Chapter 3 – The nature of geographic data

- § 1 Introduction
- § 2 The Fundamental Problems
- § 3 Spatial autocorrelation and scale
- § 4 Spatial sampling
- § 5 Distance decay
- § 6 Measuring distance effects as spatial autocorrelation
- § 7 Establishing dependence in space
- § 8 Conclusions

§ 1 Introduction

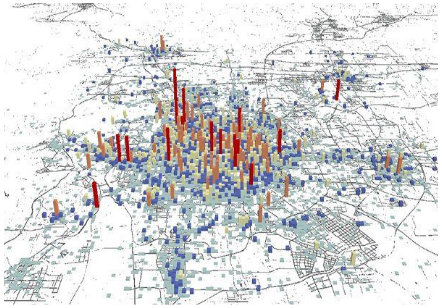
- Established three governing :
  - The representations we build in GIS are of unique *places*
  - Our representations of them are necessarily *selective* of reality, and hence incomplete
  - think of the world as either continuously varying *fields* or as an empty space littered with *objects* that are crisp and well-defined

§ 1 Introduction(cont.)

- In this chapter we build on these principles to develop a fuller understanding of the ways in which the nature of spatial variation is represented in GIS
- by asserting three further principles:
  - That *proximity* effects are key to understanding spatial variation, and to joining up incomplete representations of unique places
  - Issues of geographic *scale* and *level of detail (LOD)*are key to building appropriate representations of the world
  - Different measures of the world co-vary, and understanding the nature of *co-variation* can help us to predict

§ 2 The fundamental problem revisited

- Our behavior in geographic space often reflects past patterns of behavior
  - Variation: controlled(oscillate daily or weekly) vs uncontrolled
  - Temporal: ‘the past is the key to the present’
  - Spatial：The day-to-day operational context to our activities is very much determined by where we live and work. The longer-term strategic context may well be provided by where we were born, grew up, and went to college
- Explanation in time need only look to the past, but explanation in space must look in all directions
  - Time moves in one direction only (forward), making temporal autocorrelation one-dimensional,
  - Spatial events can potentially have consequences anywhere in two-dimensional or even three-dimensional space



§ 2 The fundamental problem revisited（cont.)

- vary smoothly across space vs exhibit extreme irregularity
  - Assessment of spatial autocorrelation can be informed by knowledge of the degree and nature of spatial *heterogeneity* – the tendency of geographic places and regions to be different from each other
  - Extreme difference of landscapes between such regions as the Antarctic, the Nile delta, the Sahara desert, or the Amazon basin
  - *Smoothness* and *irregularity* turn out to be among the most important distinguishing characteristics of geographic data
- Spatial autocorrelation helps us to build representations, but frustrates our efforts to predict
  - need to incorporate information on how two or more factors *co-vary*




§ 3 Spatial autocorrelation and scale

- Spatial autocorrelation is determined both by similarities in position, and by similarities in attributes

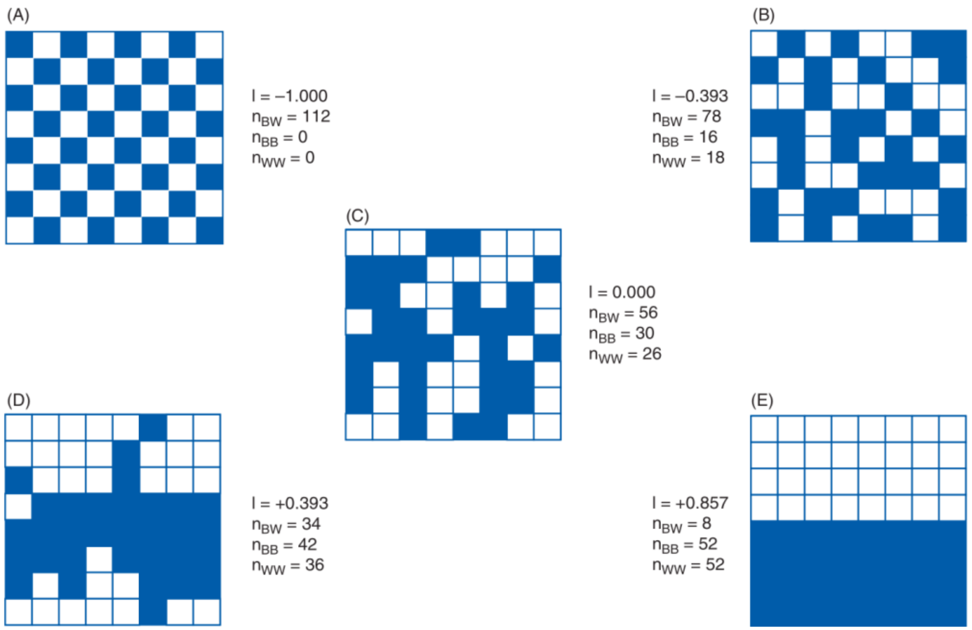


Figure 4.1 Field arrangements of blue and white cells exhibiting: (A) extreme negative spatial autocorrelation; (B) a dispersed arrangement; (C) spatial independence; (D) spatial clustering; and (E) extreme positive spatial autocorrelation. The values of the *I* statistic are calculated using the equation in Section 4.6

§ 3 Spatial autocorrelation and scale(cont.)

- The issue of sampling interval is of direct importance in the measurement of spatial auto-correlation, because spatial events and occurrences may or may not accommodate spatial structure
- In general, measures of spatial and temporal auto-correlation are scale dependent

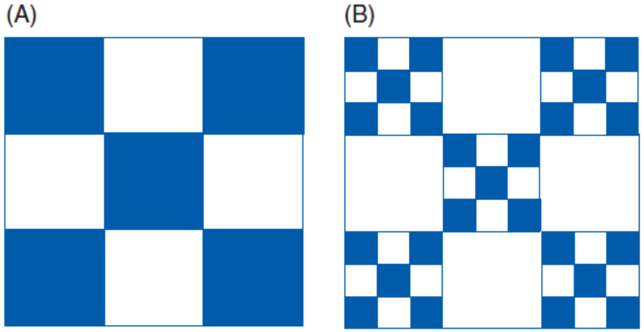
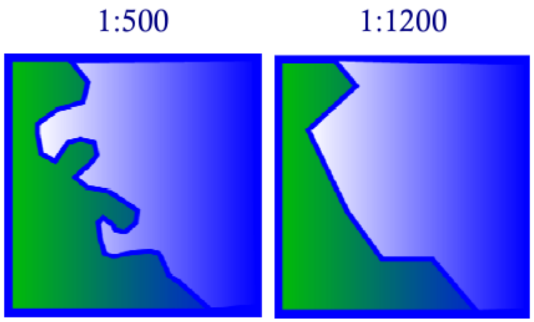
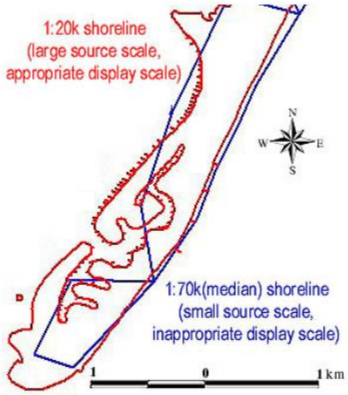


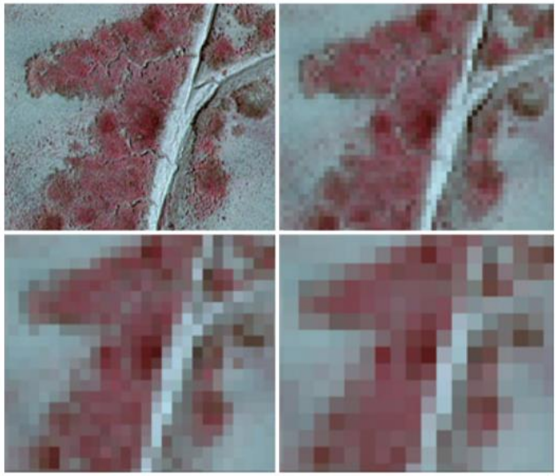
Figure 4.2 A Sierpinski carpet at two levels of resolution: (A) coarse scale and (B) finer scale

§ 3 Spatial autocorrelation and scale(cont.)

- Many meanings of *scale*
  - Scale of map, the ratio of distance on the map to distance on the ground
  - Level of detail, resolution, granularity:the size of pixel & the smallest resolvable unit
  - Extend, scope of place or project: the size of your study area & the largest resolvable unit, opposite to map scale（GE）

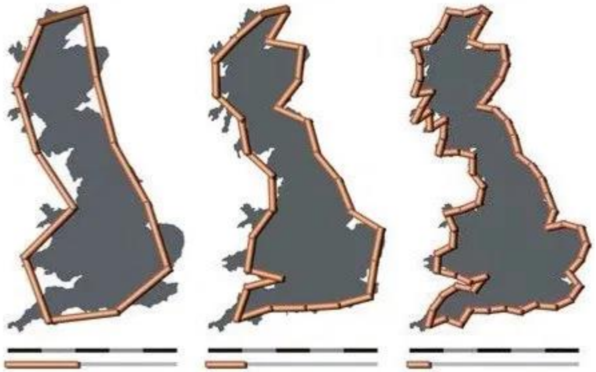


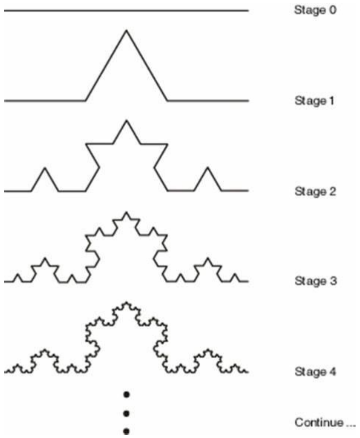




§ 3 Spatial autocorrelation and scale(cont.)

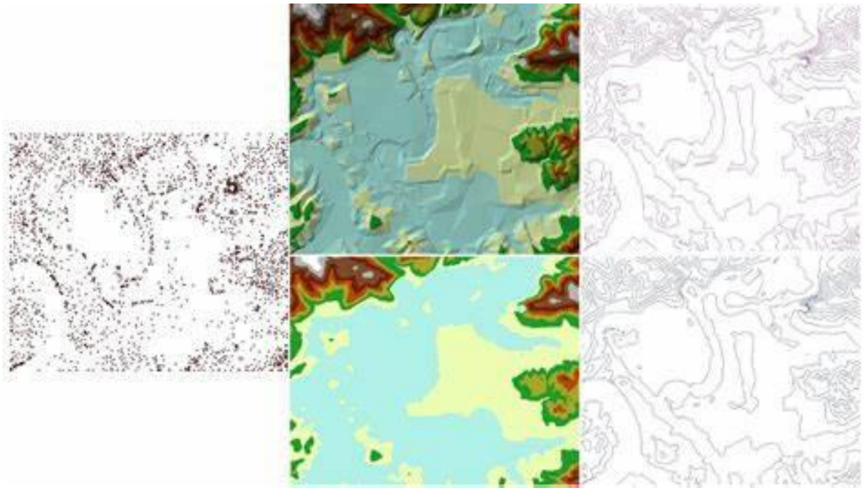
- A further important property is that of *self-similarity*
- the pattern is self-replicating at finer scales, a finer-scale representation reveals that the smallest component cells replicate the pattern of the whole area in a recursive manner
- Self-similarity is a core concept of fractals（GE-UK）





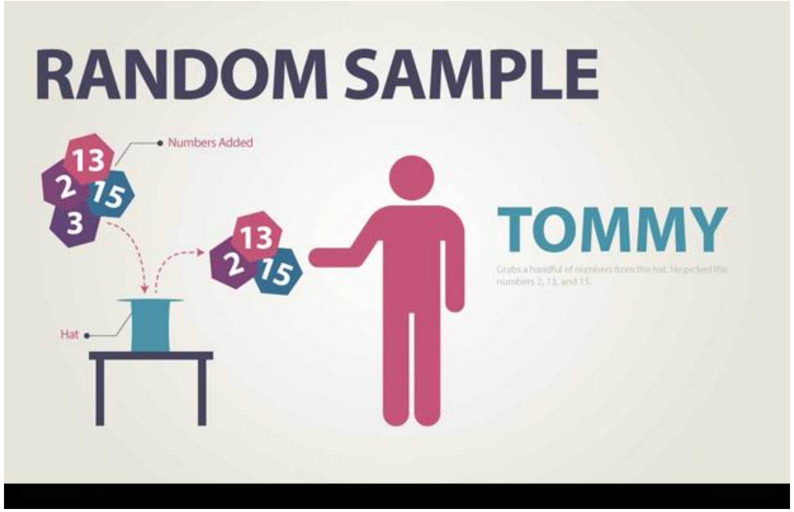
§ 4 Spatial sampling

- Any geographic representation as a kind of sample, in that the elements of reality that are retained are *abstracted* from the real world in accordance with some *overall design*
- Geographic data are only as good as the sampling *scheme* used to create them



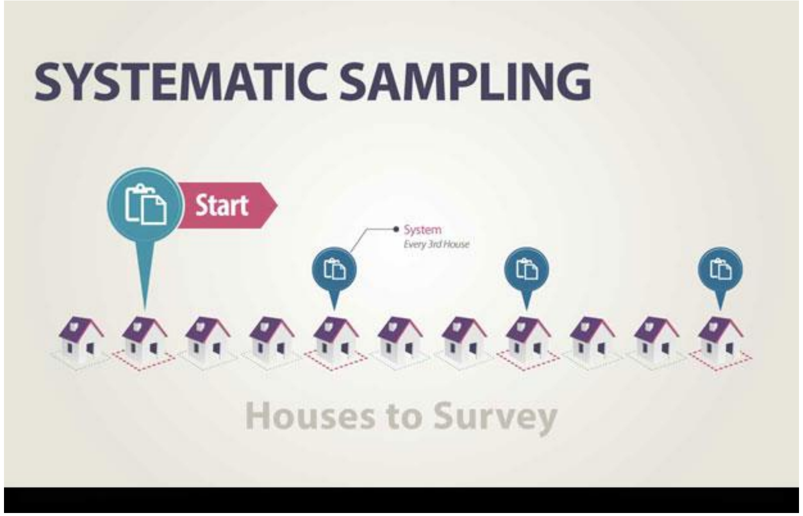
Simple Random Sampling

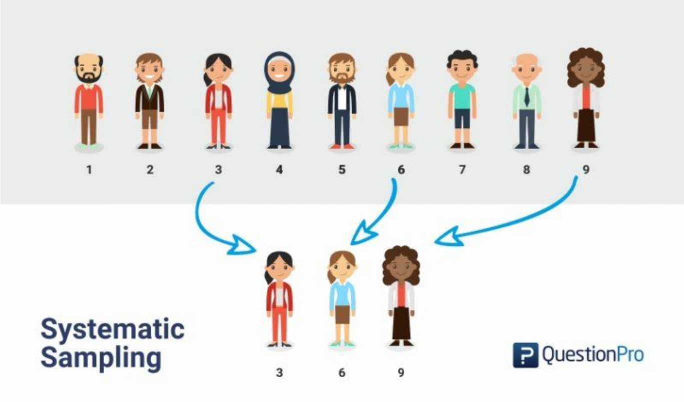
- Where we select a group of subject for study from a larger group(a population), each individual is chosen randomly and each member of the population has an equal chance of being include in the sample



Systematic Sampling

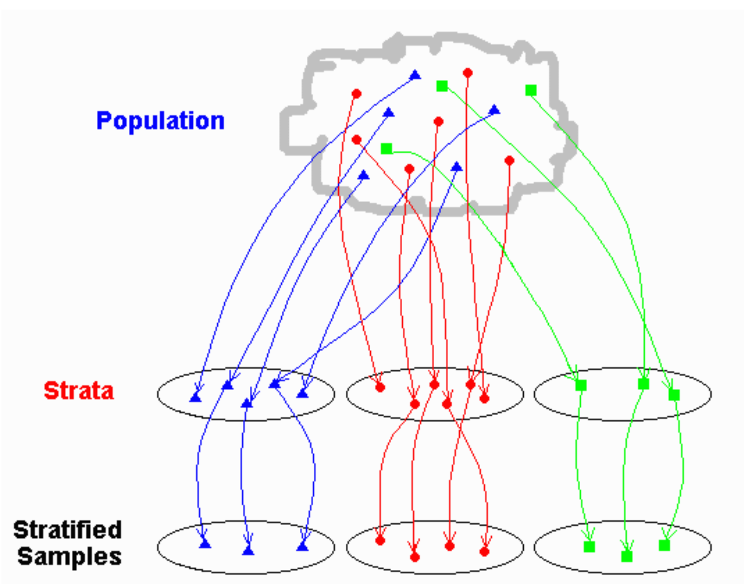
- Systematic sampling relies on arranging the target population according to some ordering scheme and then selecting elements at regular intervals through that ordered list





Stratified Sampling

- A stratified sample is a population sample that requires the population to be divided into smaller group, called ‘strata’
- Random sample can be taken from each stratum



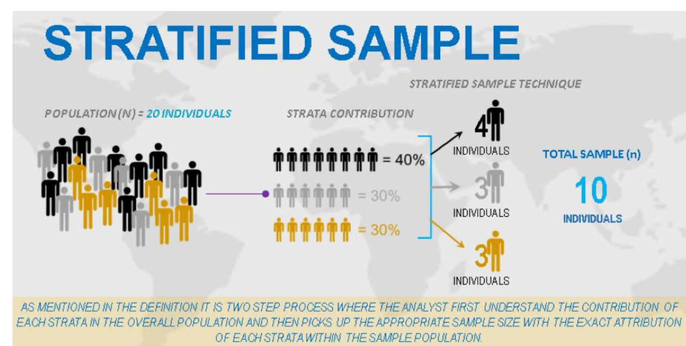
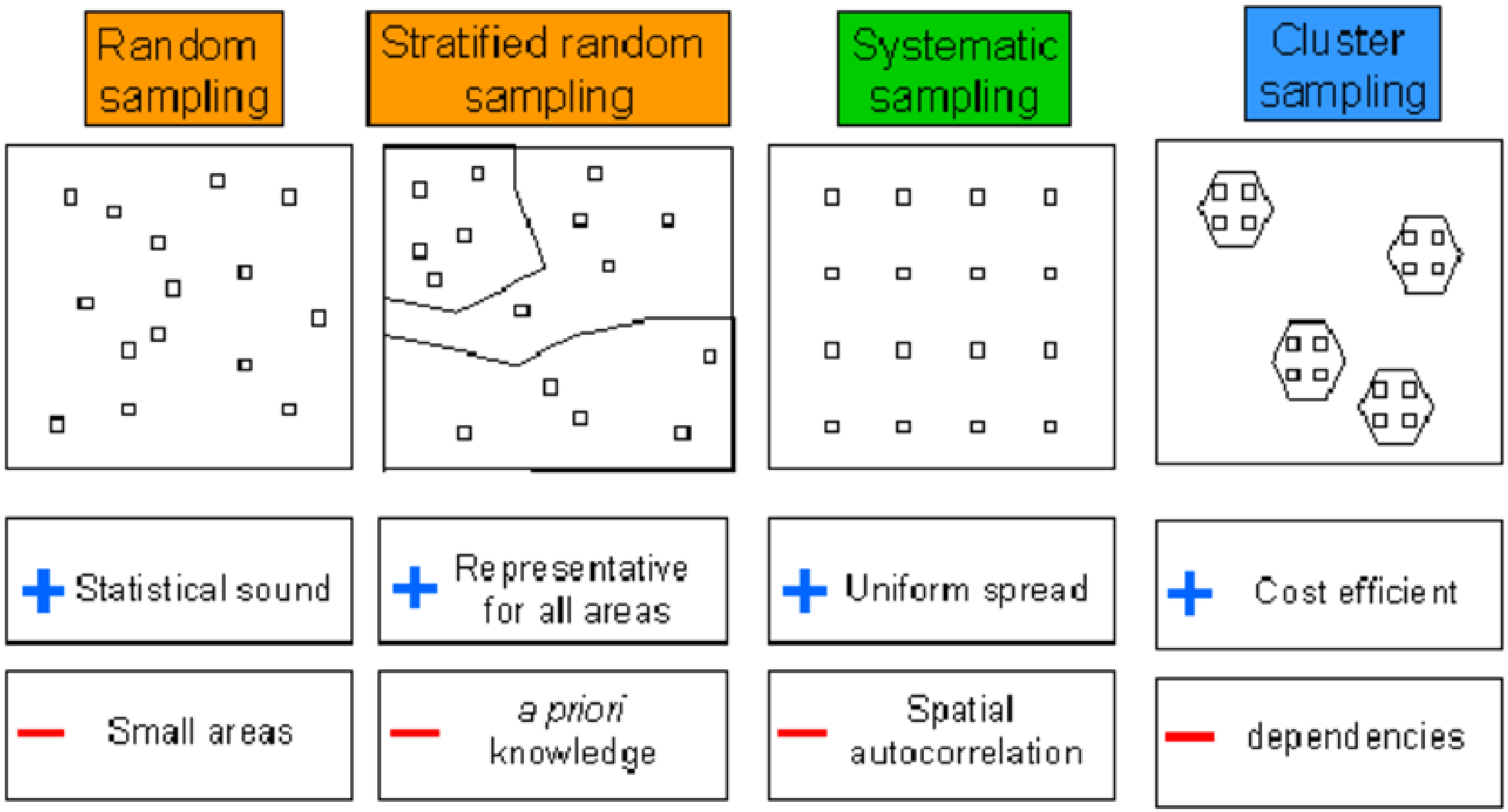


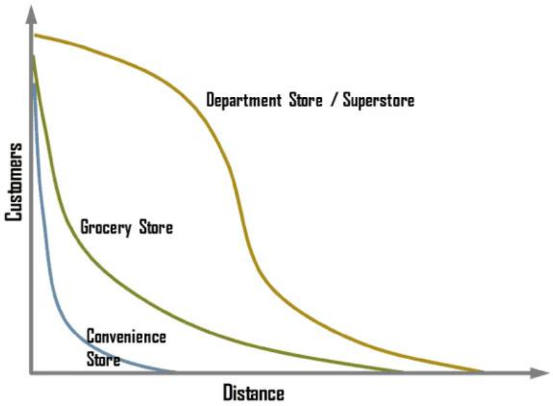
Figure: 1.12

Sampling scheme in GIS



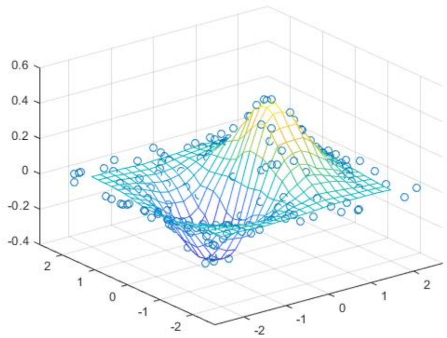
§ 5 Distance Decay

- Law of Spatial Interaction
- Distance and interaction are inversely proportional
- The shorter the distance the more likely interaction will occur
- The “friction of distance” increase with distance
- Relative distance – measured in time and cost of travel



§ 5 Distance Decay

- Interpolation: judgment is required to fill in the gaps between the observations Attenuating effect of distance between the sample observations
  - appropriate interpolation function
  - how to weight adjacent observations
- choice of function is the outcome of
  - past experience
  - fit of a particular application dataset
  - convention



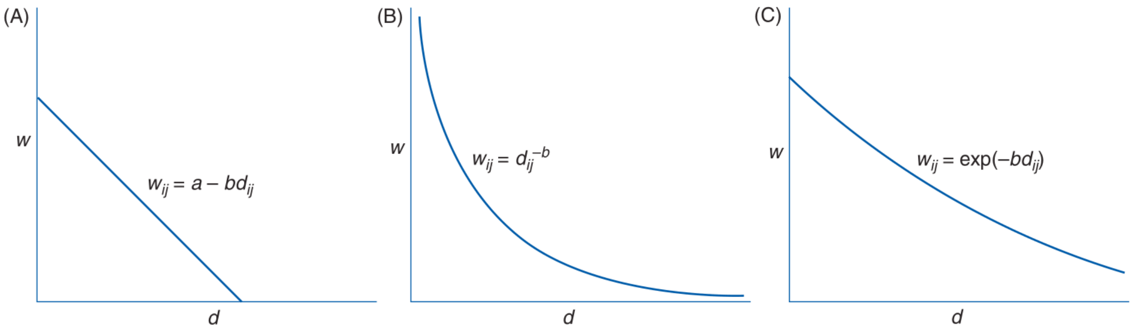


Figure 4.7 The attenuating effect of distance: (A) linear distance decay, $w_{ij} = a - bd_{ij}$; (B) negative power distance decay, $w_{ij} = d_{ij}{}^{-b}$; and (C) negative exponential distance decay, $w_{ij} = \exp(-bd_{ij})$

§ 6 Measuring distance effects as spatial autocorrelation

- An understanding of spatial structure helps us
  - to deduce a good *sampling strategy*
  - to use an appropriate means of *interpolating* between sampled points
  - to build a spatial *representation* that is fit for purpose
- in many applications we do not understand enough about geographic variability, distance effects, and spatial structure to invoke deductive reasoning.
- A further branch of spatial analysis thus emphasizes the measurement of spatial autocorrelation as an end in itself
  - This amounts to a more inductive approach to developing an understanding of the nature of a geographic dataset
- *Induction reasons from data to build up understanding, while deduction begins with theory and principle as a basis for looking at data*

§ 7 Establishing dependence in space

- In a formal statistical sense, regression analysis allows us to identify the dependence of one variable upon one or more independent variables
- The essential task of regression analysis is to identify the direction and strength of the association implied by this equation

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + \cdots + b_K X_{iK} + \varepsilon_i$$

- where b 1 through b K are termed regression *parameters*, which measure the direction and strength of the influence of the independent variables X 1 through X K on Y. b 0 is termed the *constant* or *intercept* term

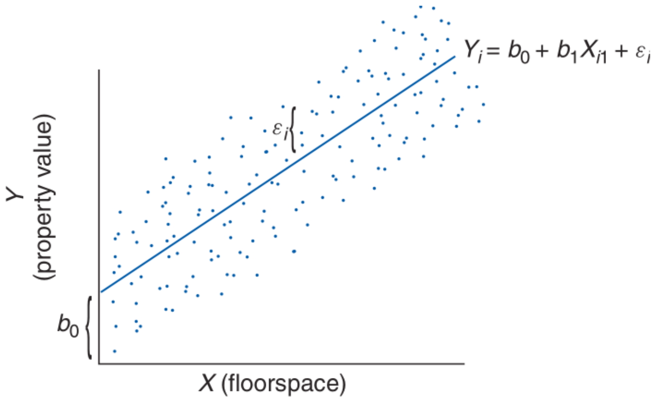


Figure 4.14 The fit of a regression line to a scatter of points, showing intercept, slope and error terms

§ 7 Establishing dependence in space(cont.)

- In a formal statistical sense, regression analysis allows us to identify the dependence of one variable upon one or more independent variables
- The essential task of regression analysis is to identify the direction and strength of the association implied by this equation

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + \cdots + b_K X_{iK} + \varepsilon_i$$

- where b 1 through b K are termed regression *parameters*, which measure the direction and strength of the influence of the independent variables X 1 through X K on Y. b 0 is termed the *constant* or *intercept* term

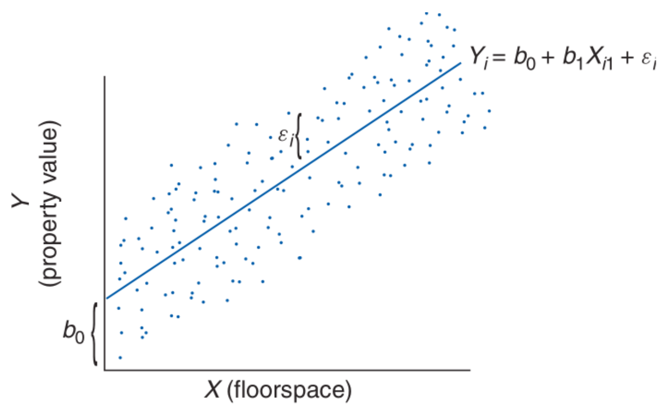


Figure 4.14 The fit of a regression line to a scatter of points, showing intercept, slope and error terms

§ 7 Establishing dependence in space(cont.)

- Regression analysis allows us to identify the dependence of one variable upon one or more independent variables
- The essential task of regression analysis is to identify the direction and strength of the association implied by this equation

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + \cdots + b_K X_{iK} + \varepsilon_i$$

- where b_1 through b_K are termed regression *parameters*, which measure the direction and strength of the influence of the independent variables X_{i1} through X_{iK} on Y. b_0 is termed the *constant* or *intercept* term

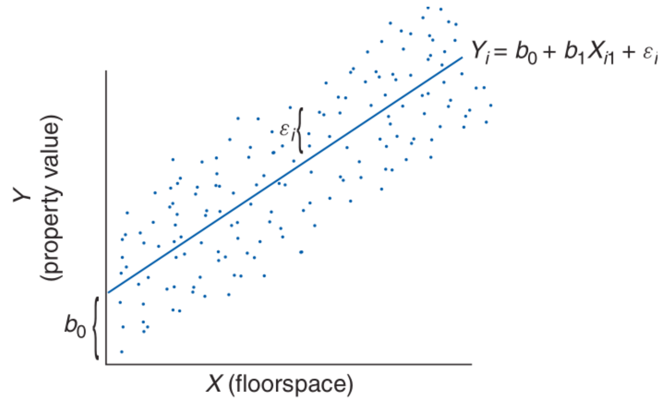


Figure 4.14 The fit of a regression line to a scatter of points, showing intercept, slope and error terms

§ 7 Establishing dependence in space(cont.)

- A straight is the easiest assumption to make and analyze, but it may not be the correct one
- Generalization is the process of reasoning from the nature of a sample to the nature of a larger group
- The assumption of zero spatial autocorrelation that is made by many methods of statistical inference is in direct contradiction to Tobler's Law
- Two maps of different phenomena over the same area would always reveal some similarities – how to keep no intercorrelation between the independent variables

c

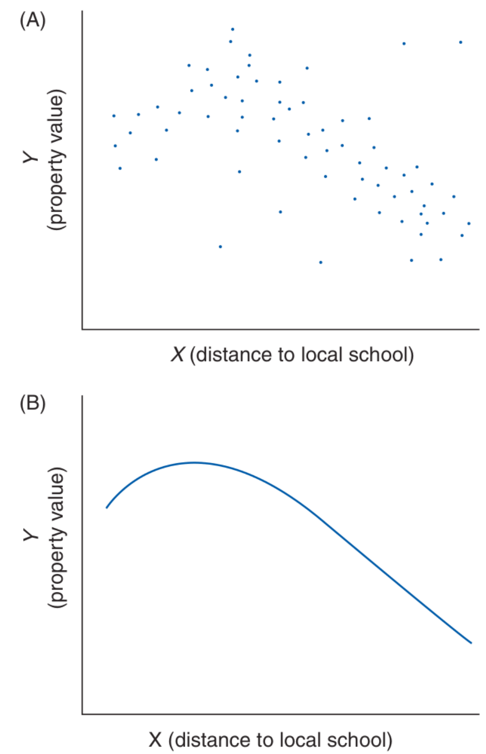


Figure 4.15 (A) A scatterplot and (B) hypothetical relationship between distance to local school and domestic property value

§ 8 Conclusions

- Spatial is special – that geographic data have a unique *nature*
- Tobler's Law presents an elementary general *rule* about spatial *structure*, and a starting point for the *measurement and simulation* of spatially autocorrelated structures
- Assists us in devising appropriate spatial *sampling schemes* and creating improved *representations*, which tell us still more about the real world and how we might represent it

§ 8 Conclusion(cont.)

- A goal of GIS is often to establish *causality* between different geographically referenced data, and the multiple *regression model* potentially provides one means of relating spatial variables to one another, and of inferring from samples to the properties of the populations from which they were drawn
- Statistical techniques often need to be *recast* in order to accommodate the special properties of spatial data, and regression analysis is no exception in this regard
- Spatial data provide the *foundations* to operational and strategic applications
- An understanding of the *nature* of spatial data allows us to use *induction* (reasoning from observations) and *deduction* (reasoning from principles and theory) alongside each other to develop effective spatial representations that are safe to use