CSISS WORKSHOP

Introduction to Spatial Pattern Analysis in a GIS Environment

Introduction to Pattern Statistics:

The Role of Distance

Arthur Getis

An Introduction to Pattern Statistics

- Nearest Neighbors
  - The CSR hypothesis
  - Clark/Evans and modification
  - Cuzick and Edwards and controls
- All events
  - k function
  - Weighted k function
  - Comparative k functions

Nearest Neighbors

- The CSR Assumptions
  1. All possible sites are equally likely to receive a point

  2. The placement of a point is independent of the placement of all other points

- Quadrats or distances
- The Poisson Distribution

$$P(x) = \lambda^x e^{-\lambda} / x! \quad \text{For } x=0,1,2,...$$

Clark/Evans and Modification

- Distance-based
- Finds expected distance to nearest neighbor in a CSR pattern: [E(d)]
- $E(d) = 0.5\ [(A/N)]^{0.5} + [0.0514 + 0.041/\ (N)^{0.5}]\ B/N$

  and

  $\text{Var(mean d)} = 0.070\ A/N^2 + 0.037\ B\ [A/(N^5)]^{0.5}$
- $Z = [(\text{observed mean d}) - E(d)] / [\text{Var(mean d)}]^{0.5}$

  where A = area, N = total number of points, B = length of the perimeter

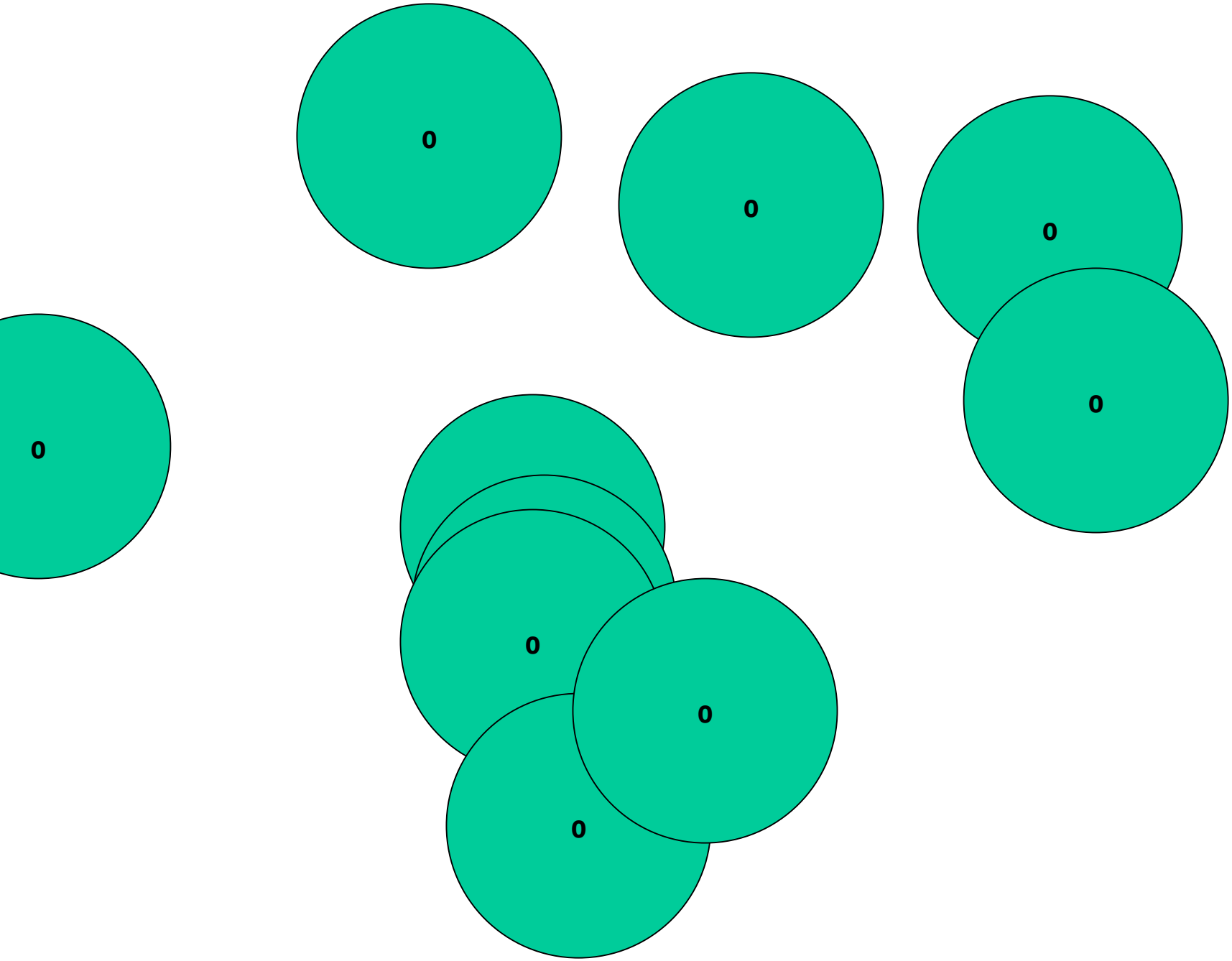
Cuzick and Edwards and Controls (k nearest neighbors)

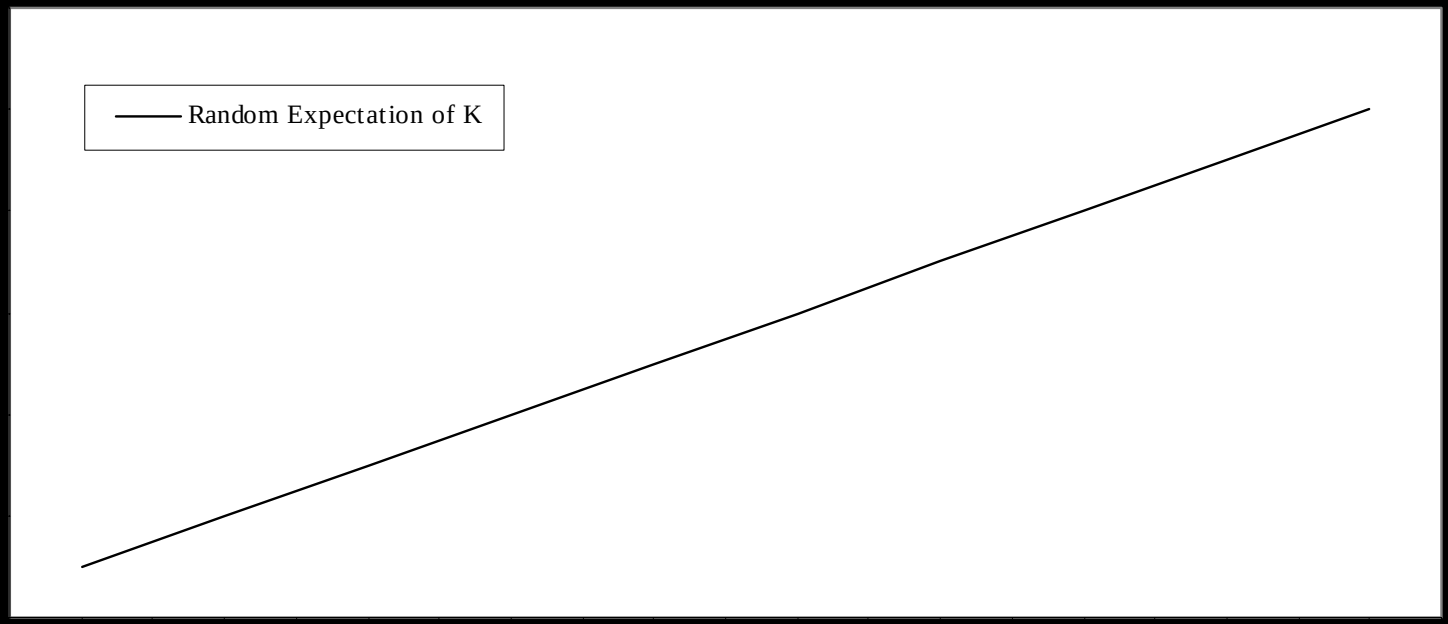
- A method for detecting spatial clustering for populations with non-uniform density.
- Label cases as x_i and controls as y_i
- Counts the number of cases (x_i) among the k nearest neighbors (x_i and y_i) to each case.
- Finds the theoretical distribution by permutation.
- Asymptotically normal. Provides test: the locations of the cases and controls follow a non-homogeneous Poisson process.

K Function Analysis: A Global Statistic

- $L(d) = \{(A[\Sigma\Sigma K(d_{ij})]/ \pi N(N-1)\}^{½}$

- where $K(d_{ij})$ is the number of pairs of points within d of *i*, and *A* is the area of the region under study.
- Used to discern the clustering pattern of the specified variable within the entire study area.
- An output file gives a table showing *L* values for each distance *(d)* increment. *E[L(d)]* in a random distribution is *d*.
- Used to give access to controls.


0
0
0
0
0
0
0
0


Random Expectation of K

K-Function

K-function is also called second-order analysis to indicate that the focus is on the variance, or second moment, of pairs of interevent distances. It considers all combinations of pairs of points. It compares the number of observed pairs with the expectation at all distances based on a random spatial distribution of points. The density of points, the borders, and the size of the sample are taken into consideration.

Input

1. The input data file, which should contain N rows of X, Y coordinates, and W values (a column of 1s).
2. The maximum distance that you want to use. The statistically unbiased maximum distance is less than the circumradius of the study area, or one-half of the length of the shortest side of a rectangular study area.
3. The number of increments.
4. The number of permutations for creating the confidence envelope.
5. The output file.

Analysis

K-function analysis is a test of the hypothesis of CSR. The expected value of L(d) is d. The confidence interval in this analysis is generated by examining the specified number of permutations of randomly generated patterns of N points over the whole study area. If for any distance, the observed L(d) falls above or below the expected L(d) the null hypothesis of CSR can be rejected at an appropriate level of significance. The level of significance is determined by the confidence envelope. An observed L(d) below the envelope indicates that the points are dispersed at that distance, whereas an observed above the envelope indicates that clustering is present at that distance.

Formula

$$L(d) = \sqrt{\frac{A\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} k(i,j)}{\pi N(N-1)}} \quad [1]$$

where:

A is the study area,

N is the number of points

Formula

$$L(d) = \sqrt{\frac{A\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} k(i,j)}{\pi N(N-1)}} \quad [1]$$

where:

A is the study area,

N is the number of points

d is the distance

$\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} k(i,j)$ is the number of j points within distance d of all i points

$k(i,j)$ is the weight, which is estimated by

a) If no edge corrections,

$k(i,j) = 1$ in case d(i,j) ≤ d

$k(i,j) = 0$ otherwise

b) If a point i is closer to one boundary than it is to point j, the border correction is employed

$$k(i,j) = \left[1 - \frac{\cos^{-1}\frac{e}{d(i,j)}}{\pi}\right]^{-1} \quad [2]$$

where e is the distance to the nearest edge.

K-function for *Aedes aegypti* pupae