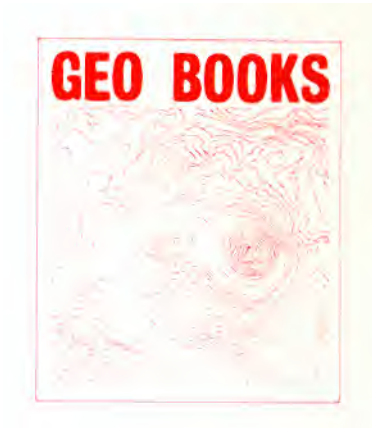
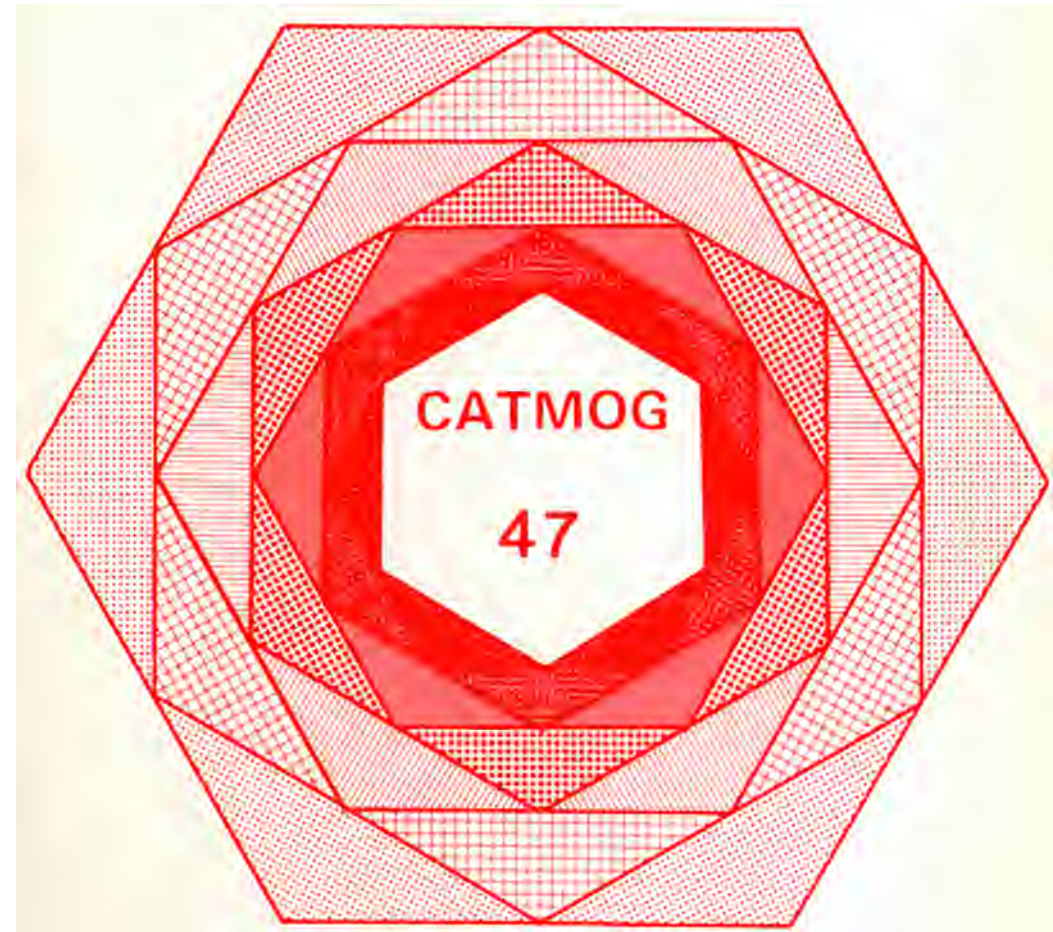


ISSN 0306-6142
ISBN 0-86094-223-6

Published by Geo Books, Norwich
Primed by Hutchins C Sons, Norwich

SPATIAL AUTOCORRELATION

Michael F. Goodchild



CATMOG - Concepts and Techniques in Modem Geography

CATMOG has been created to fill in a teaching need in the field of quantitative methods in undergraduate geography courses. These texts are admirable guides for teachers, yet cheap enough for student purchase as the basis of glasswork. Each book is written by an author currently working with the technique or concept he describes.

1. Introduction to Markov chain analysis - L. Collins
2. Distance decay in spatial interactions - P.J. Taylor
3. Understanding canonical correlation analysis - D. Clark
4. Some theoretical and applied aspects of spatial interaction shopping models - S. Openshaw
5. An introduction to trend surface analysis - D. Unwin
6. Classification in geography - R.J. Johnston
7. An introduction to factor analysis - J.B. Goddard & A. Kirby
8. Principal components analysis - S. Daultrey
9. Causal inferences from dichotomous variables - N. Davidson
10. Introduction to the use of logit models in geography - N. Wrigley
11. Linear programming: elementary geographical applications of the transportation problem - A. Hay
12. An introduction to quadrat analysis (2nd edition) - R.W. Thomas
13. An introduction to time-geography - N.J. Thrift
14. An introduction to graph theoretical methods in geography - K.J. Tinkler
15. Linear regression in geography - R. Ferguson
16. Probability surface mapping. An introduction with examples and FORTRAN programs - N. Wrigley
17. Sampling methods for geographical research - C.J. Dixon & B. Leach
18. Questionnaires and interviews in geographical research - C.J. Dixon & B. Leach
19. Analysis of frequency distributions - V. Gardiner & G. Gardiner
20. Analysis of covariance and comparison of regression lines - J. Silk
21. An introduction to the use of simultaneous-equation regression analysis in geography - D. Todd
22. Transfer function modelling: relationship between time series variables - Pong-wai Lai
23. Stochastic processes in one dimensional series: an introduction - K.S. Richards
24. Linear programming: the Simplex method with geographical applications - James E. Killen
25. Directional statistics - G.L. Gaile & J.E. Burt
26. Potential models in human geography - D.C. Rich
27. Causal modelling: the Simon-Blalock approach - D.G. Pringle
28. Statistical forecasting - R.J. Bennett
29. The British Census - J.C. Dewdney
30. The analysis of variance - J. Silk
31. Information statistics in geography - R.W. Thomas
32. Centrographic measures in geography - A. Kellerman
33. An introduction to dimensional analysis for geographers - R. Haynes
34. An introduction to 0-analysis - J. Beaumont & A. Gatrell

(continued inside back cover)

SPATIAL AUTOCORRELATION

Michael F. Goodchild
University of Western Ontario
London,
Ontario,
Canada
N6A 5C2

August, 1986

Table of Contents

List of Figures

List of Appendices

Goodchild, Michael F.
Spatial autocorrelation. _____
(Concepts and techniques in modern geography, ISSN 0306-6142; 47)
1. Geography_____Statistical methods
2. Autocorrelation (Statistics)
3. Spatial analysis (Statistics)
I. Title II. Series
519.5'37 G70.3

ISBN 0-86094-223-6

INTRODUCTION

In its most general sense, spatial autocorrelation is concerned with the degree to which objects or activities at some place on the earth's surface are similar to other objects or activities located nearby. Its existence is reflected in the proposition which Tobler (1970) has referred to as the "first law of geography: everything is related to everything else, but near things are more related than distant things." It is impossible for a geographer to imagine a world in which spatial autocorrelation could be absent: there could be no regions of any kind, since the variation of all phenomena would have to occur independently of location, and places in the same neighbourhood would be as different as places a continent apart. Gould (1970, pp. 443-444, also quoted in Cliff and Ord, 1981, p. 8) expresses the same idea:

> Why we should expect independence in spatial observations which are of the slightest intellectual interest or importance in geographic research I cannot imagine. All our efforts to understand spatial pattern, structure and process have indicated that it is precisely the lack of independence - the interdependence - of spatial phenomena that allows us to substitute pattern, and therefore predictability and order, for chaos and apparent lack of interdependence of things in time and space.

Spatial autocorrelation can be interpreted as a descriptive index, measuring aspects of the way things are distributed in space, but at the same time it can be seen as a causal process, measuring the degree of influence exerted by something over its neighbours. This duality echoes the nature of the literature on spatial autocorrelation: despite the importance of the concept, there are several very distinct traditions to be examined in the course of this book.

The spatial analytic tradition of geography is concerned with the study and interpretation of the various types of features found on the earth's surface. To some extent the methods of analysis are similar whether the features of interest are associated with economic or social activities or the physical environment. Point pattern analysis (for reviews see Ripley, 1981; Rogers, 1969a, b; Rogers 1974; Getis and Boots, 1978) is a good example of a set of techniques which has been applied to almost the full range of geographic phenomena over the past three decades. Spatial autocorrelation is another, although it has developed more slowly. Although the concept is a very general one, and of fundamental importance, it remains somewhat controversial and to some extent in conflict with the mainstream of statistical analysis, as we shall see.

Spatial analysis deals with two quite distinct types of information. On the one hand are the attributes of spatial features, which include measures such as size, value, population or rainfall, as well as qualitative variables such as name, religion or soil type. On the other hand each spatial feature has a location, which can be described by its position on a map or by various geographic referencing or coordinate systems. Many kinds of analysis look only at feature attributes without making explicit reference to location, and in this class we would have to include the vast majority of standard statistical techniques, ranging from simple tests of means and analysis of variance to multiple regression and factor analysis. Location is often involved indirectly in the sense that it is used to determine whether a case falls inside or outside the study area, but the locations of cases within the study area in no way affect the outcome of this class of analyses, and can be shuffled freely.

Other techniques, and we would probably include point pattern analysis in this set, deal with the locations of a set of features treated as members of a homogeneous class, any differences of attributes among them being ignored. So apart from the initial step of determining whether or not a feature belongs to the class of interest, the major part of the analysis is carried out on the locational information alone, without any reference to the feature attribute data.

Spatial autocorrelation is one of the relatively small set of techniques which deals simultaneously with both locational and attribute information. Location-allocation (for reviews see Hodgart, 1978; Handler and Mirchandani, 1979) and spatial interaction modelling (for reviews see Haynes and Fotheringham, 1984) also belong to this set, and it is significant that geographers have made major contributions in each of these fields. A pair of spatial features, for example two cities, may be similar or dissimilar in attributes, and their proximity will determine how similar they are in spatial location. In its broadest sense, spatial autocorrelation compares the two sets of similarities. If features which are similar in location also tend to be similar in attributes, then the pattern as a whole is said to show positive spatial autocorrelation. Conversely, negative spatial autocorrelation exists when features which are close together in space tend to be more dissimilar in attributes than features which are further apart. And finally the case of zero autocorrelation occurs when attributes are independent of location.

Figure 1 shows a simple illustration of these ideas. The features are 64 square cells arranged in the form of a chessboard, and the attributes are the two colours, black and white. Each of the five illustrations contains the same set of attributes, 32 white cells and 32 black cells, but the spatial arrangements are very different. Figure la, the familiar chess board, illustrates extreme negative autocorrelation between neighbouring cells (using the Rook's or Castle's move to define neighbour). Figure le shows the opposite extreme of positive autocorrelation, when black and white cells cluster together into homogeneous regions. The other illustrations show independent arrangements which are intermediate at intervals on a scale of autocorrelation to be discussed later. Figure lc corresponds to spatial independence, or an autocorrelation of zero, Figure lb shows a relatively dispersed arrangement and Figure ld a relatively clustered one. Note that several distinct arrangements of attributes out of the 2^{6} " possible (64!/32!32! if only arrangements with 32 white and 32 black cells are counted) may have the same spatial autocorrelation index.

The degree of spatial autocorrelation present in a pattern is very much dependent on scale. In Figure 1 the spatial autocorrelation between cells varies from one illustration to another, but in all cases the pattern is the same, perfectly homogeneous within cells. If we were to subdivide each cell into four and measure the autocorrelation between neighbouring cells the results would be quite different, although the pattern is the same. So any measurement of spatial autocorrelation must be specific to a particular scale, and a pattern can have different amounts at different scales. There are also constraints on the variation which can occur from one scale to another. For example, in order to show negative autocorrelation at the scale of the cells in Figures la or lb, it is necessary for subdivisions of the cells to be positively autocorrelated.

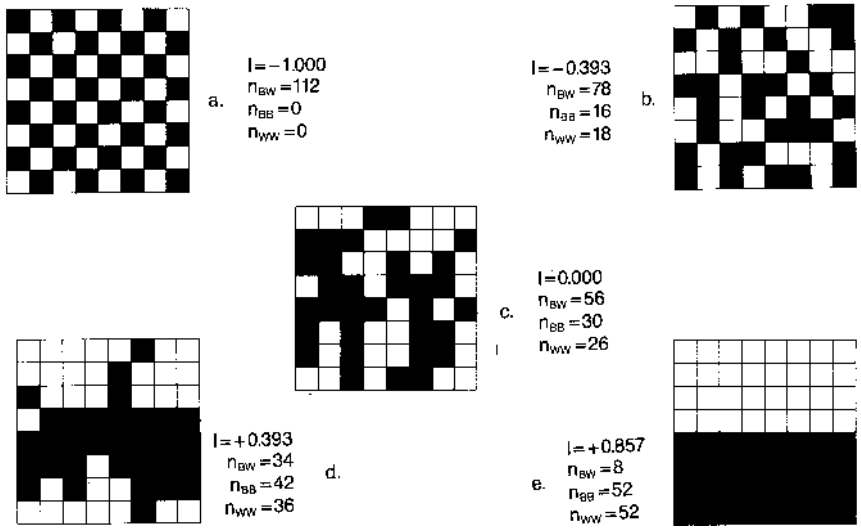


Figure 1. Varying levels of spatial autocorrelation in the pattern of 32 white and 32 black cells on an 8 by 8 raster.

What is the practical importance of spatial autocorrelation? First, as an index it provides a type of information about a spatially distributed phenomenon which is not available in any other form of statistical analysis, and which can be vital to correct interpretation. If one were forced to summarize a spatial distribution of unequal attributes in a single statistic, one would in all likelihood choose a spatial autocorrelation index, just as one would probably choose a measure of central tendency such as the mean or median to summarize a nonspatial data set. Later in the book we will look at the use of spatial autocorrelation in interpreting data on the incidence of disease, and as an example, see how it can be used to suggest possible environmental causes for various kinds of cancer. In this kind of application, spatial autocorrelation measures give a precise and objective value to something which would otherwise have to be perceived subjectively and probably inaccurately from a map.

A second important area of application follows directly from this. In looking for causes for a particular spatial distribution, it frequently happens that one variable is found to explain a pattern, but only partially. The next step is to search for other variables which might help account for the remaining variation, and this is often assisted by examining the spatial pattern of residuals. Taylor (1980) gives a worked example of this process, using precipitation figures at a number of sample points in California. His first step is to try to model precipitation as a function of altitude, latitude and distance from the coast, using multiple linear regression. These variables together cannot fully account for variation in precipitation, as the values predicted by the model are not exactly equal to the observed values at each sample point. The differences are the residuals, which are mapped to help in identifying further causative factors. In fact, they show strong positive spatial autocorrelation, those at sample points on west-facing slopes being positive, indicating actual precipitation higher than predicted, and those on the leeward side of mountain ranges being negative. If there had been no spatial autocorrelation in the residuals, the inference would have been that although explanation was not perfect, there would be little to be gained by looking for additional factors, as they would likely be complex and possibly unique to each sample station.

Third, as a measure of the process by which one place influences another, spatial autocorrelation analysis is often a necessary part of correct forecasting. To use another medical example, the incidence of an infectious disease such as 'flu in an area is best predicted not from the previous incidence of 'flu in that area, or from any of the area's characteristics, but from knowledge of the incidence of 'flu in the neighbouring areas from which it is likely to spread. The probability or rate of spreading can be measured by an appropriate index of spatial autocorrelation.

Just as spatial autocorrelation is concerned with arrangement in space, we can ask a parallel set of questions about arrangements in time. Indeed the term autocorrelation originated with the notion that for many measures, magnitude at one point in time is related to or dependent on magnitude at previous times. This area is the concern of time series analysis, which is one of the foundations of econometrics and has also attracted attention in geography. In this volume the term time autocorrelation will be used to refer to effects in the time domain, although in everyday usage the term autocorrelation normally implies time rather than spatial autocorrelation. Historically, much of the spatial autocorrelation literature has developed by extension of time autocorrelation techniques, but there are important differences and we will not take that route in this volume. Processes in time operate only in one direction, forwards in time, and in one dimension, whereas spatial processes may operate in both directions and in two or more dimensions. In addition the normal case in time series is for sampling to be at regular intervals, but this tends to be the exception for many kinds of spatial data. The field of spatial forecasting, which will be discussed at some length later, deals with processes which operate both in time and across space.

This very general introduction needs a great deal more precision before it can be very useful. The first chapter will have to define clearly what is meant by a spatial feature, and the terms near and far, and to establish a taxonomy of feature attribute types, and the result will be a class of spatial autocorrelation measures rather than one universal statistic. The second chapter focuses on the Moran and Geary indices, the subset of measures which has probably attracted the greatest attention in the geographical literature, and includes several applications. Chapter Three looks at join count statistics, which are appropriate for nominally scaled data. Finally Chapter Four looks at attempts which have been made to model autocorrelated processes.

The first text to appear devoted to the subject of spatial autocorrelation was by Cliff and Ord (1973). A much enlarged and renamed edition by the same authors appeared in 1981 (Cliff and Ord, 1981) and included a much more extensive set of example applications, besides new material on autocorrelated spatial processes. These two books are still the most useful general references to the field, although several others (Ripley, 1981; Getis and Boots, 1978; Cliff et al., 1975; Haggett, Cliff and Frey, 1977; Upton and Fingleton, 1985) include discussions of spatial autocorrelation as part of more general reviews of spatial statistics. Readers who might wish to pursue the topic at greater depth and with a more mathematical approach than that taken in this volume are recommended to begin with Cliff and Ord (1981).

1. MEASURES OF SPATIAL AUTOCORRELATION

1.1 Spatial features

The introduction referred variously to features, objects and activities distributed over space, and we should now clarify what these are and establish a formal classification scheme. To be consistent, the term spatial object will be used from now on, and the previous terms will be taken to be synonymous. For most purposes, the spatial objects of concern in any analysis correspond to measurement zones, statistical reporting areas or sampling points, and are equivalent to the cases of a statistical analysis. So for example in a study of socioeconomic conditions in cities the spatial objects would likely be census tracts, enumeration districts or some such statistical reporting areas, and their precise boundaries would have been determined by the data collection agency. The general case in human geography is for the definition of the sampled spatial objects to be outside the control of the researcher. On the other hand in the study of physical phenomena it is not unusual for the primary data collection to be part of the study, and for the researcher therefore to have control over the definition of the spatial objects. This would be true in cases where a pedologist determined the locations of soil pits, or a microclimatologist laid out his or her own rain gauges. Openshaw (Openshaw, 1977a, 1977b, 1978; Openshaw and Taylor, 1979, 1981) has written extensively on the effects of control over the definition of statistical reporting zones in human geography.

Given the range of possible research conditions, we clearly need a robust taxonomy of spatial objects. Borrowing to some extent from the spatial data handling literature, we define four types, as follows:

Points - the data of interest are attached to irregularly spaced points distributed within the study area, representing for example the locations of towns, retail stores, individual people, sample soil pits or rain gauges.

Lines - the data are attached to line features, which may or may not be connected to each other, for example road or rail links or streams.

Areas - the data are provided for statistical reporting zones such as counties or census tracts, drainage basins, forest stands or areas of homogeneous soil.

Lattices - the sampling design consists of a regular grid of points or cells, usually in the form of a square lattice. This type of spatial object is common in the analysis of topographic data and some climatic variables, and in much of the theoretical mathematical literature on spatially autocorrelated processes. Other arrangements besides square arrays are possible, including triangles and hexagons: they are collectively referred to as regular tesselations.

In many cases the type of spatial object present in the data may not correspond precisely to any real feature: an example would be the use of a point object to represent a city. However this issue is not of concern here, although it may affect the eventual interpretation of the results. The choice of an appropriate method of measuring spatial autocorrelation depends primarily on the nature of the data more than on the real phenomena represented. Similarly, we are not immediately concerned with whether the objects correspond to real spatial features, as they might in the case of watersheds represented by areas, or to arbitrarily chosen sample points.

Figure 2 shows examples of each of the four object types, with associated attributes, chosen to represent situations in which a spatial analyst might wish to measure spatial autocorrelation. The point data in Figure 2a consist of a set of statistics on the acidity of rainfall in Northeastern North America obtained at irregularly spaced measurement locations (Pack; 1980 and see also Diaconis and Efron, 1983). Over the past century or so the average height of stacks used to discharge coal smoke into the atmosphere has increased dramatically, with the result that the effects have spread over larger and larger areas. Sulphur dioxide pollution has changed as a consequence from a problem limited to heavily industrialized cities in the nineteenth century to one which now affects major global regions. We would expect this to be reflected in an increasingly positive spatial autocorrelation of rainfall acidity at the regional scale.

For line data, the example (Figure 2b) is of accident statistics on links of the Ontario provincial highway network in Southwestern Ontario (Hall, 1977). Low spatial autocorrelation in these statistics would imply local causative factors such as 'black spots' whereas strong positive autocorrelation would imply a more regional scale of variation, pointing to causative factors such as lifestyles, rural/urban and core/periphery cleavages, or climate. Area data (Figure 2c) is illustrated by the distribution of population by ethnic origin in the city of London, Ontario, as reported by census tract by Statistics Canada: this example will be used as an application in Chapter Two. Finally the case of raster objects is illustrated by Figure 2d, which shows the number of cloud-free images obtained by the Italian receiving station from the Thematic Mapper sensor over a nine-month period. Strong positive spatial autocorrelation is expected since the factors affecting the probability of cloud vary regionally over Southern Europe. However the orbit of the satellite platform is polar, so we might expect different structures of autocorrelation in the North-South and East-West directions.

The location of each type of object is described in different ways. Points can be located with a single geographic reference, a pair of grid coordinates, or latitude and longitude. Lines and areas are less straightforward, and to describe their spatial location and form requires an amount of information which depends on the complexity of the object. Finally the location of raster objects is usually determined implicitly by their position within a standard ordering, usually row by row from the top left corner (for a discussion of alternatives and their relative efficiency in the context of spatial autocorrelation see Goodchild and Grandfield, 1983).

With this taxonomy of spatial objects, we can now turn to the establishment of a parallel classification of attribute variables.

1.2 Attributes

The spatial objects in a data set can possess any number of attributes, and these are conveniently thought of in the form of a rectangular table. The rows correspond to the objects, and the columns to the different attributes or variables whose values are known for each object. This is the standard case-wise organization of data used in such common statistical analysis packages as SPSS or SAS. Each column describes the variation of a single attribute, across objects, and each row describes all of the attributes for a single object.

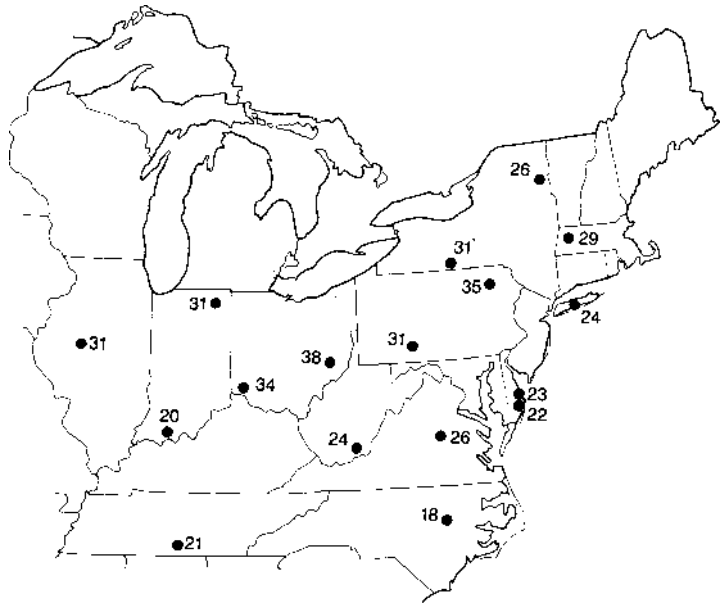


Figure 2a. Example of point objects: sulphate ion concentration in rainfall in Northeastern North America. Source: Pack (1980).

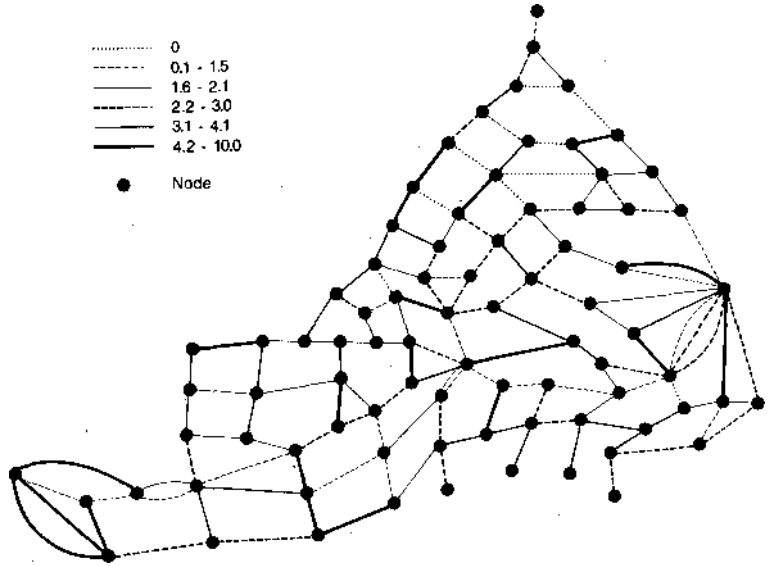


Figure 2b. Example of line objects: accident rates in the Southwestern Ontario provincial highway network. Source: Hall (1977).

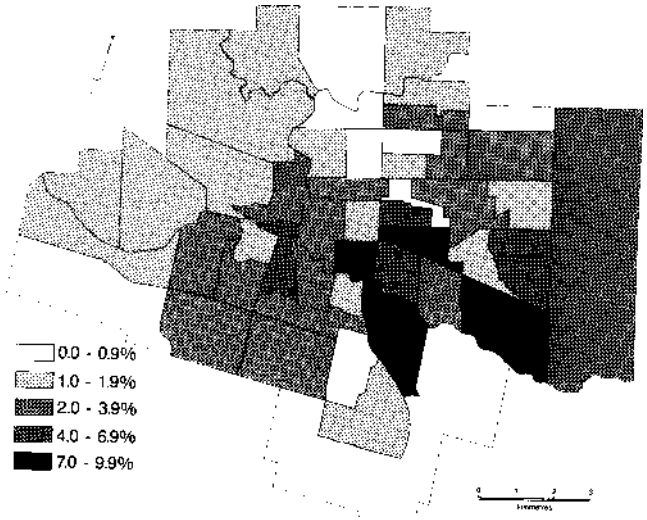


Figure 2c. Example of area objects: percent Italian by census tract, London, Ontario, 1971. Source: Statistics Canada.

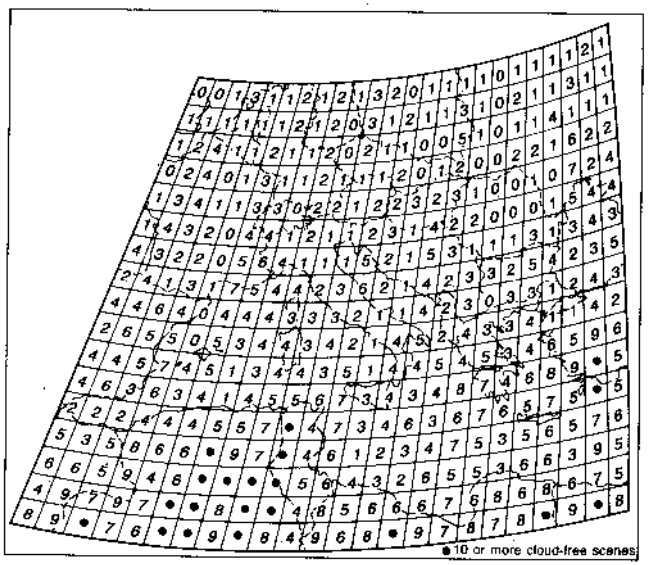


Figure 2d. Example of raster objects: number of cloud free Thematic Mapper frames acquired by Fucino (Italy) ground receiving station in 1984 (6 April - 31 December). Source: Pasco, Frei and Hsu (1985).

In general there are no restrictions on the kinds of information which can appear in the columns of an attribute table, and there are appropriate measures of spatial autocorrelation for each type. To determine the appropriate measure, we rely on the standard taxonomy of data types used in many areas of statistics:

Interval - data which establishes numeric value, so that the differences between the values given to objects have meaning. Weight, temperature, income and percent Catholic all have this property, but telephone numbers and colours do not. Much but not all interval data also has the ratio property, which gives meaning, for example, to statements such as "x is 'twice as heavy as y".

Ordinal - data which establishes rank order, but differences do not have meaning. Information on rank ordering of alternatives is ordinal but not interval, since the difference between rank 1 and rank 2 is not necessarily the same as the difference between rank 2 and rank 3.

Nominal - data which establishes category, but with no implied relationship among categories. Two telephone numbers can either be the same or different, but no other relationship is meaningful. One cannot be more or less than another, and adding, subtracting, multiplying or dividing nominal data makes no sense.

Unlike the taxonomy of objects, this one is determined not by the form of the data, but by its meaning, and hence its relationship to the real world. A sequence of digits such as 4710681 can belong in any of the three categories, depending on the meaning it conveys in a particular context.

With these two issues dealt with, we are now in a position to discuss actual measures of spatial autocorrelation, in terms of our three attribute types and four classes of spatial objects. We will always be dealing with one combination at a time, since the task is to measure the arrangement of one attribute possessed by one type of object.

1.3 Measures of spatial autocorrelation

In the introductory section we saw that spatial autocorrelation was concerned with a comparison of two types of information: similarity among attributes, and similarity of location. The ways in which the former can be measured depend on the type of data present, while the calculation of spatial proximity depends on the type of object. Finally, there will be a number of ways of comparing the two sets of information in compiling the final index. Since only a small fraction of the theoretical possibilities have actually appeared in the literature, it would be easiest to proceed by reviewing them within the general framework established thus far, rather than by examining every possible option.

To simplify things, a consistent notation will be used, as follows:

n number of objects in the sample

i, j any two of the objects

z_i the value of the attribute of interest for object i

c_{ij} the similarity of i's and j's attributes

w_{ij} the similarity of i's and j's locations, $w_{ii} = 0$ for all i.

In general, each of the measures of spatial autocorrelation has been devised to compare the set of attribute similarities c_{ij} with the set of locational similarities w_{ij}, combining them into a single index of the form

$\Sigma\Sigma\ c_{ij}w_{ij}$ with properties which lead to ready interpretation.
i j

A variety of ways have been devised for measuring spatial proximity in order to generate an appropriate matrix of w_{ij}'s. The existence of a linkage between objects is often a measure of spatial proximity, since nearby objects are more likely to be linked than distant ones. For area objects, a common boundary between areas is a simple, binary indicator of proximity but only an indirect one. Many of the early indices of spatial autocorrelation specified such binary weights matrices, but generalizations were provided by Cliff and Ord (1973) to allow the use of any measures of proximity, and these will be the forms discussed in the following sections, rather than the originals.

The length of common boundary is another possible definition of w_{ij} for area objects. Although it can be subject to uncertainty in the case of highly irregular boundaries, it is in many cases a better representation of the proximity of two areas than simple adjacency. Another option is to represent each area by a control point, in effect changing the object's type, and to measure the distances between points. As a control point one might use the geographic centroid, defined as the point about which the area would balance if it were suspended and of uniform density, but this has the disadvantage of sometimes being located outside the area (as in the case of the most southerly tract in Figure 4). The centroid location is easy to calculate if the area is represented as a polygon with known vertices, as follows (see for example Monmonier, 1982 pp. 114-118).

Let the polygon be denoted by m vertices (x_i, y_i), where i is an index assigned in clockwise order. The polygon is closed by making (x_{m+1}, y_{m+1}) identical to (x_1, y_1). The centroid coordinates are then given by:

$$X = \sum_i (y_i - y_{i+1})(x_i^2 + x_i x_{i+1} + x_{i+1}^2) / 6A \quad (1)$$

$$Y = \sum_i (x_{i+1} - x_i)(y_i^2 + y_i y_{i+1} + y_{i+1}^2) / 6A \quad (2)$$

where A is the polygon's area given by:

$$A = \sum_i (x_{i+1} - x_i)(y_{i+1} + y_i) / 2 \quad (3)$$

Alternatively, and if the appropriate information is available, one might select control points based on the distribution of density within each area.

Once distances between objects have been obtained, the weights can be set equal to some suitable decreasing function, such as a negative power,

$w_{ij} = d_{ij}^{-b}$, or negative exponential, $w_{ij} = \exp(-bd_{ij})$. In both cases b can be interpreted as a parameter which affects the rate at which weight declines with distance: a small b produces a slow decrease, and a large b a more rapid one. Examples of many of these options will be discussed later.

Gatrell (1979a) argues that in some applications it may be appropriate to base the measure of proximity on an index of social or economic interaction, rather than distance. He gives an example in which the spatial autocorrelation of the populations of major Swedish cities is measured using weights which have been derived through a multidimensional scaling.

A variety of ways have also been devised to measure the similarity of attributes, c_{ij}, in a manner which is appropriate to the types of attributes involved. For nominal data the usual approach is to set c_{ij} to 1 if i and j have the same attribute value, and zero otherwise. For ordinal data similarity is usually based on comparing the ranks of i and j, while for interval data both the squared difference $(z_i - z_j)^2$ and the product $(z_i - \bar{z})(z_j - \bar{z})$ are commonly used.

Hubert, Golledge and Costanzo (1981) pointed out that most measures of spatial autocorrelation can be reduced to an element by element comparison of two matrices, and thus fall into a much more general class of indices with a literature extending over many disciplines.

1.3.1 Geary's index (area objects, interval attributes)

Geary's index (Geary, 1954; reprinted as Geary, 1968) is a measure of spatial autocorrelation for area objects and interval data, and as such has found its most suitable applications in human geography in the analysis of data assembled for statistical reporting zones. Attribute similarity c_{ij} is calculated from the squared difference in value, that is:

$$c_{ij} = (z_i - z_j)^2 \quad (4)$$

which would be meaningful only if z were measured on an interval scale. In the original paper locational similarity was measured in a binary fashion, w_{ij} being given the value of 1 if i and j shared a common boundary, and zero otherwise. The other terms in the index ensure that the extremes occur at fixed points:

$$c = \frac{\Sigma\Sigma}{i\ j} w_{ij}\ c_{ij} \Big/ 2 \frac{\Sigma\Sigma}{i\ j} w_{ij}\sigma^2 \quad (5)$$

where:

σ^2 denotes the variance of the attribute z values, or

$$\sigma^2 = \sum_i (z_i - \bar{z})^2 / (n - 1) \quad (6)$$

$\bar{z}$ denotes the mean of the attribute z values. The value of c will be greatest when large values of w_{ij}, which correspond to pairs of areas in contact, coincide with large values of c_{ij}, or large differences in attributes.

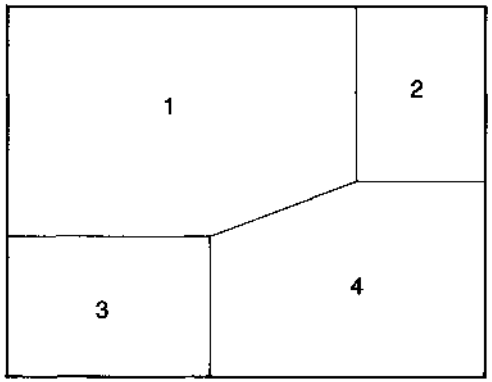


Figure 3. Example data set for calculation of the Geary and Moran coefficients.

In fact Geary arranged the index so that c would have the value of 1 when attributes are distributed independently of location, dropping below 1 when similar attributes coincide with similar locations, and above 1 otherwise. Somewhat confusingly, then, positive spatial autocorrelation corresponds to c less than 1, zero to c equal to 1, and negative to c greater than 1.

To clarify the meaning of the terms, the calculation of the Geary index can be illustrated with the simple example shown in Figure 3, using binary weights based on adjacencies.

Example calculation of the Geary index:

$n = 4$ $\quad z_1 = 3$ $\quad z_2 = 2$ $\quad z_3 = 2$ $\quad z_4 = 1$

$$w = \begin{matrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{matrix}$$

$$\bar{z} = \sum_i z_i / n = 8/4 = 2$$

$$c_{ij} = (z_i - z_j)^2 \qquad c = \begin{matrix} 0 & 1 & 1 & 4 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 4 & 1 & 1 & 0 \end{matrix}$$

$$\sum_i \sum_j w_{ij} c_{ij} = 0\times0 + 1\times1 + 1\times1 + 1\times4 + 1\times1 + 0\times0 + 0\times0 + 1\times1 + 1\times1 + 0\times0 + 0\times0 + 1\times1 + 1\times4 + 1\times1 + 1\times1 + 0\times0 = 16$$

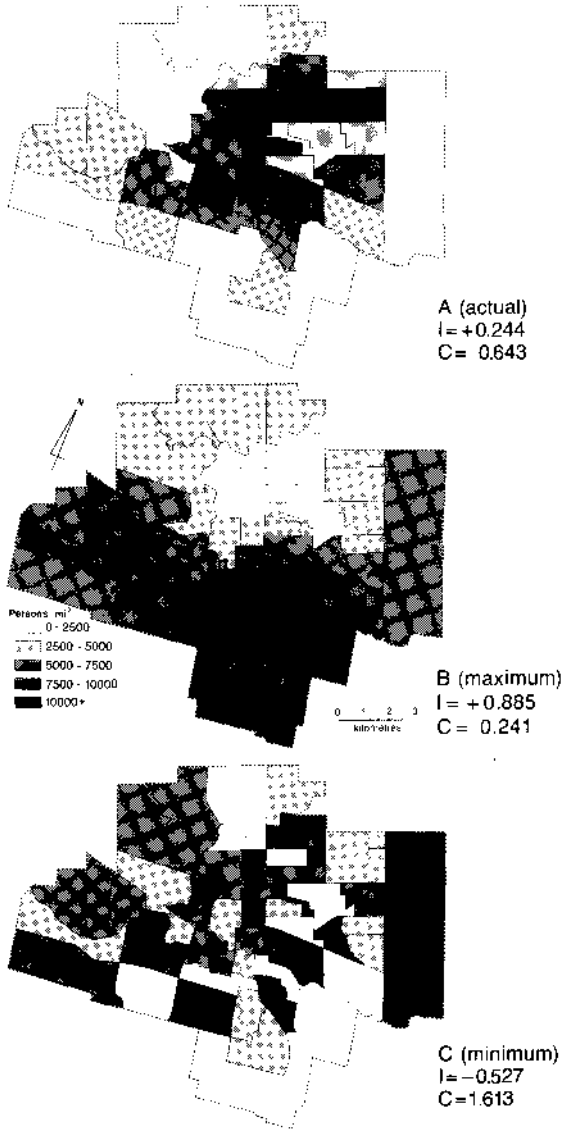


Figure 4. Alternative arrangements of the population densities reported for the 51 census tracts of London, Ontario in the 1971 census.

$$\sum_{ij} w_{ij} = 10$$

$$\sigma^2 = \sum_i (z_i - \bar{z})^2 / (n-1) = 2/3$$

$$c = \sum_{ij} w_{ij} c_{ij} / 2\sigma^2 \sum_{ij} w_{ij} = 16 / 2x(2/3)x10 = 1.200$$

To give a visual impression of the range of c values, Figure 4 illustrates a number of different arrangements of the same set of attributes over a map of areas. The figures are the population densities of the 51 census tracts of the city of London, Ontario, as reported in the 1971 census, and they have been distributed among the 51 tracts to show the actual pattern (4a), as well as artificial patterns of maximally positive (4b), and maximally negative (4c) autocorrelation. Note that each pattern is independent of the others. So although there is a tendency for a peak of population density to occur at the city centre in the actual data, peaks can occur anywhere in the city in the simulations.

We will leave the Geary index for now in order to continue with a brief review of the other autocorrelation measures which have been devised for the various object and attribute types. The next chapter will discuss the Geary index in more detail, and provide a number of applications.

1.3.2 Moran's index

Moran's index (Moran, 1948) provides an alternative to Geary's for the same data context, and in most applications both are equally satisfactory. Perhaps the only obvious advantage of one over the other is that the Moran index is arranged so that its extremes match our earlier intuitive notions of positive and negative correlation, whereas the Geary index uses a more confusing scale (but see the later discussion of this issue in Chapter Two). The Moran index is positive when nearby areas tend to be similar in attributes, negative when they tend to be more dissimilar than one might expect, and approximately zero when attribute values are arranged randomly and independently in space. These relationships are summarized in Table 1 below.

Table 1: Correspondence between Geary, Moran and conceptual scales of spatial autocorrelation

Conceptual	Geary c	Moran I
Similar, regionalized, smooth, clustered	$0 < c < 1$	$I > 0$*
Independent, uncorrelated, random	$c = 1$	$I < 0$*
Dissimilar, contrasting, checkerboard	$c > 1$	$I < 0$*

* The precise expectation is -1/(n-1) rather than 0.

The attribute similarity measure used by the Moran index makes it analogous to a covariance between the values of a pair of objects:

$$c_{ij} = (z_i - \bar{z})(z_j - \bar{z}) \quad (7)$$

where $\bar{z}$ denotes the mean of the attribute variable, as before.

Instead of two variables, x and y, c_{ij} measures the covariance between the value of the variable at one place and its value at another. This idea of autocovariance will appear in other contexts later in this volume.

The remaining terms in the Moran index are again designed to constrain it to a fixed range:

$$I = \sum_{ij} w_{ij} c_{ij} / s^2 \sum_{ij} w_{ij} \quad (8)$$

where s^2 denotes the sample variance

$$\sum_i (z_i - \bar{z})^2 / n$$

As in the Geary index, the w_{ij} terms represent the spatial proximity of i and j and can be calculated in any suitable way. For comparison, I is also calculated below for the data shown in Figure 3. The precise relationship between the two indices will be discussed further in Chapter 2, and we will also examine the question of whether the maximum and minimum values of both indices can be known, as well as issues of sampling and hypothesis testing.

Example calculation of the Moran index:

$$c_{ij} = (z_i - \bar{z})(z_j - \bar{z}) \qquad c = \begin{matrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{matrix}$$

$$\sum_{ij} w_{ij} c_{ij} = -2$$

$$s^2 = \sum_i (z_i - \bar{z})^2 / n = 2 / 4 = 0.5$$

$$I = \sum_{ij} w_{ij} c_{ij} / s^2 \sum_{ij} w_{ij} = -2 / 0.5x10 = -0.400$$

We have now defined two indices of spatial autocorrelation for area objects and interval attributes. The remaining sections of this chapter review a number of other measures for alternative data types.

1.3.3 Interval attributes; point, line and raster objects

In essence, both the Moran and Geary indices can be applied to other object types provided an appropriate method can be devised for measuring the spatial proximity of pairs of objects. One way discussed earlier of generating a suitable w_{ij} measure for area objects is to replace each one by a suitably located control point and to measure distances. The weights are then set equal to some decreasing function of distance, such as a negative power or negative exponential. This suggests a simple way of adapting either the Geary or Moran indices to points. For example, one might measure the spatial autocorrelation in the pattern of unemployment rates across cities by representing each city as a point, and basing the weights on the distances between pairs of points raised to a negative power. Another alternative would be to convert the point objects to areas by some systematic procedure, such as the generation of Thiessen or Dirichlet polygons (Boots, 1986), which partition the area into polygons each of which surrounds one point and encloses the area which is closer to that point than to any other (see Brassel and Reif, 1979 for an efficient method of generating Thiessen polygons from a point set; and see Griffith, 1982, for an example of the use of this method to measure spatial autocorrelation for point objects). The weights might then be based on the existence of, or length of, common boundary as before. In effect, points can be regarded as interchangeable with areas for both Moran and Geary measures.

In discussing line objects we need to recognize two distinct cases. First, the lines may represent links between nodal points, the problem being to measure the spatial autocorrelation present in some attribute of the nodes. In this case the c_{ij} are measures of the similarity of the attributes of each pair of nodes, and the w_{ij} are measures of the links between them. One might set w_{ij} to 1 if a direct link exists between i and j and 0 otherwise, or base w_{ij} on the link length or link capacity. In the second case the problem is to measure the spatial autocorrelation present in some attribute of the links themselves, such as transport cost per, unit length, or probability of motor vehicle accident. In this case w_{ij} is a measure of proximity between two links, and might be based on whether or not two links are directly connected, or on the distance between the centre points of two links. As in the case of point objects, then, the Geary and Moran indices provide suitable measures for interval attributes provided a little ingenuity is used in the definition of the spatial proximity measures.

For rasters (lattices) one simple way of defining the w_{ij} is to assign a value of 1 to pairs of raster cells which share a common boundary, and 0 otherwise. In some studies pairs of cells which join at a corner are also defined as adjacent and given a value of 1 (in dealing with a square raster the two alternatives are sometimes referred to as the 4-neighbour and 8-neighbour cases respectively, or Rook's case and Queen's case).

1.3.4 Ordinal attributes

A number of papers have discussed the definition of specific spatial autocorrelation indices for attributes whose scale of measurement is ordinal.

Coffey, Goodchild and MacLean (1982) were concerned with describing the spatial pattern of the Canadian settlement system, in particular the distribution of settlement sizes. Do large settlements tend to be surrounded by small settlements or by other large ones, or are settlement sizes randomly distributed? Settlements were allocated to one of five size classes, so that the attribute of interest is an ordinal measure limited to the integers 1 through 5.

The authors discuss two types of spatial autocorrelation measures of the settlement size distribution. In the first the Geary and Moran indices are calculated directly from the size classes by setting z_i equal to an integer between 1 and 5. The w_{ij} terms were set equal to 1 for pairs of settlements which shared a direct road link and 0 otherwise. In effect this approach treats the ordinal size class data as if it had interval properties, since the calculation of both Geary and Moran indices requires the taking of differences between z values.

The second set of measures was based on the observed numbers of links between settlements of different size classes. Thus n_{12} would represent the number of direct road links observed between a settlement of class 1 and one of class 2. The observed frequencies of each link type were then compared to the frequencies expected if settlement sizes are distributed randomly. In effect this approach treats the ordinal data as if it were merely nominal. Measures of this type are the topic of the next section.

While the authors proposed to deal with the ordinal case by using either interval or nominal methods, the index described by Royaltey, Astrachan and Sokal (1975) is more directly suitable for the ordinal case. Each object is given a rank based on its ordinal value, and the c_{ij} are then based on the absolute difference in ranks between each pair. The index was applied to a set of point objects, with weights set to 1 if no other point lay within a circle drawn with the pair as diameter, and 0 otherwise (this is the adjacency matrix of the Gabriel graph, (Gabriel and Sokal, 1969)). The index was further generalized by Hubert (1978). Sen and Soot (1977) discuss a number of other indices for ordinal data, again based on ranks.

1.3.5 Nominal attributes

It is convenient to think of nominal attributes as if the objects on the map were coloured using a finite set of colours. An attribute limited to two nominal classes might be thought of as a pattern of black and white. Three classes might be visualized as a distribution of the three primary colours red, green and blue. One frequently finds spatial distributions of nominal attributes referred to as k-colour maps, where k denotes the number of possible classes of the attribute. The object set being coloured might equally consist of areas, points, lines or a raster.

In dealing with nominal data we are highly constrained in the ways in which attributes can be compared, since the nature of the data allows only two notions: a pair of object attributes can be the same, or different, but no measures of difference are possible. In constructing indices, then, the c_{ij}, which measure the similarity of is and j's attributes, can take one of only two values.

Most of the measures which have been devised for nominal attributes are based on join count statistics using binary definitions for the spatial proximity measures w_{ij}. Much of the early work was in the context of raster data, where two objects can be defined as joined if they share a common boundary. The join count between colour s and colour t is defined as the number of times a cell of colour s is joined to a cell of colour t. Note that it is necessary to avoid double counting in dealing with joins between cells of the same colour. We denote the join count as n_{st}.

The spatial arrangement of colours is reflected directly in the join count statistics, so measures of spatial autocorrelation can be devised based on them. If the distribution shows positive autocorrelation then joins of the same colour will be more likely than one would expect in a random distribution of colours, and similarly negative autocorrelation will be reflected in a higher than expected incidence of joins between different colours.

Join count statistics provide a simple way of measuring spatial pattern, but they do not lead to a simple summary index or to indices analogous to the Geary or Moran measures. For this reason further discussion and illustration will be deferred to Chapter Three, where join counts will be examined in greater detail.

These ideas can be extended from rasters to points, lines and areas provided similar binary spatial proximity measures can be devised. Areas can be regarded as joined if they share a common boundary. Points can be dealt with by constructing Thiessen polygons, two points being 'joined' if they share a Thiessen edge. Or one could define two points as adjacent if either is the nearest neighbour of the other. Two lines might be regarded as adjacent if they share a common node. As before, dealing with different object types is a matter of ingenuity in the definition of spatial proximity.

1.3.6 Variograms and correlograms

We have already noted that the concept of scale is implicit in any measure of spatial autocorrelation, and that spatial patterns may possess quite different forms of autocorrelation at different scales. Scale is implicit in the definition of spatial objects, particularly if these are arbitrarily defined. For example; in a raster data set the size of the cells will affect the outcome; for a point data set the result may be affected by the density of points if these represent sample points from a larger population.

These observations suggest that it would be appropriate to make scale explicit in the measurement of spatial autocorrelation, and this can be done in several ways. First, consider cases in which w_{ij} is a binary matrix of adjacencies, as in the original Geary index. We could define a second-order adjacency as existing between two areas which were not adjacent, but separated by exactly one intervening area. In other words two areas i and j are second-order adjacent if some third area k exists such that i and k are adjacent, and k and j are adjacent, but i and j are not adjacent. The same idea can be extended up to any order (although we are limited in principle by the diameter of the adjacency graph). By measuring autocorrelation at each of these levels one could construct a correlogram showing the performance of the index across scale.

A variogram is defined in a similar fashion, but shows the variance or mean squared difference between attributes at given proximity to each other, rather than the correlation. To be consistent, in view of the nature of the two indices, we should refer to a plot of the Moran index with scale as a correlogram, and to a plot of the Geary index as a variogram.

Variograms and correlograms are most useful in the interpretation of spatial patterns of continuous distributions, in other words where it is reasonable to think of the phenomenon as having a value at every point in the space. This is most likely to be true of raster objects, which are often the result of regular sampling of a continuous variable, or points, which are often irregularly placed samples. On the other hand ared data is not usually consistent with this continuous model.

Another way of constructing a correlogram is to compute a series of Moran or Geary indices using weights matrices based on distances. Let d_{ij} represent the distance between entities i and j; in the case of area, line or raster objects the points could be suitably located control points. In order to compute proximity one might choose either the negative power or negative exponential functions, both of which are monotonically decreasing: $w_{ij} = d_{ij}^{-b}$ or $w_{ij} = \exp(-b\, d_{ij})$. In both cases the effect of increasing the value of the constant b is to increase the weight given to short distances relative to long ones. An index calculated with a large b will thus emphasize spatial variation over short distances, and a small b will emphasize variation over large distances.

Finally, one can construct correlograms and variograms by using the weights as distance filters: autocorrelation can be estimated for a certain distance range by setting weights to 1 for pairs of objects whose separation lies within the desired range and zero otherwise, or in other words by summing c_{ij} only for such pairs.

1.4 Sampling and hypothesis testing

Thus far the concern has been with devising suitable ways of measuring an intuitive property of a spatial distribution. The result will be a single value somewhere on a scale, a useful summary index of an interesting aspect of the data from which it was calculated, to be interpreted in relation to known fixed points on the scale, or in relation to values for other sets of data obtained from other places or at other times.

As with other measures in statistics, it is often useful to be able to go rather further, by making the index the basis of generalization, inference and hypothesis testing. Instead of limiting the interpretation to the actual data from which the index was calculated, it would be interesting to be able to make statements in a larger context, by generalizing from the sample which was analysed to the larger population from which it was drawn. Suppose that the Moran index has been used to measure the spatial autocorrelation present in a study area, and the result is a value of 0.2. Since we are dealing with a sample, we know that the amount of autocorrelation present in the larger population from which it was drawn is likely in the range of 0.2, but its precise value is unknown. On the one hand we might want to ask whether this value of 0.2 confirms that spatial autocorrelation is present in the larger

population from which the sample was drawn, or whether 0.2 might have arisen by chance in the sample even though autocorrelation was absent in the population. This is the hypothesis-testing context. On the other hand we might ask for limits on either side of 0.2 within which we can be reasonably confident that the true, population value lies. This is the confidence intervals context.

The literature on spatial autocorrelation provides methods for answering some of these questions, in one of two ways, depending on how one is willing to believe the sample was obtained from the population. The randomization hypothesis proposes that the population consists of all possible ways in which the observed attributes could be rearranged among the objects: the sample is just one such possible arrangement. All possible samples contain exactly the same set of attribute values, but differ only in the objects to which each value is attached. Figure 1, for example, shows three of the 64!/32!32! possible distinct samples which could be obtained by randomizing the 64 attributes in this data set.

The resampling hypothesis proposes that the attribute values assigned to any object are obtained from an infinitely large population by a random process in which each individual attribute value is drawn independently. Since the individuals are drawn independently the population must have zero spatial autocorrelation. Each sample will contain a different set of values. Furthermore it will be necessary to make assumptions about the range of values which occur in the population and their relative probabilities, since unlike the randomization hypothesis, the population is not limited to those values which occur in the sample. In the case of interval data the usual assumption will be that the data is sampled from a normal or Gaussian distribution.

The precise approach to be taken to inference and generalization will depend on the experimental context, and it is quite likely that none of the available methods will be suitable. First, the arguments over hypothesis testing versus confidence interval estimation seem rather different in the case of spatial autocorrelation measures than for more conventional indices such as the Pearson correlation coefficient. While it is easy to imagine two variables which have no influence on each other, it is relatively hard for a spatial analyst to conceive of a variable whose distribution in space is truly without spatial autocorrelation, so that the value at one place is statistically independent of the value at nearby places. It seems that virtually all spatial distributions show some form of finite spatial autocorrelation. It follows that the hypothesis testing question, of whether a population is or is not spatially autocorrelated, is not often of significant interest since the answer is rarely in doubt. To make the same point in somewhat different form, if a hypothesis test led to the conclusion that no spatial autocorrelation could be confirmed in the population, the correct interpretation would in almost all cases be that the sample was simply too small to confirm the relatively small amount present, in other words that a Type 2 statistical error had been made. It appears that for spatial autocorrelation statistics the more relevant form of inference lies in using sample statistics to estimate confidence intervals for the population rather than to test hypotheses. Unfortunately both randomization and resampling are suitable for hypothesis testing rather than confidence interval estimation.

It would be a mistake to take this argument to the point of denying the possibility of spatial independence, however. Since spatial autocorrelation is

scale-dependent, it is quite likely that for a given distribution, some scale - exists at which the distribution appears independent. The issue is rather over whether it is reasonable to adopt a hypothesis-testing approach which assumes that absence of spatial autocorrelation is the norm.

The question of whether to apply a randomization or a resampling approach in a given situation is best answered by one's conceptualization of the sampling process, or the mechanism by which the sample was drawn from the population. In most cases the set of objects used as a sample in measuring spatial autocorrelation is the entire set found within the study area boundary at some given point in time. Because of the way the indices are defined it would be difficult to take a random sample from within a given area, and much easier to take the entire population. It is therefore unreasonable to argue that the sample is a random sample of the objects found in a larger area of which the study area is a part, and equally unreasonable to argue that the sample is randomly drawn from the attributes which would exist at other points in time.

In one sense, however, it may be possible to argue that the sample is indeed randomly drawn from a population, although the latter must be strictly hypothetical. Suppose that the processes which led to the observed outcome, the values of the attributes, are conceived as extremely complex and subject to influences which we can regard as random. If the clock were set back and the system allowed to redevelop, we might argue that the result would be the same in its general properties, but different in detail, to the extent that the actual pattern can be regarded as randomly drawn from the set of all possible patterns. Resampling would be appropriate if the hypothetical values were drawn independently and were different from the real values, while randomization would be appropriate if the effect of rerunning the clock were simply to redirect the same values to other parts of the map. It seems that resampling is more reasonable in most contexts, despite the inherent disadvantage that assumptions must be made about the statistical distribution from which the hypothetical values are drawn.

In summary, the following chapters will describe two alternative approaches to hypothesis testing: resampling and randomization. Although confidence interval estimation might seem more appropriate to spatial autocorrelation, neither resampling nor randomization is useful in this context since neither proposes a population with non-zero spatial autocorrelation. Moreover, it is difficult to argue that the sampling design normally used in spatial autocorrelation studies provides a random sample of anything other than a somewhat artificially conceived hypothetical population. So although inference and generalization are conceptually useful, in practice the spatial autocorrelation statistics to be described in this volume are most often used as indices descriptive of the spatial pattern present in a limited study area. The examples will show how useful results can be obtained by observing how these measures vary through time, from area to area or from one variable to another.

2. THE GEARY AND MORAN INDICES

2.1 Introduction

Chapter One introduced the basic taxonomy of spatial objects and attribute types, and several of the most commonly used indices. The purpose of this chapter is to explore two of those, the Moran and Geary, more fully. In

particular the second section will discuss their use in hypothesis testing, under the general conceptual framework introduced in section 1.4. Subsequent sections will provide two extensive applications of the Geary index, review the applications which have been described in the literature, and explore further the relationships between the two indices and the conceptual scale of spatial autocorrelation.

2.2 Hypothesis testing

The simplest and most straightforward hypothesis to test regarding the spatial autocorrelation exhibited by a sample of n cases is that the sample was drawn from a population in which autocorrelation is zero. This null hypothesis was discussed at some length in section 1.4. As we have already seen, there are two different interpretations of the term 'drawn'. The randomization null hypothesis proposes that the sample is one randomly chosen possibility from among the n! possible arrangements of the observed attributes among the n objects, and estimates the probability that such a randomly chosen arrangement would have a spatial autocorrelation index as extreme as that observed in the real arrangement. The resampling null hypothesis proposes on the other hand that each of the n attributes was drawn independently from a population of normally distributed values: since the drawing was independent, the population must have no spatial autocorrelation by definition. It also estimates the probability that a sample of n attributes drawn in this way and assigned randomly to the n objects would exhibit a spatial autocorrelation index as extreme as the one which is observed.

The general form of both tests is the same, and closely parallels the format of most statistical hypothesis testing, so much of the material which follows in this section may be familiar to many readers. The probability that the null hypothesis could yield results as extreme as those observed is denoted by a . Since the results were indeed observed, it can also be interpreted as the probability that the null hypothesis is true, referred to as the significance level. Finally, if this probability is small, less than say 5%, the null hypothesis is rejected and we conclude that indeed spatial autocorrelation is present in the population.

The phrase "as extreme as" raises an issue which is general in hypothesis testing and deserves special attention in the context of spatial autocorrelation. In most statistical tests it is possible for the test statistic to deviate from what would be expected under the null hypothesis towards one of two extremes, referred to statistically as tails. In the case of the Moran index, the statistic might differ by being too positive or too negative. In cases like this, the researcher has the option of specifying which of the two tails are of interest. In many applications the aim is merely to show the presence of an effect without specifying which direction the effect takes: in this case both extremes or tails are of interest and a two-tail test is used. In other applications we may be looking for a particular effect which should push the statistic in one direction only. The amount by which the statistic has moved in this direction will then determine the outcome of the test. If the statistic has in fact moved in the other direction, we have no choice but to accept the null hypothesis and deny the presence of the effect.

In the case of spatial autocorrelation indices the effects which produce positive and negative autocorrelation are so different that it is hard to imagine

a researcher being interested in both of them equally. For this reason discussion which follows will assume that all tests are one-tail.

The significance a is relatively easy to calculate because the sampling distributions of both I and c are asymptotically normal under both null hypotheses (for proof see Cliff and Ord, 1973, 1981; Sen, 1976). In other words the indices calculated for random samples of size n drawn according to either null hypothesis show a distribution which is approximately normal or Gaussian as long as n is relatively large, and perfect when n is infinite. Cliff and Ord (1973, p. 21) assumed that a sample size of 16 was adequate to ensure the accuracy of the approximation, but other rules of thumb, both more and less conservative, have been suggested in the literature. By using Monte Carlo simulation Cliff and Ord (1981, pp. 53-56 and see also Cliff and Ord, 1971, 1973) were able to provide more detailed rules under which the assumption of normality can be taken to be valid for small sample sizes.

We use the subscripts N and R to denote resampling and randomization null hypotheses respectively. Note that this particular form of the resampling null hypothesis assumes that the attributes be drawn from a normal or Gaussian distribution: if this is not valid, it may be appropriate to use randomization instead.

The following expressions give the expected value and variance of I and c for samples of size n under each null hypothesis:

$$E_N(I) = -1/(n-1) \tag{9}$$

$$E_R(I) = -1/(n-1) \tag{10}$$

$$E_N(c) = 1 \tag{11}$$

$$E_R(c) = 1 \tag{12}$$

$$Var_N(I) = (n^2S_1 - nS_2 + 3S_0^2) / S_0^2(n^2-1) - E_N(I)^2 \tag{13}$$

$$Var_R(I) = n[(n^2-3n+3)S_1 - nS_2 + 3S_0^2] / (n-1)^{(3)}S_0^2 - b_2[(n^2-n)S_1 - 2nS_2 + 6S_0^2] / (n-1)^{(3)}S_0^2 - E_R(I)^2 \tag{14}$$

$$Var_N(c) = [(2S_1 + S_2)(n-1) - 4S_0^2] / 2(n+1)S_0^2 \tag{15}$$

$$Var_R(c) = \{(n-1)S_1[n^2-3n+3-(n-1)b_2] - (n-1)S_2[n^2+3n-6-(n^2-n+2)b_2]/4 + S_0^2[n^2-3-(n-1)^2b_2]\} / n(n-2)^{(2)}S_0^2 \tag{16}$$

where:

n is the number of cases

S_0 is the total of the weights matrix, $\sum_i \sum_j w_{ij}$

S_1 is $\frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2$

S_2 is $\sum_i (w_{i.} + w_{.i})^2$

$w_{i.}$ and $w_{.i}$ are the row and column totals of the weights matrix, $\sum_i w_{ij}$ and $\sum_j w_{ji}$, respectively.

b_2 is the sample kurtosis coefficient m_4/m_2^2, where m_4 is the fourth sample moment about the mean and m_2 the second, as follows:

$$m_4 = \sum_i (z_i - \bar{z})^4 / n \qquad (17)$$

$$m_2 = \sum_i (z_i - \bar{z})^2 / n \qquad (18)$$

$n^{(b)}$ denotes the product $n\ (n-1)\ (n-2)\ (n-3)\ \ldots\ (n-b+1)$

Note that b_2 is relevant only for the randomization null hypothesis, since the resampling null hypothesis assumes a normal distribution and therefore zero kurtosis.

The expressions for variance under the randomization null hypothesis depend on b_2 for both I and c, but the kurtosis is divided by n in the case of c and n^2 in the case of I. Cliff and Ord (1981) comment that this may make the Moran index preferable in many applications because it means that I will be less affected by deviations from the normal or Gaussian distribution than will c.

Having calculated the appropriate values for E and Var, the test statistic for all four cases is:

$$x = (s - E) / \sqrt{(var)} \qquad (19)$$

where s is the observed I or c index for the sample. Since x is expected to have a distribution, under either null hypothesis, which is normal with zero mean and unit standard deviation, we can obtain the significance very easily by entering a table of areas under the normal curve, using x as a z score.

The data in Figure 3 will now be used in order to illustrate the calculation of the test statistic and a, for both N and R null hypotheses and for both I and c. Of course the sample size of four makes this an inappropriate use of the method, but it may nevertheless help to illustrate the process of calculation. As a further aid a program in BASIC is included as Appendix 1. Since it requires the user to enter every element in the weights matrix, it becomes cumbersome to use for large examples, but may be of value for teaching purposes. Most of the calculations in the examples in this volume

have been made by using a computer program which operates directly on the spatial objects in a geographic information system.

Example calculation of the significance of the Geary and Moran coefficients.

$n = 4$

$z_1 = 3$

$z_2 = 2$

$z_3 = 2$

$z_4 = 1$

$\bar{z} = 2$

$$W_{ij} = \begin{matrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{matrix}$$

$S_0 = \sum_i \sum_j w_{ij} = 10$

$S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2 = 20$

$S_2 = \sum_i (w_{i.} + w_{.i})^2 = 104$

$m^2 = \sum_i (z_i - \bar{z})^2 / n = 0.5$

$m_4 = \sum_i (z_i - \bar{z})^4 / n = 0.5$

$b_2 = m_4 / m_2^2 = 2$

$E_N(I) = -1 / (n-1) = -0.333$

$E_R(I) = -0.333$

$E_N(c) = 1.0$

$E_R(c) = 1.0$

$Var_N(I) = (n^2 S_1 - nS_2 + 3S_0^2) / S_0^2(n^2-1) - E_N(I)^2$

$= (4^2 \times 20 - 4 \times 104 + 3 \times 10^2) / 10^2 \times (4^2-1) - 1/9$

$= 0.025$

$Var_R(I) = n[(n^2-3n+3)S_1 - nS_2 + 3S_0^2] / (n-1)^{(3)} S_0^2$

$- b_2[(n^2-n)S_1 - 2nS_2 + 6S_0^2] / (n-1)^{(3)} S_0^2$

$- E_R(I)^2$

$$= 4[(4^2 - 3x4 + 3)x20 - 4x104 + 3x10^2] / 3x2x1x10^2$$
$$- 2[(4^2 - 4)x20 - 2x4x104 + 6x10^2] / 3x2x1x10^2 - 1/9$$
$$= 0.022$$

$$Var_N(c) = [(2S_1+S_2)(n-1)-4S_0^2] / 2(n+1)S_0^2$$
$$= [(2x20 + 104)(4 - 1) - 4x10^2] / 2(4 + 1)x10^2$$
$$= 0.032$$

$$Var_R(c) = \{(n-1)S_1[n^2-3n+3-(n-1)b_2]$$
$$-(n-1)S_2[n^2+3n-6-(n^2-n+2)b_2] / 4$$
$$+S_0^2[n^2-3-(n-1)^2b_2]\} / n(n-2)^{(2)}S_0^2$$
$$= \{(4 - 1)x20x[4^2 - 3x4 + 3 - (4 - 1)x2]$$
$$- (4 - 1)x104x[4^2 + 3x4 - 6 - (4^2 - 4 + 2)x2] / 4$$
$$+ 10^2[4^2 - 3 - (4 - 1)^2x2]\} / 4x2x1x10^2$$
$$= 0.035$$

$$x = (s - E) / \sqrt{(var)}$$

N hypothesis, Moran's I: $x = (-0.400 + 0.333)/\sqrt{(0.025)} = -0.424$

$\alpha = 0.337$

R hypothesis, Moran's I: $x = (-0.400 + 0.333)/\sqrt{(0.022)} = -0.452$

$\alpha = 0.326$

N hypothesis, Geary's c: $x = (1.200 - 1.000)/\sqrt{(0.032)} = 1.119$

$\alpha = 0.131$

R hypothesis, Geary's c: $x = (1.200 - 1.000)/\sqrt{(0.035)} = 1.069$

$\alpha = 0.142$

The final outcome of the hypothesis test will be determined by the value of the probability α, that is, the probability that the null hypothesis is true given the observed sample. The smaller α, the less likely the null hypothesis, and the greater the confidence with which we can reject it. The acceptable levels of α are those adopted in most statistical tests; for scientific purposes a level of 5% or a chance of 1 in 20 has become the conventional standard. Thus if α is less than .05 we conclude that the null hypothesis can be rejected; we can be confident that there is no spatial autocorrelation in the population from which the sample was drawn.

0,1

A type 1 error results during a statistical test when a null hypothesis is rejected when in fact it is true. Because it deals in probabilities, no statistical test can ever produce a certain result, and when a null hypothesis is rejected following the calculation of a low α, in effect the researcher is accepting a probability equal to α that his or her decision is wrong. The chance of a type 1 error is thus equal to α.

A type 2 error occurs when the null hypothesis is accepted when in fact it is false. The chance of this occurring will depend on the sample size, since a small sample is more likely to lead to a type 2 error than a large one, and also on the degree to which the null hypothesis is false: if it is almost true, the chance of a type 2 error is clearly higher. As we saw in the first chapter, it is hard to believe that any distribution is truly lacking in spatial autocorrelation, and therefore a type 2 error is almost inevitable in cases where a test leads to the acceptance of the null hypothesis. This point and its implications for interpretation will be explored further in the example applications later in this chapter.

2.3 Properties of the indices

The Moran and Geary indices were introduced in Chapter 1 as mathematical expressions yielding results which could be related to the conceptual scale of spatial autocorrelation. We saw that the Geary scale was essentially an inversion of the Moran and conceptual scales, and that the Geary scale assigned zero autocorrelation a value of 1. In this section the properties of the two indices will be explored a little further.

2.3.1 Expected values

The expected values E listed in section 2.2 for the two indices give the values we expect to find by averaging the respective indices for samples drawn under the null hypothesis and therefore from a population with zero spatial autocorrelation. In both cases it makes no difference whether the resampling or randomization forms are used. The Geary index expectation is exactly 1. However, the expectation of the Moran index is not exactly 0, but slightly negative at -1/(n-1). For large samples this is asymptotic to 0, but for small samples the difference becomes quite large. So unfortunately it is not precisely correct to identify 0 on the Moran scale with independence or with absence of spatial autocorrelation on the conceptual scale.

2.3.2 Maxima and minima

In order to interpret a given index it may be useful to know not only the fixed point corresponding to independence but also the upper and lower limits of the scale. We might ask the question in two ways, corresponding to the randomization and resampling null hypotheses respectively. Under randomization the problem is to find two arrangements of the given attributes, one yielding the highest possible index and the other the lowest possible. Under resampling the problem is to select both the attributes and the arrangements which will yield the two extreme patterns, given the distribution from which the attributes are believed to be drawn.

Unfortunately answers are difficult to obtain in both cases, and the locations of the extreme points on both I and c scales depend on the particular

details of the weights matrix and attribute values. Although the definition of the Moran index is similar to that of the Pearson correlation coefficient, I can exceed 1.0 for certain matrices of weights and samples of attribute values. The patterns shown in Figures 4b and 4c are the extremes under randomization: no arrangements have more extreme spatial autocorrelation indices. Yet in neither case are the values of the Moran or Geary index convenient round numbers. In summary, there are no fixed extreme points on either scale. So only one fixed point, independence, is available for purposes of interpretation. This is somewhat limiting, but reasonable given the nature of spatial autocorrelation.

2.3.3 Complementarity

The Geary and Moran indices are conceptually complementary, suggesting that there might be some simple mathematical relationship between them. Goodchild (1980) notes that the equation defining I can be manipulated to the following form:

$$I = n\sum_i ((z_i-\bar{z})^2 \sum_j w_{ij}) / \sum\sum_{ij} w_{ij} \sum_i (z_i-\bar{z})^2$$

$$- n\sum\sum_{ij} (z_i-z_j)^2 w_{ij} / 2 \sum\sum_{ij} w_{ij} \sum_i (z_i - \bar{z})^2 \qquad (20)$$

where the second term is similar but not identical to the Geary index. The first term is 1 if the weights matrix is standardized, $w_{.i} = w_{i.} = 1$ for all i. So a degree of complementarity exists, as we noted earlier on conceptual grounds, but the degree of mathematical complementarity is never perfect.

2.4 Example applications

We now turn to two applications of the Geary and Moran indices to real data. Besides illustrating typical applications of the indices as descriptive measures of spatial pattern, the examples also draw attention to some of the difficulties and limitations commonly found in interpretation. The first example is an analysis of patterns of ethnic group distribution in a Canadian city, and the changes and trends which have occurred through time in response to such social processes as immigration and assimilation into the mainstream of Canadian society. The second example uses statistics on mortality due to cancer of various types over a six year period to draw some limited conclusions regarding possible causal factors and indicators of risk.

2.4.1 Ethnic group distributions in London, Ontario, 1961-1971

This example is drawn from the work of Khondakar (1981). The data to be examined consist of the numbers of persons of different ethnic origins living in London, Ontario in the census years 1961 and 1971. Ontario cities have received large numbers of immigrants over the past one hundred years, initially from Britain, but increasingly from other parts of Europe and now from all parts of the world. The postwar period of the late 1940s and 1950s saw large numbers of Dutch, German, Italian, Polish and Ukrainian arrivals.

The questions to be asked in the analysis of the data are largely concerned with the concept of ethnic assimilation. Some groups have moved rapidly into the mainstream of Canadian society, while others have taken rather longer, remaining socially distinct for a significant period. This social distinction will be reflected in uneven spatial distribution of ethnic groups over the city.

The degree to which any variable is uniformly distributed in a set of statistical reporting zones, such as the 51 census tracts in the city, can be measured readily using the Gini coefficient. Suppose the value of the variable is known for each tract, and is expressed as a percentage of the total city. Let the tracts be ordered on the basis of this variable, from the tract which has the lowest percentage of the total to that with the highest. With n tracts, if the distribution is uniform then all percentages will be equal to 100/n%. Plot the data in cumulative form on a graph, with the vertical axis ranging from 0% to 100%, and the horizontal axis from 1 to n. Draw a diagonal line from the lower left to the upper right corners of the graph, representing a uniform distribution. The Gini coefficient is then simply the area between this diagonal line and the sample curve, scaled so that a value of 1 indicates total concentration of the variable in one tract, and 0 uniform distribution over all tracts.

Table 2 shows the Gini coefficients for each ethnic group for the two years of the study. The British, French and German, older groups derived from northwestern Europe, show substantial decreases in concentration between 1961 and 1971. Much less change occurred in the Ukrainian, Italian and Polish groups, all of which remained highly concentrated, and by implication less assimilated.

Table 2: Gini and Geary coefficients for ethnic groups in London, Ontario, 1961 and 1971

Ethnic Group	1961 Gini	1971 Gini	1961 Geary	1971 Geary
British	0.366	0.141	1.047	0.976
French	0.212	0.163	0.839	0.895
German	0.175	0.120	0.955	1.025
Hungarian	no data	0.324	no data	0.982
Italian	0.394	0.410	0.864	0.876
Dutch	0.245	0.254	0.885	0.996
Polish	0.377	0.343	0.912	0.727
Scandinavian	0.222	0.294	1.060	0.959
Ukrainian	0.394	0.366	0.856	0.974

The Gini coefficient pays no attention to the spatial arrangement of the reporting zones, and identical results would be obtained if the attribute values were redistributed over the spatial objects. Thus we have no way of knowing whether the zones with high percentages of one group are scattered over the city, or concentrated in a single region. However this is precisely the aspect of the pattern which can be captured using a spatial autocorrelation coefficient.

Table 2 also shows the Geary coefficients calculated for the same data, using unstandardized binary weights based on the adjacencies between census tracts. The results show several effects which were not apparent in the Gini coefficients. First, since the British and German groups show very little

autocorrelation at both dates, the changes in concentration between 1961 and 1971 indicated by the Gini coefficients must have occurred with almost no change in the spatial pattern. This would suggest that the change occurred by substantial dilution of the 1961 concentrations.

The Dutch and Ukrainian groups show a major increase in the Geary index from 1961 to 1971 while the Gini coefficients remained constant. This suggests a rearrangement from a single region of high concentration in 1961 to a scatter of equally concentrated but isolated tracts in 1971. New immigrants often move into clearly defined regions of cheap housing, but after some years are able to move outward into areas of higher socioeconomic status. In the cases of these two groups this move outward and upward does not coincide with substantial assimilation. The Polish group remained concentrated, but became more strongly spatially autocorrelated in 1971. The 1960s saw high rates of Polish immigration to London, and the small, scattered community which had existed in 1961 had by 1971 become concentrated and localized on the southern edge of the core.

This example illustrates the use of the Geary index as a simple, summary description of a spatial distribution, augmenting the information available from the aspatial Gini coefficient. The study from which this example was drawn examined the question of scale in greater detail by computing both coefficients by Enumeration Area, a reporting unit with roughly one tenth the population of a census tract. It is then possible to compare the degree of concentration and regionalization of each group at each scale. Furthermore, since the two sets of zones are hierarchically related, we can compute the indices for the Enumeration Areas separately for each tract, and look at the variation in pattern from one tract to another. Concentration appears to be much higher in some types and ages of housing than in others.

2.4.2 Cancer mortality in Southern Ontario

The raw data for this second study (Boost, 1979) consist of numbers of deaths recorded for various forms of cancer for each of the counties of Southern Ontario for the period 1971-6. Cancer rates vary from place to place in response to a number of factors. Besides straightforward environmental effects such as air and water quality, we would expect lifestyle factors to vary spatially and to influence mortality for certain cancers. Occupational hazards are often significant factors, as are various components of diet. The scale of variation in mortality will be affected by and tend to mirror the scale of variation of the cause, but will also be affected by scales of human movement. The incidence of cancer attributable to working in a particular place will not be limited to the place alone, but will occur over the whole area from which workers commute. Similarly movement of patients from residence to hospital for long term treatment may create another form of spatial averaging.

Raw mortality rates can be calculated from these data in the form of ratios per 1000 population, for each year and county and for each major form of cancer. But although these rates show clear variation, they are difficult to interpret because age distributions are not uniform across the study area, and some areas are therefore inherently more at risk than others. We can remove the effects of age and sex of the population, and thus intrinsic risk, by computing the province-wide mortality rates for each age and sex category, and then expressing each county's mortality as a percentage of what would have

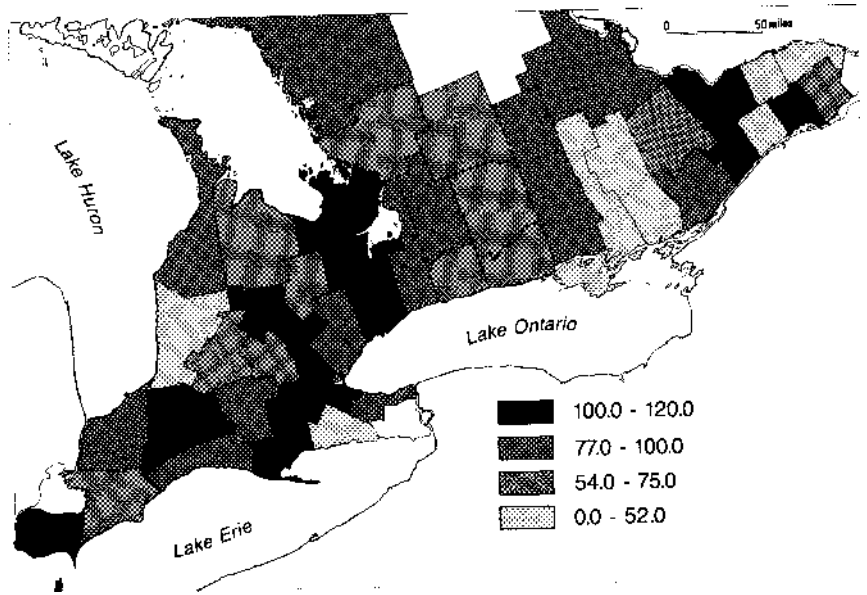


Figure 5. SMR's for cancer of the lung by county, Southern Ontario, 1971.

been expected if the county had behaved in exactly the same way as the whole province. These adjusted rates are known as Standardized Mortality Ratios (SMRs) and are shown in Figure 5. More precisely, the SMRs are given by:

$$SMR_{ijt} = 100\ O_{ijt} / (\sum_{k} D_{ikt}\ P_{jkt} / P_{kt}) \qquad (21)$$

where:

O_{ijt} is the observed no. of deaths due to cause i in county j in year t,

D_{ikt} is the no. of deaths in the province in age sex category k due to cause i in year t,

P_{jkt} is the population in county j in age sex category k in year t,

P_{kt} is the provincial population in age sex category k in year t.

The SMRs shown in Figure 5 indicate a quite different pattern of concentration from the raw mortalities. Similar SMRs are used by Cliff and Ord (1981) in their examples of bronchitis infection in London, 1959-1963.

The SMRs can be tested for significance by comparing the observed number of deaths with the number expected if the county is no different from the province as a whole in a standard chi-square test. Of the 34 counties shown, only eight have SMRs significantly different from 100. This is partly due to the sample sizes, since the numbers of cases for each county, cause and year are often small. Furthermore, some of the eight may result from type 1 errors, and there may be type 2 errors present in the 26 cases not judged significant, making the results difficult to interpret. However we have not yet made use of the information available from the spatial arrangement of county SMRs.

Table 3 shows the Geary indices calculated for each year and for each of the six cancer types studied. The types listed are those for which the Geary c index was significantly different from 1 at the .01 significance level on the N (resampling) null hypothesis. We would argue here that the individual county values are subject to a large number of factors whose combined effects can be regarded as random, and the SMRs show a distribution which is approximately normal, making the N null hypothesis reasonable in this case. Cancers of the lung and breast show a significant persistence in space in all years, suggesting that the causative factors are similarly persistent in space between counties.

Persistence in time may also be a useful indicator, since the long periods associated with the development of many cancers would suggest that high SMR's should persist for many years. Table' 4 shows the Spearman rank correlation coefficients for each pair of years for lung cancer SMR's. Correlations are not high, which is not surprising given the sample sizes, but are positive in all cases but one, and the majority are significant at the 0.05 level. To some extent, then, the anomalies which are observed in SMR's in certain counties are persistent from one year to the next, as well as being persistent from one county to the next. So although only eight individual lung cancer SMR's were significant in 1971, the existence of significant correlations in both space and time suggests that causes can be found at the county scale.

Perhaps the most instructive point of this example application lies in the importance of scale in interpretation. The spatial scale over which spatial autocorrelation is observed to operate suggests causative factors which operate over the same spatial scale. For example, it is unlikely that the effects of a single point source of water contamination would be reflected in the statistics of adjacent counties, and therefore in the degree of spatial autocorrelation between counties. On the other hand a health problem attributable to urban lifestyle would persist in all of the urbanized region around Toronto, would be significantly absent in the periphery, and would be revealed in strong spatial autocorrelation at the county level. Similar observations apply to duration of persistence in the time dimension.

Table 3: Spatial autocorrelation for Standard Mortality Ratios, Ontario, 1971-1976

	Geary's c	z-score
Types for 1971		
Digestive	.5254	-3.390
Intestines	.5776	-2.971
Lung	.5866	-2.908
Prostate	.5930	-2.863
Stomach	.5992	-2.819
Breast	.6642	-2.362
Types for 1972		
Digestive	.3884	-4.331
Lung	.4874	-3.606
Breast	.5198	-3.378
Intestines	.5531	-3.144
Types for 1973		
Lung	.3993	-4.226
Intestines	.4975	-4.535
Breast	.5525	-3.148
Digestive	.5900	-2.885
Types for 1975		
Lung	.4459	-3.577
Breast	.5514	-3.025
Digestive	.5587	-2.849
Types for 1976		
Breast	.3102	-4.454
Lung	.3959	-3.900
Intestine	.4834	-3.335
Digestive	.5067	-3.184
Stomach	.5837	-2.688

Values listed are significant using a one-tailed test at the .01 level against a critical z-score of 2.33.

Table 4: Spearman's rank correlation coefficients, Lung Cancer SMR's, Southern Ontario 1971-1976

	1971	1972	1973	1974	1975	1976
1971	1.00	0.31*	0.27*	0.29*	0.00	0.44*
1972		1.00	0.13	0.11	0.20	0.32*
1973			1.00	0.48*	0.11	0.37*
1974				1.00	0.35*	0.42*
1975					1.00	0.23
1976						1.00

*Denotes significant at the .05 level

2.5 Other Applications in the Literature

Cliff and Ord (1973, 1981) used the Moran index to reexamine the diffusion simulation model of Hagerstrand, as applied to the spread of improved pasture subsidy among Swedish farmers. They also give an example of the use of the Moran index to study the incidence of cholera from contaminated water supplies in London boroughs in 1854.

Geary's original application of the c index (Geary, 1954, reprinted in Geary, 1968) was to the analysis of patterns of agricultural, social and economic statistics across Irish counties.

Gatrell (1979b) examined population patterns in Southern Germany in an application of spatial autocorrelation statistics in the context of central place theory, while Bannister (1975) used the Moran index and various weighting schemes to expose the spatial structure of settlement in Southern Ontario.

Griffith (1982) used both I and c in an analysis of point objects in order to investigate the space-time structure of the Manitoba grain handling system.

Jumars, Thistle and Jones (1977) describe the use of both I and c, together with weights based on inverse square distances, in analyzing patterns of species abundances. Jumars (1978) gives an interesting three dimensional application to patterns of microfauna in the San Diego trench. Similar applications are described by Sokal and Oden (1978a, b) in a comprehensive review of the use of spatial autocorrelation measures in biology.

Hodder and Orton (1976) discuss the use of spatial autocorrelation measures in archaeology, and use I as a Measure of pattern in the spatial distribution of various types of ancient material.

3. JOIN COUNT STATISTICS

3.1 Measures

Join count statistics were introduced in Chapter 1 as a means of measuring spatial autocorrelation for nominal attributes. Since the classes of a nominal variable can be thought of as colours, it is common to refer to a distribution of two classes in terms of black and white objects, and of more than two classes as a k-colour map.

A join count refers to the number of cases of adjacent objects on the map. Counts could be made for joins between any pair of colours, or between objects of the same colour: for example, for a binary attribute we can count the number of BB, WW and BW joins. We can also generalize the notion of a join count to the case of weights. If each pair of objects i and j is given a weight w_{ij}, defined as for the Moran and Geary indices, then the BB count, for example, would be defined as the sum of the weights for all BB pairs. Counting joins is then precisely equivalent to summing binary adjacency weights. In the discussion which follows the most general case will be given first. Simplifications will follow in most cases where adjacencies are used as weights or where the objects are raster cells, or where there are only two colours.

3.2 Tests of Significance

Let n_r denote the observed number of objects with colour r, and let p_r denote the probability that a cell has colour r. The n's will be relevant in the case of randomization, where the observed colours are hypothetically redistributed over the objects. On the other hand the p's will be important in evaluating the resampling null hypothesis, which proposes that each object's attribute is obtained by random sampling from a parent population having unlimited numbers of attributes of each colour. In the interval case, resampling was assumed to occur from a normally distributed parent population, but in the nominal case each attribute is assumed to be obtained by a single trial under estimated probabilities. In most cases the p's would be estimated from the observed proportions of each colour present in the sample, but it is possible that other sources of information might exist.

In the following sections the notation used will be similar to that in Chapter 2. The subscript R will refer to the randomization null hypothesis and N to resampling. The expressions given below were first obtained by Moran (1946, 1947) and Krishna Iyer (1947, 1948, 1949a,b, and see also Decay, 1968).

3.2.1 Joins of the same colour

The index of interest in this case is the number of joins observed between objects of the same colour r. This is equivalent to the number of BB joins (or WW joins) in the binary case. As before, E refers to the expectation and Var to the variance.

$$E_N = S_0 p_r^2 / 2 \tag{22}$$

$$E_R = S_0 n_r^{(2)} / (2 n^{(2)}) \tag{23}$$

$$Var_N = \{S_1 p_r^2 + (S_2 - 2S_1) p_r^3 + (S_1 - S_2) p_r^4\}/4 \tag{24}$$

$$Var_R = \{S_1 n_r^{(2)}/n^{(2)} + (S_2 - 2S_1) n_r^{(3)}/n^{(3)} + (S_0^2 + S_1 - S_2) n_r^{(4)}/n^{(4)}\}/4 - E_R^2 \tag{25}$$

If weights are binary, the S terms, which are derived from the weights matrix, simplify as follows, and these relationships apply also to the following section:

S_0 = 2A where A is the count of all unique joins

S_1 = 4A

S_2 = $4 \sum_i B_i^2$ where B_i denotes the number of objects adjacent to object i.

3.2.2 Joins of different colours

Here the index of concern is the observed number of joins between objects coloured r and s.

$$E_N = S_0 p_r p_s \quad (26)$$

$$E_R = S_0 n_r n_s / n^{(2)} \quad (27)$$

$$Var_N = \left(2S_1 p_r p_s + (S_2 - 2S_1) p_r p_s (p_r + p_s) + 4(S_1 - S_2) p_r^2 p_s^2 \right) / 4 \quad (28)$$

$$Var_R = \left(2S_1 n_r n_s / n^{(2)} + (S_2 - 2S_1) n_r n_s (n_r + n_s - 2) / n^{(3)} + 4(S_0^2 + S_1 - S_2) n_r^{(2)} n_s^{(2)} / n^{(4)} \right) / 4 - E_R^2 \quad (29)$$

The expressions for E_R and Var_R reduce to the well-known ones for the Runs test of a binary sequence (Siegel, 1956) when the weights are binary adjacencies and the objects are arrayed along a single dimension.

More complex expressions for the total number of joins between counties of different colours are given in Cliff and Ord (1981, pp. 19-20)

3.2.3 Hypothesis testing

Given the above expressions for expectations and variances of the join count statistics under both null hypotheses, it is a relatively straightforward matter to calculate the test statistic using the methods outlined in Chapter 2. Like the Moran and Geary statistics, the join counts are asymptotically normally distributed under both null hypotheses, and so identical methods can be used.

Because it is possible to make separate tests of each join count statistic, it is possible to compare the degree of spatial autocorrelation revealed by the arrangement of one colour over the map with that exhibited by one or more of the others. The join count test thus has the potential to reveal somewhat more than a test of I or c, although a loss of information is inevitable in reducing data from an interval to a nominal scale.

An example calculation and test using join count statistics is shown below to clarify the meaning of each of the terms in the equations, for the data shown in Figure 6. In addition the BB, WW and BW join counts are shown for each of the patterns in Figure 1. Note that the BB and WW counts are highest in those patterns having positive spatial autocorrelation, and the BW count is highest when spatial autocorrelation is negative. A BASIC program for calculating join count statistics is included in Appendix 2.

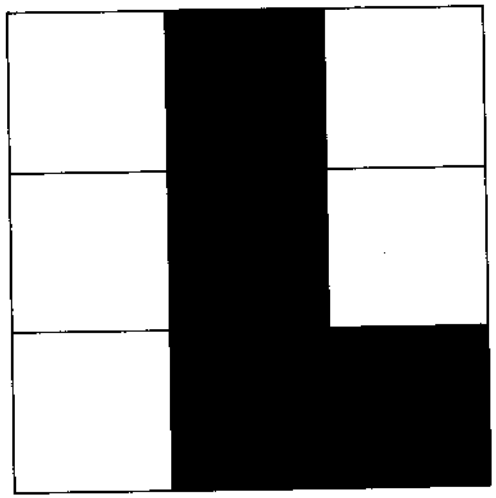


Figure 6. Example data set for calculation of join count statistics.

Example calculation for join count statistics

$n = 9$

$n_B = 4$ $\quad n_{BW} = 6$

$n_W = 5$ $\quad n_{BB} = 3$

$p_B = 0.444$ $\quad n_{WW} = 3$

$p_W = 0.556$

$S_0 = 24$ $\quad S_1 = 48$ $\quad S_2 = 272$

Joins of the same colour:

$E_N = S_0 p_r^2 / 2$

black: $E_N = 24 \times 0.444^2 / 2 = 2.37$

white: $E_N = 24 \times 0.556^2 / 2 = 3.71$

$E_R = S_0 n_r^{(2)} / 2n^{(2)}$

black: $E_R = 24 \times 4 \times 3 / 2 \times 9 \times 8 = 2.00$

white: $E_R = 24 \times 5 \times 4 / 2 \times 9 \times 8 = 3.33$

$$Var_N = [S_1 p_r^2 + (S_2 - 2S_1) p_r^3 + (S_1 - S_2) p_r^4] / 4$$

black: $Var_N = [48 \times 0.444^2 + (272 - 2\times48) \times 0.556^3 + (48 - 272) \times 0.444^4] / 4$

$= 4.05$

white: $Var_N = [48 \times 0.556^2 + (272 - 2\times48) \times 0.556^3 + (48 - 272) \times 0.556^4] / 4$

$= 5.91$

$$Var_R = (S_1 n_r^{(2)}/n^{(2)} + (S_2 - 2S_1) n_r^{(3)}/n^{(3)} + (S_0^2 + S_1 - S_2) n_r^{(4)}/n^{(4)})/4 - E_R^2$$

black: $Var_R = [48\times4\times3 / 9\times8 + (272 - 2\times48)\times4\times3\times2 / 9\times8\times7 + (24^2 + 48 - 272)\times4\times3\times2\times1 / 9\times8\times7\times6] / 4 - 2.00^2$

$= 0.794$

white: $Var_R = [48\times5\times4 / 9\times8 + (272 - 2\times48)\times5\times4\times3 / 9\times8\times7 + (24^2 + 48 - 272)\times5\times4\times3\times2 / 9\times8\times7\times6] / 4 - 3.33^2$

$= 0.952$

Joins of different colours:

$E_N = S_0 p_r p_s = 24\times0.444\times0.556 = 5.93$

$E_R = S_0 n_r n_s / n^{(2)} = 24\times4\times5 / 9\times8 = 6.67$

$Var_N = [2S_1 p_r p_s + (S_2 - 2S_1) p_r p_s (p_r + p_s) + 4(S_1 - S_2) p_r^2 p_s^2] / 4$

$= [2\times48\times0.444\times0.556 + (272 - 2\times48)\times0.444\times0.556\times(0.444 + 0.556) + 4\times(48 - 272)\times0.444^2\times0.556^2] / 4$

$= 3.13$

$$Var_R = [2S_1 n_r n_s/n^{(2)} + (S_2 - 2S_1) n_r n_s (n_r + n_s - 2)/n^{(3)} + 4(S_0^2 + S_1 - S_2) n_r^{(2)} n_s^{(2)}/n^{(4)}] / 4 - E_R^2$$

$= [2\times48\times4\times5 / 9\times8 + (272 - 2\times48)\times4\times5\times(4+5-2) / 9\times8\times7 + 4\times(24^2 + 48 - 272)\times4\times3\times5\times4 / 9\times8\times7\times6] / 4 - 6.67^2$

$= 2.38$

$x = (s - E)/\sqrt{(var)}$

Black-white joins:	resampling	$x = (6 - 5.93)/\sqrt{(3.13)} = 0.042$
	randomization	$x = (6 - 6.67)/\sqrt{(2.38)} = -0.432$
Black-black joins:	resampling	$x = (3 - 2.37)/\sqrt{(4.05)} = 0.313$
	randomization	$x = (3 - 2.00)/\sqrt{(0.79)} = 1.122$
White-white joins:	resampling	$x = (3 - 3.71)/\sqrt{(5.91)} = -0.289$
	randomization	$x = (3 - 3.33)/\sqrt{(0.95)} = -0.342$

3.3 Applications

Cliff and Ord (1981) use join count statistics in their analysis of the Hagerstrand simulation model to supplement the Moran index. Join counts are also used in their analysis of the incidence of measles in Cornwall, and mortality due to bronchitis in London and chest diseases in Wales. The Cornish measles data are also described in Cliff and Haggett (1979). A more general discussion of the use of spatial autocorrelation in analyzing epidemics can be found in Cliff and Haggett (1984).

An interesting perspective on join counts is provided by Olson (1975) who discusses the visual perception of autocorrelated spatial patterns.

4. AUTOCORRELATED PROCESSES

4.1 Simulations

The emphasis in this volume to this point has been almost exclusively on the measurement and description of spatial autocorrelation rather than on the processes which generate it. This emphasis is deliberate, partly because discussion of process tends to be more technical, complex and diifficult, and partly because of the relatively poorly developed level of our understanding of spatially autocorrelated processes.

A simple mental experiment will illustrate the difficulty. Suppose we wished to simulate a topographic surface by generating a raster or grid of simple elevations. To simulate the spatial autocorrelation present to some degree in all topographic surfaces, the generating process would have to ensure that each point in the representative raster is correctly autocorrelated with each of its neighbours, four in number in a regular square array. It is clearly impossible to generate such a surface by sequentially assigning elevations row by row from the top, since one could ensure independence from the previous elevation in the row, or the elevation immediately above in the previous row, but not both simultaneously. In fact there has been considerable debate in the literature over appropriate methods of generating reasonable-looking topographic surfaces (see Mandelbrot, 1977, 1982).

The process used to generate the various arrangements of attribute Values in Figures 1 and 4 bears no relationship to any process which might operate in the real world, but was devised merely for the generation of

patterns with known spatial autocorrelation (Goodchild, 1980). A set of attribute values is generated, and allocated randomly to the objects. Pairs of objects are then selected at random, and their attributes are swapped if doing so would bring the spatial autocorrelation of the entire arrangement closer to the prescribed target.

At this point, we can now turn to a discussion of the attempts which have been made to model processes operating to produce autocorrelation in real data. Because of the complexity of the field the treatment will be necessarily descriptive and superficial, but references to more complete discussions will be given for the interested reader. For a discussion of autocorrelated processes in the context of simulation see Haining, Griffith and Bennett (1983).

4.2 Spatial Processes

An attribute variable can show spatial autocorrelation in its arrangement across spatial objects either because neighbouring objects influence each other directly, so that the value at one place is caused directly by values at neighbouring places (autocorrelation), or because the value at each place is determined by some other variable at the same place which is itself autocorrelated. For example, we do not know whether the Italian ethnic group in the city of London, Ontario (Figure 2c) shows autocorrelation at the Census tract level because people of Italian origin are attracted to areas near other people of the same ethnic group, or because the factor which attracts them is also present in other areas with large Italian populations.

Cliff and Ord (1981, pp. 141ff.) refer to the two interpretations as interactive and reactive respectively. Suppose the process which generated the attributes is solely reactive. Then if all of the causative factors can be found and modelled, the residuals from this model will be completely lacking in spatial autocorrelation. So if a test for autocorrelation of residuals in negative, and if we are willing to ignore the possibility of a Type 2 error, we can conclude that the modelling effort is successful and that the assumption of no interactive effects is valid. On the other hand if a test for autocorrelation is positive, we will be unable to resolve whether the source is interactive effects, or reactive effects not included in the model. So we can in general resolve the ambiguity between interaction and reaction only in those cases where perfect models can be obtained.

Whittle (1954) proposed the first spatial autoregressive model for the interactive case as follows:

$$z_i = \sum_j r_{ij} z_j + e_i \qquad (39)$$

where the e_i's are uncorrelated error terms and the r_{ij}'s are coefficients. A simple scheme would be to set r_{ij} equal to $\rho\ w_{ij}$ where the w_{ij}'s have their usual meaning and ρ is a constant. In a raster with binary weights based on adjacency this would allow the value in each cell to be an average of the values of its neighbours, distorted by a random error term:

$$z_{ij} = (z_{i-1,j}+z_{i,j-1}+z_{i+1,j}+z_{i,j+1})/4 + e_{ij}$$

where z_{ij} and e_{ij} denote the attribute value and error in cell (i,j) respectively.

Given a suitable model of interactive process, the next step would be to calibrate the model using observed data, in other words to attempt to find the value of the constant ρ which gives the best fit between the model and observation, and to obtain the values of the residuals e_i. The least squares method centres on the squared differences between the predictions of the model and the observed values of the z_i's, and finds the model which will reduce the total of these squared differences to the minimum. Maximum likelihood focuses on the likelihood that a given model produced the observed z_i's, and finds the model which is most likely.

Unfortunately neither approach leads to simple results in the case of the autoregressive model. Whittle notes that the least squares estimators are inconsistent, although Ord (1975) provided a modified least squares method which yields consistent, if inefficient, estimators. And the mathematics necessary to obtain the maximum likelihood estimators are too difficult except in certain simple cases.

An alternative to the autoregressive model is the method of moving averages developed by Haining (1978):

$$z_i = \bar{z} + e_i + \sum_j r_{ij} e_j \qquad (40)$$

where all terms have their previous meanings. In effect, the process consists of a constant mean disturbed by autocorrelated errors. Again the model is not easy to calibrate.

The literature on spatial processes has grown rapidly in recent years. Further information can be found in Cliff and Ord (1981, Chapter 6) and Ripley (1981 pp. 88-95). Many of the methods are extensions of one-dimensional methods developed initially in the study of time series.

4.3 Autocorrelation in Residuals

The previous section discussed models of processes which were purely interactive. We now turn to cases of mixtures of interactive and reactive processes, and to models of reactive processes where not all of the causative factors are known. In either case we might construct and calibrate a model of the form:

$$z_i = b_0 + \sum_k b_k x_{ik} + e_i \qquad (30)$$

where the b's are constants to be determined and the x_{ik}'s represent the values of causative attributes present at each object and to which the z_i's react. If not all of the causative factors are included, or if interactive processes are also present, the residuals or e.'s will be spatially autocorrelated. The b's would be obtained by ordinary least squares regression.

One of the assumptions made in a large number of traditional statistical tests. including regression analysis, is that the observations are independent of each other. In this particular technique it is assumed that the residuals e_i are mutually independent and sampled from the same parent distribution. Many

standard packages for regression allow the user to test for autocorrelation of the residuals in time by computing a test statistic based on the sequence of observations, but it is rare for any test to be made of spatial autocorrelation.

The importance of the assumption is quite easy to see. Suppose all of the variables in the model, including the z_i's, show strong, positive spatial autocorrelation. If a new sample is obtained from a point close to one of the existing samples, we know that the values of all of the variables will be close to those at the existing point. In a sense, then, this is not truly new data, and the task of predicting the new data is not as difficult as it would have been had the new point been located a long way from any previous point. We can visualize the effect of spatial autocorrelation as reducing the apparent number of degrees of freedom, so that the model appears more powerful than it really is.

Autocorrelation can appear in the residuals for several reasons: because a linear model has been used for a nonlinear relationship; because one or more of the causal variables have been omitted; or because both reactive and interactive processes are present. If present, autocorrelation leads to bias in the estimation of residual variance and therefore in R^2 and other measures of the success of the model, and inefficiency in the estimation of the regression coefficients, the b's.

Cliff and Ord (1981, pp. 200 ff.) discuss a test statistic based on the Moran index which can be used to test for spatial autocorrelation among the residuals of a regression model, and provide methods of testing its statistical significance. If spatial autocorrelation is found to be present, one useful strategy would be to manipulate the mix of causal variables in the model to try to reduce it, and Cliff and Ord illustrate this using an example. Geary (1954) argued that absence of spatial autocorrelation among the residuals would indicate complete success: if we assume that all causal variables are spatially autocorrelated, then it would indicate that no causative variables would remain to be found. However it seems more likely that the correct inference would be that a type 2 error had been made: there are causes not yet included in the model, but their influence is so weak that any spatial autocorrelation induced in the residuals is insufficient to lead to the rejection of a null hypothesis of no spatial autocorrelation.

Another strategy would be to delete observations selectively, particularly those close to others and therefore likely to contribute to autocorrelation. This is somewhat drastic, however, as it diminishes the data available. Finally, one might modify the form in which the independent variables appear in the model by powering or taking logs, particularly if it is suspected that the cause of the autocorrelation is the use of a linear model when a nonlinear one is needed.

5. CONCLUSIONS

The emphasis in this volume has been on the use of measures of spatial autocorrelation, particularly the Moran, Geary and join count statistics, in order to measure certain significant aspects of spatial pattern. As noted earlier, one can argue that spatial autocorrelation is the most significant aspect of the arrangement of attributes over the objects present in geographical space.

This is only one of the traditions of the spatial autocorrelation literature, and references to others have been made from time to time, although constraints of space and mathematical complexity have not allowed more detailed discussion. One approach has been to regard spatial autocorrelation as a problem which violates the assumptions of many standard statistical tests, and to attempt to devise modifications of those tests which might be used where appropriate. There was limited discussion of this in relation to regression in Chapter 4, and the literature also contains analyses of the t test of means and lengthy discussion of the effect of spatial autocorrelation on the calibration of spatial interaction models (see Cliff and Ord, 1981 for review).

A second approach to autocorrelation derives from spectral analysis, since there is a simple relationship between the spectrum and the autocorrelation function of a one or two dimensional series. Spectral analysis has found applications in several branches of geography, particularly in climatology. More generally, however, its applications have been limited by two factors: first, there tend to be few repetitive, cyclical aspects of patterns in space; and second, spatial phenomena tend to exhibit trend components. Both factors suggest that the variogram may be more suitable as a descriptive device for spatial distributions than the spectrum. Geographers have made use of correlograms and variograms in a number of areas, many of which would make very interesting additions to this volume if more space were available. Kriging is a method of spatial interpolation developed in the mining industry (Krige, 1951) and by the French school of geostatistics (Matheron, 1965; David, 1977; Olea, 1974; Delfiner and Delhomme, 1975) based on the principle that the interpolated data should show the same variogram, in effect the same structure of spatial autocorrelation, as that observed between the known data points. Variograms are also a basic tool in the simulation of terrain by fractal processes (see the striking illustrations in Mandelbrot, 1977, 1982).

The third basic approach to spatial autocorrelation which has been touched on at various points, particularly in Chapter Four, is its analysis as a parameter of process, rather than a description of form or arrangement. Again space and mathematical complexity have been limiting factors, as well as the technical difficulties of calibration. This is clearly an area of great potential for spatial modelling.

Geographers are trained to see spatial autocorrelation as the rule rather than the exception in the phenomena they study, and so it comes as something of a shock to realize the extend to which traditional statistical methods rely on assuming its absence: Gould (1970) expresses this sentiment very clearly in the passage quoted in the introduction. Yet our texts still largely reflect the traditional statistical view. Perhaps the real test of the coming of age of statistical geography will be the design of an introductory quantitative course which takes spatial autocorrelation as the rule rather than the exception, and builds systematically from that premise.

References

Bannister, G., 1975. Population change in Southern Ontario. Annals, Association of American Geographers, 65, 177-188.

Boost, B.A., 1979. A search for true anomalies in the distributions of site-specific cancers, Ontario, 1971-76. Unpublished MA dissertation, Department of Geography, University of Western Ontario.

Boots, B. N., 1986. Voronoi Polygons. CATMOG 45, Geo Books, Norwich.

Brassel, K. and D. Reif, 1979. A procedure to generate Thiessen polygons. Geographical Analysis, 11, 289-303.

Cliff, A.D. and J.K. Ord, 1971. Evaluating the percentage points of a spatial autocorrelation coefficient. Geographical Analysis, 3, 51-62.

Cliff, A.D. and J.K. Ord, 1973. Spatial Autocorrelation. Pion, London.

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

Cliff, A.D. and P. Haggett, 1979. Geographical aspects of epidemic diffusion. In: N. Wrigley, editor, Statistical Applications in the Spatial Sciences. Pion, London, 5-44.

Cliff, A.D. and J.K. Ord, 1981. Spatial Processes: Models and Applications. Pion, London.

Cliff, A.D. and P. Haggett, 1984. Island epidemics. Scientific American, 250(5), 138-147.

Coffey, W., M.F. Goodchild and L.C. MacLean, 1982. Randomness and order in the topology of settlement systems. Economic Geography, 58, 20-28.

Dacey, M.F., 1968. A review on measures of contiguity for two and k-color maps. In: B.J.L. Berry and D.F. Marble, editors, Spatial Analysis: A Reader in Statistical Geography, Prentice-Hall, Englewood Cliffs, N.J., 479-495.

David, M., 1977. Geostatistical Ore Reserve Estimation. Elsevier, New York.

Delfiner, P. and J.P. Delhomme, 1975. Optimal interpolation by Kriging. In: J.C. Davis and M.J. McCullagh, editors, Display and Analysis of Spatial Data, Wiley, New York 96-114.

Diaconis, P. and B. Efron, 1983. Computer intensive methods in statistics. Scientific American, 248(5), 116-131.

Pasco, U. Frei and A. Hsu, 1985. Thematic Mapper: operational activities and sensor performance at ESA/Earthnet. Photogrammetric Engineering and Remote Sensing, 51, 1299-1314.

Gabriel, K.R and R.R. Sokal, 1969. A new statistical approach to geographic variation analysis. Systematic Zoology, 18, 259-278.

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

Gatrell, A.C., 1979a. Autocorrelation in spaces. Environment and Planning A, 11, 507-516.

Gatrell, A.C., 1979b. The autocorrelation structure of central place populations in Southern Germany. In: N. Wrigley, editor, Statistical Applications in the Spatial Sciences, Pion, London 111-126.

Geary, R.C., 1954. The contiguity ratio and statistical mapping. Incorporated Statistician, 5, 115-141.

Geary, R.C., 1968. The contiguity ratio and statistical mapping. In: B.J.L. Berry and D.F. Marble, editors, Spatial Analysis: A Reader in Statistical Geography, Prentice-Hall, Englewood Cliffs, N.J. 461-478.

Getis, A. and B. Boots, 1978. Models of Spatial Processes. Cambridge University Press.

Goodchild, M.F., 1980. Simulation of autocorrelation for aggregate data. Environment and Planning A, 12, 1073-1081.

Goodchild, M.F. and A.W. Grandfield, 1983. Optimizing raster storage: an examination of four alternatives. Proceedings of the Sixth International Symposium on Automated Cartography (AutoCarto VI), Ottawa 1 400-407.

Gould, P.R., 1970. Is statistix inferens the geographical name for a wild goose?. Economic Geography, 46, 439-448.

Griffith, D., 1982. Dynamic characteristics of spatial economic systems. Economic Geography, 58, 177-196.

Haggett, P., A.D. Cliff and A. Frey, 1977. Locational Analysis in Human Geography, (2nd edition, 2 volumes) Arnold, London.

Haining, R.P., 1978. The moving average model for spatial interaction. Transactions and Papers, Institute of British Geographers, (New Series) 3, 202-225.

Haining, R.P., D.A. Griffith and R.J. Bennett, 1983. Simulating two-dimensional autocorrelated surfaces. Geographical Analysis, 15, 247-253.

Hall, J.R.W., 1977. The geography of motor vehicle collisions: an investigation of the Southwestern Ontario highway network. Unpublished BA dissertation, Department of Geography, University of Western Ontario.

Handler, G.Y. and P.B. Mirchandani, 1979. Location on Networks: Theory and Algorithms. MIT Press, Cambridge, Mass.

Haynes, K.E. and A.S. Fotheringham, 1984. Gravity and Spatial Interaction Models. Scientific Geography Series Vol. 2, edited by G.I. Thrall, Sage, Beverly Hills, Calif.

Hodder, I. and C. Orton, 1976. Spatial Analysis in Archaeology. Cambridge University Press.

Hodgart, R.L., 1978. Optimising access to public facilities: a review of problems, models and methods of locating central facilities. Progress in Human Geography, 2, 17.

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

Hubert, L.J., R.G. Golledge and C.M. Costanzo, 1981. Generalized procedures for evaluating spatial autocorrelation. Geographical Analysis, 13, 224-233.

Jumars, P.A., D. Thistle and M.L. Jones, 1977. Detecting two dimensional spatial structure in biological data. Oecologia, 28, 109-123.

Jumars, P.A., 1978. Spatial autocorrelation with RUM: vertical and horizontal structure of a bathyal benthic community. Deep-Sea Research, 25, 589-604.

Khondakar, N., 1981. Some changes in socioeconomic and spatial characteristics of ethnic groups in London, Ontario, 1961-1971. Unpublished PhD dissertation, Department of Geography, University of Western Ontario.

Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa, 52, 119-139.

Krishna Iyer, P.V.A., 1947. Random association of points on a lattice. Nature, 160, 714.

Krishna Iyer, P.V.A., 1948. Random association of points on a lattice. Nature, 162, 333.

Krishna Iyer, P.V.A., 1949a. Random association of points on a lattice. Journal of the Indian Statistical Association, 2, 141-160.

Krishna Iyer, P.V.A., 1949b. The first and second moment of some probability distributions arising from points on a lattice and their applications. Biometrika, 36, 135-141.

Mandelbrot, B.B., 1977. Fractals: Form, Chance and Dimension. Freeman, San Francisco.

Mandelbrot, B.B., 1982. The Fractal Geometry of Nature. Freeman, San Francisco.

Matheron, G., 1965. Les Variables Regionalisees et Leur Estimation. Editions Masson, Paris.

Monmonier, M.S., 1982. Computer-Assisted Cartography: Principles and Prospects. Prentice-Hall, Englewood Cliffs, New Jersey.

Moran, P.A.P., 1946. Random associations on a lattice. Nature, 158, 521.

Moran, P.A.P., 1947. Random associations on a lattice. Proceedings of the Cambridge Philosophical Society, 43, 321-328.

Moran, P.A.P., 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society, Series B 10, 243-251.

Olea, R.A., 1974. Optimal contour mapping using Universal Kriging. Journal of Geophysical Research 79 695-702.

Olson, J.M., 1975. Autocorrelation and visual map complexity. Annals, Association of American Geographers, 65, 189-204.

Openshaw, S., 1977a. A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. Transactions, Institute of British Geographers, (New Series) 2, 459-472.

Openshaw, S., 1977b. Optimal zoning systems for spatial interaction models. Environment and Planning A, 9, 169-184.

Openshaw, S., 1978. An empirical study of some 'zone-design criteria. Environment and Planning A, 10, 781-794.

Openshaw, S. and P.J. Taylor, 1979. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: N. Wrigley, editor, Statistical Applications in the Spatial Sciences, Pion, London.

Openshaw, S. and P.J. Taylor, 1981. The modifiable areal unit problem. In: N. Wrigley and R.J. Bennett, editors, Quantitative Geography: A British View, Routledge, London.

Ord, J.K., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association, 70, 120-126.

Pack, D.H., 1980. Precipitation chemistry patterns: a two network data set. Science, 208, 1143-1144.

Ripley, B.D., 1981. Spatial Statistics. Wiley, New York.

Rogers, A., 1969a. Quadrat analysis of urban dispersion. (1) Theoretical techniques. Environment and Planning, 1, 47-80.

Rogers, A., 1969b. Quadrat analysis of urban dispersion. (2) Case studies of urban retail systems. Environment and Planning, 1, 155-171.

Rogers, A., 1974. Statistical Analysis of Spatial Dispersion. The Quadrat Method. Pion, London.

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

Sen, A.K., 1976. Large sample-size distribution of statistics used in testing for spatial autocorrelation. Geographical Analysis, 9, 175-184.

Sen, A.K. and S. Soot, 1977. Rank tests for spatial correlation. Environment and Planning A, 9, 897-905.

Siegel, S., 1956. Nonparametric Statistics for the Behavioural Sciences, Wiley, New York.

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

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

Taylor, P.J., 1980. A pedagogic application of multiple regression analysis. Geography, 65, 203-212.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography, 46 (Supplement), 234-240.

Upton, G.J.G. and B. Fingleton, 1985. Spatial Data Analysis by Example. Volume I: Point Pattern and Quantitative Data. Wiley, Toronto.

Whittle, P., 1954. On stationary processes in the plane. Biometrika, 41, 434-449.

APPENDIX I

BASIC Program for calculating Gerry and Moran statistics and significance tests.

```
10 REM'Example program to calculate Moran and Geary statistics
20 REM'Written for the IBM PC BASIC, but should be compatible
30 REM'with many other BASICs.
40 REM'
50 REM'The user is prompted to enter the weights matrix and z
60 REM'vector element by element.
70 REM'
80 DIM Z(50),W(50,50),SWIDOT(50),SWDOTI(50)
90 REM'NMAX gives the maximum number of objects allowed
100 NMAX=50
110 PRINT
120 INPUT "Number of objects: ",N
130 PRINT
140 IF N<=NMAX THEN 180
150 PRINT N;"is too many"
160 GOTO 120
170 REM'Enter the values for each object
180 FOR I=1 TO N
190 PRINT "Value for object";I;
200 INPUT Z(I)
210 NEXT I
220 REM'Find the mean and variances
230 SUM1=0
240 SUM2=0
250 FOR I=1 TO N
260 SUM1=SUM1+Z(I)
270 SUM2=SUM2+Z(I)↑2
280 NEXT I
290 ZBAR=SUM1/N
300 SSQ=SUM2/N-ZBAR↑2
310 SIGSQ=(SUM2-ZBAR*SUM1)/(N-1)
320 PRINT
330 PRINT "Mean value: ";ZBAR
340 PRINT "Sample variance: ";SSQ;"  std deviation: ";SQR(SSQ)
```



```
350 PRINT "Population variance: ";SIGSQ;"  std deviation: ";SQR(SIGSQ)
360 REM'Zero the weights matrix then request data entry
370 FOR I=1 TO N
380 FOR J=1 TO N
390 W(I,J)=0
400 NEXT J
410 NEXT I
420 PRINT
430 PRINT "Enter the weights matrix row by row"
440 PRINT "Only nonzero entries need be entered"
450 PRINT "To finish entering each row type 0 when asked for next column"
460 FOR I=1 TO N
470 PRINT
480 PRINT "Weights matrix row";I
490 INPUT "Enter column number for next nonzero weight or 0: ",J
500 IF J=0 THEN 610
510 REM'Check the column number is valid
520 IF J<20 OR J>2N THEN 570
530 REM'Check the diagonal is not entered
540 IF J=I THEN 590
550 INPUT "Enter weight: ",W(I,J)
560 GOTO 490
570 PRINT "Error in column number"
580 GOTO 490
590 PRINT "Diagonal entries in the weights matrix must be zero"
600 GOTO 490
610 PRINT
620 PRINT "Weights entered for row";I;":"
630 FOR J=1 TO N
640 PRINT J,W(I,J)
650 NEXT J
660 INPUT "Enter 0 to continue this row or 1 to move to next row: ",FLAG
670 IF FLAG=0 THEN 490
680 NEXT I
690 REM'Find various sums for the weights matrix
700 FOR I=1 TO N
710 SWIDOT(I)=0
720 SWDOTI(I)=0
730 NEXT I
740 S0=0
750 S1=0
760 S2=0
770 IDENOM=0
780 CDENOM=0
790 FOR I=1 TO N
800 FOR J=1 TO N
810 SWIDOT(I)=SWIDOT(I)+W(I,J)
820 SWDOTI(J)=SWDOTI(J)+W(I,J)
830 S0=S0+W(I,J)
840 S1=S1+(W(I,J)+W(J,I))↑2
850 IDENOM=IDENOM+(Z(I)-ZBAR)*(Z(J)-ZBAR)*W(I,J)
860 CDENOM=CDENOM+(Z(I)-Z(J))↑2*W(I,J)
870 NEXT J
880 NEXT I
890 REM'Calculate indices
900 MORAN=IDENOM/(SSQ*S0)
```

```
910 GEARY=CDENOM/(2*SIGSQ*S0)
920 PRINT
930 PRINT "Coefficients:"
940 PRINT "Moran's I: ";MORAN
950 PRINT "Geary's c: ";GEARY
960 FOR I=1 TO N
970 S2=S2+(SWIDOT(I)+SWDOTI(I))↑2
980 NEXT I
990 S1=S1/2
1000 REM'Calculate higher moments of z
1010 M2=SSQ
1020 M4=0
1030 FOR I=1 TO N
1040 M4=M4+(Z(I)-ZBAR)↑4
1050 NEXT I
1060 M4=M4/N
1070 B2=M4/M2↑2
1080 REM'Print out these results
1090 PRINT
1100 PRINT "Intermediate statistics:"
1110 PRINT "S0: ";S0
1120 PRINT "S1: ";S1
1130 PRINT "S2: ";S2
1140 PRINT "m2: ";M2
1150 PRINT "m4: ";M4
1160 PRINT "b2: ";B2
1170 PRINT
1180 INPUT "Hit any key to continue",A$
1190 REM'Calculate expectations
1200 EI=-1/(N-1)
1210 EC=1
1220 PRINT
1230 PRINT "Expectations: "
1240 PRINT "Resampling, Moran: ";EI
1250 PRINT "Randomization, Moran: ";EI
1260 PRINT "Resampling, Geary: ";EC
1270 PRINT "Randomization, geary: ";EC
1280 REM'Calculate variances
1290 VARNI=(N↑2*S1-N*S2+3*S0↑2)/(S0↑2*(N↑2-1))-EI↑2
1300 VARRI=N*((N↑2-3*N+3)*S1-N*S2+3*S0↑2)-B2*((N↑2-N)*S1-2*N*S2+6*S0↑2)
1310 VARRI=VARRI/((N-1)*(N-2)*(N-3)*S0↑2)-EI↑2
1320 VARNC=((2*S1+S2)*(N-1)-4*S0↑2)/(2*(N+1)*S0↑2)
1330 VARRC=(N-1)*S1*(N↑2-3*N+3-(N-1)*B2)
1340 VARRC=VARRC-1/4*(N-1)*S2*(N↑2+3*N-6-(N↑2-N+2)*B2)
1350 VARRC=(VARRC+S0↑2*(N↑2-3-(N-1)↑2*B2))/(N*(N-2)*(N-3)*S0↑2)
1360 PRINT
1370 PRINT "Variances: "
1380 PRINT "Resampling, Moran: ";VARNI
1390 PRINT "Randomization, Moran: ";VARRI
1400 PRINT "Resampling, Geary: ";VARNC
1410 PRINT "Randomization, Geary: ";VARRC
1420 REM'Calculate standard deviates
1430 SNI=(MORAN-EI)/SQR(VARNI)
1440 SRI=(MORAN-EI)/SQR(VARRI)
1450 SNC=(GEARY-EC)/SQR(VARNC)
1460 SRC=(GEARY-EC)/SQR(VARRC)
1470 PRINT
```



```
1480 PRINT "Standard deviates:"
1490 PRINT "Resampling, Moran: ";SNI
1500 PRINT "Randomization, Moran: ";SRI
1510 PRINT "Resampling, Geary: ";SNC
1520 PRINT "Randomization, Geary: ";SRC
1530 PRINT
1540 PRINT "Finished"
1550 END
```

APPENDIX II

BASIC program for calculating join count statistics and significance tests.

```
10 REM'Example program to calculate join count statistics
20 REM'Written for the IBM PC BASIC, but should be compatible
30 REM'with many other BASICs.
40 REM'
50 REM'The user is prompted to enter the adjacency matrix and z
60 REM'vector element by element.
70 REM'
80 DIM Z(50),W(50,50)
90 REM'NMAX gives the maximum number of objects allowed
100 NMAX=50
110 PRINT
120 INPUT "Number of objects: ",N
130 PRINT
140 IF N<=NMAX THEN 180
150 PRINT N;"is too many"
160 GOTO 120
170 REM'Enter the colours for each object
180 PRINT "Enter the colours as integers in the range 1 to 50"
190 PRINT "for example 1 for white and 2 for black"
200 PRINT
210 FOR I=1 TO N
220 PRINT "Colour for object";I;
230 INPUT Z(I)
240 NEXT I
250 PRINT
260 REM'Count the frequency of each colour
270 ACCUM=0
280 FOR COLOUR=1 TO 50
290 NR=0
300 FOR I=1 TO N
310 IF Z(I)<>COLOUR THEN 330
320 NR=NR+1
330 NEXT I
340 IF NR=0 THEN 360
350 PRINT "Colour";COLOUR;"count";NR
360 ACCUM=ACCUM+NR
370 IF ACCUM=N THEN 400
380 NEXT COLOUR
390 REM'Zero the adjacency matrix
400 FOR I=1 TO N
410 FOR J=1 TO N
420 W(I,J)=0
430 NEXT J
```

```
440 NEXT I
450 PRINT
460 PRINT "Enter the objects to which each object is joined"
470 PRINT "To finish entering each object's joins, type 0 when asked"
480 PRINT "for next object.  Each join need be entered once only"
490 FOR I=1 TO N
500 PRINT
510 PRINT "Joins for object";I
520 INPUT "Enter next joined object or 0: ",J
530 IF J=0 THEN 650
540 REM'Check the column number is valid
550 IF J<0 OR J>N THEN 610
560 REM'Check the diagonal is not entered
570 IF J=I THEN 630
580 W(I,J)=1
590 W(J,I)=1
600 GOTO 520
610 PRINT "Error in column number"
620 GOTO 520
630 PRINT "An object cannot join itself"
640 GOTO 520
650 PRINT
660 PRINT "Joins entered for object";I;":"
670 FOR J=1 TO N
680 IF W(I,J)=0 THEN 700
690 PRINT J
700 NEXT J
710 PRINT "Enter 0 to continue this object, 1 to move to next"
720 INPUT "or 2 to delete a join: ",FLAG
730 IF FLAG=0 THEN 520
740 IF FLAG<>2 THEN 790
750 INPUT "Join to delete: ",J
760 W(I,J)=0
770 W(J,I)=0
780 GOTO 650
790 NEXT I
800 REM'Find various sums for the adjacency matrix
810 PRINT
820 A=0
830 B=0
840 FOR I=1 TO N
850 ROWSUM=0
860 FOR J=1 TO N
870 IF W(I,J)=0 THEN 890
880 ROWSUM=ROWSUM+1
890 NEXT J
900 B=B+ROWSUM↑2
910 A=A+ROWSUM
920 PRINT "Object";I;"has";ROWSUM;"joins"
930 NEXT I
940 A=A/2
950 PRINT
960 PRINT "Count of unique joins: ";A
970 S0=2*A
980 S1=4*A
990 S2=4*B
```



```
1000 REM'Print out these results
1010 PRINT
1020 PRINT "Intermediate statistics:"
1030 PRINT "S0: ";S0
1040 PRINT "S1: ";S1
1050 PRINT "S2: ";S2
1060 PRINT
1070 INPUT "Hit any key to continue: ",A$
1080 REM'Get a pair of colours
1090 PRINT
1100 PRINT "Enter a pair of colours for analysis or 0 to quit"
1110 INPUT "Number of first colour: ",R
1120 IF R=0  THEN 1770
1130 INPUT "Number of second colour: ",S
1140 IF R=S THEN 1170
1150 PRINT "Statistics used are for joins of different colours"
1160 GOTO 1180
1170 PRINT "Statistics used are for joins of same colour"
1180 REM'Count objects and joins
1190 NR=0
1200 NS=0
1210 NRS=0
1220 FOR I=1 TO N
1230 IF Z(I)<>R THEN 1250
1240 NR=NR+1
1250 IF Z(I)<>S THEN 1270
1260 NS=NS+1
1270 FOR J=1 TO N
1280 IF W(I,J)=0 THEN 1310
1290 IF Z(I)<>R OR Z(J)<>S THEN 1310
1300 NRS=NRS+1
1310 NEXT J
1320 NEXT I
1330 PR=NR/N
1340 PS=NS/N
1350 IF R<>S THEN 1370
1360 NRS=NRS/2
1370 PRINT
1380 PRINT "Number of joins found: ";NRS
1390 PRINT "Number of colour";R;"objects is";NR
1400 PRINT "Probability estimate for colour";R;"is";PR
1410 IF R=S THEN 1440
1420 PRINT "Number of colour";S;"objects is";NS
1430 PRINT "Probability estimate for colour";S;"is";PS
1440 REM'Calculate statistics
1450 IF R=S THEN 1540
1460 EN=S0*PR*PS
1470 ER=S0*NR*NS/(N*(N-1))
1480 VARN=(2*S1*PR*PS+(S2-2*S1)*PR*PS*(PR+PS)+4*(S1-S2)*PR↑2*PS↑2)/4
1490 VARR=2*S1*NR*NS/(N*(N-1))+(S2-2*S1)*NR*NS*(NR+NS-2)/(N*(N-1)*(N-2))
1500 VARR=VARR+4*(S0↑2+S1-S2)*NR*(NR-1)*NS*(NS-1)/(N*(N-1)*(N-2)*(N-3))
1510 VARR=VARR/4-ER↑2
1520 GOTO 1600
1530 PRINT
1540 EN=S0*PR↑2/2
1550 ER=S0*NR*(NR-1)/(2*N*(N-1))
```

```
1560 VARN=(S1*PR↑2+(S2-2*S1)*PR↑3+(S1-S2)*PR↑4)/4
1570 VARR=S1*NR*(NR-1)/(N*(N-1))+(S2-2*S1)*NR*(NR-1)*(NR-2)/(N*(N-1)*(N-
2))
1580 VARR=VARR+(SO↑2+S1-S2)*NR*(NR-1)*(NR-2)*(NR-3)/(N*(N-1)*(N-2)*(N-3))
1590 VARR=VARR/4-ER↑2
1600 PRINT
1610 PRINT "Expectations:"
1620 PRINT "Resampling: ";EN
1630 PRINT "Randomization: ";ER
1640 PRINT
1650 PRINT "Variances:"
1660 PRINT "Resampling: ";VARN
1670 PRINT "Randomization: ";VARR
1680 SN=(NRS-EN)/SQR(VARN)
1690 SR=(NRS-ER)/SQR(VARR)
1700 PRINT
1710 PRINT "Standard deviates:"
1720 PRINT "Resampling: ";SN
1730 PRINT "Randomization: ";SR
1740 PRINT
1750 INPUT "Hit any key to continue: ",A$
1760 GOTO 1090
1770 PRINT
1780 PRINT "Finished"
1790 END
```


35. The agricultural census - United Kingdom and United States - G. Clark
36. Order-neighbour analysis - G. Aplin
37. Classification using information statistics - R.J. Johnston & R.K. Semple
38. The modifiable areal unit problem - S. Openshaw
39. Survey research in underdeveloped countries - C.J. Dixon & B.E. Leach
40. Innovation diffusion: contemporary geographical approaches - G. Clark
41. Choice in field surveying - Roger P. Kirby
42. An introduction to likelihood analysis - Andrew Pickles
43. The UK census of population 1981 - J.C. Dewdney
44. Geography and humanism - J. Pickles
45. Voronoi (Thiessen) polygons - B.N. Boots
46. Goodness-of-fit statistics - A. S. Fotheringham & D. C. Knudsen
47. Spatial autocorrelation - Michael F. Goodchild
48. Introductory matrix algebra - Keith Tinkler
49. Spatial applications of exploratory data analysis - David Sibley

This series is produced by the Study Group in Quantitative methods, of the Institute of British Geographers.
For details of membership of the Study Group, write to the Institute of British Geographers, 1 Kensington Gore, London SW7 2AR, England.

The series is published by:

Geo Books,
Regency House,
34 Duke Street,
Norwich NR3 3AP,
UK,

to whom all other enquiries should be addressed.