SPATIAL POINT PATTERN ANALYSIS

1. Examples of Point Patterns

We begin by considering a range of point pattern examples that highlight the types of statistical analyses to be developed. These examples can be found in ARCMAP map documents that will be discussed later.

1.1 Clustering versus Dispersion

Consider the following two point patterns below. The first represents the locations of redwood seedlings in a section of forest.[1] This pattern of points obviously looks too

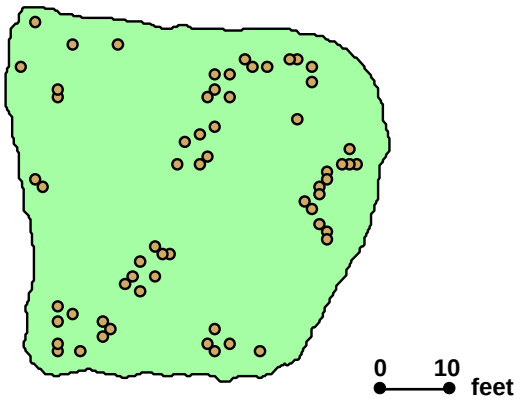

Fig.1.1. Redwood Seedlings

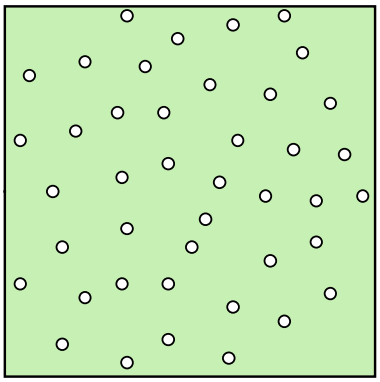

Fig.1.2. Redwood Seedlings

clustered to have occurred by chance. The second microscope slide.[2] While this pattern may look more random than the redwood seedlings, it is actually much too *dispersed* to have occurred by chance.[3] This can be seen a bit more clearly by including the cell walls, shown schematically in Figure 1.3 to the right. This additional information shows that indeed there is a natural spacing between these cells, much like the individual cells of a beehive. [The cell walls were actually constructed schematically in ARCMAP by using the "Voronoi Map" option in the *Geostatistical Analyst* extension of ARCMAP. But this process is a reasonable depiction of the actual

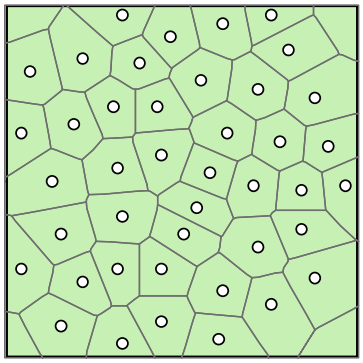

Fig.1.3. Cell Walls

[1] This data first appeared in Strauss (1975), and is the lower left-hand corner of his Figure 1 (which contains 199 redwood seedlings).
[2] This data first appeared in Ripley (1977), where it relates to an interesting biological problem regarding the process of cell division, posed by Dr. Francis Crick (of "Crick and Watson" fame).
[3] The term "dispersion" is sometimes called "uniformity" in the literature. Here we choose the former.

cell-packing process]. So the key question to be addressed here is how we can distinguish these patterns *statistically* in a manner that will allow us to conclude that the first is "clustered" and the second is "dispersed" – without knowing anything else about these patterns.

The approach adopted here is to begin by developing a statistical model of *purely random* point patterns, and then attempt to test each of these patterns against that statistical model. In this way, we will be able to conclude that the first is "significantly more clustered than random" and the second is "significantly more dispersed than random".

1.2 Comparisons between Point Patterns

Figures 1.4 and 1.5 below show the locations of abandoned houses in central Philadelphia for the year 2000.[4] The first shows those abandonments for which the owner's residence is off site, and the

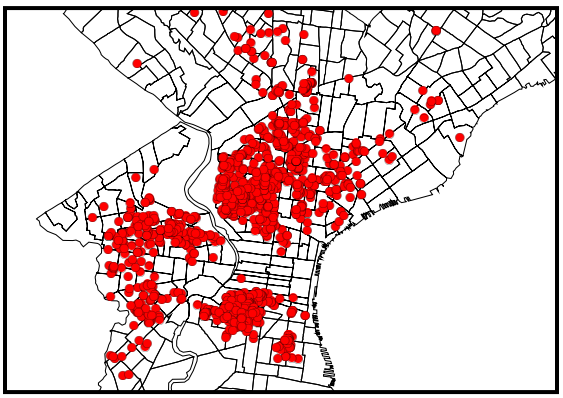

Fig.1.4. Off-Site Owners

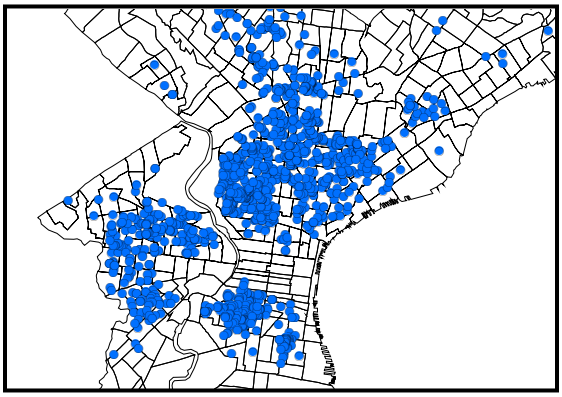

Fig.1.5. On-Site Owners

second shows properties for which the owner's residence is on site. If off-site ownership tends to reflect abandoned rental properties, while on-site ownership reflects abandoned residences, then one might hypothesize that different types of decisions were involved: abandoning a rental property might be more directly an economic decision than abandoning one's home. However, these patterns look strikingly similar. So one may ask whether there are any statistically significant differences between them.

Notice that there appears to be significant clustering in each pattern. But here it is important to emphasize that one can only make this judgment by comparing these

[4] This data was obtained from the *Neighborhood Information System* data base maintained by the Cartographic Modeling Lab here on campus, http://www.cml.upenn.edu/. For further discussion of this data see Hillier, Culhane, Smith and Tomlin (2003).

patterns with the *pattern of all housing* in this area. For example, there are surely very few houses in Fairmont Park, while there are many houses in other areas. So here it is important to treat the pattern of overall housing as the relevant *reference pattern* or "backcloth" against which to evaluate the significance of any apparent clusters of abandoned houses.

A second comparison of point patterns is given by an example from [BG] (p.80,129-132). This example involves a study of lung and larynx cancer cases in Lancashire county, England during the period 1974-1983.[5] The specific data set is from the south-central area of Lancashire county, shown by the red area in Figure 1.6. An enlargement of this region is shown in Figure 1.7 below, where the population of blue dots are *lung cancers* during that period, and the smaller population of red dots are *larynx cancers*. Here the smaller areal subdivisions shown are *parishes* [also called civil parishes (cp)] and correspond roughly in scale to our census tracts.

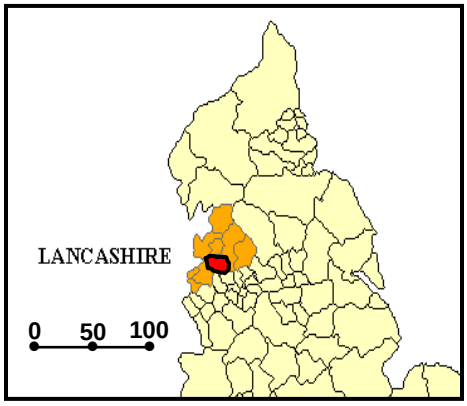

Fig.1.6. Lancashire County

Here again it should be clear that clustering of such cancer cases is only meaningful relative to the distribution of population in this area. The population densities in each parish are shown in Figure 1.8 below.

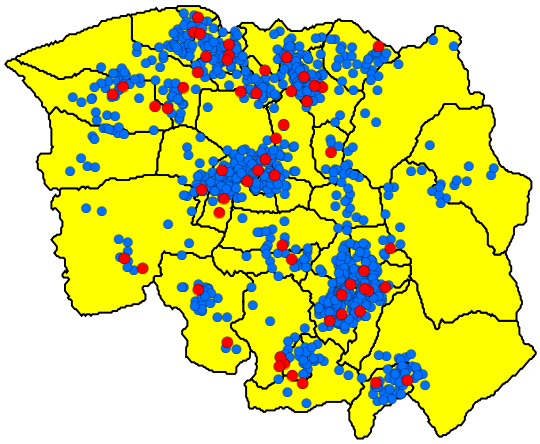

Fig.1.7. Larynx and Lung Cases

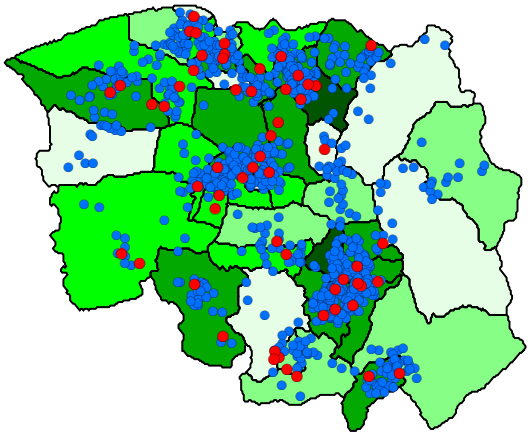

Fig.1.8. Population Backcloth

[5] This data first appeared in the paper by Diggle, Gatrell and Lovett (1990) which is included as Paper 12 "Larynx Cancer" in the Reference Materials on the class web page.

An examination of these population densities reveals that the clustering of cases in some of the lower central parishes is now much less surprising. But certain other clusters do not appear to be so easily explained. For example the central cluster in the far south appears to be in an area of relatively sparse population. This cluster was in fact the center of interest in this particular study. An enlargement of this southern portion in Figure 1.9 below indicates that a large incinerator[6] is located just upwind of this cluster of cases.[7]

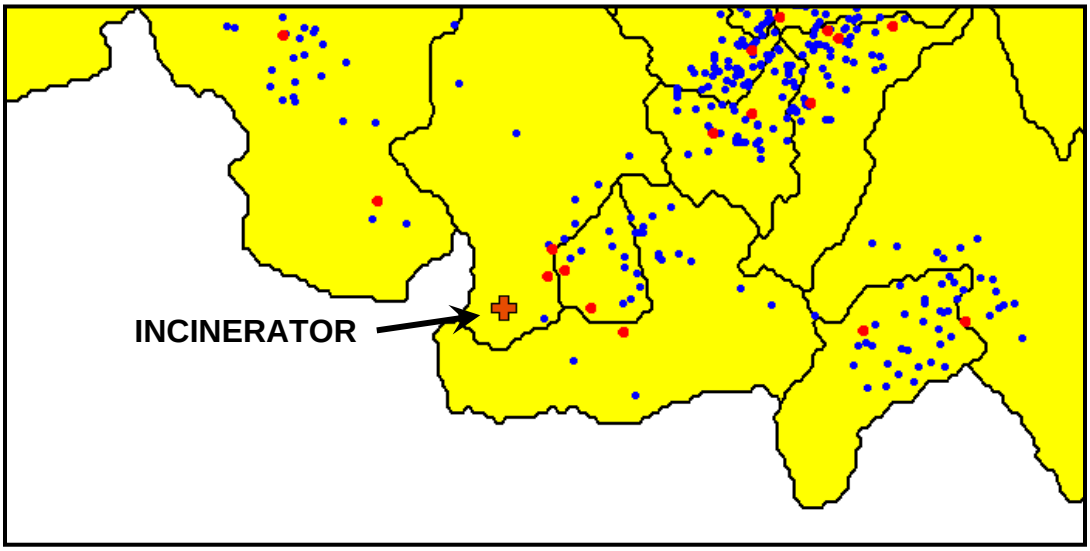

Fig.1.9. Incinerator Location

Moreover, an examination of the composition of this cluster suggests that there are significantly more *larynx* cases present than one would expect, given the total distribution of cases shown in Figures 1.7 and 1.8 above. This appears to be consistent with the fact that large airborne particles such as incinerator ash are more likely to lodge in the larynx rather than the lungs. So there is some suspicion that this incinerator may be a significant factor contributing to the presence of this particular clustering of cases.

To analyze this question statistically, one may ask how likely it is that this could simply be a coincidence. Here one must model the likelihood of such *local clustering* patterns.

[6] According to Diggle, Gatrell and Lovett (1990), this incinerator burned industrial wastes, and was active during the period from 1972-1980.
[7] Prevailing winds are from the Atlantic ocean to the west, as seen in Figure 1.6 above.