinatorial optimization lines, a general strategy has been suggested (Hubert, 1980) for locating and seriating only a part of a proximity matrix that appears to be most consistent with the basic underlying spatial model. This latter technique can assist in identifying subsets of an object set that can be seriated well and those subsets that are not represented satisfactorily along a continuum.

As one example of particular importance we note that the topic of criminal mobility could define one of the more interesting applications for unidimensional scaling in the criminal justice area. For example, based on movement data from place of residence to place of the committed crime, we may wish to rate a set of geographical areas in terms of criminal attractivity, with the possible goal of comparing these rates over different crime types, age groups,

and so on. Typically, the basic data are flow statistics for a set of n localities defined by the number of people, m_{ij}, who travel from region i to j. Our aim is to model these data in terms of the distances among the localities and their assumed placement along an attractivity continuum.

Following Tobler's (1979) lead, the simplest model we consider is defined in terms of the skew-symmetric matrix $q^*_{ij} = m_{ij} - m_{ji}$, which specifies the degree to which j atracts more from i than it exports. We assume that these statistics conform to a model defined as

$$q^*_{ij} \simeq \frac{A_j - A_i}{d_{ij}} ,$$

where d_{ij} is the distance between locations i and j and $A_1, A_2, \ldots, A_n$ define attractivities along a single dimension. Obviously, since distances are typically known, our analysis task is to estimate the n attractivities, which in turn scales the n localities according to attractivity along a continuum.

Tobler (1979) discusses in detail two major approaches to the estimation of attractivities. The first is called the potential method in which the A_i's are given implicitly by the matrix equation

$$\begin{bmatrix} \sum_{i \neq 1} \frac{1}{d_{i1}} & -\frac{1}{d_{12}} & -\frac{1}{d_{13}} & \cdots & -\frac{1}{d_{1n}} \\ -\frac{1}{d_{21}} & \sum_{i \neq 2} \frac{1}{d_{i2}} & -\frac{1}{d_{23}} & \cdots & -\frac{1}{d_{2n}} \\ \vdots & & & & \\ -\frac{1}{d_{n1}} & -\frac{1}{d_{n2}} & -\frac{1}{d_{n3}} & \cdots & \sum_{i \neq n} \frac{1}{d_{in}} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix} = \begin{bmatrix} \sum_i q^*_{i1} \\ \sum_i q^*_{i2} \\ \vdots \\ \sum_i q^*_{in} \end{bmatrix} .$$

Since this system is not of full rank, Tobler suggests letting $A_1 = 0$ and deleting the first row and column of the coefficient matrix. The system can then be solved by inverting the reduced coefficient matrix and using it as a premultiplier on the right-hand vector. The second procedure may generate different estimates since it is based on minimizing the least-squares criterion:

$$\sum_{i,j} (q^*_{ij} - \frac{A_j - A_i}{d_{ij}})^2. \tag{2}$$

Again, a system of equations similar to that given above is generated that is not of full rank. Based on the coefficient matrix having diagonal entries

$$\sum_{i=1}^{n} (\frac{1}{d^2_{ki}} + \frac{1}{d^2_{ik}})$$

and off-diagonal entries

$$-(\frac{1}{d^2_{ik}} + \frac{1}{d^2_{ki}}),$$

and the right-hand vector defined by an arbitrary entry

$$\sum_{\substack{i=1 \\ i\neq k}}^{n} (\frac{q^*_{ik}}{d_{ik}} - \frac{q^*_{ki}}{d_{ki}}),$$

the deletion option for $A_1 \equiv 0$ can then be used to obtain a closed-form solution through matrix inversion.

Although Tobler's discussion is very elegant, seemingly minor modifications in the way the model is stated will eventually lead to several useful simplifications. In particular, since we assume q^*_{ij} should be "close" to $(A_j - A_i)/d_{ij}$ and the d_{ij}'s are known, it should also be true that $d_{ij}\, q^*_{ij}$ is "close" to $(A_j - A_i)$. Thus, the original gradient notion characterized as a division

by d_{ij} is redefined by a distance weighting of the observed skew-symmetric proximities. Continuing in this way it should also be true that $(d_{ij}q^*_{ij})^2$ is "close" to $(A_j - A_i)^2$. Our problem is now reduced to fitting the entries in a symmetric matrix $\{(d_{ij}q^*_{ij})^2\}$ by a squared distance matrix $\{(A_j - A_i)^2\}$. This latter task can be approached by the type of eigenvector analysis introduced earlier. Or, if we assume $|d_{ij}q^*_{ij}| \approx |A_j - A_i|$, the strategies developed by Defays (1978) and De Leeuw and Heiser (1977) could be followed.

By taking explicit advantage of a model equivocation, we can concentrate on the skew-symmetric proximities $t_{ij} \equiv d_{ij}q^*_{ij}$, which are supposedly defined by the simple differences between attractivities. For example, a least-squares loss-function would minimize

$$\sum_{i,j} (t_{ij} - (A_j - A_i))^2,$$

which is equal to a criterion weighted by the squared distances d^2_{ij}:

$$\sum_{i,j} d^2_{ij}(q^*_{ij} - (\frac{(A_j - A_i)}{d_{ij}})^2.$$

The least-squares measure used by Tobler in (2) is similar in general form to this latter expression but is unweighted.

The redefinition of the estimation problem to use t_{ij} may seem trivial but it leads immediately to several convenient results. For example, the least-squares estimate of A_i subject to the constraint that $\sum_{i=1}^{n} A_i = 0$ is given by $t_{\cdot i}/n$. This same estimate is also obtained by the potential method, and consequently, both methods lead to the same solution in this context. From a slightly more general perspective, suppose we fit a matrix of the form $\{C_j - C_i\}$ to our arbitrary skew-symmetric matrix $\{t_{ij}\}$ by maximizing the

correlation between the corresponding entries. Again, the solution is obtained when C_i is defined as $t_{.i}/n$, and when used to define the matrix $\{C_j - C_i\}$, these values induce a correlation of

$$\sqrt{\frac{2}{n}\frac{\sum_j\left[\sum_i t_{ij}\right]^2}{\sum_j\sum_i t_{ij}^2}}$$

All of these last results are very close to some work by Noether (1960) on paired comparison scaling.

It should be apparent that many different approaches could be developed for estimating the attractivities $A_1, \ldots, A_n$ from a skew-symmetric matrix t_{ij} or q^*_{ij} merely by varying the explicit form of the model used and the loss function. This arbitrariness is troublesome since minor variations can dramatically affect the final estimation process (see Noether, 1960). To provide some hedge, and as we have suggested before, it may be appropriate to rely only on the absolute-value data in the estimation stage and use the sign data as a strategy for validating the order of the estimates along the continuum. Sign information has the nice property of being independent of the form of the gradient model being used as long as the signs are governed by differences in attractivities.

References

Ager, J.W., and Brent, S.B. An index of agreement between a hypothesized partial order and an empirical rank order. Journal of the American Statistical Association, 1978, 73, 827-830.

Baker, F.B., and Hubert, L.J. Applications of combinatorial programming to data analysis: Seriation using asymmetric proximity measures. British Journal of Mathematical and Statistical Psychology, 1977, 30, 154-164.

Bowman, V.J., and Colantoni, C.S. Majority rule under transitivity constraints. Management Science, 1973, 19, 1029-1041.

Coombs, C.H. The scale grid: Some interrelations of data models. Psychometrika, 1956, 21, 313-329.

_____. A Theory of Data. New York: Wiley, 1964.

Defays, D. A short note on a method of seriation. British Journal of Mathematical and Statistical Psychology, 1978, 31, 49-53.

De Leeuw, J., and Heiser, W. Convergence of correction-matrix algorithms for multidimensional scaling. In Lingoes, J.C. (ed.), Geometric Representations of Relational Data. Ann Arbor, Michigan: Mathesis Press, 1977.

Greenberg, M.G. A method of successive cumulations for scaling of pair-comparison preference judgments. Psychometrika, 1965, 30, 441-448.

Guttman, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. Psychometrika, 1968, 33, 459-506.

Hall, K.M. An r-dimensional quadratic placement algorithm. Management Science, 1970, 17, 219-229.

Hubert, L.J. Seriation using asymmetric proximity measures. British Journal of Mathematical and Statistical Psychology, 1976, 29, 32-52.

_____. Analyzing proximity matrices: The assessment of internal variation in combinatorial structure. Journal of Mathematical Psychology, in press, 1980.

Hubert, L.J., and Schultz, J.V. Maximum likelihood paired-comparison ranking and quadratic assignment. Biometrika, 1975, 62, 655-659.

_____. Quadratic assignment as a general data analysis strategy. British Journal of Mathematical and Statistical Psychology, 1976, 29, 190-241.

Kendall, M.G. Rank Correlation Methods. New York: Hafner, 1970.

Merchant, D.K., and Rao, M.R. Majority decisions and transitivity: Some special cases. Management Science, 1976, 23, 125-130.

Morse, P.M. Optimal linear ordering of information items. Operations Research, 1972, 20, 741-751.

Noether, G.E. Remarks about a paired comparison model. Psychometrika, 1960, 25, 357-367.

Tobler, W.R. Estimation of attractivities from interactions. Environment and Planning A, 1979, 11, 121-127.

Proximity Matrix Reorganization and Hierarchical Clustering*

Lawrence J. Hubert, Reginald G. Golledge,
G.D. Richardson

The University of California, Santa Barbara

* Partially supported by LEAA/NILCJ 79-NI-AX-0057 and NSF SOC-7728227.

Proximity Matrix Reorganization and Hierarchical Clustering

Introduction

In recent years various substantive journals in the social and behavioral sciences have published a number of empirical studies that rely on the newer data reduction techniques of hierarchical clustering (HC). For one example in the field of criminal justice the reader is referred to Megargee and Bohn's (1979) extensive typology of criminals based on MMPI profiles. In much of this work, it is common for a researcher to present the results of a cluster analysis in a form produced more or less directly by the output of some standard computer program. In fact, certain representations have now become so well-known in the literature that a novice might easily confuse what is a rather arbitrary graphic device with the operation of the method itself. Our interest will be in one such strategy for reporting the results of a HC that can be defined by a particular reorganization of the proximity matrix used in the original analysis. This reordering scheme, besides being a very convenient way for the researcher to report the results of a HC, leads to a basic theoretical connection between the clustering process and several areas of current interest in operations research that study the partitioning of "flow" matrices. In turn, the relationship to flow suggests a very interesting commonality between HC and another data analysis problem concerned with unidimensional scaling or seriation. It is this latter relationship that can be exploited as an effective data analysis scheme.

The basic intent of what we have to say appears in several related (but informal) discussions, such as Sneath and Sokal (1973), Johnson (1972), Forsyth and Katz (1946), or as far back as Cattell (1944). The strategy to be reviewed and made more precise requires a preliminary

reorganization of the original matrix of proximity values to "fit" as close as is possible a particularly convenient form, and secondly, an inspection of the reordered matrix to suggest a reasonable HC that may have a convincing substantive justification. Our interest is not to define another new clustering method touted to be the best, but instead, to suggest how an investigator can organize and rely on his/her own intuition in a relatively efficient manner. Even though it may be easier to depend on an arbitrary computer program that analyzes a proximity matrix in some mysterious way and accept the results of such an analysis without question, this mechanical process may also limit the possibility of examining a given data set in its entirety and could conceivably lead to inappropriate interpretations. Our goal is similar to that now being stressed in areas such as location-allocation modelling (Adrian, 1979; Church, 1980); i.e., to reemphasize the importance of insight and commonsense and to develop a procedure that could help this judgemental process along. Since cluster analysis methods typically give precise answers to data analysis tasks that are not very well defined, it should be standard policy to reorganize the manifest proximity matrix itself as a way of verifying the presence of an assumed clustering pattern.

In areas such as criminal justice, the importance of developing simple procedures for presenting the results of research in a form that is close to the raw data cannot be overstated. Typically, the audience for which such research has serious policy implications does not have the same methodological sophististication as the original investigator. Thus, the research conclusions could either be dismissed since the analysis is not very understandable, or at the other extreme, accepted too uncritically. We consider it a virtue to reduce raw data as little as possible and yet have it in a form that facilitates a substantive argument. And it appears that the use of seriation as an adjunct to HC has this latter characteristic.

Background

Hierarchical Clustering (HC)

As a very brief summary of the purpose behind HC, suppose S is a set of n objects, $\{O_1, O_2, \ldots, O_n\}$, and $s(\cdot,\cdot)$ is some symmetric non-negative (proximity) function defined on SxS that specifies the relationship between object pairs. For example, $s(O_i, O_j)$ may denote some type of correlation or distance measure between O_i and O_j. Given this latter index of correspondence, a specific HC procedure, such as the well-known complete-link technique (cf., Johnson, 1967), will reproduce a sequence of partitions of S, denoted by $\mathcal{L}_0, \mathcal{L}_1, \ldots, \mathcal{L}_K$. The first partition $\mathcal{L}_0$ contains n object classes each including a single element of S, $\mathcal{L}_K$ contains a single object class encompassing all of the members of S, and $\mathcal{L}_{k+1}$ is constructed by uniting certain of the subsets from $\mathcal{L}_k$. Typically, K will equal n-1 and only two subsets in $\mathcal{L}_k$ will be united at any one time to form the new subset $\mathcal{L}_{k+1}$. For convenience, we assume the proximity function $s(\cdot,\cdot)$ is defined in such a way that large proximity values are assigned to dissimilar object pairs.

Given a sequence of partitions $\mathcal{L}_0, \mathcal{L}_1, \ldots, \mathcal{L}_K$, a second function $d(\cdot,\cdot)$ can be constructed on SxS that has the properties of an ultrametric. Formally, let

(1) $d(O_i, O_j) = \min \{k \mid O_i \text{ and } O_j \text{ are first placed together in a single subset in } \mathcal{L}_k\}$,

then, using this definition, $d(\cdot,\cdot)$ can be shown to satisfy the following four properties, the last of which is clearly the most restrictive:

i) $d(O_i, O_j) \geq 0$;

ii) $d(O_i, O_j) = 0$ if and only if $O_i = O_j$;

iii) $d(O_i, O_j) = d(O_j, O_i)$;

iv) $d(O_i, O_j) \leq \max \{d(O_i, O_k), d(O_k, O_j)\}$.

It should be apparent that any strictly monotone increasing function of $d(\cdot,\cdot)$ that maps zero to zero would also be an ultrametric.

If the original proximity function $s(\cdot,\cdot)$ has the strict monotone property mentioned above with respect to $d(\cdot,\cdot)$, then $s(\cdot,\cdot)$ is itself an ultrametric and the sequence of partitions $\mathcal{L}_0, \mathcal{L}_1, \ldots, \mathcal{L}_K$ represents a perfect structure for $s(\cdot,\cdot)$. An example is given by the matrix of Table 1. This proximity function satisfies the four properties for an ultrametric and corresponds to the following sequence of partitions:

$\mathcal{L}_0 = \{\{O_1\}, \{O_2\}, \{O_3\}, \{O_4\}, \{O_5\}, \{O_6\}, \{O_7\}\}$

$\mathcal{L}_1 = \{\{O_1, O_2, O_3\}, \{O_4\}, \{O_5\}, \{O_6\}, \{O_7\}\}$

$\mathcal{L}_2 = \{\{O_1, O_2, O_3, O_4\}, \{O_5\}, \{O_6\}, \{O_7\}\}$

$\mathcal{L}_3 = \{\{O_1, O_2, O_3, O_4\}, \{O_5, O_6, O_7\}\}$

$\mathcal{L}_4 = \{\{O_1, O_2, O_3, O_4, O_5, O_6\}\}$

Table 1 here

For this set of partitions, the function $d(\cdot,\cdot)$ is given in Table 2, and as should be clear after inspection, is monotone with respect to $s(\cdot,\cdot)$.

Table 2 here

Most proximity functions that would be encountered in actual data analysis situations will not satisfy an ultrametric condition perfectly. Nevertheless,

TABLE 1

Illustrative Proximity Function Having
the Ultrametric Property

	O_1	O_2	O_3	O_4	O_5	O_6	O_7
O_1	0	2.30	2.30	4.15	6.35	6.35	6.35
O_2	2.30	0	2.30	4.15	6.35	6.35	6.35
O_3	2.30	2.30	0	4.15	6.35	6.35	6.35
O_4	4.15	4.15	4.15	0	6.35	6.35	6.35
O_5	6.35	6.35	6.35	6.35	0	6.01	6.01
O_6	6.35	6.35	6.35	6.35	6.01	0	6.01
O_7	6.35	6.35	6.35	6.35	6.01	6.01	0

TABLE 2

Ultrametric Function $d(\cdot,\cdot)$ Derived From
the Sequence of Partitions Given in the
Text for the n=7 Illustration

	O_1	O_2	O_3	O_4	O_5	O_6	O_7
O_1	0	1	1	2	4	4	4
O_2	1	0	1	2	4	4	4
O_3	1	1	0	2	4	4	4
O_4	2	2	2	0	4	4	4
O_5	4	4	4	4	0	3	3
O_6	4	4	4	4	3	0	3
O_7	4	4	4	4	3	3	0

any HC method will still attempt to locate a "good" sequence of partitions usually by some set of heuristic rules, based on whatever proximities are available. For our purposes, it is important to note that the resultant sequence of partitions produced by any HC procedure could still be used to obtain a function $d(\cdot,\cdot)$ as in (1). Furthermore, the degree to which $d(\cdot,\cdot)$ is monotone with respect to the original function $s(\cdot,\cdot)$ could be formalized by some index and used to measure the extent to which the imperfect patterning of the original proximity values matches the perfect ultrametric structure of the integers assigned by $d(\cdot,\cdot)$. For a more complete development of this goodness-of-fit notion, the reader is referred to Baker and Hubert (1975).

Partitioning matrices

Given an arbitrary proximity function $s(\cdot,\cdot)$ and the sequence of partitions constructed by some HC, one rather well-known graphic technique used in representing the analysis rearranges the original proximity matrix in a very specific way (cf. BMDP program manual: Dixon, 1977, p. 621-632). In particular, suppose a HC is carried out and the defining ultrametric function $d(\cdot,\cdot)$ is obtained. If the values assigned by $d(\cdot,\cdot)$ are organized into the form of an nxn matrix $\underset{\sim}{D} = \{d(O_i, O_j)\}$, then there exists an ordering (not unique) of the rows and simultaneously the columns of $\underset{\sim}{D}$ that will put the reordered matrix into a form with the following properties:

If $\underset{\sim}{D}^*$ is an appropriate rearrangement of $\underset{\sim}{D}$, then

i) $\underset{\sim}{D}^*$ can be partitioned as

$$\underset{\sim}{D}^* = \left[\begin{array}{c|c} \underset{\sim}{D}_{11}^* & \underset{\sim}{D}_{12}^* \\ \hline \underset{\sim}{D}_{21}^* & \underset{\sim}{D}_{22}^* \end{array}\right],$$

where all the elements of D_{12}^* and D_{21}^* are equal to the single largest element of D;

ii) the submatrices D_{11}^* and D_{22}^* are once again partitionable as in (i);

iii) the partitioning process can be repeated until all the resulting submatrices are of order 1.

Given the reorganized matrix D^*, the original proximity matrix, say $P = \{s(O_i, O_j)\}$, can also be restructured by using the same row and column ordering. If for notational purposes this reordered matrix is denoted by P^*, then the degree to which P^* does not have the exact same partition structure as did D^* will give some indication of adequacy for the hierarchical clustering represented by the ultrametric function $d(\cdot,\cdot)$. From a more practical point of view, it is of interest to note that many of the standard programs for hierarchical clustering, such as the routines given in BMDP (see Dixon, 1977) and by Johnson (1967), automatically provide an object ordering to transform D to D^*. Consequently, an inspection of a reordered proximity matrix provides no extra computational burden for the applied user.

Although we will not go into any detail, the general type of partitioning operation performed on D^* is also discussed in electrical engineering under the title of "principle partitioning", and specifically, in stating an equivalence between a terminal capacity matrix for a flow network and what are essentially ultrametric conditions on the entries in a matrix (the proximities are keyed in the opposite direction for flow, but the basic ideas are the same. For a more complete discussion, see Frank and Frisch, 1971). It should also be noted that many orderings can exist for transforming D to D^* and our discussion only assumes that any one such ordering is desired. For instance, if K equals n-1 in the original sequence of partitions, then 2^{n-1} possible row (and column) orderings of the objects in S would produce a D^* of the necessary form.

Object sequencing

Once the ultrametric values assigned by $d(\cdot,\cdot)$ are organized into a matrix that has the D^* form, the entries themselves display an interesting pattern that has great importance in unidimensional scaling. Specifically, the entries in D^* are monotonically nondecreasing when moving away from the main diagonal within a row or within a column. A condition of this general type is called an anti-Robinson property and is discussed in detail in the seriation literature when characterizing a "perfect" ordering along a single dimension (cf. Kendall, 1971).

To be more precise, suppose the n objects in S are placed in some order at the n integer positions 1, 2,..., n along a continuum. As a notation, a permutation, $\rho(\cdot)$, on the first n integers will be used to specify the given orderings; thus, position 1 is occupied by $O_{\rho(1)}$, position 2 is occupied by $O_{\rho(2)}$, and so on. As a further convention, any ordering used in defining D^* will be denoted by $\rho^*(\cdot)$.

Now, given a proximity matrix $\{s(O_i, O_j)\}$, the seriation problem involves locating a permutation $\rho(\cdot)$ and reordering the proximity matrix using $\rho(\cdot)$ that will satisfy the anti-Robinson property as "close" as possible. In other words, if D denotes the reorganized proximity matrix, then the ith row and column of D_ρ contains the object labeled $\rho(i)$. As an index of "closeness" to the anti-Robinson property, we choose what is essentially an unnormalized form of a correlation statistic:

$$\Gamma(\rho) = \sum_{i,j} s(O_{\rho(i)}, O_{\rho(j)}) \, |i-j| \, .$$

Thus, larger values of $\Gamma(\rho)$ supposedly denote "better" sequencings. For a more complete discussion of this measure as a goodness-of-fit index, the reader is referred to Hubert and Schultz (1976) and Szczotka (1972). (Although

we will rely on $\Gamma(\cdot)$ as our index-of-fit for a given object sequencing throughout, we note that many other possibilities also exist. For example, several statistics that depend only on the order of the proximity values are discussed in Hubert (1978)). Returning to the reorganized matrix $\underset{\sim}{D}^*$ and treating $d(\cdot,\cdot)$ as a proximity function, the index $\Gamma(\cdot)$ reaches its global maximum for any of the permutations $\rho^*(\cdot)$. A more formal statement of this fact could be developed by first converting the optimization task into the one-dimensional module placement problem discussed by Lawler (1975) and the details are left to the reader.

An informal approach to HC

Using the background material of the previous sections as a motivation, suppose we begin with an arbitrary (not necessarily ultrametric) proximity function $s(\cdot,\cdot)$ and we wish to find a HC that "fits" the function well. The standard analysis procedure would require the use of some single clustering method, such as the complete-link technique, and selecting the resultant sequence of partitions as the appropriate HC. A more ideal but obviously infeasible alternative would require a search among all possible HC's of the object set S, and then based on the correspondence between $s(\cdot,\cdot)$ and the induced ultrametric $d(\cdot,\cdot)$, a choice of some final HC for the purpose of a later substantive interpretation. Obviously, this search strategy is computationally impossible except for very trivial object sets. However, if the search is limited to only certain HC's that are probably the most adequate from other considerations, the reliance on a single constructive HC procedure can be avoided, or at least augmented by some further study of the patterning of the given proximity values.

To effect a limited search within all possible HC's and based on the connections between HC's and seriations through the anti-Robinson condition, suppose we use the initial proximity function $s(\cdot,\cdot)$ to sequence

the objects in S along a continuum based on the index $\Gamma(\cdot)$. Secondly, for some permutation, say $\rho'(\cdot)$, that leads to a very large value for an index such as $\Gamma(\cdot)$, those HC's that are defined by partitioning the proximity matrix reordered by $\rho'(\cdot)$ could then be evaluated in more detail. In particular, certain representations could be chosen merely from substantive considerations or on the basis of some further index of fit between $s(\cdot,\cdot)$ and the reconstructed ultrametric. In general, once the matrix $\{s(O_i, O_j)\}$ is reorganized to fit the anti-Robinson condition as close as possible, then a secondary partitioning procedure (possibly judgemental) can be implemented to elicit the particular type of structure desired.

The notion of using a condition basic to unidimensional scaling for effecting a cluster analysis may seem rather obvious. However, some theoretical work by Holman (1972) and Buneman (1970) would suggest that clustering and unidimensional scaling are inherently incompatible. An ultrametric requires a higher dimensional space for a perfect imbedding even though our context relys on a somewhat contrary connection (from an anti-Robinson condition) between a single dimension representation and the property of being an ultrametric. In fact, since an independently constructed HC should produce a satisfactory seriation when the matrix of proximities is appropriately reordered to reflect the induced partitioning, seriation procedures based on clustering have also been suggested in the literature (see Schuler and Ulrich, 1972).

Some Operational Details

For a given proximity matrix of moderate size, say greater than 12, the task of finding a globally optimal permutation maximizing $\Gamma(\cdot)$ is computationally very difficult. Consequently, various heuristics have been suggested in the literature that seek local optima through a set of limited operations that

attempt to better a given permutation. As one very powerful scheme, suppose some random permutation $\rho_1(\cdot)$ is used as a starting point and an attempt is made to transform $\rho_1(\cdot)$ to $\rho_2(\cdot)$ by performing the pairwise interchange of objects that maximizes the increase in $\rho(\cdot)$. Continuing, $\rho_3(\cdot)$ is constructed from $\rho_2(\cdot)$ in the same way, and so on until no pairwise interchange can increase the chosen index, i.e., until a local optimum, say $\rho'(\cdot)$, has been achieved. Other operations could then be tried to modify $\rho'(\cdot)$ and increase the index; for example, an insertion of a set of k consecutively ordered objects between two others or at either endpoint. Repeating this process over and over, a local optima is again achieved and the final permutation, say $\rho''(\cdot)$, is resubjected to the pairwise operation, and so on. Some final permutation will be located that is both "pairwise" and "insertion" locally optimal.

It should be obvious that many different variations of this general heuristic scheme could be tried -- including altering the type of local operations performed and how they are sequenced, changing the number of starting permutations used, and so on. However, the use of the pairwise and insertion operations just described appear to work very well and will be the method used in the later numerical example. For a simple dynamic programming approach to maximizing $\rho(\cdot)$ that is feasible when n is less than, say, 12, the reader is referred to Lawler (1975).

Numerical example

As an example of how a reordered matrix appears as generated by the complete link procedure in BMDP, the upper triangular portion of Table 3 presents a set of proximity values obtained between the 15 largest SMSA's whose statistical areas have remained unchanged for the last 10 years. Alphabetically these are: Baltimore (1), Boston (2), Chicago (3), Dallas-Fort Worth (4), Detroit (5), Honolulu (6), Los Angeles-Long Beach (7), Minneapolis-St. Paul (8), Nassau-Suffolk N.Y. (9), New York (10), Philadelphia (11), Pittsburgh (12), San Fransisco-Oakland (13),

St. Louis (14), Washington, D.C. (15). These proximities were obtained by calculating the Euclidean distance between each SMSA pair using their profiles over the seven index crimes for 1977: murder and non-negligent manslaughter, forcible rape, robbery, aggravated assault, burglary, larceny-theft, and motor vehicle theft. Each crime rate in the profile is a rate per 100,000 standardized to mean zero and variance 1 over the 15 SMSA's. As mentioned, the order of the cities in Table 3 is generated by the procedure used in the standard BMDP program in conjunction with a complete-link clustering; the partitioning represented in the matrix is defined according to the induced ultrametric listed in the lower triangular portion of the matrix.

TABLE 3 HERE

Using the interchange heuristic discussed earlier and 10 random starts, only one local optimum was identified, which was used to reorder the Table 3 matrix to the form given in Table 4. Although the pattern of entries in

TABLE 4 HERE

Table 4 is far from satisfying a perfect anti-Robinson condition, the general trend is clear. In fact, if the final cross-product index is normalized to obtain the Pearson product-moment correlation between the reorganized proximity matrix and the perfect target, a rather substantial value of .87 is achieved. More to the point, the reorganized Table 4 matrix can be used to identify partitions of the 15 SMSA's that are "better" than that obtained with the complete-link method even when we use as our criterion the same subset diameter condition the complete-link procedure is attempting to optimize in a heuristic manner. For instance, by using the order of the rows in Table 4, 14 different two-group partitions can be identified by merely splitting the

TABLE 3

Proximity Matrix (Upper Triangle) Among 15 SMSA's
as Reordered by the Complete-link BEDP Procedure -
the Induced Ultrametric is Represented in the Lower Triangle

	1	4	13	3	14	15	5	6	7	10	2	8	9	11	12
1	X	2.84	2.76	2.83	1.94	2.39	2.21	3.45	3.09	3.76	4.63	3.67	4.82	3.68	4.97
4	8	X	1.79	3.13	2.32	2.53	2.70	2.36	3.55	4.78	5.38	3.55	4.78	4.32	5.78
13	8	5	X	3.68	2.73	3.07	2.11	2.83	2.63	3.76	5.02	4.07	5.46	4.89	6.33
3	11	11	11	X	1.56	1.78	2.39	2.35	4.49	4.67	3.55	2.61	3.18	2.25	3.60
14	11	11	11	4	X	1.47	1.70	1.85	3.30	4.00	3.85	2.61	3.73	2.37	3.89
15	11	11	11	4	1	X	2.42	2.32	4.32	4.83	4.45	2.06	3.24	2.44	3.80
5	11	11	11	7	7	7	X	2.04	2.41	2.91	3.95	3.74	4.95	3.68	5.13
6	11	11	11	7	7	7	6	X	3.52	4.45	4.57	3.38	4.49	3.29	4.76
7	13	13	13	13	13	13	13	13	X	2.87	5.26	5.42	6.78	5.27	6.64
10	13	13	13	13	13	13	13	13	9	X	5.31	6.15	7.15	5.79	7.07
2	14	14	14	14	14	14	14	14	14	14	X	3.80	4.04	3.74	4.25
8	14	14	14	14	14	14	14	14	14	14	12	X	1.64	2.19	3.01
9	14	14	14	14	14	14	14	14	14	14	12	3	X	2.57	2.73
11	14	14	14	14	14	14	14	14	14	14	12	10	10	X	1.55
12	14	14	14	14	14	14	14	14	14	14	12	10	10	2	X

TABLE 4

Proximity Matrix Among 15 SMSA's
as Reordered by the Seriation Heuristic

	10	7	13	4	1	5	6	14	15	3	8	11	2	9	12
10	X	2.87	3.76	4.78	3.76	2.91	4.45	4.00	4.83	4.67	6.15	5.79	5.31	7.15	7.07
7	2.87	X	2.63	3.55	3.09	2.41	3.52	3.30	4.32	4.49	5.42	5.27	5.26	6.78	6.64
13	3.76	2.63	X	1.79	2.76	2.11	2.83	2.73	3.07	3.68	4.07	4.89	5.02	5.46	6.33
4	4.78	3.55	1.79	X	2.84	2.70	2.36	2.32	2.53	3.13	3.55	4.32	5.38	4.78	5.78
1	3.76	3.09	2.76	2.84	X	2.21	3.45	1.94	2.39	2.83	3.67	3.68	4.63	4.82	4.97
5	2.91	2.41	2.11	2.70	2.21	X	2.04	1.70	2.42	2.39	3.74	3.68	3.95	4.95	5.13
6	4.45	3.52	2.83	2.36	3.45	2.04	X	1.85	2.32	2.35	3.38	3.29	4.57	4.49	4.76
14	4.00	3.30	2.73	2.32	1.94	1.70	1.85	X	1.47	1.56	2.61	2.37	3.85	3.73	3.89
15	4.83	4.32	3.07	2.53	2.39	2.42	2.32	1.47	X	1.78	2.06	2.44	4.45	3.24	3.80
3	4.67	4.49	3.68	3.13	2.83	2.39	2.35	1.56	1.78	X	2.61	2.25	3.55	3.18	3.60
8	6.15	5.42	4.07	3.55	3.67	3.74	3.38	2.61	2.06	2.61	X	2.19	3.80	1.64	3.01
11	5.79	5.27	4.89	4.32	3.68	3.68	3.29	2.37	2.44	2.25	2.19	X	3.74	2.57	1.55
2	5.31	5.26	5.02	5.38	4.63	3.95	4.57	3.85	4.45	3.55	3.80	3.74	X	4.04	4.25
9	7.15	6.78	5.46	4.78	4.82	4.95	4.49	3.73	3.24	3.18	1.64	2.57	4.04	X	2.73
12	7.07	6.64	6.33	5.78	4.97	5.13	4.76	3.89	3.80	3.60	3.01	1.55	4.25	2.73	X

matrix into two parts, i.e., {{New York}, {Los Angeles-Long Beach,..., Pittsburgh}}, {{New York, Los Angeles-Long Beach}, {San Francisco-Oakland,..., Pittsburgh}}, and so on. In this example, the complete-link result for two groups happens to be equivalent to one of these splits: {{New York,..., Chicago}, {Minneapolis-St. Paul, ..., Pittsburgh}} and the maximum diameter (i.e., the largest proximity within a subset) over these two subsets is 4.83. However, the immediately preceding partition constructed from Table 4, {{New York, ..., Washington, D.C.}, {Chicago, ..., Pittsburgh}}, has exactly this same maximum diameter and the three partitions that precede this one have a slightly smaller maximum diameter of 4.78. In short, we can use the same optimization criterion that the complete-link method is attempting to satisfy and find better partitions according to this index by simply inspecting the matrix reordered by our seriation heuristic. For our purposes, the question of a best substantive interpretation can be left unanswered since our concern is only to suggest that the choice of a single best result constructed by some given clustering method could be more or less arbitrary. At the very least, it would seem prudent to inspect a reordered matrix to verify whatever substantive interpretations we would like to obtain from the use of a clustering strategy.

Discussion

The idea of using a seriation of the object set prior to looking for a specific clustering reappears continually in the literature although in many disguised forms. The key to recognizing this general paradigm is by the presence of some object ordering before a final clustering is given. Obvious examples would include Hartigan's (1975) leader algorithms, Matula's sequential graph coloring schemes (Matula, Marble, and Isaacson, 1972), Fisher's (1958) single variable clustering, and Szczotka's (1972) notion of an admissible partition. Implicitly or explicitly all of these methods rely on an object

ordering, typically as an initial organizing step prior to a final clustering based on partitioning the reordered matrix in some particular way, e.g., seriating the "break points" that define the possible subsets of a partition.

We do not wish to advocate the superiority of seriation over a particular method of HC or conversely. Instead, our aim is to point out their complementary nature and how clustering and seriation could be used together to justify a specific analysis. Looking at one's data in ways that could suggest alternative interpretations may seem to be a very obvious tactic. Unfortunately, it is easily forgotten when a scheme is available that promises to give a single best answer and without the ambiguity that is usually attached to a more intuitive data analysis strategy.

REFERENCES

1. Adrian, C.J., "Modelling the sequencing-location-allocation of public facilities," Proceedings, 49th ANZAAS Conference, Auckland University, New Zealand, 1979.

2. Baker, F.B. and L.J. Hubert, "Measuring the power of hierarchical cluster analysis," Journal of the American Statistical Association, 1975, 70, pp. 31-38.

3. Buneman, P., "An ultrametric is just Euclidean," unpublished manuscript, University of Edinburgh, 1970.

4. Cattell, R., "A note on correlation clusters and cluster search methods," Psychometrika, 1944, 9, pp. 169-184.

5. Church, R.L., "On determining many close to optimal configurations for single and multiple objective location problems," European Journal of Operations Research, 1980 (in press).

6. Dixon, W.J. (ed.), Biomedical Computer Programs, P-Series, Berkeley: University of California Press, 1977.

7. Fisher, W.D., "On grouping for maximum homogeneity," Journal of the American Statistical Association, 1958, 53, pp. 789-798.

8. Frank, H. and I.T. Frisch, Communication, Transmission, and Transportation Networks, Reading, Mass.: Addison-Wesley, 1971.

9. Forsyth, E. and L. Katz, "A matrix approach to the analysis of sociometric data: Preliminary report," Sociometry, 1946, 9, pp. 340-347.

10. Hartigan, J.A., Clustering Algorithms, New York: Wiley, 1975.

11. Holman, E.W., "The relation between hierarchical and Euclidean models for psychological distances," Psychometrika, 1972, 37, pp. 417-423.

12. Hubert, L.J., "Generalized proximity function comparisons," British Journal of Mathematical and Statistical Psychology, 1978, 31, pp. 179-192.

13. Hubert, L.J. and J.V. Schultz, "Quadratic assignment as a general data analysis strategy," British Journal of Mathematical and Statistical Psychology, 1976, 29, pp. 190-241.

14. Johnson, L., Jr., "Introduction to imaginary models for archaeological scaling and clustering," in D.L. Clarke (ed.), Models in Archaeology, London: Methuen, 1972.

15. Johnson, S.C., "Hierarchical clustering schemes," Psychometrika, 1967, 32, pp. 241-254.

16. Kendall, D.G., "Seriation from abundance matrices," in F.R. Hodson, et al. (eds.), Mathematics in the Archaeological and Historical Sciences, Edinburgh: Edinburgh University Press, 1971.

17. Lawler, E.L., "The quadratic assignment problem: A brief review," in B. Roy (ed.), Combinatorial Programming: Methods and Applications, Dordrecht, Holland: D. Reidel, 1975.

18. Megargee, E.I. and M.J. Bohn, Jr., Classifying Criminal Offenders, Sage Publications: Beverly Hills, California, 1979.

19. Matula, D., G. Marble, and J. Isaacson, "Graph coloring algorithms," in R.C. Reid (ed.), Graph Theory and Computing, New York: Academic Press, 1972.

20. Schuler, D.M. and E.G. Ulrich, "Clustering and linear placement," Design Automation Workshop, 1972, 9, pp. 50-56.

21. Sneath, P.H.A. and R.R. Sokal, Numerical Taxonomy, San Francisco: W.H. Freeman, 1973.

22. Szczotka, F.A., "On a method of ordering and clustering of objects," Zastosowania Mathematyki, 1972, 13, pp. 23-33.

Generalized Procedures for Evaluating Spatial Autocorrelation

by

L.J. Hubert, R.G. Golledge,

& C.M. Costanzo

The University of California, Santa Barbara

* Partial support supplied by LEAA/NILCJ 79-NI-AX-0057 & NSF SOC-77-28227.

L.J. Hubert is a Professor of Education at U.C.S.B.: R.G. Golledge is a Professor of Geography and C.M. Costanzo a graduate student in geography at U.C.S.B.

Generalized Procedures for Evaluating Spatial Autocorrelation

Introduction

Given a set S containing n geographical units, spatial autocorrelation (SA) refers to the relationship between some variable observed in each of the n localities and a measure of geographical proximity defined for all n(n-1) pairs chosen from S. Statistical methods developed for indexing this correspondence have traditionally been associated with the field of geography and more specifically with the subfield concerned with the assessment of spatial pattern ([6], [8], [12], [17]). The techniques for analyzing the effects of geographical proximity, however, are really very general, and when interpreted appropriately they offer a valuable set of inference strategies in many other disciplines and for problems that are far removed from a concern with spatial phenomena. Some of these broader implications have now been recognized in a few areas such as biology ([37], [38]) and sociology [2], but undoubtedly, more didactic papers will appear once the wide-spread applicability of these techniques becomes well-known.

Turning this assessment around, it is also true that work done in other fields such as epidemiology ([27], [28], [29]) and psychology[26] has significant implications for extending the SA methodology itself. We will emphasize this latter perspective throughout the paper and review a general context that includes as special cases many of the approaches used in the literature. Although some of these connections have already been suggested indirectly ([20], [22], [37]), there appears to be no single published source that seriously reviews the general field.

Since the literature on SA is rather extensive, a complete review of the area is well beyond the boundaries of this paper. Although we do give

a very brief overview of several SA ideas that will be of importance later on, we assume a basic familiarity with the seminal monograph by Cliff and Ord [8]. The reader interested in pursuing the topic more thoroughly should refer to this monograph, the papers cited earlier, and those by Cliff, Haggett, Ord, Bassett, and Davies [10]; Dacey [13]; Royaltey, Astrachan, and Sokal [31]; Winsborough, Quarantelli and Yutzy [40]; Campbell, Kruskal, and Wallace [7]; and Freeman [16]. Furthermore, we ignore the problems debated by Sen [32], Cliff and Ord [9], Sen [33], and Sen and Soot [34] regarding asymptotic normality for several of the well-known SA measures. Much of this newer distributional work can be considered peripheral since our emphasis will be on an alternative randomization model. For some possible connections between these asymptotic results and the development here, the recent paper by Shapiro and Hubert [36] may be helpful, particularly since it is based on work not usually cited in the geographical literature.

Finally, it should be noted that most of the discussion below introducing generalizations of the SA concept already appears in some form in the psychological and biomedical literature. Consequently, we can merely refer to other published papers for a more complete development. The titles in this other literature usually do not convey the possible applications to SA; therefore, it is important to have a single source that helps the interested reader develop the necessary relationships. We also refer the reader to the forthcoming second edition of Cliff and Ord's monograph [8]. This latter source independently uses an approach to SA similar in general form to the one presented here although the details are different.

Spatial Autocorrelation

Descriptive problems

As mentioned earlier, SA is concerned with the relationship between geographical proximity for a set of n localities, $\{O_1, O_2, \ldots, O_n\}$, and some variable measured on each of the single elements in the set S. The observed variables are denoted by $x_1, x_2, \ldots, x_n$, and the notion of geographical proximity is expressed through a (possibly asymmetric) n x n weight matrix $\underset{\sim}{W}$ in which the entry in the i^{th} row and j^{th} column, W_{ij}, represents the relationship between O_i and O_j. As a technical convenience, larger weights are assigned to pairs that supposedly are "more related"; for example, W_{ij} could be 1 if O_i and O_j were adjacent localities and 0 otherwise; or more generally, W_{ij} could specify the inverse of the actual physical distance from O_i to O_j. Since "self-weights" are irrelevant, the main diagonal in $\underset{\sim}{W}$ is assumed to consist of all zeros, thus, $W_{ii} = 0$ for all i.

The presence of SA implies that the pattern of weights in $\underset{\sim}{W}$ is related to the variate values $x_1, x_2, \ldots, x_n$. Traditionally, two measures of this relationship have been used -- Moran's I statistic and Geary's c coefficient [8]:

$$I = \left[\frac{n}{\sum_{i,j} W_{ij}}\right] \frac{\sum_{i,j} W_{ij} (x_i - \bar{x}) (x_j - \bar{x})}{\sum_{i} (x_i - \bar{x})^2}$$

$$c = \left[\frac{n-1}{2 \sum_{i,j} W_{ij}}\right] \frac{\sum_{i,j} W_{ij} (x_i - x_j)^2}{\sum_{i} (x_i - \bar{x})^2},$$

where $\bar{x} = \left[1/n\right] \sum_{i} x_i$. Both measures are normalized cross-product coefficients in which each weight W_{ij} is compared either to $(x_i - \bar{x})(x_j - \bar{x})$ or to $(x_i - x_j)^2$. Based on expectations that can be derived from the inference model of the next section, positive SA is obtained when I is greater than $-\frac{1}{(n-1)}$ or when c is less than 1; negative SA is obtained when I is less than $-\frac{1}{(n-1)}$ or when c is greater than 1.

Although I and c are based on normalizations that standardize the measures in particular ways, the crucial quantity in both is a cross-product statistic that relates the weight matrix $\underset{\sim}{W}$ to some function on the variates $x_1, x_2, \ldots, x_n$. In Moran's I, the important cross-product statistic is

$$\sum_{i,j} W_{ij} (x_i - \bar{x}) (x_j - \bar{x})$$

and for Geary's c we use

$$\sum_{i,j} W_{ij} (x_i - x_j)^2 .$$

Viewed in a slightly different fashion, the weight matrix $\underset{\sim}{W}$ is compared to a second matrix that we will denote by $\underset{\sim}{C}$. For Moran's I, the entry in the i^{th} row and j^{th} column of $\underset{\sim}{C}$, C_{ij}, is defined as $(x_i - \bar{x})(x_j - \bar{x})$, and for Geary's c, $C_{ij} = (x_i - x_j)^2$. In both cases, since $W_{ij} = 0$ for $i = j$ and all products of the form $W_{ij} C_{ij}$ are zero when $i = j$, we can also assume without loss of generality that $C_{ij} = 0$ for $i = j$. As a very brief summary, the general problem of indexing spatial autocorrelation can be phrased in terms of comparing two matrices $\underset{\sim}{C}$ and $\underset{\sim}{W}$ using a cross-product statistic $\Gamma = \sum_{i,j} C_{ij} W_{ij}$. This raw index is then subject to various standardizations to produce a final descriptive measure. (We might note that the idea of matrix comparison

does not correspond to any of the usual matrix algebra operations such as taking an inverse, matrix multiplication, and so on. This distinction between matrix comparison and typical matrix algebra operations should be kept in mind to avoid any possible confusion).

What we have just reviewed very tersely could be expanded in much greater detail. In fact, several of the references cited in the introduction do just that and show how to define a number of different special cases of Γ when the numerical variables $x_1, x_2, \ldots, x_n$ are from various levels of measurement. All such measures, however, rely on a matrix comparison through Γ plus a more or less arbitrary standardization that can safely be ignored since it remains invariant under the specific inference model we choose to work with. In short, the notion of matrix comparison should not be viewed as an alternative to the traditional ways of assessing SA but merely as a generalization that allows additional flexibility. Although both the usual I and c measures essentially form special cases of the cross-product index Γ, other types of matrix comparison, as we will see, can also be framed in the same context.

Inference problems

Even though we may be able to calculate rather easily the value for some SA measure, the task still remains to assess the relative size of the descriptive index compared to some chance mechanism. Because of its generality, we emphasize what has become known as the randomization model in which Γ is evaluated with respect to a reference distribution constructed by randomly permuting the rows and columns of $\underset{\sim}{C}$ together (or what is equivalent, the rows and columns of $\underset{\sim}{W}$). If the matrix $\underset{\sim}{C}$ is constructed from the sequence $x_1, x_2, \ldots, x_n$, and k is the number of distinct entries in this set and n_i the frequency of the i^{th} value, then an equally likely distribution over the n! indices will produce an equally likely distribution over at most $n!/\sum_{i=1}^{k} n_i$ distinct values. Even in

this latter set, however, some of the Γ statistics may be the same depending on the structure of $\underset{\sim}{W}$. Ideally, n! possible values of Γ would be generated (not necessarily distinct) by relabeling the rows and simultaneously the columns of $\underset{\sim}{C}$ in all n! possible ways. The tabled frequency distribution of these Γ's is then used in the same way as the exact sampling distribution for any test statistic, i.e., we reject the null model of randomness if the observed cross-product statistic is sufficiently extreme with respect to the constructed reference distribution.

The computational burden that complete enumeration imposes is usually so great that various approximations must be pursued. For completeness, we mention four different approaches that have been suggested in the literature:

(a) Since exact moments of the randomization distribution can be derived, crude Chebyshev or Cantelli bounds may be appropriate if the observed statistic is large. Specifically, when $Z = [\Gamma - E(\Gamma)]/\sqrt{V(\Gamma)}$, then a two-tailed significance level is at most $1/Z^2$ and a one-tailed level at most $1/(Z^2+1)$. General formulas for $E(\Gamma)$ and $V(\Gamma)$ are available in Mantel [27] for arbitrary $\underset{\sim}{C}$ and $\underset{\sim}{W}$ matrices.

(b) Although not always appropriate, normal approximations to Z are sometimes justified by rigorous convergence theorems subject to certain regularity conditions on $\underset{\sim}{C}$ and $\underset{\sim}{W}$ ([1], [34], [36]). Some recent results, however, suggest that care must be taken in assuming the triviality of the regularity requirements. It is not true that they can safely be ignored in most practical applications [30].

(c) Higher-order moments (third and fourth) have been derived for indices such as Γ when both $\underset{\sim}{C}$ and $\underset{\sim}{W}$ are symmetric. Consequently, various curve fitting procedures could be used to approximate the complete enumeration accurately [35].

(d) Complete enumeration can be approximated by random sampling and the significance level estimated by the number of generated Γ's that are as extreme or more extreme than the observed index (for example, if we assume an upper one-tailed test is appropriate, we count those Γ's that are as large or larger than the observed index). This topic is discussed in more detail by Besag and Diggle [5], Edgington [15], and Cliff and Ord [8].

The use of a randomization model offers a number of distinct advantages over the usual alternatives that rely on distributional assumptions for the observations $x_1, \ldots, x_n$. For example, the randomization model can be used for a wide variety of rather complex SA statistics that would be very difficult to handle in a classical parametric framework, e.g., those that require more than pairwise functions on S x S. Second, accuracy estimates for the approximations obtained by random sampling can be dealt with rigorously by standard non-parametric techniques, such as confidence intervals on percentile points or Kolmogorov-Smirnov bounds on the exact cumulative distribution function calculated from sample distributions. Under the typical distributional models for $x_1, x_2, \ldots, x_n$ and even though it is known that a particular statistic is asymptotically normal, bounds on the degree of accuracy for such an approximation are very difficult to obtain. Third, the normalizations that define the statistics such as I and c are typically invariant under randomization. Thus, the much simpler cross-product terms can be dealt with separately in making inferences regarding the significance of any given statistic. Finally, the randomization model is very easy to explain to a layman whereas the more classical distributional theory approaches are usually more "mysterious". In fact, even in the most rigorous papers that deal with distributional assumptions, very subtle errors appear that could invalidate the stated theorems. For instance, it is <u>not</u> true that the asympototic normality for two random variables auto-

matically insures the asympototic normality for their sum; such a result, however, has been implicitly relied on in at least one proof for the asymptotic normality of Geary's c ([32], p. 181).

<u>Numerical Example</u>

As a very simple illustration of the spatial autocorrelation concepts just introduced and of what a complete enumeration would look like, suppose the set S consists of 4 localities $\{O_1, O_2, O_3, O_4\}$ and the weight matrix is of the form

$\underset{\sim}{W} =$

	O_1	O_2	O_3	O_4
O_1	0	1	2	5
O_2	4	0	4	6
O_3	1	2	0	3
O_4	3	5	6	0

representing some asymmetric pattern of proximities among the four localities. If the variable measured on each of the four objects has values $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 4$, then for the Geary c coefficient we would define a matrix $\underset{\sim}{C}$ as follows:

$\underset{\sim}{C} =$

	O_1	O_2	O_3	O_4
O_1	0	1	4	9
O_2	1	0	1	4
O_3	4	1	0	1
O_4	9	4	1	0

and a raw cross-product statistic Γ between $\underset{\sim}{W}$ and $\underset{\sim}{C}$ of 148. If we fix the matrix $\underset{\sim}{C}$ as is and randomly permute the rows and simultaneously the columns of $\underset{\sim}{W}$ in all 4! = 24 possible ways (or equivalently, permute the rows and columns of $\underset{\sim}{C}$ with $\underset{\sim}{W}$ fixed), the following distribution over Γ would be obtained:

Row and column permutation of W	Γ
3241, 3421, 1243, 1423	108
2431, 2341, 1432, 1342	124
2143, 3142, 2413, 3412	132
1234, 4231, 4321, 1324	148
3124, 3214, 4213, 4123	156
4132, 4312, 4314, 2134	172

The proportion of indices in this distribution that are as large or larger than the observed index of 148 is 12/24. This value defines the significance level (we assume that negative SA is of interest, and thus, because of the way c is keyed, an upper one-tailed test is appropriate). Exactly the same distribution would be obtained if $\underset{\sim}{W}$ were held fixed and the rows and columns of $\underset{\sim}{C}$ permuted.

Although a significance level can be given for the raw index itself, it may also be appropriate to obtain a final descriptive measure by normalizing Γ to define c : $\frac{(4-1)}{(2[42])} \cdot \frac{(148)}{5} \approx 1.06$. This value is very close to the expectation under randomness. The I measure could be handled in the same manner by generating an exact reference distribution for the defining cross-product measure.

Generalized SA Measures

Using the general cross-product statistic $\Gamma = \sum_{i,j} W_{ij} C_{ij}$, there are a number of very obvious generalizations of the standard indices for SA. In much of the previous work (e.g., [8]), the emphasis was in developing SA indices that could accomodate an arbitrary weight matrix $\underset{\sim}{W}$. However, very specific forms of what we refer to as the $\underset{\sim}{C}$ matrix were required. In particular, the values in $\underset{\sim}{C}$ were assumed to be rather simple functions of the univariate observations $x_1, x_2, \ldots, x_n$ obtained for each of the n elements in S (for example, the number of milch cows in a set of areal units). It should now be apparent that C_{ij} could also be a more comprehensive measure based on, say, vectors of observations for O_i and O_j, e.g., Mahalanobis distances, Euclidean distances, or correlations. The randomization inference strategy remains the same since the multivariate data on each object are first reduced to a single numerical value for each object pair. Other extensions could be developed for the concordance context discussed by Hubert [24], in which the degree of internal concordance among a set of matrices, say, $\underset{\sim}{C}_1, \ldots, \underset{\sim}{C}_K$, could be evaluated. Here, a sum over all pairwise indices for the K matrices could be obtained and tested using the same type of inference model that justifies Kendall's coefficient of concordance. Alternatively, the K matrices could be compared to a single weight matrix $\underset{\sim}{W}$, and possibly a single target matrix fitted to the set. These problems can be rephrased as a combinatorial optimization task - for which an extensive literature exists (for example, see the references in [24]). All of these extensions, however, are rather immediate; consequently, it may be of greater interest to sketch several other variations that move beyond the cross-product measures that depend on simple pairwise functions S x S.

Replacing cross-products by 4-place functions

The raw cross-product measure Γ defines the relationship between two matrices through a specific function on each pair of corresponding entries in $\underset{\sim}{W}$ and $\underset{\sim}{C}$. In other words, if we consider the object pair O_i and O_j, the similarity between the two entries W_{ij} and C_{ij} is defined by the product $W_{ij}\,C_{ij}$ and forms part of the overall index Γ. As one generalization, suppose we replace such products by 4-place functions of $q(\cdot,\cdot,\cdot,\cdot)$, where the first two arguments relate to the i^{th} row and j^{th} column of $\underset{\sim}{W}$ and the last two arguments relate to the k^{th} row and l^{th} column of $\underset{\sim}{C}$. Thus, if $q(i,j,k,l)$ is defined as $W_{ij}\,C_{kl}$, then our earlier statistic Γ is equivalent to $\sum_{i,j} q(i,j,i,j)$. The first and last two arguments in $q(\cdot,\cdot,\cdot,\cdot)$ are both i and j since the observed index is concerned with a function of entries in the two matrices that are in comparable positions. More generally, we denote $\sum_{i,j} q(i,j,i,j)$ as Λ and apply the same type of randomization model followed for Γ. This inference strategy is discussed in detail by Hubert [25] including exact formulas for the first two moments of Λ.

As an example of how Λ could be used, suppose that the elements in $\underset{\sim}{W}$ and $\underset{\sim}{C}$ take on a small number of values, possibly on a nominal scale, and our concern is with the number of times an entry in $\underset{\sim}{W}$ is equal to its corresponding entry in $\underset{\sim}{C}$. If we define $q(i,j,k,l) = 1$ when $W_{ij} = C_{kl}$ and 0 otherwise, the index Λ counts the number of such perfect matches. As a second illustration, suppose the elements in $\underset{\sim}{W}$ and $\underset{\sim}{C}$ are reasonably commensurable and we wish to define a measure of SA through the size of the absolute differences between corresponding entries. If $q(i,j,k,l)$ is specified as $|W_{ij} - C_{kl}|$, then Λ is $\sum_{i,j} |W_{ij} - C_{ij}|$. Clearly, the number of possible choices for $q(\cdot,\cdot,\cdot,\cdot)$ is limitless and the spatial autocorrelation index can be tailored specifically for the aims of a given research project and data set.

SA measures that only depend on order

Several of the more interesting generalizations of the traditional SA statistics are to measures that only require the order of the entries in $\underset{\sim}{W}$ and $\underset{\sim}{C}$. Prior to developing an explicit example to illustrate such generalizations, we concentrate on two paradigms -- the first is based on object triples and the second on object quadruples.

Assuming that only the order of the entries in $\underset{\sim}{C}$ and $\underset{\sim}{W}$ are meaningful (or of immediate interest), suppose a cross-product statistic Ω is defined by

$$\Omega = \sum_{i,j,k} a(i,j,k)\; b(i,j,k),$$

where
$$a(i,j,k) = \begin{cases} 1 \text{ if } i,j,k \text{ are distinct and } W_{ij} < W_{ik}; \\ 0 \text{ otherwise,} \end{cases}$$

$$b(i,j,k) = \begin{cases} 1 \text{ if } i,j,k \text{ are distinct and } C_{ij} > C_{ik}; \\ 0 \text{ otherwise.} \end{cases}$$

Intuitively, since large values in $\underset{\sim}{W}$ usually denote closeness in some geographical space and small values in $\underset{\sim}{C}$ indicate closeness in some variable space, Ω counts the number of consistent object triples between $\underset{\sim}{C}$ and $\underset{\sim}{W}$. Here, a triple is consistent if and only if the weight from i to j is less than the weight from i to k and an opposite order exists for C_{ij} and C_{ik}. It should be noted that comparisons are based within the rows of $\underset{\sim}{W}$ and $\underset{\sim}{C}$ only, and consequently, no across row comparisons are carried out. Therefore, since each row can be defined by its own metric, only the entries within a single row need be commensurable. The index Ω gives a way of comparing two conditional proximity matrices in the sense of Coombs [11].

In the very simple 4 x 4 illustration given earlier, there are $n(n-1)(n-2) = 4 \cdot 3 \cdot 2 = 24$ distinct triples and both the functions $a(\cdot, \cdot, \cdot)$ and $b(\cdot, \cdot, \cdot)$ take on the value of 1 ten times. Since there are five triples that generate 1's for both functions {(3, 1, 2), (3, 1, 4), (4, 1, 2), (4, 1, 3), and (4, 2, 3)}, the index Ω has the value 5. For example, the triple (3, 1, 2) is consistent since $W_{31} < W_{32}$ and $C_{31} > C_{32}$.

Alternatively, if across row comparisons are meaningful, it may be more appropriate to use this added information in defining a different index θ:

$$\theta = \sum_{i,j,k,l} r(i,j,k,l)\, s(i,j,k,l)$$

where

$$r(i,j,k,l) = \begin{cases} 1 \text{ if } i \neq j,\ k \neq l \text{ and } W_{ij} < W_{kl}; \\ 0 \text{ otherwise,} \end{cases}$$

$$s(i,j,k,l) = \begin{cases} 1 \text{ if } i \neq j,\ k \neq l \text{ and } C_{ij} > C_{kl}; \\ 0 \text{ otherwise.} \end{cases}$$

With this interpretation, we count the number of consistent quadruples in the same manner that Ω counted consistent triples. Now, however, comparisons are carried out across rows. In the 4 x 4 illustration, for instance, there are $[n(n-1)]^2 = [4 \cdot 3]^2 = 144$ quadruples. The functions $r(\cdot, \cdot, \cdot)$ and $s(\cdot, \cdot, \cdot)$ take on the common value of 1 for 17 of these, and thus, $\theta = 17$. For example, the quadruple (1, 4, 2, 4) is consistent since $W_{14} < W_{24}$ and $C_{14} > C_{24}$.

The randomization model remains the same as before for both Ω and θ. The matrix $\underset{\sim}{W}$ is merely held fixed and the rows and columns of $\underset{\sim}{C}$ are reorganized in all possible ways. The interested reader is referred to Hubert [23] for an extended discussion of other 3 and 4 place cross-product statistics.

An Application

As a simple numerical illustration of the ideas given earlier, we will use the data for 26 counties of Eire originally given by Geary [19] and reanalyzed by Cliff and Ord [8]. The two major variables are the number of milch cows per 1,000 acres of crop and pasture in 1952 and the town and village populations as a percentage of total population in 1951. The contiguity matrix for the counties given by Cliff and Ord [8, p. 53] specifies the weight matrix $\underset{\sim}{W}$.

Defining the entries in $\underset{\sim}{C}$ by the absolute value of the difference between the values for two counties on a given variable and using the mean and variance formulas given by Mantel [27], we obtain the following results:

	Milch Cows	Population
Observed Γ =	3024.00	1721.59
$E(\Gamma)$ =	4679.33	2259.80
$\sqrt{V(\Gamma)}$ =	441.851	221.556
Z =	-3.75	-2.43

In both cases, positive spatial autocorrelation exists (defined by relatively small index values) and the results are generally consistent with those obtained by Cliff and Ord [8] based on the c and I statistics. Similarly, using the three-place measure Ω defined earlier we would obtain indices of 1481 and 1369 for milch cows and popualtion, respectively. Both indices are significant at the usual (upper-tails) levels obtained from a Monte Carlo testing strategy based on samples of 99 permutations (assuming the observed value of an index is another observation drawn at random -- see the distribution of Table 1). In both cases, the 0-1 structure of $\underset{\sim}{W}$ implies that there are

2344 distinct triple comparisons for which $W_{ij} < W_{ik}$. Out of this number, the index Ω indicates how many also have the property that $C_{ij} > C_{ik}$: 1481 for milch cows and 1369 for population. Again, these results are similar to those relying on other indices even though a very different form of index is now being considered.

Table 1 here

Table 1

Approximate Permutation Distribution for the Ω Index Using Geary's Data - Sample Size of 99

Ω	Cumulative Frequencies	
	Milch Cows	Population
1000	3	2
1020	4	3
1040	8	5
1060	15	8
1080	20	16
1100	30	21
1120	34	26
1140	43	33
1160	51	39
1180	59	52
1200	66	64
1220	72	74
1240	86	83
1260	89	90
1280	92	90
1300	94	91
1320	97	94
1340	98	95
1360	98	97
1380	99	98
1400	99	99

Discussion

Autocorrelation measures, as traditionally used by geographers, describe the pattern of an observed variate over a map system and imply something about the predictability of the map or the structure. By generalizing spatial autocorrelation measures using the randomization model as a base, a number of advantages accrue over the classical models based on specific distributional assumptions for the data. In general, the model proposed in this paper generates an immediately assessible inference paradigm for situations that would be very difficult to handle in a classical parametric framework. In fact, it may well be that standard tests of significance are at times inappropriate for classical SA measures. By using a randomization model and the complete enumeration process, significance levels appropriate for SA measures can be obtained.

In geography it is common to distinguish between two different approaches to spatial autocorrelation. One is tied to expressing spatial autocorrelation in a lagged form and depends on calculating and expressing covariances between different data values at different lagged distance or directional lengths. The second approach examines spatial autocorrelation in terms of the influence each observation is assumed to have on other observations. Our emphasis is more in line with the second approach rather than the first. Regardless of which approach is used, similar types of problems face the individual attempting to assess spatial autocorrelation and similar problems face researchers attempting to use and extend the procedure. Gattrell [18] states the first of these as the need to specify alternate forms of the "distance" concept that provides the base for the calculation of spatial autocorrelation effects. At least on the variate side of the problem we have shown how spatial autocorrelation can be generalized to such alternate

measures, including Mahalanobis distances, correlation coefficients, and any other arbitrarily defined indices of proximity. More general distance measures can be handled directly by defining the weight matrix appropriately, e.g., Mahalanobis distances in some generalized multidimensional space.

The extension of SA measures to data structures that are nominal or ordinal produces indices comparable with standard SA measures without the need to adhere to the stricter parametric assumptions necessary to generate an inference model for those standard measures. Basing the spatial autocorrelation index on randomization overcomes one of the more critical problems currently being ced by geographers. For example, Haining [21] argues that with respect to the Cliff, et al. study of measles data for Southwest England [10] the absence of information on the joint sampling distribution of the average correlations, together with the small sample sizes involved and the generally insignificant values assumed by the correlations, tend to cast doubt on their interpretation of the measles epidemic as having a central place type diffusion structure. Developing an autocorrelation index using the randomization model clearly overcomes the first of these deficiencies, although it does not solve the second problem - that of defining a satisfactory model base for the interpretation of results.

The use of spatial autocorrelation in geography to compare observed and theoretical or expected map patterns has in the past been limited by the problems involved in measuring the degree of departure from randomness. As Dacey [14] and Cliff and Ord [8] have found out, rejecting a hypothesis of randomness based on Poisson models cannot be taken as indicative of apparent contagion. While Besag [3, 4] has examined this problem in more detail, the inference problems raised in the geographic studies can be approached using randomization procedures to construct a reference distribution against which to measure the magnitude of deviations. Thus by using such an index and

reaching a stage where a hypothesis of randomness is rejected, the researcher may feel more at ease assuming that the patterns examined are being produced by similar processes.

In summary, the approach we have described for measuring SA has at least four major advantages. The first relates to the generality of the paradigm and the fact that many different measures, even those tailor-made for specific substantive problems, can all be placed under one common framework and tested for significance using the same type of randomization argument. It is somewhat inappropriate, however, to view the general notion of matrix comparison as a competitor to the traditional way of handling SA tasks since special cases of matrix comparison have been used for some time. Nevertheless, there is an obvious inherent value in offering alternatives that may be more suited for particular research applications than the I and c indices. An obvious example would be in our ability to deal with more than one variable at a time in assessing SA through a multivariate measure of distance, defining the entries in the matrix $\underset{\sim}{C}$. Second, the randomization strategy itself can be approached through Monte Carlo sampling, bypassing the optimistic use of asymptotic distributional results of possibly unknown accuracy. These latter large sample size results are very spotty and do not cover all the SA statistics that could be defined in our framework. Third, by placing SA into a larger matrix comparison structure, an obvious pedagogical advantage is achieved. This is analogous to the perspective provided by understanding the general linear model even though the special cases of analysis of variance and regression may continue to be the most popular alternatives as implemented by routines that are specialized from the more general structure. Fourth, once a comprehensive framework is understood, further work on the framework itself immeidately suggests many associated results that are pertinent to a class of measures. Thus, once the commonality of analysis tasks are recognized, there is an obvious broader purpose taken on by the research enterprise.

REFERENCES

1. Abe, O., 1969, "A central limit theorem for the number of edges in the random intersection of two graphs," Annals of Mathemetical Statistics, 40, 144-151.

2. Althauser, R.P., Burdick, D.S., and Winsborough, H.H., 1966, "The Standardized contiguity ratio," Social Forces, 45, 237-245.

3. Besag, J.E., 1972, "On the correlation structure of some two dimensional stationary processes," Biometrika, 59, 43-48.

4. Besag, J.E., 1974, "Spatial interaction and the statistical analysis of lattice systems," Journal of the Royal Statistical Society B, 36, 192-236.

5. Besag, J.E., and Diggle, P.J., 1977, "Simple Monte Carlo tests for spatial pattern," Applied Statitsics, 26, 327-333.

6. Brandsma, A.S. and Ketellapper, R.H., 1979. "A biparametric approach to spatial autocorrelation," Environment and Planning A, 11, 51-58.

7. Campbell, D.T., Kruskal, W.H., and Wallace, W.P., 1966, "Seating aggregation as an index of attitude," Sociometry, 29, 1-15.

8. Cliff, A.D. and Ord, J.K., 1973, Spatial Autocorrelation, London: Pion (2nd Edition, forthcoming, 1981).

9. Cliff, A.D. and Ord, J.K, 1977, "Large sample-size distribution of statistics used in testing for spatial correlation: A comment," Geographical Analysis, 9, 297-299.

10. Cliff, A.D. and Haggett, P., Ord, J.K., Bassett, K.A., and Davies, R.B., 1975, Elements of Spatial Structure, Cambridge: Cambridge University Press.

11. Coombs, C.H., 1979, "Data and scaling theory," in G. Menges (Ed.) Handbook of Mathematical Economic Sciences, Vol. 2, Dusseldorf: West Deutscher Verlag.

12. Curry, L., 1966, "A note on spatial association," The Professional Geographer, 18, 97-99.

13. Dacey, M.F., 1968a, "A review on measures of contiguity for two and k-color maps," in B.J.L. Berry and D.F. Marble (Eds.) Spatial Analysis, Englewood Cliffs, N.J.: Prentice-Hall.

14. Dacey, M.F., 1968b, "An empirical study of the areal distribution of houses in Puerto Rico," Transactions IBG, 45, 51-69.

15. Edgington, E.S., 1969, Statistical Inference: The Distribution-free Approach, New York: McGraw-Hill.

16. Freeman, L.C., 1978, "Segregation in social networks," Sociological Methods and Research, 6, 411-429.

17. Gatrell, A.C., 1977, "Complexity and redundancy in binary maps," Geographical Analysis, 9, 29-41.

18. Gatrell, A.C., 1979, "Autocorrelation in spaces," Environment and Planning A, II, 507-516.

19. Geary, R.C., 1954, "The contiguity ratio and statistical mapping," The Incorporated Statistician, 5, 115-145.

20. Glick, B.J., 1979, "Tests for space-time clustering used in cancer research," Geographical Analysis, 11, 202-208.

21. Haining, R., 1980, "Spatial autocorrelation problems," in D.J. Herbert and R.J. Johnston (Eds.), Geography and the Urban Environment, New York: John Wiley & Sons.

22. Hubert, L.J., 1978a, "Nonparametric tests for patterns in geographic variation: Possible generalizations," Geographical Analysis, 10, 86-88.

23. Hubert, L.J., 1978b, "Generalized proximity function comparisons," British Journal of Mathematical and Statistical Psychology, 31, 179-192.

24. Hubert, L.J., 1979a, "Generalized concordance," Psychometrika, 44, 135-142.

25. Hubert, L.J., 1979b, "Matching models in the analysis of cross-classifications," Psychometrika, 44, 21-41.

26. Hubert, L.J. and Subkoviak, M.J., 1979, "Confirmatory inference and geometric models," Psychological Bulletin, 86, 361-370.

27. Mantel, N., 1967, "The detection of disease clustering and a generalized regression approach," Cancer Research, 27, 209-220.

28. Mantel, N. and Bailar, J.C., 1970, "A class of permutational and multinomial tests arising in epidemiological research," Biometrics, 26, 687-700.

29. Mantel, N. and Valand, R.S., 1970, "A technique of nonparametric multivariate analysis," Biometrics, 26, 547-558.

30. Mielke, P.W., 1978, "Clarification and appropriate inferences for Mantel and Valand's nonparametric multivariate analysis technique," Biometrics, 34, 277-282.

31. Royaltey, H.H., Astrachan, E., and Sokal, R.R., 1975, "Tests for patterns in geographic variation," Geographical Analysis, 7, 369-395.

32. Sen, A., 1976, "Large sample-size distribution of statistics used in testing for spatial correlation," Geographical Analysis, 8, 175-184.

33. Sen, A., 1977, "Large sample-size distribution of statistics used in testing for spatial correlation: A reply," Geographical Analysis, 9, 300.

34. Sen, A.K., and Sööt, S., 1977, "Rank tests for spatial correlation," Environment and Planning A, 9, 897-903.

35. Siemiatycki, J., 1978, "Mantel's space-time clustering statistic: Computing higher moments and a comparison of various data transforms," Journal of Statistical Computation and Simulation, 7, 13-31.

36. Shapiro, C.P. and Hubert, L., 1979, "Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions," The Annals of Statistics, 4, 788-794.

37. Sokal, R.R., 1979, "Testing statistical significance of geographical variation patterns," Systematic Zoology, 28, 227-232.

38. Sokal, R.R. and Oden, N.L., 1978a, "Spatial autocorrelation in biology 1: Methodology," Biological Journal of the Linnean Society, 10, 199-228.

39. Sokal, R.R. and Oden, N.L., 1978b, "Spatial autocorrelation in biology 2: Some biological implications and four applications of evolutionary and ecological interest," Biological Journal of the Linnean Society, 10, 229-249.

40. Winsborough, H.H., Quarantelli, E.L., and Yutzy, D., 1963, "The similarity of connected observations," American Sociological Review, 28, 977-983.

Inference Models for Roll-Call Cohesion Measures*

Lawrence J. Hubert**

and

Reginald G. Golledge

* Partial support provided by LEAA/NILCJ Grant #79-NI-AX-0057 and NSF-GSOC-77-28227

** Lawrence J. Hubert is a Professor of Education and Reginald G. Golledge is a Professor of Geography at the University of California, Santa Barbara

Inference Models for Roll-Call Cohesion Measures

Introduction

Measures of legislative roll-call cohesion have been widely used in the political science literature to index the degree of solidarity manifested by a definable subgroup of voters (see MacRae, 1970, for a review). Unfortunately, the most popular measures have also been plagued by the lack of inference models that can take into account the degree of cohesion present in the larger unit containing the subgroup. For example, Rice's well-known measure of subgroup cohesion is defined (up to a multiplicative constant) by the absolute difference between the proportions of a positive and a negative vote. Chance cohesion is supposedly indicated by a 50-50 split (i.e., each proportion has a value of 1/2) and deviations from the 50-50 split are interpreted as evidence for an increased cohesion among the subgroup members. This equally-likely baseline is assumed even though the complete population may not have a perfect 50-50 split. Consequently, it is possible for the larger population to demonstrate more cohesion than the subgroup, even though the latter by itself may appear to be highly cohesive when evaluated against the null assumption of a 50-50 split. A similar problem would exist in performing goodness-of-fit tests in contingency

tables if marginal information were ignored and expected values were obtained from an equally-likely assumption on the cell probabilities.

The problem of defining inference models for measures of roll-call cohesion has been recognized in the literature. For instance, as one way of developing a more reasonable inference structure, Born and Nevison (1975) introduced a probability measure based on the cumulative distribution of votes in the more inclusive body. Although this is a step in the appropriate direction, there are at least three limitations on the Born and Nevison approach. First of all, since the probability measure requires a rather sophisticated understanding of probability theory and deviates markedly from the justification behind the indices typically used in the literature, acceptance of the new statistic may be very slow in coming, particularly since extensive tables and/or the use of specifically designed computer programs are required for its calculation. Secondly, the probability measure is essentially limited to votes that have 2-values (Aye and Nay) and voting options with k alternatives are not easily incorporated within the paradigm. Finally, the Born-Nevison measure is really a significance level, and therefore, it is heavily dependent on the size of the subgroup being considered. For example, two subgroups with the same values on a more traditional index of cohesion could also have very different probability values depending on the sizes of the two subgroups. It would seem more appropriate to consider a

statistic that has properties similar to the Pearson correlation coefficient; i.e., an index that would give an indication of the absolute level of cohesion irrespective of sample size but have an associated significance statement that was dependent on the number of observations, or in our case, on the number of voters in the subgroup as well as in the more inclusive body. Using these three concerns as our motivation, it is relatively straightforward to carry out the original Born-Nevison goal of providing a suitable inference model for a measure of cohesion through the well-known percentage voting agreement measure.

Background

To provide some introductory notation, we assume that a body of n individuals $P = \{I_1, I_2, \ldots, I_n\}$ has voted on an issue and the cohesion of a subgroup S of size m is of interest. If there are k voting options, then the percentage voting agreement measure is defined as (Born and Nevison, 1975):

$$\Gamma = \sum_{j=1}^{k} m_j(m_j - 1)/m(m - 1),$$

where m_j is the number of individuals in S using voting alternative j, $1 \leq j \leq k$, and $m = \sum_{j=1}^{k} m_j$. In other words, Γ is the ratio of the number of pairs of individuals in S who vote in the same way to the total number of pairs in S.

Since Γ is a descriptive statistic and should be interpreted in relation to the distribution of votes in P, the null model we assume conjectures that S was constructed at random from P. Thus, the distribution of Γ under this null model can be obtained from the enumeration of Γ over all $\binom{n}{m}$ subsets of size m that could be formed from P. Increased cohesion is typically of interest; consequently, if the observed value of Γ is sufficiently large with respect to this distribution, we declare the raw measure significant.

Since complete enumeration is very time consuming and must be redone for each separate test of Γ, it is convenient to rely on exact mean and variance parameters and an approximate normality (when m is not too large or small) for evaluating the size of Γ (Bloemena, 1964). In particular, the literature gives (Hubert and Levin, 1977)

$$E(\Gamma) = \sum_{j=1}^{k} M_j(M_j - 1)/n(n - 1),$$

where M_j is the number of voters in the population P using alternative j and $n = \sum_{j=1}^{k} M_j$. The variance can be calculated exactly using the expression

$$V(\Gamma) = \left[\frac{1}{n(n-1)}\right]^2 A_1 + \frac{2}{n(n-1)\, m(m-1)} A_3$$

$$+ \frac{4(m-2)}{n(n-1)(n-2)\, m(m-1)} [A_2 - A_3]$$

$$+ \frac{(m-2)\,(m-3)}{n(n-1)(n-2)(n-3)\, m(m-1)} [A_1 - 4A_2 + 2A_3],$$

CONTINUED

2 OF 3

where

$$A_1 = \left[\sum_{j=1}^{k} M_j(M_j - 1) \right]^2 ;$$

$$A_2 = \sum_{j=1}^{k} M_j(M_j - 1)^2 ;$$

$$A_3 = \sum_{j=1}^{k} M_j(M_j - 1) \quad .$$

Finally, the Z-statistic

$$Z = (\Gamma - E(\Gamma))/\sqrt{V(\Gamma)}$$

could be referred to the standard normal distribution to find the significance level that should be attached to Γ .

It might also be noted that the probability of obtaining a Z-value as large or larger than the one observed corresponds closely in role to the probability measure used by Born and Nevison. The raw index Γ , however, can also be given a very simple descriptive interpretation that is independent of its significance level; i.e., the proportion of pairs in the subgroup that vote in the same way. As we will see, this descriptive capability extends to much more general measures as well.

There is one modification of the Γ measure that should be mentioned that involves how the self-pairs in a subgroup are counted, i.e., pairs that are defined by the same voter with

himself. For example, suppose there are 2 voting options and a group of size 4 splits 2 and 2 giving a Γ of 1/3. A group of 40 splitting 20 and 20, however, would generate a larger Γ of 19/39. Consequently, even though the splits are the same in both cases, the size of the group being evaluated produces different values since the number of self-pairs that are not counted differ in each case. If a correction is desired, the self-pairs could be included and a new index, say Γ^* , defined as

$$\Gamma^* = \frac{\sum_{j=1}^{k} m_j^2}{m^2}$$

Thus, in the small example given above, Γ^* would be 1/2 in both instances. In general, the mean and variance of Γ^* are immediate since Γ^* is a simple linear transformation of Γ :

$$\Gamma^* = \left(\frac{m-1}{m}\right) \Gamma + \frac{1}{m} \quad .$$

Specifically,

$$E(\Gamma^*) = \left(\frac{m-1}{m}\right) E(\Gamma) + \frac{1}{m} ,$$

and

$$V(\Gamma^*) = \left(\frac{m-1}{m}\right)^2 V(\Gamma) \quad .$$

The major disadvantage of this correction is that $E(\Gamma^*)$ now depends on m whereas $E(\Gamma)$ did not.

In summary, the percentage voting agreement measure can be evaluated directly against what could be expected from the distribution of votes in the population. If we reject the conjecture that S was constructed at random (by having the Z-value be positive and large), the obvious substantive implication is that the subgroup has some voting cohesion over and above what could be expected merely by forming a group of this size at random from the larger population.

There are numerous extensions of this same model that can provide strategies for evaluating the relative size of a host of voting correspondence measures. For example, suppose we have two disjoint groups S_1 and S_2 containing r and s individuals in each. The statistic

$$\Lambda = \sum_{j=1}^{k} m_{1j}m_{2j}/rs$$

is the ratio of the number of pairs of individuals between the two groups who vote the same to the total number of pairs. Here, m_{1j} and m_{2j} are the number of individuals in S_1 and S_2, respectively, who use the j^{th} voting option. Thus, $\sum_{j=1}^{k} m_{1j} = r$ and $\sum_{j=1}^{k} m_{2j} = s$. Assuming S_1 and S_2 are formed at random (and without replacement) from P, the expectation of Λ is the same as that for Γ; i.e., $\sum_{j=1}^{k} M_j(M_j - 1)/n(n - 1)$, but the variance is now

$$V(\Lambda) = -\left[\frac{1}{n(n-1)}\right]^2 A_1 + \frac{1}{n(n-1)rs} A_3 + \frac{(r+s-2)}{n(n-1)(n-2)rs} [A_2 - A_3] + \frac{(r-1)(s-1)}{n(n-1)(n-2)(n-3)rs} [A_1 - 4A_2 + 2A_3] .$$

The Z-statistic could be formed as before and its significance level approximated by using the standard normal distribution.

As we will see in the next section, one general procedure for comparing two matrices includes both Γ and Λ as special cases. More importantly, this comparison strategy suggests an enormous variety of other analysis options for assessing voting patterns that cannot be handled in any convenient way with more traditional statistical methods.

Some Extensions

As mentioned above, both the Γ and Λ indices introduced in the previous section can be obtained as special cases of a more general matrix comparison strategy. To begin with, we let $\underset{\sim}{Q}$ and $\underset{\sim}{C}$ denote two $n \times n$ matrices with zeros along the main diagonal. If the entries in the i^{th} row and j^{th} column are denoted by q_{ij} and c_{ij}, respectively, a general index between the two matrices can be defined by the cross-product statistic

$$\theta = \sum_{i,j} q_{ij} c_{ij} .$$

Now, if $\underset{\sim}{Q}$ is defined in the following schematic form

$$\underset{\sim}{Q} = \begin{array}{c|ccc} & M_1 & M_2 & \cdots & M_k \\ \hline M_1 & \begin{matrix}0 & 1\\ 1 & 0\end{matrix} & 0 & \cdots & 0 \\ M_2 & 0 & \begin{matrix}0 & 1\\ 1 & 0\end{matrix} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ M_k & 0 & 0 & \cdots & \begin{matrix}0 & 1\\ 1 & 0\end{matrix} \end{array} \qquad (1)$$

where the 1's and 0's represent generic entries in the indicated sections of the matrix, then θ becomes Γ when $\underset{\sim}{C}$ is defined as

$$\underset{\sim}{C} = \begin{array}{c|cc} & m & \\ \hline m & \begin{matrix}0 & \frac{1}{m(m-1)}\\ \frac{1}{m(m-1)} & 0\end{matrix} & 0 \\ & 0 & 0 \end{array} \qquad (2)$$

Again, the 0's and $\frac{1}{m(m-1)}$ denote generic elements.

In a similar fashion, Λ is obtained when $\underset{\sim}{C}$ is specified as

$$\underset{\sim}{C} = \begin{array}{c|ccc} & r & s & \\ \hline r & 0 & \frac{1}{2rs} & 0 \\ s & \frac{1}{2rs} & 0 & 0 \\ & 0 & 0 & 0 \end{array}$$

using $\frac{1}{2rs}$ as a generic entry.

To develop an inference structure based on the general measure θ, suppose we fix the matrix $\underset{\sim}{C}$ as is and evaluate for all $n!$ possible ways of reorganizing the rows and simultaneously the columns of $\underset{\sim}{C}$ (or equivalently, we could fix $\underset{\sim}{C}$ and reorder $\underset{\sim}{Q}$). If $\underset{\sim}{C}$ has the specific structures given above in (2) and (3), this permutation model generates the appropriate null distribution for both Γ and Λ. For example, when $\underset{\sim}{C}$ is defined as in (2), the $n!$ possible reorderings of $\underset{\sim}{C}$ induce $\binom{n}{m}$ subsets of the n objects in P (each subset is duplicated the same number of times). Similarly, when $\underset{\sim}{C}$ is defined as in (3), the $n!$ possible reorderings of

$\underset{\sim}{C}$ induce an equally-likely distribution over all possible pairs of subsets with r and s objects in each.

Since formulas for the first two moments of the general index θ are available in the literature even when $\underset{\sim}{Q}$ and $\underset{\sim}{C}$ are asymmetric (Mantel, 1967; Hubert and Schultz, 1976), the expectation and variance for Γ and Λ can be derived as special cases of very comprehensive expressions. More significantly, the matrices $\underset{\sim}{Q}$ and $\underset{\sim}{C}$ can also be redefined to obtain indices that are very specific and/or are directed toward different aspects of analyzing roll-call cohesion. For instance, (a) the $\underset{\sim}{Q}$ matrix could be defined as a sum over a set of roll calls and $\underset{\sim}{C}$ defined as in (2) and (3). Here, we are interested in assessing cohesion for more than a single issue. (b) The matrix $\underset{\sim}{C}$ could represent the results of a second roll-call or set of roll-calls. Now, our interests would be in assessing the correspondence between two sets of roll-calls. With this latter interpretation, particularly when both $\underset{\sim}{C}$ and $\underset{\sim}{Q}$ represent partitions of the form given in (1), we have a way of analyzing political cleavages as defined by Rae and Taylor (1970). (c) The simultaneous cohesion within more than one party (whether they exhaust P or not) can be developed by merely extending (2) to include more than one "block" of non-zero values along the main diagonal. Similarly, cross-connections between more than two parties could be evaluated by extending (3) in the obvious way. (d) If $\underset{\sim}{Q}$ is given as in (1), the $\underset{\sim}{C}$ matrix could be defined by some proximity measure between voters, e.g., the absolute

differences between age, years of education, and so on. Our concern here would be in relating voting pattern to a non-dichotomous variable. (e) Finally, since there is a basic duality between voters and issues, the $\underset{\sim}{Q}$ and $\underset{\sim}{C}$ matrices could represent the relationships among issues (e.g., for each pair of issues we could calculate as a proximity measure the proportion of voters who give the same response option), and $\underset{\sim}{C}$ could be used to evaluate whether a structure exists among the issues; e.g., whether there is subset homogeneity or an apparent partitioning. Since the options are developed at least implicitly in related psychological contexts, the reader is referred to this literature (e.g., Hubert and Schultz, 1976). In addition, extensions to sets of matrices and/or to nonmetric matrix comparisons that only require the order of the entries in $\underset{\sim}{Q}$ and $\underset{\sim}{C}$ are available in Hubert (1978; 1979). For some comments on possible problems with asympototic normality for these more general statistics, the reader should consult the recent paper by Mielke (1978).

REFERENCES

Blomena, A.R. (1964). Samoling From a Graph. Amsterdam: Mathematisch Centrum.

Born, R. and Nevison, C. (1975). "A probabilistic analysis of roll call cohesion measures," Political Methodology 2: 517-528.

Hubert, L.J. (1978). "Generalized proximity function comparisons," British Journal of Mathematical and Statistical Psychology 31: 179-192.

_____. (1979). "Generalized concordance," Psychometrika, 44: 135-142.

_____ and Levin, J.R. (1976). "Evaluating object set partitions: Free-sort analysis and some generalizations," Journal of Verbal Learning and Verbal Behavior 15: 459-470.

_____. (1977). "Inference models for categorical clustering," Psychological Bulletin 84: 878-887.

_____ and Schultz, J.V. (1976). "Quadratic assignment as a general data analysis strategy," British Journal of Mathematical and Statistical Psychology 29: 190-241.

MacRae, D. (1970). Issues and Parties in Legislative Voting. New York: Harper & Row.

Mantel, N. (1967). "The detection of disease clustering and a generalized regression approach," Cancer Research 27: 209-220.

Mielke, P.W. (1978). "Clarification and appropriate inferences for Mantel and Valand's nonparametric multivariate analysis technique," Biometrics 34: 277-282.

Rae, D. and Taylor, M. (1970). The Analysis of Political Cleavages. New Haven: Yale University Press.

SOME COMMENTS ON
NON EUCLIDEAN MENTAL MAPS

by

Reginald G. Golledge
Department of Geography
&
Lawrence J. Hubert
Department of Education

University of California,
Santa Barbara

* Research for this project was supported by a grant from the Academic Senate of the University of California, Santa Barbara, and the paper was first prepared for the International Geographical Union, Working Group on Systems Analysis & Mathematical Models meeting in Lagos, Nigeria (July 1978). Edwin Gustafson, Nathan Gale and G. Donald Richardson developed and prepared the illustrations used. Acknowledgement is due to Waldo Tobler for his constructive advice and for the use of his computer program "DARCY" which computed the distorted grid and street systems.

Introduction

For a science that claims to have space as its central theme, geography is at best myopic. Throughout the greater part of their existence, geographers have represented spatial phenomena in simple Euclidean space. While a few brave souls have ventured into the non-Euclidean realm (e.g., Bunge, 1962; Tobler, 1972, 1976; Lau, 1978), there is a remarkable reluctance within the discipline as a whole to think of spatial concepts in anything but Euclidean terms. Our models use Euclidean measures of distance, the bulk of our maps are constructed on a Euclidean coordinate system, and out theories of spatial organization are summarized cartographically and mathematically in Euclidean terms.

The widespread acceptance of Euclidean geometry as the most appropriate for representing space, predisposes a certain type of perspective on the world. Within the Euclidean framework, space is conceived as being isotropic - that is, the same geometric relations hold in all parts of the space. A second important concept is that of parallelism - that is, parallels do not converge. Accepting these concepts readily allows us to implement perhaps the best known and most widely used formula in the discipline of geography, that of measuring inter-point distances in N-dimensional Euclidean spaces. This formula is:

$$D_{ij} = \left[\sum_k \left(\left| x_{i,k} - x_{j,k} \right| \right)^r \right] \frac{1}{r} \qquad (1)$$

where: D_{ij} is the distance between two arbitrarily defined points i and j;

$x_{i,k}$ is the coordinate for point i on the k^{th} dimension;

r is the exponent to which displacements in any dimension are taken in the particular distance formulation (i.e., the Minkowskian Metric).

While this general Minkowskian formula is well known, almost invariably the Euclidean (r = 2) is preferred to other Minkowskian metrics such ar r = 1 (the Manhattan or city block metric) or $r = \infty$ (the dominance metric or SUP-metric) (see Krause, 1975). In this latter metric, the distance between any pair of points is defined as the longest side of the right-angled triangle constructed in the space about the points (i.e., $d_{ij} = \max_k(x_{ik}, x_{jk})$).

The Euclidean metric is in many ways a convenient one for abstracting from reality and summarizing it in the form of a map, which for geography is the commonly used model. For many years we have accepted the fact that most maps come on flat sheets of paper and geographers have spent considerable time and ingenuity on procedures for transforming occurrences on the surface of the earth into appropriately represented schema on flat sheets of paper.

In contrast to this, if one examines the perceptual mechanism of the human eye, then it becomes obvious that divergences from the Euclidean metric are evident even at the local scale. If physical space is Euclidean then perceptual space must be non Euclidean and even the simple task of recognizing a straight line as straight must be learned (see Roberts and Suppes, 1967). Admittedly, one can argue that the curvature of the earth is such that it can be interpreted as being locally Euclidean, but it is also evident that if one looks along a railway track stretching along a plain, the parallels of the track appears to converge. If one looks at a series of telephone lines stretching away into the distance, the poles become smaller. Distance itself is conceived by a process of intellectual synthesis that involves an equalibration of both motoric experiences and visual effects. When one perceives objects becoming smaller or parallels converging, it is assumed that a greater amount of effort is required to reach such objects. Thus we are faced with a fundamental conflict between our senses and our knowledge that we need to resolve on a day by day basis. However, we make very little attempt to constructively incorporate the essential differences into our day by day experiences. At the very least, one can argue that the internal geometry of the world of perception and cognition may have very few of the attributes of Euclidean geometry and may more likely be represented in non-Euclidean formats (e.g., see Luneburg, 1950; Zautinsky, 1959; Blank, 1961; Leeuwenberg and Buffart, 1978; and Indow, 1979).

If this is the case, then the adoption of the Euclidean perspective for our maps, our models, and our theories, involves not only complex abstraction

from the real world but a conscious distortion of perceptual data. As an example of this, imagine an individual traversing most North American cities. He is forced to move for the most part in right angle steps but most frequently has to provide information in terms of straight line distances. "Distance" then becomes an experience which should not necessarily lead to a response that is best represented in Euclidean terms. If, therefore, we can claim that there is nothing inherently Euclidean about space or about spatial concepts developed and used by geographers, why should we not pay an increasing amount of attention to non-Euclidean spaces? For the balance of this paper, we would like to address ourselves to presenting a selection of alternative spaces which seem appropriate for use in the context of cognitive mapping and its relation to spatial behavior.

Some General Properties of Spaces

General space is an abstract set possessing a topology. Given such a definition, one can expect that the space is a set over which concepts of continuity and proximity may be given meaning. For an abstract set E with elements p, q it is possible to specify a set of correspondence rules which define a measure on ordered pairs, denoted (pq). Such a distance space can be called semi-metric if (i) the measure (pq) is a non-negative real number; (ii) (pq) = 0 if and only if p = q, and (iii) there is symmetry such that (pq) = (qp). (For a more detailed discussion of the development of geometry and some comments on General spaces, see Spiro & Noshiro, 1966; Golos, 1968; or Sklar, 1977).

Semi-metric spaces in general are characterized by two undesirable properties: a) the discontinuity of the distance function, and b) "unnatural" distance properties. With respect to (a), given the possibility for discontinuous distance functions, one may finish up with an indefinable topology so that the space as a whole cannot be characterized and may be identified only

in local terms. With regard to the second of these properties, one of the "unnatural" distance properties may relate to the concept of continuity in such a space. Continuity may be cnsured if the following characterization holds.

Let p,q be elements of semi-metric space S. If, for any two sequences p_n, q_n of elements of this semi-metric space, lim p_n = p and lim q_n = q implies the lim p_n q_n = pq, then the distance function is said to be continuous at p,q. The distance function is continuous on S if and only if it is continuous at each point pair of S. Imagine a series of polygons inscribed in this space such that, as the number of sides of the polygons approaches infinity, the distances between consecutive vertices approaches zero (i.e., implying that in the limit the distance between two distinct points goes to zero), then this leads to a contradiction of our intuitive feelings about distance. To avoid the problem of unusual distance properties (particularly those relating to continuity), the independence of the distance function-point assignment must be eliminated. To do this, we must make the distance assignment of a given point pair dependent on the distance assignments for previous point pairs. One common way is to impose the triangle inequality in addition to the assumptions of identity and symmetry. Stated formally, the requirements would then be:

1) if p = q, then (pq) = 0
2) if p ≠ q, then (pq) > 0
3) (pq) = (qp)
4) for any p, q, r, (pq) + (qr) ≥ (pr)

Any semi-metric space satisfying the four criteria is a metric space, i.e., the space is positive definite (criteria 1 & 2), symmetric (criteria 3) and satisfies the triangle inequality (criteria 4). Given some of these

characteristics of distances in a general metric or semi-metric space, it is possible to develop examples of distance relations for a variety of such metric spaces (see Love and Morris, 1972; and Gower, 1977).

Let us now turn to a discussion of some critical properties of mental maps that strongly suggest the use of non-Euclidean modes of representation as their most appropriate form.

Isotropy, Incompleteness, and Curvature

Space may not be isotropic. If isotropy does not hold, then the rotation of a set of objects about any given point will change distance relationships even though relative location itself is not affected. Thus in contrast to the Euclidean case where an orthogonal rotation of axes will maintain constant inter-point distance relationships, such a rotation in a Manhattan space, for example, will change the coordinate values and the distances. While a Manhattan space is metric, it is not isotropic since distance depends on angular orientation. Using a dominance metric, distance once again is related to the coordinate system and distance values vary depending on axis location.

Within a given population, one might expect that, because of differences in physiological and motoric skills, differences in exposure to sociological, psychological, and educational experiences, and differences in other personal, functional, and structural variables, some proportion of the population will more readily be described by one or another of these three general sub-classes of the Minkowskian metric. In a sample of 60 residents in the city of Columbus, Ohio (Golledge and Rayner, 1975) approximately 64% of individuals made inter-point distance judgments among selected pairs of 48 locations which were "best" represented in Euclidean space (i.e., when "best" was defined in terms of multi-dimensional scaling (MDS) "STRESS" statistics). Approximately 18% of the population each give distance judgments which conformed more to Manhattan or to dominance metrics than to Euclidean. In a later study (Spector, 1978), 121 out of 153 individuals gave information which produced correlations of greater than .8

between configurations of points reconstructed from their scaling of the two-dimensional inter-point distance relationships among the points and a representation of the same point set in two dimensional Euclidean space. Considering this evidence it is tempting to argue that at the scale of intra-urban analysis the use of a Euclidean metric can be supported for the construction of mental maps of such environments. But at this stage we still don't know whether a different form of geometry may be even more appropriate. For example, reappraisal of the data used in Spector's study produced an abrupt reversal of his results. Taking the MDS-derived configurations in a range of Minkowskian metrics for the 48 points as a base, the entire set of inter-point distances were calculated, then ranked, and a Spearman rank correlation coefficient between this set of distances and the set of objective inter-point distances was calculated. Distances in each point set were calculated using Manhattan, Euclidean, and Dominance metrics (see equation 1). The results indicated that the Euclidean metric has the highest correspondence between subjective and objective configurations only 18.7% of the time; city block and dominance metrics provided best fits for 49.6% and 31.7% of the subjects, respectively. Other tests, using regression procedures and clustering methods confirmed the significance of the city block metric and the poorer fit of the Euclidean (Richardson, 1979). In light of these conflicting results it would seem appropriate to speculate on some of the different forms of geometry that may be suitable for investigating the structure of cognitive representations.

Space is described through the use of the concept of curvature perhaps more frequently than through the use of the concept of straight lines. In Euclidean geometry a fundamental assumption is that of parallelism - i.e., that parallel lines do not meet. In spherical metrics, parallel lines not only eventually meet but in some curved spaces (such as hyperbolic space) they diverge. The perceptual experiences that individuals have when travelling

through space characteristically reflects a spherical geometry (Luneberg, 1950; Indow, 1974, 1979). In a sense, then, one acts as though while traversing a surface, the slope of the normal to the surface successively changes in a constant manner. With a positive change in a Cartesian coordinate and if the slope of the normal successively increases, the plane can be described in terms of spherical space; but if the slope successively decreases, the space is said to be hyperbolic.

We previously suggested that our concern with space should be at least in part tied to the way that individuals experience space. To adequately represent cognitive images, we need mental maps that truly reflect the internal geometry of people's perceptions (see Wood, 1978). While we have just described the case of a space of constant curvature (i.e., a spherical space), Riemann showed us that curvature need not be constant and that a geometry based on distance formulation may incorporate the idea that at a local level, or over very small displacements, a space of non-constant curvature can be closely approximated by a Euclidean space. Riemann utilized the concept of a tensor - i.e., a magnitude which can be used to transform a given set of points of an arbitrarily defined coordinate system to a new coordinate system. The definition of a general distance tensor on a plane is:

$$ds^2 = gxdx^2 + 2gxydxdy + gydy^2 \qquad (5)$$

where ds is a small distance displacement,

dx and dy are displacements along two arbitrary axes x and y, and

gx, gy, gxy is the weighting or tensor applied to each term of the squared elements of the distance displacement.

This fundamental tensor is frequently referred to as a covariant tensor of the second rank. Obviously, when one is concerned with examining a configuration composed of relative locations based on individual subjective estimates of location and proximity,

the nature of the set of tensors used to map the subjective judgments is extremely important. Some of the implications of the Riemann tensor that are particularly significant relate to the fact that space can be stretched or compressed as if it were elastic at different locations. The implication for this is obvious. If one takes a uniform set of points and stretches them differentially depending on location in the point set and direction from some particular origin, then distances between adjacent points can be either very small or very large. What this in turn implies is that if distance evaluations made by a subject and incorporated into his/her mental map are not based on the Euclidean formulation, then there is some warpage or disturbance of this space in which the judgments are made, and a Riemann tensor may be more appropriate for describing the occurrence of breaks or discontinuities in any mapping that is produced. If an individual uses notions of curved space in making his evaluations of proximities, then fitting his judgments to a Euclidean surface can only lead to seeming inconsistencies or "error" when we evaluate his/her judgments in the context of Euclidean space.

One of the critical differences between Euclidean geometries and geometries based on curved spaces is simply that the curved spaces are bounded whereas the Euclidean spaces are theoretically infinite. In a spherical space, no matter what direction one travels one should always be able to return to the origin point. This is clearly not the case in Euclidean space.

Perhaps the most innovative contributions concerning the use of non-Euclidean geometries for representing mental maps is contained in a recent article by Tobler on the "Geometry of Mental Maps" (Tobler, 1976). Even considering the relatively large scale mappings typical of urban areas, Tobler examines the possibility of using Riemannian tensors to transform what is normally regarded as a Euclidean urban space to preserve most of the critical topological features that dominate empirical examples of mental maps. The application of a Riemann tensor of course maintains the essential topological properties that can be recovered through a

variety of simple experimental designs based on subjective evaluations of inter-point distances or controlled sketching procedures. However, it distorts some of the directional, distance, and areal relationships that exist in the Euclidean representation. Tobler also suggests that the use of incomplete experimental designs and methods such as paired and triadic comparisons may produce holes, folds, cracks, and tears in the fabric of the cognitive image because of "fuzziness" with respect to the subject's conceptualization of place location. In other words, he suggests that the Euclidean plane be replaced with a topological manifold that for some individuals may be somewhat fluid and may vary in its structure either from time to time or as one changes the major origin node of the individual. Given the ephemeral nature of such a manifold, Tobler suggests that the wholesale adoption of purely metric assumptions may be unrealistic for the construction of mental maps.

Since folds or tears in the manifold representing an individual's knowledge of an area vary with the knowledge content of space, a third dimension (such as familiarity)should be used as a weight in the construction of the maps. Figures 1 and 2 show samples of individual cognitive configurations of 48 locations in Columbus, and familiarity ratings of the same points (subject #141 and #217). When examining such map pairs for the 126 subjects in the Columbus sample, it became quite obvious that major distortions in the cognitive configuration are correlated with the areas of lowest familiarity. Recent work by Rivizzigno (1976), Spector (1978) and Golledge and Spector (1978) has shown clear relationships between the size of location errors for places subjectively located in an urban environment, distance from the home place (i.e., the key primary node in the individual's spatial structure), and familiarity with the places.

An increase in the fuzziness component of location error as distance from a central origin increases, makes the geometric description of a mental map quite complex (e.g., see MacKay and Zinnes, 1978; Gale, 1980). Admittedly, one can summarize the trends of the distance distortion by the simple method

of transforming grids [pioneered once again by Tobler (1965) (see Fig. 3)], but a more feasible suggestion [once again from the work of Tobler (1976)] is to describe the mental map in terms of a small set of tensors that summarize the varying nature of the transformation between map locations and images of those locations (see Fig. 4). This would mean that the mathematical model describing the geometry of such a surface would have all the attributes of a tensor probability surface. To date, the closest existing descriptor along these lines is summarized in Tobler's recent work on directional components of flows in which he embeds a vector field on an isotropic Riemann manifold (Tobler, 1977; 1980). Applications of this work to cognitive maps can be seen in Gale (1980) and Gale and Golledge (1980).

From this initial discussion, there are a number of critical points that have been raised concerning the geometry of mental maps. First is the fundamental question of whether or not mental maps should be constructed as isotropic spaces. A second and equally important question is whether or not such a map can be rotated, scale transformed, or reflected without disturbing the distance relationship among key points on it. A third question concerns whether the most appropriate space for the construction of mental maps is an infinite plane or a bounded curved space. Following this is the question of whether the space has constant curvature or whether the curvature varies considerably from place to place or from sector to sector on the map. A question intuitively raised by almost everyone concerned with mental mapping is simply whether such maps have <u>any</u> metric properties whatsoever or whether they are only locally metric. As a first step towards answering some of these questions, we suggest that a variety of non-Euclidean metrics be considered for representing cognitive information. We start by presenting some of the features of perhaps the simplest curved spaces which may be appropriate for examining mental maps - i.e., Riemann spaces with constant curvature (see Ahlfors and Sario, 1960).

A Riemann space with coordinate system $(x_1, x_2, \ldots, x_n)$ has a metric representation in terms of local derivatives. As a quadratic form it is represented as:

$$ds^2 = \sum_{ij} g_{ij}\, dx_i\, dx_j \qquad (6)$$

The g_{ij} coefficients must satisfy the following conditions:

1. Each g_{ij} is a real single valued function of the coordinates and possesses continuous partial derivatives.
2. The axiom of symmetry holds (i.e., $g_{ij} = g_{ji}$)
3. $\det \{g_{ij}\} \neq 0$.

The quadratic form determines both the local and global geometric structure of a Riemannian space. What is of most critical importance as far as mental mapping is concerned, however, is that the most important parameter of a Riemann space is its curvature at each point. The Gaussian or total curvature at any point is defined as a product of the maximum and minimum curvature of all the geodesics passing through the point - (i.e., the maximum and minimum of all the curves of minimum distance passing through the point). The Riemann space itself is a hypersurface or manifold <u>S</u> in a higher dimension space <u>T</u>. The tangent space at points X in S separates <u>T</u> into two regions. The curvature of a geodesic through X arbitrarily is called "positive" if its center of curvature is in one region, and "negative" if its center is in the other region. Riemann spaces with constant curvature are called elliptic or spherical if the parameter of curvature <u>K</u> > 0, hyperbolic if <u>K</u> < 0, and Euclidean if K = 0.

While the quadratic form completely specifies any Riemann space, our familiarity with Euclidean spaces sometimes makes it more convenient to specify a Riemann space by a global equation. Lindman and Caelli (1978, p. 91) specify the form of such an equation as:

$$\sum_k r_k^2 = \frac{1}{K} \quad , \tag{7}$$

where

r_k = the coordinates of the higher dimension space T;

K = the Gaussian or total curvature.

The equation given by Lindman and Caelli represents the elliptic hyperspace of $K > 0$. They point out, however, that when at least one squared term of the quadratic is negative then the space is hyperbolic with $K < 0$. When the manifold S is embedded in the higher dimension space, the distances in the two spaces are related by the following equations (Lindman and Caelli, 1978, p. 92):

$$q_{ij} = 2/(K)^{\frac{1}{2}} \left[\sin \frac{(K^{\frac{1}{2}} d_{ij})}{2}\right] \qquad K > 0 \tag{8}$$

$$q_{ij} = d_{ij} \qquad K = 0 \tag{9}$$

$$q_{ij} = 2/(|K|)^{\frac{1}{2}} \left[\sinh (|K|^{\frac{1}{2}} d_{ij})\right] \qquad K < 0 \tag{10}$$

In the above formulation, d_{ij} is the distance arc measured in the manifolds, and q_{ij} is the distance measured in the space in which the manifold is embedded. Note particularly that when the space is complex, the q_{ij} may not be metric, but the d_{ij} will be metric if the complex space has exactly one imaginary dimension.

Some Alternate Metrics

In his discussion of a range of metrics suitable for representing structure in similarity data, Shepard (1974, p. 405) describes a hierarchy with increasing degrees

of generality. At the base of this hierarchy are the specific Minkowskian r metrics which include the Euclidean, Manhattan or city block, and Dominance metric (see equation 1). Each of these relates to metrics developed in isotropic infinite spaces. At a similar specific level are the hyperbolic and spherical spaces which are derivates of spaces of constant curvature. A space of constant curvature is seen as a special case of a General Riemann space. A constant curvature space is both curved and locally Euclidean, but unlike the General Riemann space it is also isotropic. This last condition holds because the unit spheres on which the space of constant curvature are based are uniformly spherical. A similar level to the General Riemann space is that space in which a General Minkowski metric applies; such a space is isotropic, flat and locally non-Euclidean. The isotropic and flat characteristics follow from the fact that the spheres underlying it are of constant size and shape. The General Riemann space is curved and locally Euclidean with ellipsoid unit spheres. Next in this hierarchy of generality is a Finsler space; this presupposes a continuous underlying coordinate space with its own intrinsic dimensionality. This space is curved, locally Minkowskian, but the unit spheres change continuously with location. The General Riemann space is a derivative of Finsler space.

The most general of all spaces in Shepard's classification are general metric and semi-metric spaces in which the spaces are defined solely in terms of inter-point distances; they do not imply a specific dimensionality and do not implicitly embed coordinate systems within them. Postulates underlying the general metric space are too general by themselves to allow the development of rigid theoretical geometrical constructions. Shepard suggests that the most appropriate space for the representation of cognitive information is one which is not too particular and maintains a level of structure, yet allows for non-trivial investigation and conclusion. This would imply that the fundamental metric axioms of

triangularity, symmetry, and positivity and identity at least have to be maintained within the fabric of the space. Such a space must be finitely compact or boundedly infinite: in other words, a finite dimensioned space such as this one cannot be extended without increasing its dimensionality. Further, such a space should be convex and incorporate the notion of betweenness: i.e., given two distinct points X and Z then it is possible for a point Y to exist between those two locations. Given these brief notions, one can immediately see that both Riemann spaces and those spaces suited to the use of all Minkowskian metrics are particular subclasses of a general metric space.

Conclusions and Speculations

If we examine configurations of points that have been generated using inter-point distance estimation or "a set of points located in a space which is defined solely in terms of inter-point distances, then we are confronted with a space which implies neither a specific dimensionality nor is there implicitly embedded within it a coordinate system. If we can further find key nodes in this particular space, we might imagine that as distance increased from each key node, the probability of a fold, crack, tear, hole, or other warpage of the space would increase consideraly. Assume further that the key nodes are not uniformly spaced. If one were then to construct a set of Thiessen polygons for this set of non-uniformly spaced nodes, there would be considerable variations in the distance of the edges of the polygons from the key nodes (Fig. 5). One might further expect that, in those areas of each polygon which are most distant from the key node, information about the area may be least and the probability of distortion or warpage might increase except along the dominant gradient or link path between adjacent nodes. Using conventional ideas from probabilistic market area analysis and the recent suggestions of Tobler (1976) concerning non-constant

warpages and the development of a probability surface for distortions in "mental maps", one can visualize a mental map of the previous set of locations with the major distortions or warpings occurring away from the major nodes and the primary paths that connect them, and increasing in probability of occurrence in the more inaccessable or distant paths of each Thiessen polygon (Fig. 6). The result would be to produce a map which associates points or key locations with error probability surfaces. The probability of warping would, therefore, be non-constant as direction changes from key nodes, or as distance increases from the key node to different edge segments of the Thiessen polygon.

The building of such a probability surface is simple in Euclidean space, but the specification of the surface parameters is somewhat more complicated even in a simple Riemann space of constant curvature. Detailed examination of over 200 individual configurations of 48 locations for the city of Co[illegible]bus has shown that location errors differs considerably across the individual maps and that there is definitely both a directional and distance component to the distribution of errors (e.g., see Rivizzigno, 1976; Spector, 1978; Golledge and Spector, 1978;Gale, 1980)(Fig. 4. shows location errors for subject 100). What is more, this error surface undulates depending on the activity pattern of the individual, for as information about different segments of the environment increases, the probability of maintaining a constant location error diminishes. At various time periods, therefore, the manifold in which the points are located can be warped differentially (e.g., see Fig. 7 - an example of the street map of Fig. 4 warped to fit the cognitive configuration obtained for subject 100). If one were to obtain a cross-section through time of a series of these manifolds, one should be able to recreate a history of the main repetitive components of an individual's spatial behavior for that time period.

Obviously, the first thing to be done is to attempt to define the appropriate parameters which describe the Riemann space in which a number of subjective

configurations exist. Once this has proved to be a feasible operation, then expanding the work to cover manifolds produced at different stages of the environmental learning process and recreating histories of spatial behaviors associated with each manifold would seem to be an intriguing direction for constructive use of current work on mental maps.

An alternate way of envisioning the warped manifolds suggested in the previous paragraph is to introduce the concept of a mean information field with holes. Figure 8, for example, shows a standard grid with a series of familiarity measures allocated to each grid cell. As one can see, there are two holes in the mean information field where zero information is recorded, and two major peaks - one towards the S.W. corner of the field and another towards the N.W. corner. If one contoured the mean information field, the holes would stand out in the two dimensional Euclidean representation of it. However, if we collapsed the field so as to eliminate the holes, the configuration that resulted would more readily be described in non-Euclidean terms. Such a warped field more closely approximates the sketch maps drawn by individuals with incomplete information about test environments; the consequent shortening of distances across places with low information levels and the exaggeration of distances where information is consistently high mirrors many of the types of distortions recovered from individual configurations of urban areas in earlier works (Golledge and Rayner, 1975; Golledge, Rivizzigno, and Spector, 1976; Rivizzigno, 1976; Spector, 1978).

As a further step in an attempt to define the types of metrics most suitable for the representation of subjective configurations of places, current work at U.C. Santa Barbara is aimed at defining configurations of places in Riemann spaces such that an index of fit between subjective and objective configurations mapped onto the same space can be obtained. In general, it would appear that questions related to the suitability of representing mental maps in metric spaces

need to be answered before much confidence can be placed in widespread use of such maps in conventional geographic work.

REFERENCES

1. Ahlfors, L.V. and Sario, L., 1960, Riemann Surfaces, Princeton, New Jersey: University Press.

2. Blank, A.A., 1961, "Curvature of binocular visual space," Journal of the Optical Society of America, 51, 335-339.

3. Bunge, W., 1962, Theoretical Geography, Lund; Lund Studies in Geography Series C General and Mathematical Geography #1.

4. Gale, N., 1980, "An analysis of spatial variation in the distortion and fuzziness of cognitive maps", unpublished MA Thesis; Department of Geography, U.C.S.B.

5. Gale, N. and Golledge, R.G., 1981, "Distortion and fuzziness components of error in cognitive maps" paper to be read at AAG meetings, Los Angeles.

6. Golledge, R.G. and Rayner, J.N. (Eds.), 1975, Cognitive Configurations of the City: Volume 1, Volume 2 (1977), Columbus, Ohio: Department of Geography and the Ohio State University Research Foundation.

7. Golledge, R.G., Rivizzigno, V., and Spector, A.N., 1976, "Learning about a city: Analysis by multidimensional scaling" in R.G. Golledge and G. Rushton (Eds.), Spatial Choice and Spatial Behavior, Columbus, Ohio: Ohio State University Press, 95-119.

8. Golledge, R.G. and Spector, A.N., 1978, "Comprehending the urban environment: Theory and practice," Geographical Analysis, 9, 403-426.

9. Golos, E.B., 1968, Foundations of Euclidean and Non-Euclidean Geometry, New York: Holt, Rinehart and Winston, Inc.

10. Gower, J.C., 1977, "The analysis of asymmetry and orthogonality" in J.R. Barra, et al. (Eds.), Recent Developments in Statistics, New York: North Holland Publishing Co., 109-123.

11. Indow, T., 1974, "On geometry of frameless binocular perceptual space," Psychologia, 17, 50-63.

12. Indow, T., 1979, "Alleys in visual space," Journal of Mathematical Psychology, 19, 221-258.

13. Krause, E.F., 1975, Taxicab Geometry, Menlo Park, California: Addison-Wesley.

14. Lau, J., 1978, "On taxicab geometry and geography," Ph.D. Dissertation, University of Michigan, Ann Arbor, Michigan.

15. Leeuwenberg, E.J.L. and Buffart, H.F.J.M., 1976, Formal Theories of Visual Perception, N.Y.: John Wiley & Sons.

16. Ley, D. and Samuels, M. (Eds.), 1978, Humanistic Geography, Chicago: Maaroufa Press.

17. Lindman, H. and Caelli, T., 1978, Constant curvature Riemannian scaling," Journal of Mathematical Psychology, 17, 89-109.

18. Love, R.F. and Morris, J.G., 1972, "Modelling inter-city road distances by mathematical functions," Operations Research Quarterly, Vol. 23, #1, 61-71.

19. Luneberg, R.K., 1950, "The metric of binocular visual space," Journal of the Optical Society of America, 40, 637-642.

20. Richardson, G.D., 1979, The Appropriateness of Using Various Minkowskian Metrics for Representing Cognitive Maps Produced by Nonmetric Multidimensional Scaling, M.A. Thesis, Department of Geography, University of California, Santa Barbara.

21. Rivizzigno, V.L., 1976, Cognitive Representations of an Urban Area, unpublished Ph.D. Dissertation, Department of Geography, Ohio State University.

22. Roberts, F.S. and Suppes, P., 1967, "Some problems in the geometry of visual perception," Synthese, 17, 173-201.

23. Sario, L. and Noshiro, K., 1966, Value Distribution Theory, Princeton, New Jersey: Van Nostrand Co., Inc.

24. Shepard, R.N., 1974, "Representation of structure in similarity data - problems and prospects," Psychometrika, 39, 373-422.

25. Sklar, L., 1977, Space, Time and Spacetime, Berkeley, University of California Press (paperback edition).

26. Spector, A.N., 1978, An Analysis of Spatial Imagery, unpublished Ph.D. Dissertation, Department of Geography, Ohio State University.

27. Stevens, A. and Coupe, P., 1978, "Distortions in judged spatial relations," Cognitive Psychology, Vol. 10, 422-437.

28. Tobler, W.R., 1965, "Computation of the correspondence of geographic patterns," Papers and Proceedings of the Regional Science Association, 15, 131-139.

29. Tobler, W.R., 1976, "The geometry of mental maps" in R.G. Golledge and G. Rushton (Eds.), Spatial Choice and Spatial Behavior, Columbus, Ohio: Ohio State University Press, 69-82.

30. Tobler, W.R., 1977, Bidimensional Regression, Department of Geography, University of California at Santa Barbara.

31. Tobler, W.R., 1980 - Personal Communication - August 18th.

32. Wood, D., 1978, "Introducing the cartography of reality" in D. Ley and M. Samuels (Eds.), Humanistic Geography, Chicago: Maaroufa Press, 207-219.

33. Zautinsky, E.M., 1959, "Spaces with non-symmetric distance," Memoirs of the American Mathematical Society, 34.

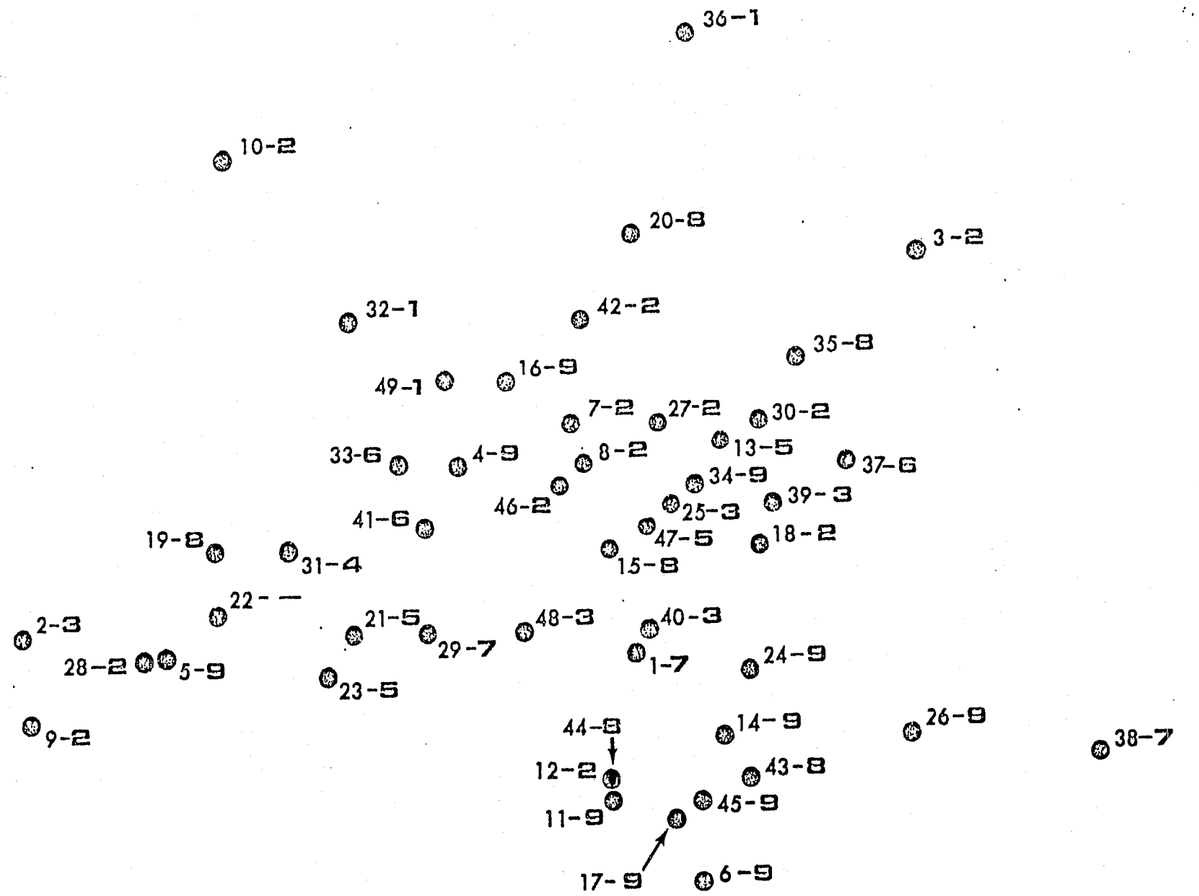

36-1
10-2
20-8
3-2
32-1
42-2
35-8
49-1
16-9
7-2
27-2
30-2
13-5
33-6
4-9
8-2
37-6
34-9
46-2
39-3
25-3
41-6
47-5
18-2
19-8
31-4
15-8
22-—
2-3
21-5
48-3
40-3
29-7
1-7
24-9
28-2
5-9
23-5
44-8
14-9
26-9
38-7
9-2
12-2
43-8
11-9
45-9
17-9
6-9

FIG. 1

(186)

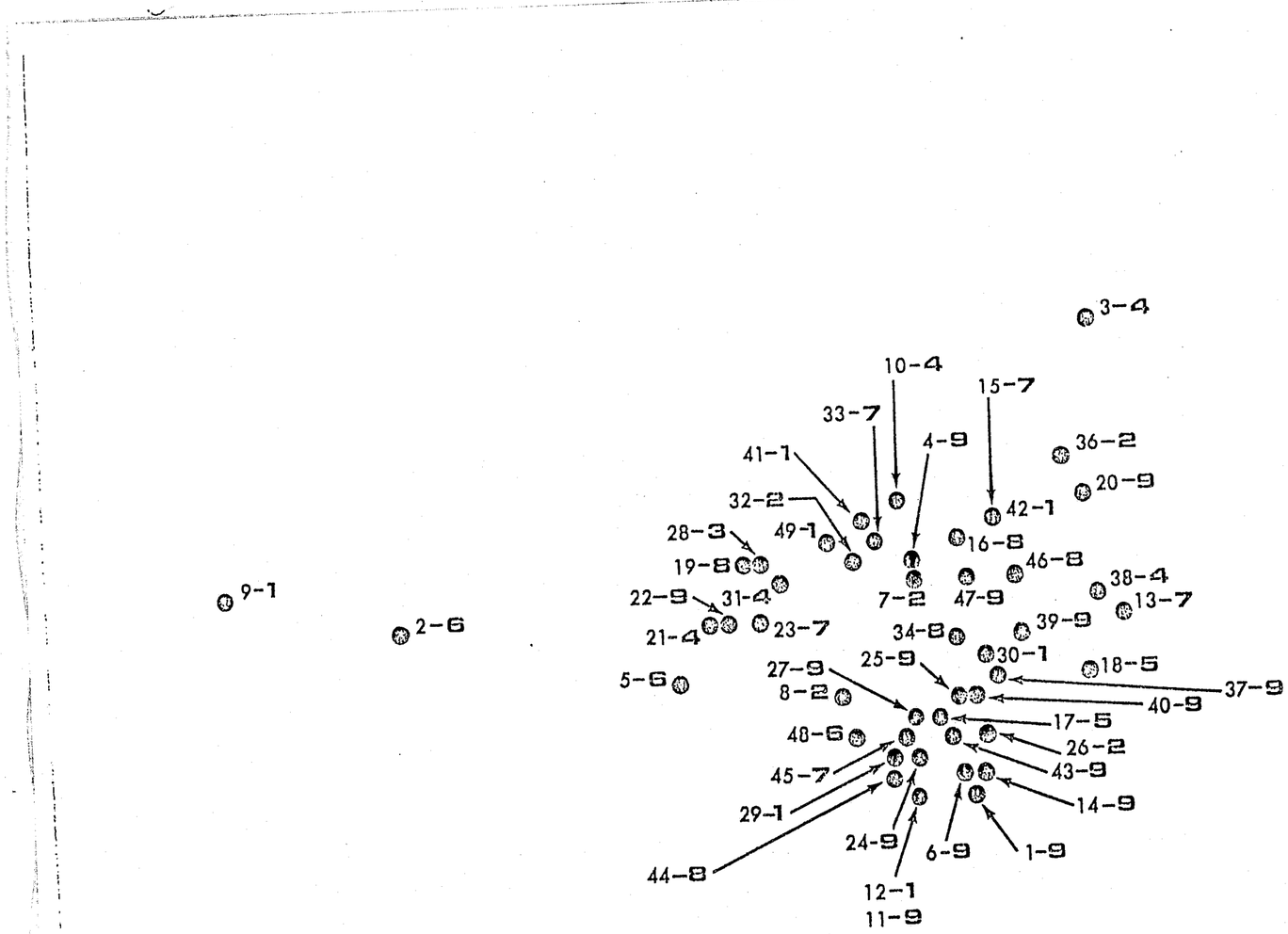
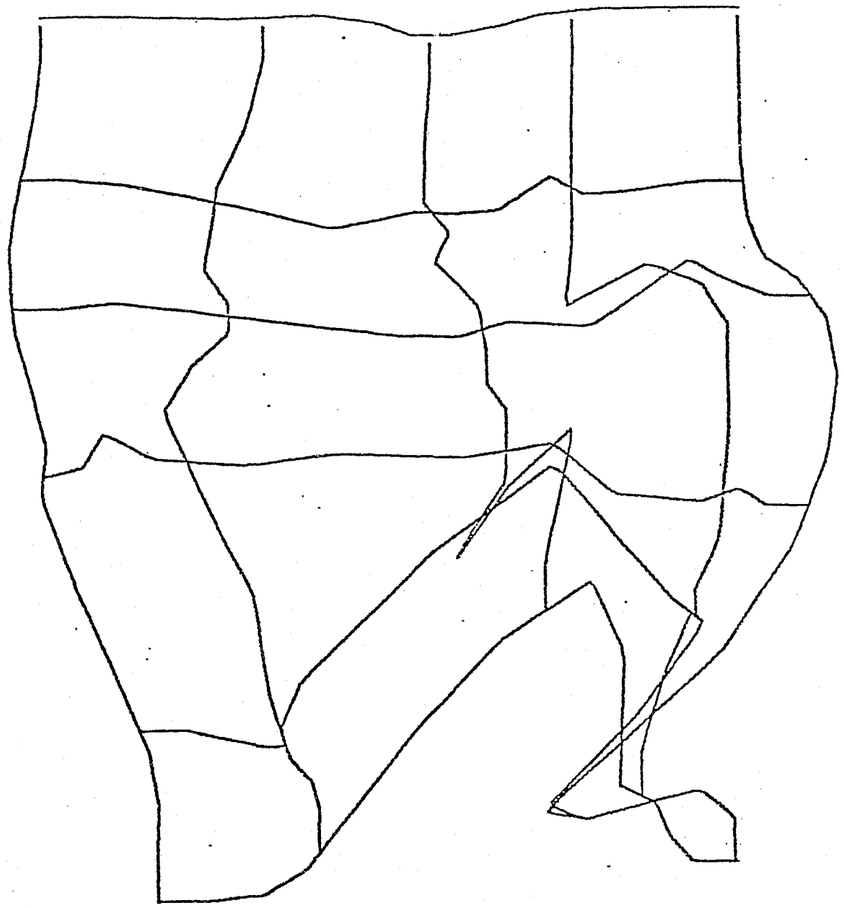

3-4
10-4
33-7
15-7
41-1
4-9
36-2
20-9
32-2
42-1
49-1
28-3
16-8
19-8
46-8
38-4
9-1
22-9
31-4
7-2
47-9
13-7
2-6
21-4
23-7
34-8
39-9
30-1
18-5
5-6
27-9
25-9
37-9
8-2
40-9
17-5
48-6
26-2
45-7
43-9
29-1
14-9
24-9
6-9
1-9
44-8
12-1
11-9

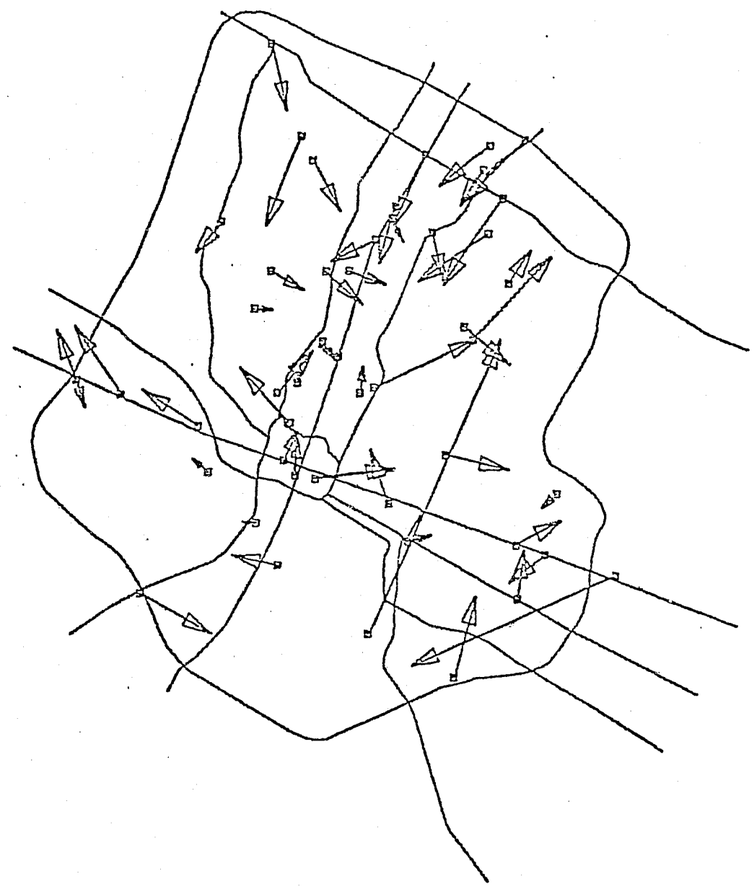
FIG. 2
(187)



COGNITIVE CONFIGURATION OF COLUMBUS - SUBJECT 100

THE TRANSFORMED IMAGE

FIG. 3



COGNITIVE CONFIGURATION OF COLUMBUS - SUBJECT 100

DIFFERENCE BETWEEN THE IMAGES (AFTER AN AFFINE FIT)

FIG. 4

(a) Thiessen Polygons

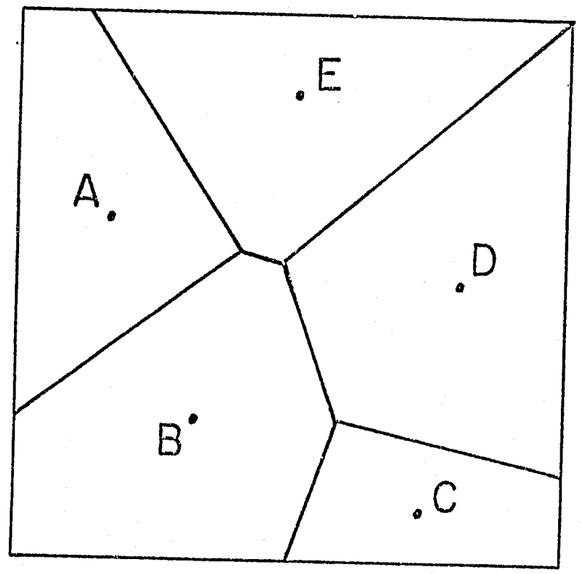
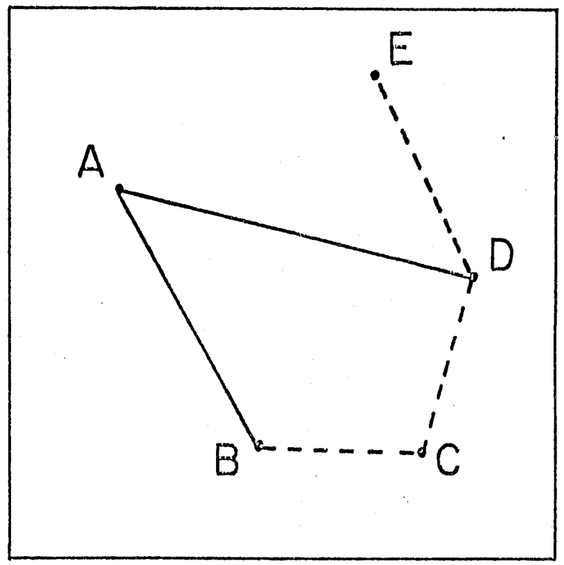


FIG. 5(a)

(190)

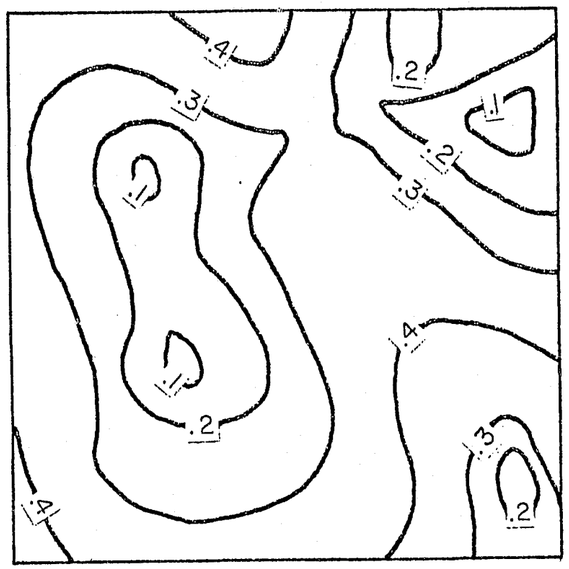
(j) Networks



Major link
Minor link

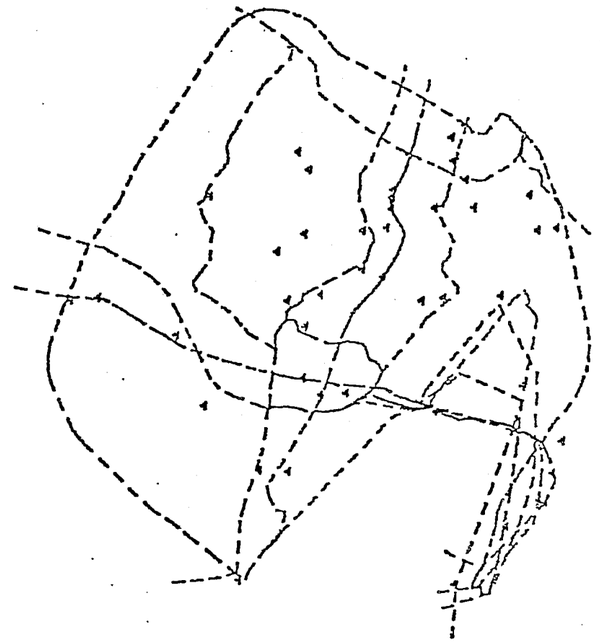
FIG. 5(b)
(191)

(b) Possible Error
Probability Surface



(192)

FIG. 6



COGNITIVE CONFIGURATION OF COLUMBUS - SUBJECT 100

THE TRANSFORMED IMAGE

FIG. 7

(c) Subjective M.I.F.

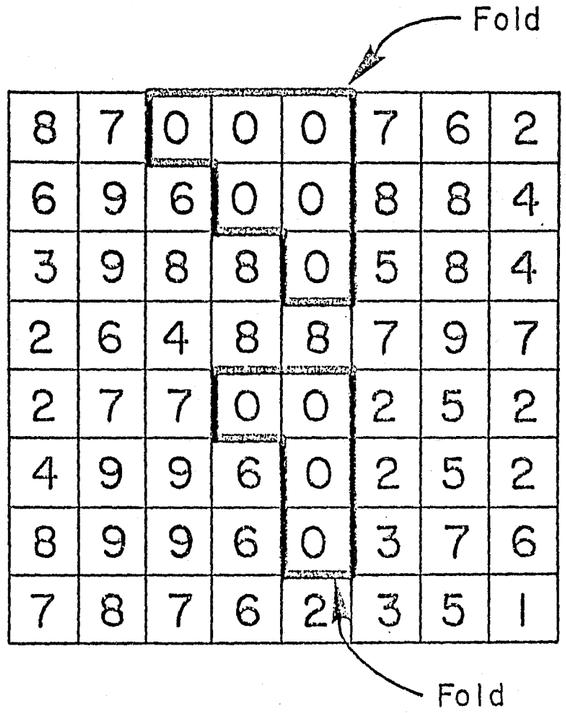


FIG. 8

8) GENERAL REFERENCES

Baker, F.B. and Hubert, L.J. Inference procedures for ordering theory. Journal of Educational Statistics, 1977, 2 217-233 (a).

Baker, F.B. and Hubert, L.J. Applications of combinatorial programming to data analysis: Seriation using asymmetric proximity measures. British Journal of Mathematical and Statistical Psychology, 1977, 30, 154-164 (b).

Besag, J. and Diggle, P.J. Simple Monte Carlo tests for spatial pattern. Applied Statistics, 1977, 26, 327-333.

Blalock, H. Causal Inferences in Nonexperimental Research. Chapel Hill: University of North Carolina Press, 1964.

Blalock. H. Aggregation and measurement error. Social Forces, 1971, 50, 151-165.

Blin, J.M. and Whinston, A.B. Discriminant functions and majority voting. Management Science, 1975, 22, 557-566.

Box, G.E.P. Science and statistics. Journal of the American Statistical Association, 1976, 71, 791-799.

Brown, L.A., Odland, J. and Golledge, R.G. Migration, functional distance and the urban hierarchy. Economic Geography, 1970, 46 (3), 472-485.

Carroll, J.D. and Chang, J.J. Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. Psychometrika, 1970, 35, 283-319.

Carter, R.L. The criminal's image of the city. Unpublished Ph.D dissertation, University of Oklahoma, 1974.

Clark, W.A.V. and Avery, K.L. The effects of data aggregation in statistical analysis. Geographical Analysis, 1976, VIII (4), 428-438.

Clark, W.A.V. and Avery, K.L. Aggregation and statistical analysis: A comment. Geographical Analysis, 1978, X (2), 194-195.

Cliff, A.D. and Ord, J.K. Spatial Autocorrelation. London: Pion, 1973 (revised, 1981).

Curry, L. A note on spatial association. The Professional Geographer, 1966, 18, 97-99.

Duncan, O.D. and Davis, B. An alternative to ecological correlation. American Sociological Review, 1953, 18, 665-666.

Dywer, P.S. The mean and standard deviation of the distribution of group assembly sums. Psychometrika, 1964, 29, 397-408.

Edgington, E.S. Statistical Inference: The Distribution-free Approach. New York: McGraw-Hill, 1969.

Ehrenberg, A.S.C. On sampling from a population of rankers. Biometrika, 1952, 39, 82-87.

Golledge, R.G. Multidimensional analysis in the study of environmental behavior and environmental design. In I. Altman and J. Wohlwill (Eds.), Research on human behavior, Vol. II. Plenum Press, 1978, 1-42.

Golledge, R.G., Rivizzigno, V.L. and Spector, A. Learning about urban environments: Analysis by M.D.S. In R.G. Golledge (Ed.), Cognitive configurations of the city, Vol. I. Department of Geography and OSURF, Ohio State University, 1975.

Golledge, R.G. and Rushton, G. Multidimensional scaling: Review and geographic applications. Technical Report #10, AAG Commission on College Geography, 1972, 1-71.

Golledge, R.G. and Rushton, G. Spatial Choice and Spatial Behavior. Columbus, Ohio: Ohio State University Press, 1976.

Golledge, R.G. and Spector, A. Comprehending the urban environment: Theory and practice. Geographical Analysis, 1978, X (4), 403-426.

Goodman, L.A. Ecological regressions and behavior of individuals. American Sociological Review, 1953, 18, 663-664.

Goodman, L.A. Some alternatives to ecological correlation. American Journal of Sociology, 1959, 64, 610-625.

Hannan, M.T. Aggregation and Disaggregation in Sociology. Lexington, Mass.: D.C. Heath, 1971 (a).

Hannan, M.T. Problems of aggregation. In Causal Models in Social Sciences, H. Blalock (Ed.). Chicago: Aldine, Atherton, Inc., 1971 (b).

Hannan, M.T. Approaches to the Aggregation Problem. Technical Report #46, Laboratories for Sociological Research, Standford University, 1972.

Harries, K.D. The Geography of Crime and Justice. New York: McGraw-Hill, 1974.

Hays, W.L. A note on average tau as a measure of concordance. Journal of the American Statistical Association, 1960, 55, 331-341.

Hildebrand, D.K., Laing, J.D., and Rosenthal, H. Prediction Analysis of Cross-Classifications. New York: Wiley, 1977.

Hope, A.C.A. A simplified Monte Carlo significance test procedure. Journal of the Royal Statistical Society, Series B, 1968, 30, 582-598.

Hubert, L.J. Seriation using asymmetric proximity measures. British Journal of Mathematical and Statistical Psychology, 1976, 29, 32-52.

Hubert, L.J. Nominal scale response agreement as a generalized correlation. British Journal of Mathematical and Statistical Psychology, 1977, 30, 98-103.

Hubert, L.J. Nonparametric tests for patterns in geographic variation: Possible generalizations. Geographical Analysis, 1978, 10, 86-88 (a).

Hubert, L.J. Generalized proximity function comparisons. British Journal of Mathematical and Statistical Psychology, 1978, 31, 179-192 (b).

Hubert, L.J. Matching models in the analysis of cross-classifications. Psychometrika, 1979, 44, 21-41 (a).

Hubert, L.J. The comparison of sequences. Psychological Bulletin, 1979, 86, 1098-1106 (b).

Hubert, L.J. Generalized concordance. Psychometrika, 1979, 44, 135-142 (c).

Hubert, L.J. and Baker, F.B. Analyzing distinctive features. Journal of Educational Statistics, 1977, 2, 79-98 (a).

Hubert, L.J. and Baker, F.B. The comparison and fitting of given classifications schemes. Journal of Mathematical Psychology, 1977, 16, 233-253 (b).

Hubert, L.J. and Baker, F.B. Analyzing the multitrait-multimethod matrix. Multivariate Behavioral Research, 1978, 13, 163-179 (a).

Hubert, L.J. and Baker, F.B. Applications of combinatorial programming to data analysis: The traveling salesmen and related problems. Psychometrika, 1978, 43, 81-91 (b).

Hubert, L.J. and Baker, F.B. Evaluating the conformity of sociometric measures. Psychometrika, 1978, 43, 31-41 (c).

Hubert, L.J. and Baker, F.B. Evaluating the symmetry of a proximity matrix. Quality and Quantity, 1979, 13, 77-84.

Hubert, L.J. and Levin, J.R. A general statistical framework for assessing categorical clustering in free recall. Psychological Bulletin, 1976, 83, 1072-1080 (a).

Hubert, L.J. and Levin, J.R. Evaluating object set partitions: Free-sort analysis and some generalizations. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 459-470 (b).

Hubert, L.J. and Levin, J.R. Inference models for categorical clustering. Psychological Bulletin, 1977, 84, 878-887.

Hubert, L.J. and Schultz, J.V. Maximum likelihood paired-comparison ranking and quadratic assignment. Biometrika, 1975, 62, 655-659.

Hubert, L.J. and Schultz, J.V. Quadratic assignment as a general data analysis strategy. British Journal of Mathematical and Statistical Psychology, 1976, 29, 190-241 (a).

Hubert, L.J. and Schultz, J.V. A note on seriation and quadratic assignment. Classification Society Bulletin, 1976, 3, 16-24 (b).

Hubert, L.J. and Subkoviak, M.J. Confirmatory inference and geometric models. Psycholgical Bulletin, 1979, 86, 361-370.

Kaiser, H.F. A second generation little jiffy. Psychometrika, 1970, 35, 401-415.

Kendall, M.G. Rank Correlation Methods (4th edition). New York: Hafner, 1970.

Lawler, E.L. The quadratic assignment problem: A brief review. In B. Roy (Ed.), Combinatorial programming: Methods and applications. Dordrecht: Reidel, 1975, 351-360.

Lee, Y. and Egan, F. The geography of urban crime: The spatial pattern of serious crime in the city of Denver. Proceedings, AAG, 1972, 59-64.

Lyerly, S.B. The average Spearman rank correlation coefficient. Psychometrika, 1952, 17, 421-428.

Mantel, N. The detection of disease clustering and a generalized regression approach. Cancer Research, 1967, 27, 209-220.

Mielke, P.W., Berry, K.J., and Johnson, E.S. Multi-response permutation procedures for a priori classification. Communications in Statistics--Theory and Methods, 1976, 5, 1409-1424.

Mirkin, B.G. Approximation problems in relation space and the analysis of nonnumeric methods. Automation and Remote Control, 1974, 35, 1424-1431.

Nettler, G. Explaining Crime. New York: McGraw-Hill, 1974.

Page, E.B. Ordered hypotheses for multiple treatments: A significance test for linear ranks. Journal of the American Statistical Association, 1963, 58, 216-230.

Peet, R. The geography of crime: A political critique. The Professional Geographer, 1975, 27, 277-280.

Phillips, D.P. A prologue to the geography of crime. Proceedings, AAG, 1972, 86-91.

Phillips, D.P. Risk-related crime rates and crime patterns. Proceedings, AAG, 1973, 221-224.

Puri, M.L. and Sen, P.K. Nonparametric Methods in Multivariate Analysis. New York: Wiley, 1971.

Pyle, G.F. The Spatial Dynamics of Crime. Chicago: University of Chicago, Department of Geography Research Paper #159, 1974.

Schucany, W.R. and Frawley, W.H. A rank test for two group concordance. Psychometrika, 1973, 38, 249-258.

Schultz, J.V. and Hubert, L.J. A nonparametric test for the correspondence between two proximity matrices. Journal of Educational Statistics, 1976, 1, 59-67.

Shapiro, C.P. and Hubert, L.J. Asymptotic normality of permutation statistics derived from weighted sums of bivariate functions. Annals of Statistics, 1979, 7, 788-794.

Simpson, D. and Tsukui, J. The fundamental structure of input-output tables: An international comparison. Review of Economic Statistics, 1965, 47, 434-446.

Smit, B. On aggregation and statistical analysis. Geographical Analysis, 1978, X (2), 190-194.

Tukey, J.W. The future of data analysis. Annals of Mathematical Statistics, 1962, 33, 1-67.

Younger, D.H. Minimum feedback arc sets for a directed graph. IEEE Transactions on Circuit Theory, 1963, 10, 238-245.

PART II

Selected Computer Programs

Summary of Selected Computer Programs

1) Programs used in paper "Aggregation in Data Tables: Implications for Evaluating Criminal Justice Statistics"*

a) Program to correlate each row of a ranked matrix with a perfect pattern, then to find the average correlation and its associated Z-statistic.

b) Same as above except used for raw data, not ranks.

c) Computation of aggregated correlation coefficients between a standardized data matrix and a perfect pattern.

d) Multiple group concordance.

e) Hierarchical clustering based on R*A and R*B.

f) Pre and post aggregation correlational analysis.

* Programs available on request from PI's

2) Program XBLOK used in paper #2 "Assessing Homogeneity in Cross-Classified Proximity Data"*

PROBLEM-- Given an arbitrary proximity matrix which is cross-classified according to two dimensions, the problem is to assess the salience of each dimension individually.

ALGORITHM-- The basic strategy is to block on the level of one dimension while assessing saliency in the other dimension.

Suppose that N profiles are defined over a set of T variables, and that the N profiles are cross-classified by m levels of factor A and n levels of factor B, so that N = mn. Any measure of proximity, such as correlations or Euclidean distances, may be used to produce a symmetric N x N proximity matrix. One appropriate ordering of the N profiles would be to have the n levels of factor B repeated for each of the m levels of factor A down the rows and across the columns of the proximity matrix. The main diagonal of the matrix would contain measures of similarity between same levels of both factors. There would also be m blocks (submatrices) of size n x n straddling the main diagonal and these blocks would contain measures of proximity between same levels of factor A; however, the off-diagonal elements of these blocks would all be measures of proximity between different levels of factor B. Similarly, there would be m(m - 1) blocks away from the main diagonal which contain proximities that are all between different levels of factor A. And the main diagonal elements of these off-diagonal blocks would

each be a measure of similarity between same levels of factor B.

Thus by focussing attention on those latter elements, homogeneity within the levels of factor B may be assessed without regard to the confounding effects of homogeneity within the levels of factor A. In particular, those elements may be summed to produce an index of homogeneity within factor B and this index, called Γ obs, may be compared to its expected value under the assumption of randomness to determine its statistical significance.

The significance tests utilized are of two varieties and both are non-parametric in nature. Within each level of factor A, the n! possible reorderings of the levels of factor B are assumed equally likely. Since there are m levels of A, this implies $(n!)^m$ equally likely such realizations of the complete matrix. Each such realization leads to a value for Γ, and when tabled, these values form a reference distribution to evaluate the size of the observed index. In the usual interpretation, the hypothesis of randomness (blocking on factor A) would be rejected if Γ obs were sufficiently extreme with respect to the reference distribution; the proportion of realizations giving values of Γ as great or more extreme than Γ obs would be reported directly as the significance level.

An alternative test is also available. First note that Γobs is an index calculated for a partition of the N objects into m groups of n objects each, and the overall index is the sum of the elements corresponding to the n same levels of factor B within groups corresponding to different levels of factor A. If the sum of the diagonal elements of one such block is denoted as Γ_{uv},

then the first two moments of an arbitrary value of this form is given in the literature (Puri and Sen, 1971) as:

$$E\ (\Gamma_{uv}) = 1/n \sum_i \sum_j d_{ij} \ ;$$

$$V\ (\Gamma_{uv}) = 1/(n(n-1))\ \{(1/n)\ Q_1 - (Q_2 + Q_3) + nQ_4\}\ .$$

where $Q_1 = \left(\sum_i \sum_j d_{ij} \right)^2$;

$Q_2 = \sum_j \left(\sum_i d_{ij} \right)^2$;

$Q_3 = \sum_i \left(\sum_j d_{ij} \right)^2$;

$Q_4 = \sum_i \sum_j d^2_{ij}$.

Moreover, since the separate indices of the form Γ_{uv} are independent in pairs, the first two moments of Γ are:

$$E\ (\Gamma) = \sum_{u>v} E\ (\Gamma_{uv}) \ ;$$
$$V\ (\Gamma) = \sum_{u>v} V\ (\Gamma_{uv}) \ ,$$

the associated Z-statistic for Γ obs is:

$$Z = \frac{\Gamma \text{obs} - E\ (\Gamma)}{\sqrt{V\ (\Gamma)}}$$

and a simple Cantelli bound conservatively assures that the significance level for any form of reference distribution whatsoever will be no larger than $1/(Z^2 + 1)$.

INPUT

Deck Make-up:

1. System Cards
2. Source or object program deck
3. System cards
4. Control Card 1
5. Control Card 2
6. Control Card 3
7. Control Card 4
8. Data cards
9. System Delimiter Card

Card Format:

Control Card 1 -- (I4,D12.0)

cols. 1-4 Number of random permutations wanted (right adjust)

cols. 5-16 Integer random permutation seed

Control Card 2 -- (4A4,I4)

cols. 1-16 Name of factor A

cols. 17-20 Number of levels (m) of factor A (right adjust)

Control Card 3 -- (4A4,I4)

cols. 1-16 Name of factor B

cols. 17-20 Number of levels (n) of factor B (right justify)

Control Card 4 -- (18A4)

cols. 1-72 Variable format for a logical row of input (must be F or E format)

Data Cards --

The data consists of an arbitrarily-defined symmetric proximity matrix. The entire matrix should be input, and it should be ordered such that each level of factor A contains all n levels of factor B. (note: $mn \leq 100$)

COMPUTATION-- As the matrix is symmetric, all computations are performed below the main diagonal. Four general descriptive averages are computed first. These include: 1) the average proximity among elements corresponding to same level of factor A and different level of factor B; 2) the converse average; 3) the average proximity among elements corresponding to different levels of both factors; and 4) the average proximity in the matrix as a whole, excluding the main diagonal.

Then Γ obs for factor B is computed by summing the diagonal elements of all the lower diagonal blocks of the matrix. Following that is the computation of E (Γ) and V (Γ), which in turn allow the calculation of the Z-statistic and a Cantelli significance level.

A reference distribution for Γ is then produced by randomly permuting the rows and separately the columns of each of the below-diagonal blocks of the matrix and recalculating the index. This is done as many times as is requested by control card 1.

The matrix is then reordered such that the m levels of factor A are contained within each level of factor B. The same computations are repeated, except that now homogeneity within the levels of factor A are being assessed while blocking on factor B.

OUTPUT-- The proximity matrix is printed out in exactly the same format as it was input. This requires two points of caution on input: 1) Column one should be left blank on input if the output will be to a line printer, since this column is omitted by the printer; 2) One extra space per value must be allocated for a decimal point on output if the data is input in F format without decimal points.

Following the proximity matrix is output a table containing the four descriptive averages discussed at the beginning of the previous section.

The statistics for assessing homogeneity within factor B while blocking on factor A are then outputted. These include Γobs, E (Γ), V (Γ) and the standard deviation of Γ, the Z-value, and the Cantelli significance level.

Finally the reference distribution of random indices which has been sorted from least to greatest is printed.

Having completed the assessment of homogeneity in factor B, the matrix is reordered to assess the same in factor A. The matrix is printed out in its reordered from and the statistics pertaining to factor A are pointed out in the same sequence as were those pertaining to factor B.

LIMITS AND COMPATABILITY-- The program is designed to operate on any symmetric proximity matrix of order 100 x 100 or less. It requires a region in core of approximately 140k bytes. The sample run of a 6 x 6 matrix required 2.7 seconds of CPU time, while the original application on a 30 x 30 matrix required 4.5

seconds of CPU time. Both were run on an ITEL AS-6, which is comparable to an IBM 360/70.

The program is written in IBM FORTRAN, and has been executed on the level H-extended compiler. It must be linked with IMSL subroutine GGUBS for the purpose of random number generation.

The input and output devices may vary. While they are set at units 5 and 6, respectively, those unit numbers may be easily changed by altering the data statement at the front of the program.

* Program listings available on request from PI's

* QAP is currently not fully exportable and is specific to the U.C.S.B. Itel AS-6. Those interested in following its development to exportable stage should contact Professor L.J. Hubert, Graduate School of Education, U.C.S.B.

PART III: APPENDICES

A) Personnel

The following personnel were associated with the project either as investigators or as consultants.

1) Reginald G. Golledge, Professor of Geography, U.C.S.B. (Co-PI)
2) Lawrence J. Hubert, Professor of Education, U.C.S.B. (Co-PI)
3) Waldo Tobler, Professor of Geography, U.C.S.B.
4) Richard Church, Assoc. Professor of Civil Engineering, University of Tennessee
5) G. Donald Richardson, Geography Graduate Student, U.C.S.B.
6) C.M. Costanzo, Geography Graduate Student, U.C.S.B.
7) Scott Davis, Geography Graduate Student, U.C.S.B.
8) Trish Foshi, Geography Graduate Student, U.C.S.B.
9) T. Kenney, Education Graduate Student, U.C.S.B.
10) Patricia Fenwick, Senior Typist-Clerk

The project was administered through the Social Processes Research Insitute (SPRI) at U.C.S.B. under the management of Patricia Griffith, Administrative Assistant.

B) List of papers presented at conferences

1) "Aggregation in Data Tables: Implications for Criminal Justice Statistics" presented at NIJ Conference on Research Methodologies and Evaluation Methods, Baltimore, March 1980; at the Association of American Geographers, annual conference, Louisville, April 1980; and at the Society for Multivariate Experimental Psychology, Los Angeles, November, 1979.

2) "Assessing Homogeneity in Cross-Classified Proximity Data" presented at I.G.U. Commission on Quantitative Methods, Nagoya, Japan, August 1980; and the Department of Psychology, University of Illinois, Urbana, December, 1979.

3) "Unidimensional Seriation: Implications for Evaluating Criminal Justice Statistics" presented at the International Geographical Union Congress, Tokyo, Japan, September 1980.

4) "A Heuristic Method for Analyzing Proximity Data" presented at the Annual Meetings, Regional Science Association, Milwaukee, November 1980.

C) Papers published and submitted for publication

1) "Aggregation in Data Tables: Implications for Evaluating Criminal Justice Statistics," Environment and Planning A, 1981, in press.

2) "Assessing Homogeneity in Cross-Classified Proximity Data," Geographical Analysis, 1981, in press.

3) "Unidimensional Seriation: Implications for Evaluating Criminal Justice Statistics," Environment and Planning A, 1981, in press.

4) "Generalized Procedures for Evaluating Spatial Autocorrelation," Geographical Analysis, 1981, in press.

5) "Some Comments on Non-Euclidean Mental Maps," Environment and Planning A, 1981, in press.

6) "Inference Models for Roll-Call Cohesion Measures" submitted to Quantity and Quality, November 1980.

7) "Proximity Matrix Reorganization and Hierarchical Clustering" submitted to Environment and Planning A, November, 1980.

D) Projected MA theses and Ph.D. dissertations

1. C.M. Costanzo - "Aggregation and Spatial Autocorrelation Effects in Crime Data: The Case of Homicide Rates"

MA thesis, Department of Geography, U.C.S.B. (projected completion date: Spring 1980).

2. T. Kenney - "Methodological Investigations in Non-Parametric Data Analysis Strategies"

Ph.D. dissertation, Graduate School of Education, U.C.S.B. (projected completion date: Spring 1982).