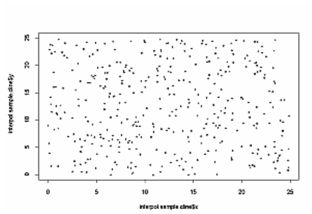
Nearest neighbor



- a form of point pattern analysis
- computational process involves the measurement of distances between points
- a coordinate system is created and the horizontal (X) coordinate and the vertical (Y) coordinates for the points are recorded
- for each point the nearest neighbor is determined
- distances can be derived as straight line Pythagoran form or measured directly

- once the observed distances are found we can compare the mean observed distance to a hypothesized distance
- clustered - the theoretical distance d_E is zero since the distance between each point and its nearest neighbour would be zero
- the general formula is: $$\bar{d}_E = \frac{\beta_1}{\sqrt{\frac{n}{A}}}$$

where β_1 is available from a table to follow

regular

$$\bar{d}_E = \frac{1}{\sqrt{\frac{n}{A}}}$$

Where n/A is the density

random

$$\bar{d}_E = \frac{0.5}{\sqrt{\frac{n}{A}}}$$

hexagonal lattice

$$\bar{d}_E = \frac{1.0746}{\sqrt{\frac{n}{A}}}$$

- the test is similar in form to a t test
  - the test statistic is

$$c = \frac{\bar{d}_O - \bar{d}_E}{SE_{\bar{d}}}$$

$\bar{d}_O$ is mean of observed nearest neighbour distances

$\bar{d}_E$ is expected mean of nearest neighbour distances for an arrangement

$SE_{\bar{d}}$ is the standard error of the mean nearest neighbour distances

- c is a normal standard deviate like Z
- In this case β_2=.26136 (see table) so

$$SE_{\bar{d}} = \frac{\beta_2}{\sqrt{\frac{n^2}{A}}}$$

$$SE_{\bar{d}} = \frac{0.26136}{\sqrt{\frac{n^2}{A}}}$$

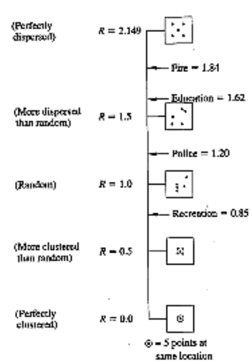
- the standard error is exactly analogous to the ordinary standard error of the mean, which is

$$\sigma_x = \frac{s}{\sqrt{n}}$$

- c is a standard normal deviate, like Z, so significance is determined by reference to the cumulative normal frequency table,
- so if α=.05, c_c=1.96
- direct comparison of results from different problems or different regions is difficult

- to overcome this there is a standardized nearest neighbor index R
- Where $\bar{d}_E$ is calculated for random situation

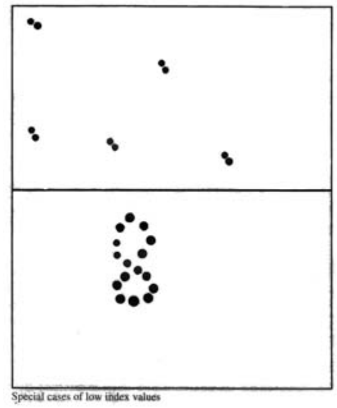
$$R = \frac{\bar{d}_O}{\bar{d}_E}$$



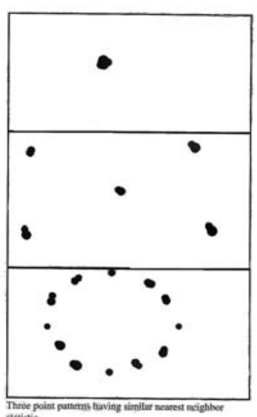
Continuum of R Values in Nearest Neighbor Analysis: Baltimore Services Displayed

problems

- the procedure as it stands suffers from serious drawbacks
- 1) measuring distance only to the closest nearest neighbor can result in observed mean distance values d_E that are not logically consistent
- to get around this problem the approach can be modified to take the average distance from k closest points
- k is called the order



Special cases of low index values



Three point patterns having similar nearest neighbor statistic

- for test statistics for randomness this table can be used

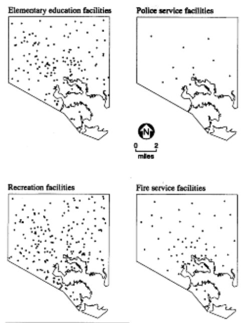
order		
	$ß_1$	$ß_2$
1	0.5000	0.2613
2	0.7500	0.2722
3	0.9375	0.2757
4	1.0937	0.2775
5	1.2305	0.2784
6	1.3535	0.2789

- 2) values of test statistic are affected by the size of the area used in the calculations
  - this is called the boundary problem
  - details of the solutions to this problem is beyond the context of this course
  - but there are 4 solutions

- a) if the surface is a rectangle or square make it into a *torus* (donut shaped) and then measure distances
- b) a *disregard strategy* - only use distances that are less than between point i and the boundary of the study area
- c) *buffer zone* - delimit the study area as a portion of a much larger area, measurements are only made to points within the study area
- d) use of *Donnelly's correction factors* in formulas - cannot be used for irregularly shaped areas

example

- four community services for Baltimore are shown in map



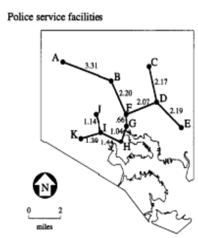
Location of Selected Public Facilities in Baltimore, Maryland (*Source*: Baltimore City Planning Commission, Dept. of Planning, 1972, 1990.)

- many public services should be dispersed relatively equally to ensure equal access by the populace
- One could ask: are the existing sites of the service providing such equal access?
- Lets look at police service provision
- First we need to find the nearest neighbors for the police stations

---

- Lets look at police service provision
- First we need to find the nearest neighbors for the police stations
- The mean nearest neighbor distance is: 1.63

$$\bar{d}_o = \frac{\sum dist}{n} = \frac{17.97}{11} = 1.63$$



Nearest Neighbor Distances for Police Service Facilities in Baltimore, Maryland

---

- calculate the random nearest neighbor distance:

$$\bar{d}_E = \frac{.50}{\sqrt{\frac{11}{80.86}}}$$

- area is given as 80.86,
- 0.26136 is from the table for the randomness test

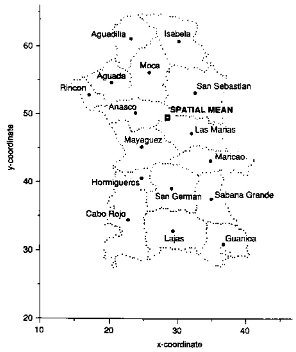
$$SE_{\bar{d}} = \frac{0.26136}{\sqrt{\frac{121}{80.06}}} = 0.21$$

- p=.1038

$$c = \frac{\bar{d}_o - \bar{d}_E}{SE_{\bar{d}}} = \frac{1.63 - 1.36}{.21} = 1.26$$

Example

- Determine if milk production in Puerto Rico is randomly distributed. Use the following map and table as your input.



Municipio	Area	X	Y	neighbour	distance
Aguada	30.21	20.0314	54.6399	Rincon	3.44
Aguadilla	35.57	23.1881	61.1065	Moca	5.74
Anasco	40.05	23.8571	50.1807	Mayaguez	5.18
Cabo Rojo	72.35	22.6322	34.4176	Hormigueros	6.36
Guanica	36.52	36.6477	30.8942	Sabana Grande	6.76
Hormigueros	11.16	24.7229	40.4245	Mayaguez	4.67
Isabela	55.47	29.103	60.8053	Moca	6.38
Lajas	60.23	29.4894	32.815	San German	6.23
Las Marias	47.03	32.1001	47.2462	Maricao	5.02
Maricao	36.85	34.9884	43.1383	Las Marias	5.02
Mayaguez	56.95	24.8361	45.0923	Hormigueros	4.67
Moca	50.43	25.8709	56.0336	Aguadilla	5.74
Rincon	14.14	17.0815	52.8758	Aguada	3.43
Sabana Grande	35.24	35.0721	37.4654	Maricao	5.67
San German	53.9	29.2389	39.0352	Hormigueros	4.72
San Sebastian	70.8	32.6112	53.1334	Las Marias	5.91
	706.9				84.94

- Test for randomness

$$\bar{d}_E = \frac{.50}{\sqrt{16/706.9}} = 3.3234 \qquad \bar{d}_o = \frac{84.94}{16} = 5.3088$$

$$SE_{\bar{d}} = \frac{.26136}{\sqrt{16^2/706.9}} = .4343 \qquad c = \frac{5.3088 - 3.3234}{0.4343} = 4.5715$$

- Is it random?