# 空間自相關
# Spatial Autocorrelation

授課教師：溫在弘
E-mail: wenthung@ntu.edu.tw
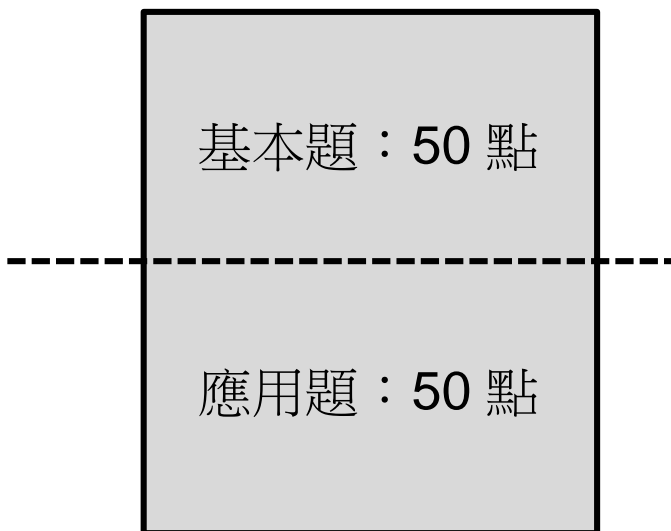
# 期末考(Take-home Exam)的計算方式

## Self-learning track
### (總分：50 分)

基本題

答對率：40 分以上
學期等第：C-

## Regular track
### (總分：100 點)

基本題：50 點

- - - - - - - - - - - - - - - - - - - - -

應用題：50 點

基本題答對率：35 點(含) ～ 40 點 → 學期等第：C-

基本題答對率：41 點(含) ～ 44 點 → 學期等第：C

基本題答對率：45 點(含) 以上　→ 學期等第：C＋

# 期末考分數的點數轉換方式

**考卷：100 點**

**全班總分：5000 分**

$$原始分數 = 獲得點數 \times \frac{5000}{全班總點數}$$

舉例：

某生獲得 60 點、全班總點數：4000點

→5000 分 / 4000 點 ＝ 1.25 分 / 點

→該生的(原始)分數 60 x 1.25 ＝ 75

→依照第一週公告的分數調整規則（第6高分=90分），進行調整計畫

# 本週課程：空間自相關的觀念與計算

## Moran's I coefficient

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

$N$: no. of spatial units

$w_{i,j}$ : a matrix of spatial weights

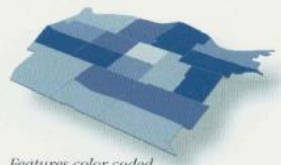$W = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}$ (sum of all $w_{i,j}$ )

# 教科書的重點研讀章節內容

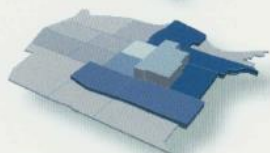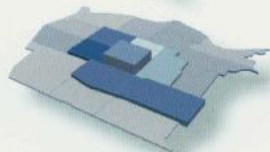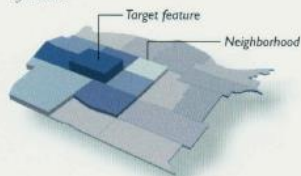TEXT_Spatial_Weights.pdf          TEXT_Pattern.of.Feature.Value.pdf



Mitchell (2005), The ESRI Guide to GIS Analysis, Volume 2: Spatial Measurements and Statistics, ESRI Press

# Global and Local Measures

- ***Global* Measures**

  - ❑ A single value which applies to the entire data set

    - The same pattern occurs over the entire geographic area

    - An average for the entire area

- ***Local* Measures**

  - ❑ A  value calculated for <u>each</u> observation unit

    - Different patterns or processes may occur in different parts of the region

    - A unique number for each location

# Global Analysis Methods 全域分析的方法

- **Point data *without* attributes**

  - ❑ Quantrat Analysis

  - ❑ Nearest Neighbor Methods

    - K-order Nearest Neighborhood Analysis (NNA), G and F functions

  - ❑ Ripley's K-function: K(d) and L(d)

- **Point/Polygon data *with* attributes**

  - ❑ Definition of Neighborhoods or Spatial Structures

  - ❑ Spatial Autocorrelation Index

    - **Moran's I** and Geary's C Ratio

  - ❑ Spatial Concentration Index

    - **General G-statistic**
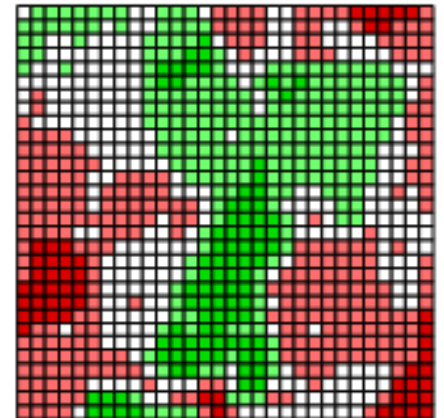
# 空間關連性與相依性
# Spatial Relationship and Dependency

Tobler's *First Law of Geography* (1970)：

Everything is related to everything else,

but near things are more related than distant things.

Dispersed ⟵———————————————⟶ Clustered

(-) ← Spatial Autocorrelation → (+)

Tobler W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2): 234-240.
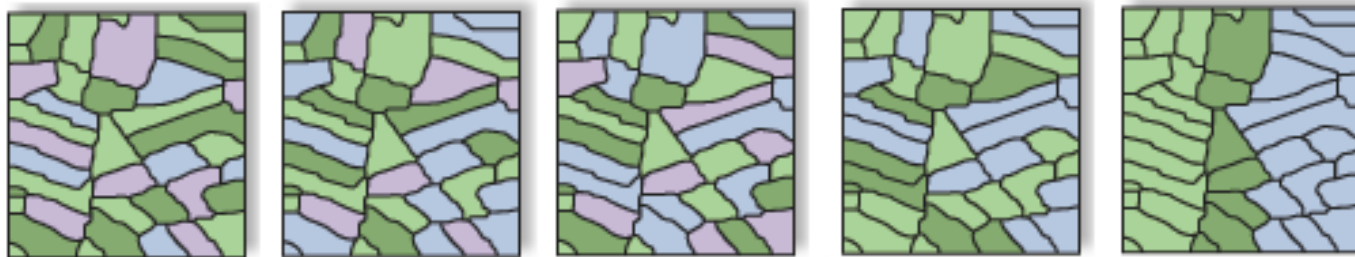
# 無所不在的空間自相關（空間相依的特性）

2000年
人口密度
(人數/平方公里)
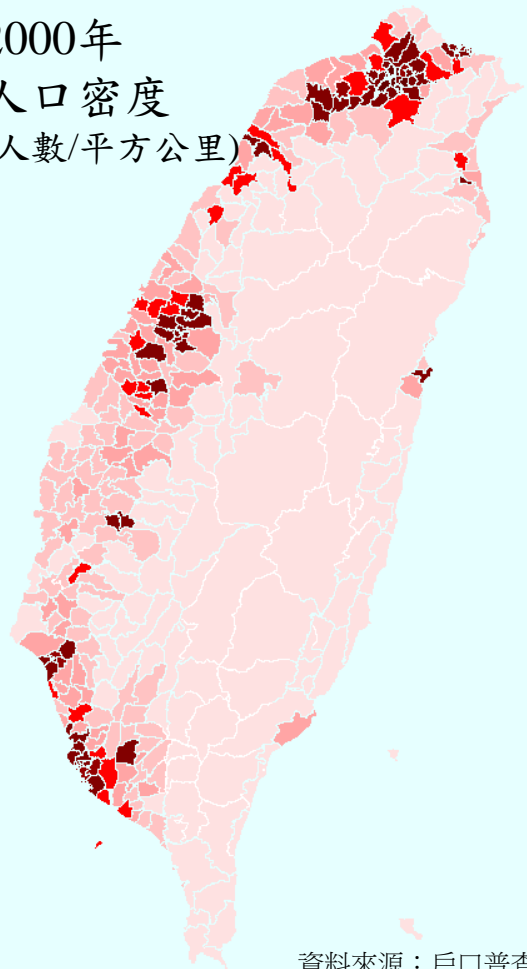
2003年
人均所得
(千元/人)

2004年民進黨
得票率 (%)

資料來源：戶口普查(主計處)
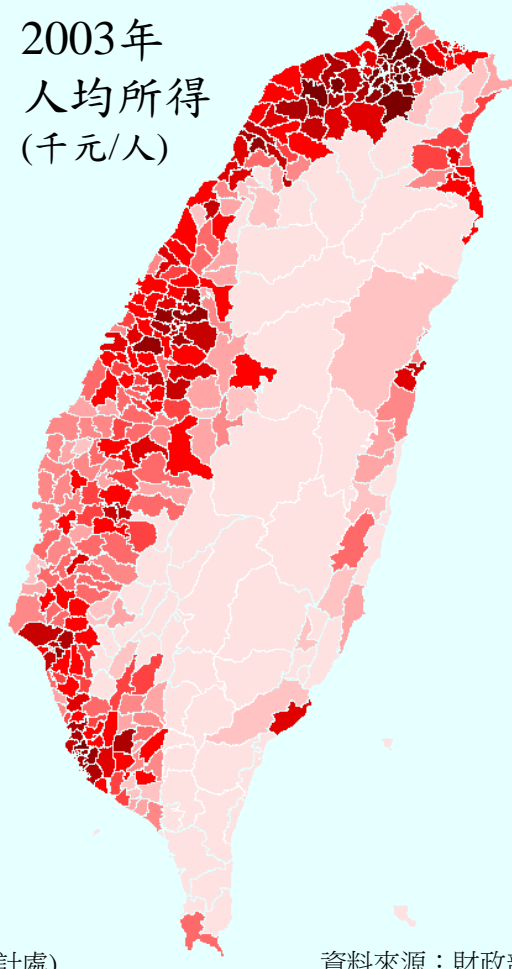
資料來源：財政部
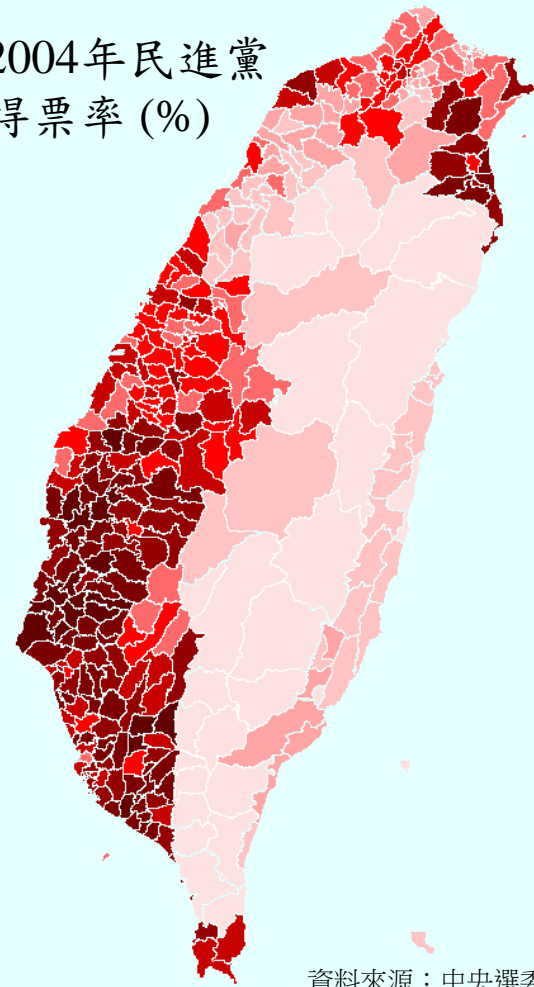
資料來源：中央選委會

# Definition of Spatial Autocorrelation

■ Measurement of the *similarity of attributes* among spatial units within their *neighborhood.*

(-) ← *Spatial Autocorrelation* → (+)
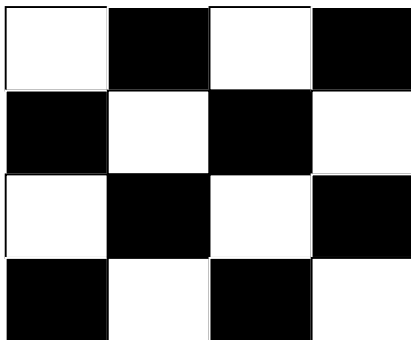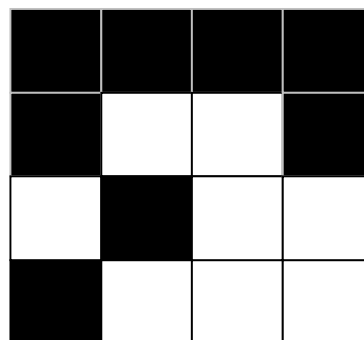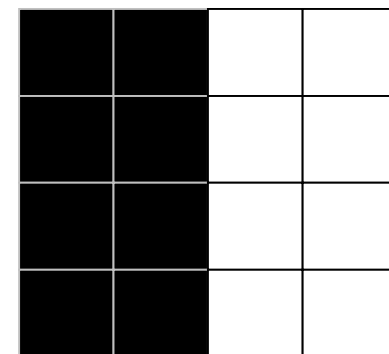


Dispersed            Random            Clustered

# How we define the Neighborhood ?

1. Spatial adjacency

   ❑ Physically contacted with each others

2. Distances between the centroids

# 1. Spatial Adjacency

■ **Rook's**

❑ Units that shares common boundary with length greater than zero

■ **Queen's**

❑ Units that have common vertex are also included

(e.g. unit 067 in figure below)

*Rook's*

| A | B | C |
|---|---|---|
| D | X | E |
| F | G | H |

*Queen's*

| A | B | C |
|---|---|---|
| D | X | E |
| F | G | H |
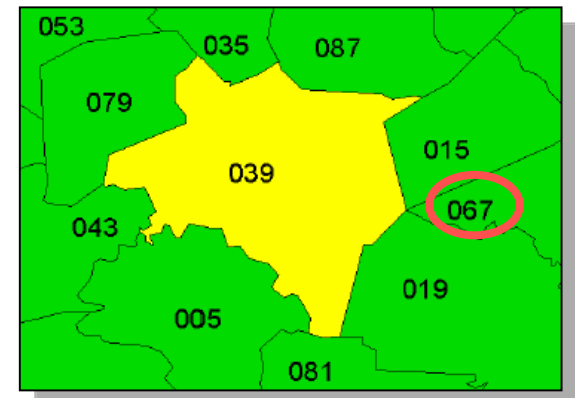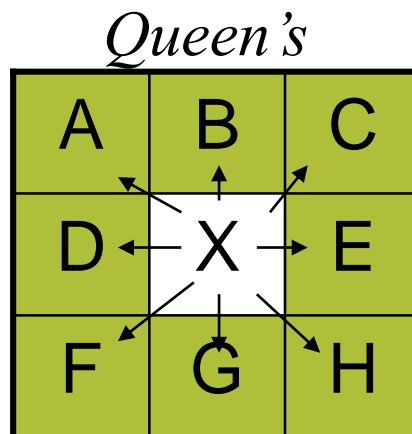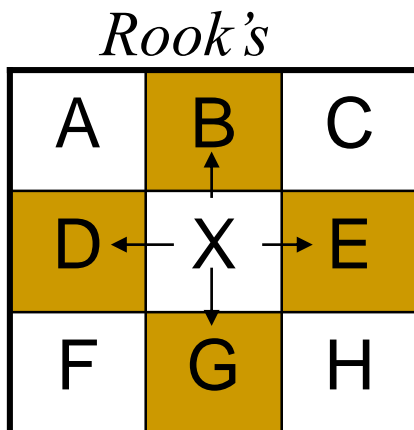
# Order of Neighborhood

- 1st order

  - Immediate neighbor

  - Defined by Rook's or Queen's criteria

- 2nd order

- Higher order
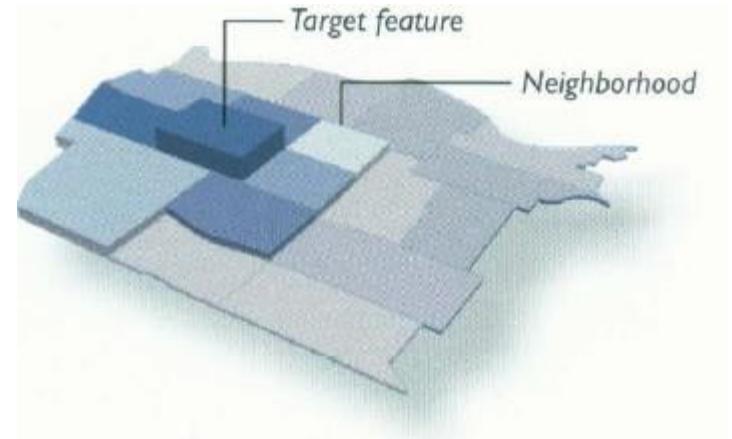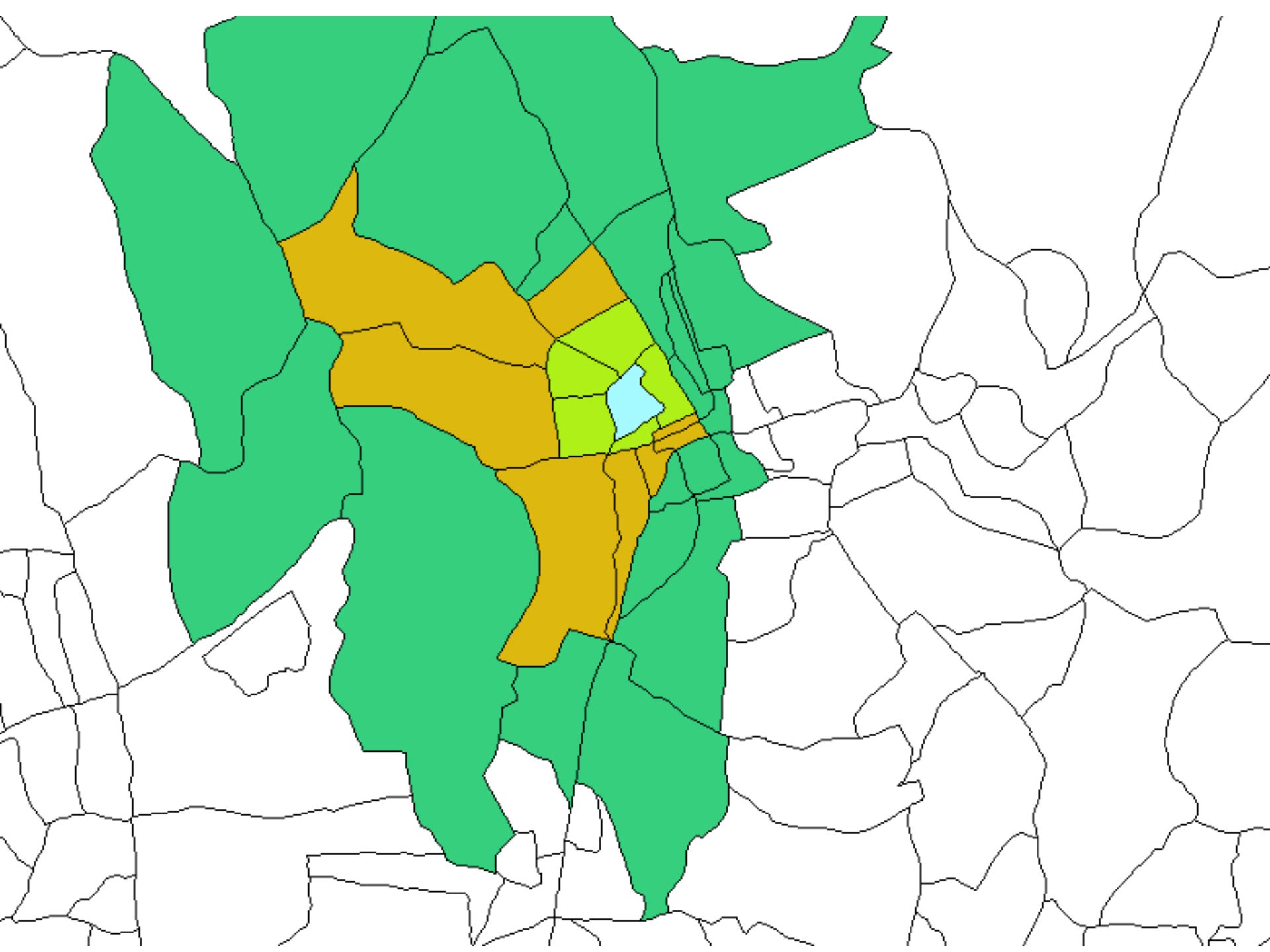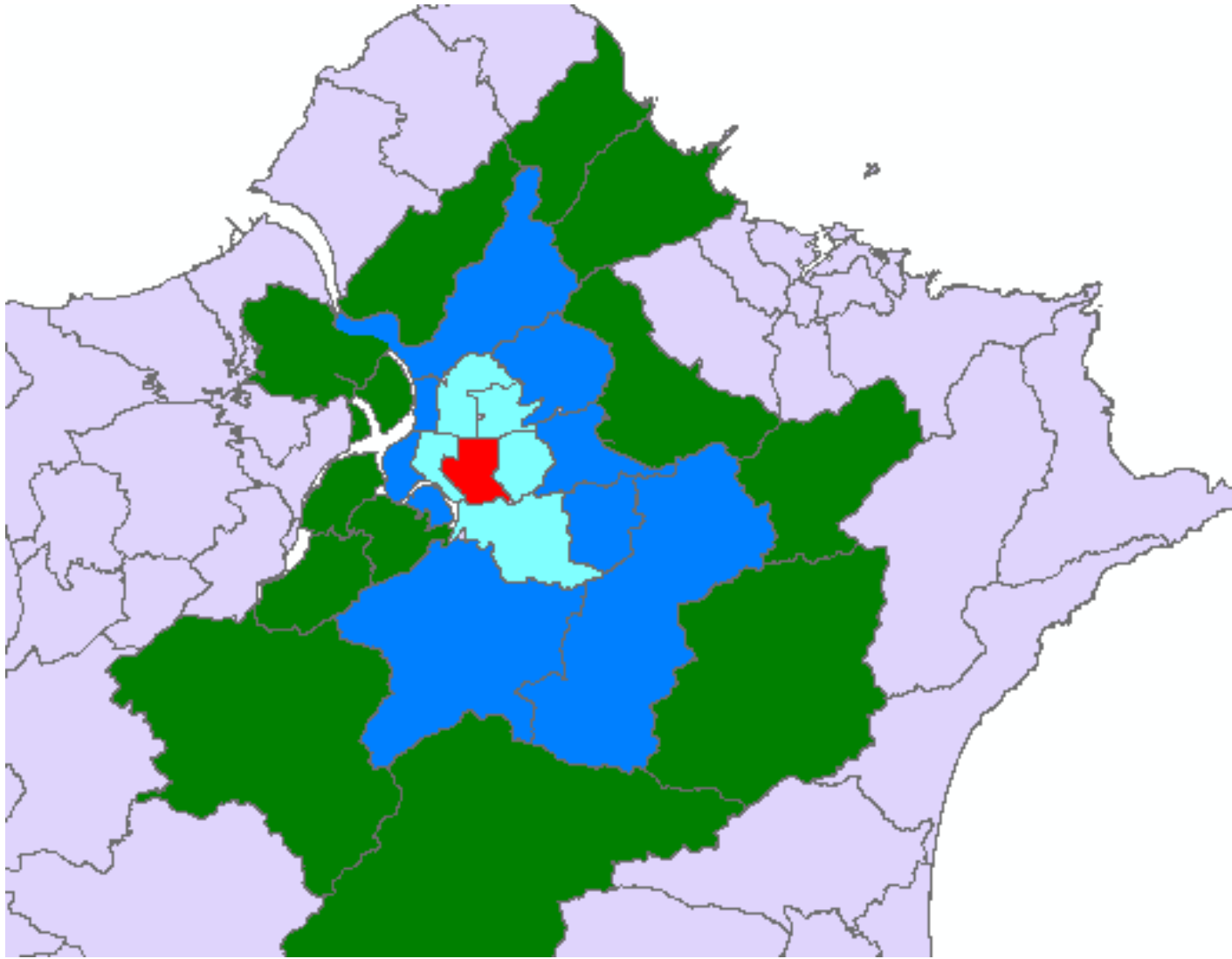
# Binary Connectivity Matrix

- Symmetrical     $C_{ij} = C_{ji}$

- Values on diagonal are zeros

- Raw sum  $C_i = \Sigma C_{ij}$

  - The number of  neighbors of unit I


- Same as connectivity matrix for network

- Not efficient for large numbers of objects

  - Redundant storage

  - Mostly zeros

  - Another way : Sparse Matrix (稀疏矩陣)

# Sparse matrix

| ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | Sum |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 長安里 | 文化里 | 中心里 | 中庸里 | 豐年里 | 溫泉里 | 大同里 | 清江里 | | | | | | | | 7 |
| 中心里 | 泉源里 | 開明里 | 林泉里 | 中庸里 | 長名里 | 溫泉里 | | | | | | | | | 6 |
| 中庸里 | 中和里 | 開明里 | 文化里 | 智仁里 | 中心里 | 長名里 | | | | | | | | | 6 |
| 開明里 | 中和里 | 泉源里 | 智仁里 | 中心里 | 中庸里 | | | | | | | | | | 5 |
| 文化里 | 智仁里 | 中庸里 | 豐年里 | 長名里 | 大同里 | | | | | | | | | | 5 |
| 大同里 | 文化里 | 豐年里 | 長名里 | 清江里 | 中央里 | | | | | | | | | | 5 |
| 中和里 | 泉源里 | 開明里 | 智仁里 | 中庸里 | | | | | | | | | | | 4 |
| 泉源里 | 中和里 | 開明里 | 林泉里 | 中心里 | | | | | | | | | | | 4 |
| 智仁里 | 中和里 | 開明里 | 文化里 | 中庸里 | | | | | | | | | | | 4 |
| 溫泉里 | 林泉里 | 中心里 | 長名里 | 清江里 | | | | | | | | | | | 4 |
| 清江里 | 長名里 | 溫泉里 | 大同里 | 中央里 | | | | | | | | | | | 4 |
| 林泉里 | 泉源里 | 中心里 | 溫泉里 | | | | | | | | | | | | 3 |
| 豐年里 | 文化里 | 大同里 | | | | | | | | | | | | | 3 |
| 中央里 | 大同里 | 清江里 | | | | | | | | | | | | | 2 |

# Binary Connectivity Matrix

| ID | 中和里 | 泉源里 | 開明里 | 文化里 | 智仁里 | 林泉里 | 中心里 | 中庸里 | 豐年里 | 長名里 | 溫泉里 | 大同里 | 清江里 | 中央里 | Sum |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 中和里 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| 泉源里 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| 開明里 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| 文化里 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 5 |
| 智仁里 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| 林泉里 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 3 |
| 中心里 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 6 |
| 中庸里 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 6 |
| 豐年里 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 3 |
| 長安里 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 7 |
| 溫泉里 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 4 |
| 大同里 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 5 |
| 清江里 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 4 |
| 中央里 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 2 |

# Sparse matrix 稀疏矩陣的儲存方式

Index

| | | | | | |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 3 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 6 | 0 | 0 |
| 0 | 0 | 9 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 12 | 0 |

Sparse Matrix

| 5 | 6 | 4 |
|---|---|---|
| 1 | 1 | 3 |
| 2 | 3 | 6 |
| 3 | 2 | 9 |
| 4 | 4 | 12 |

Line #1: 這個矩陣是5X6矩陣，非零元素有4個

Line #2：記錄其位置的列索引、行索引與儲存值

# Stochastic Matrix

- Equally weighted for neighbors

  - $W_{ij} = C_{ij} / C_i$

- 又稱做 Row-standardized matrix (列標準化矩陣)

  - 考慮每一個相鄰的object 的影響量

# Binary Connectivity Matrix

| ID | 中和里 | 泉源里 | 開明里 | 文化里 | 智仁里 | 林泉里 | 中心里 | 中庸里 | 豐年里 | 長名里 | 溫泉里 | 大同里 | 清江里 | 中央里 | Sum |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 中和里 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| 泉源里 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| 開明里 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| 文化里 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 5 |
| 智仁里 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| 林泉里 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 3 |
| 中心里 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 6 |
| 中庸里 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 6 |
| 豐年里 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 3 |
| 長安里 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 7 |
| 溫泉里 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 4 |
| 大同里 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 5 |
| 清江里 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 4 |
| 中央里 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 2 |

# Stochastic Weighted Matrix

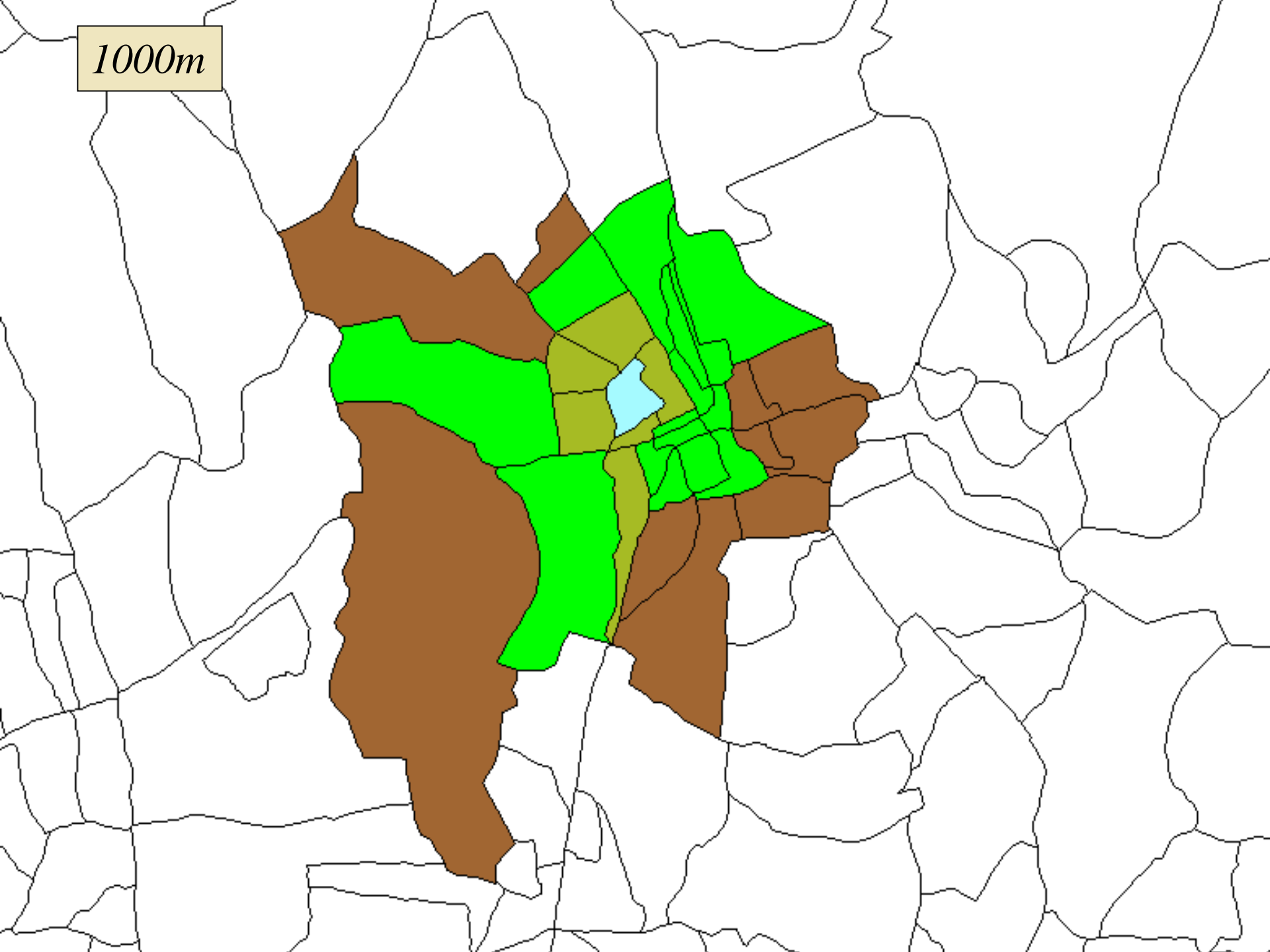| ID | 中和里 | 泉源里 | 開明里 | 文化里 | 智仁里 | 林泉里 | 中心里 | 中庸里 | 豐年里 | 長名里 | 溫泉里 | 大同里 | 清江里 | 中央里 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 中和里 | 0.00 | 0.25 | 0.25 | 0.00 | 0.25 | 0.00 | 0.00 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 泉源里 | 0.25 | 0.00 | 0.25 | 0.00 | 0.00 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 開明里 | 0.20 | 0.20 | 0.00 | 0.00 | 0.20 | 0.00 | 0.20 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 文化里 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.20 | 0.20 | 0.20 | 0.00 | 0.20 | 0.00 | 0.00 |
| 智仁里 | 0.25 | 0.00 | 0.25 | 0.25 | 0.00 | 0.00 | 0.00 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 林泉里 | 0.00 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.33 | 0.00 | 0.00 | 0.00 | 0.33 | 0.00 | 0.00 | 0.00 |
| 中心里 | 0.00 | 0.17 | 0.17 | 0.00 | 0.00 | 0.17 | 0.00 | 0.17 | 0.00 | 0.17 | 0.17 | 0.00 | 0.00 | 0.00 |
| 中庸里 | 0.17 | 0.00 | 0.17 | 0.17 | 0.17 | 0.00 | 0.17 | 0.00 | 0.00 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 |
| 豐年里 | 0.00 | 0.00 | 0.00 | 0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.33 | 0.00 | 0.33 | 0.00 | 0.00 |
| 長安里 | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 | 0.00 | 0.14 | 0.14 | 0.14 | 0.00 | 0.14 | 0.14 | 0.14 | 0.00 |
| 溫泉里 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.00 | 0.00 | 0.25 | 0.00 | 0.00 | 0.25 | 0.00 |
| 大同里 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.20 | 0.00 | 0.00 | 0.20 |
| 清江里 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.25 | 0.25 | 0.00 | 0.25 |
| 中央里 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.50 | 0.00 |

# 2. Distances

- Distance decay

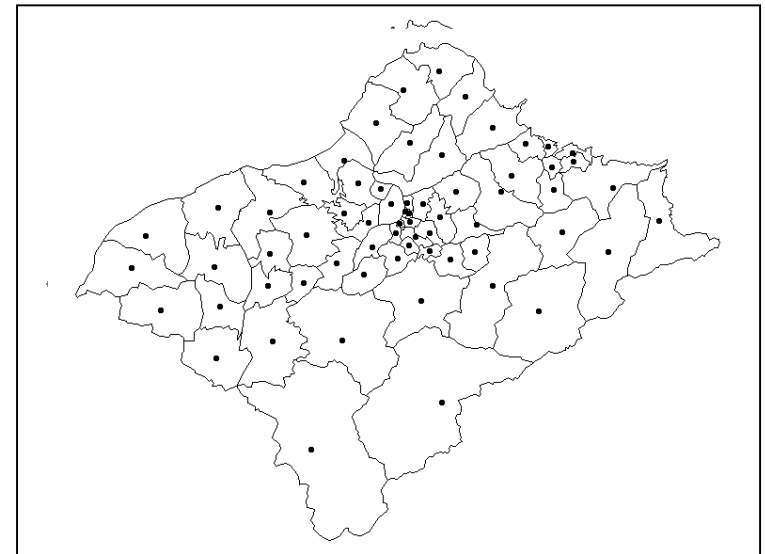  - $W_{ij} = 1 / d_{ij}$

  - $W_{ij} = 1 / d^2_{ij}$

# Centroid Distances

- Distance between centroids

- Centroid : geometric center of the polygon

  - Affected by the shape of the polygon

  - May be located outside the polygon

# *Spatially Weighted Matrix*

# *Using Centroid Distances*



| ID | 中和里 | 泉源里 | 開明里 | 文化里 | 智仁里 | 林泉里 | 中心里 | 中庸里 | 豐年里 | 長安里 | 溫泉里 | 大同里 | 清江里 | 中央里 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 中和里 | 0 | 1359 | 1602 | 2229 | 1948 | 2334 | 1913 | 2142 | 3149 | 2530 | 2763 | 3169 | 3238 | 3375 |
| 泉源里 | 1359 | 0 | 2033 | 2966 | 2657 | 1737 | 2002 | 2602 | 3716 | 2824 | 2764 | 3618 | 3337 | 3758 |
| 開明里 | 1602 | 2033 | 0 | 974 | 684 | 1477 | 506 | 585 | 1692 | 930 | 1224 | 1633 | 1645 | 1812 |
| 文化里 | 2229 | 2966 | 974 | 0 | 309 | 2335 | 1309 | 607 | 952 | 932 | 1519 | 1073 | 1607 | 1316 |
| 智仁里 | 1948 | 2657 | 684 | 309 | 0 | 2103 | 1075 | 478 | 1204 | 903 | 1444 | 1265 | 1644 | 1496 |
| 林泉里 | 2334 | 1737 | 1477 | 2335 | 2103 | 0 | 1029 | 1747 | 2683 | 1678 | 1310 | 2461 | 1869 | 2508 |
| 中心里 | 1913 | 2002 | 506 | 1309 | 1075 | 1029 | 0 | 741 | 1790 | 829 | 856 | 1640 | 1381 | 1762 |
| 中庸里 | 2142 | 2602 | 585 | 607 | 478 | 1747 | 741 | 0 | 1115 | 450 | 966 | 1049 | 1209 | 1238 |
| 豐年里 | 3149 | 3716 | 1692 | 952 | 1204 | 2683 | 1790 | 1115 | 0 | 1005 | 1480 | 324 | 1203 | 522 |
| 長安里 | 2530 | 2824 | 930 | 932 | 903 | 1678 | 829 | 450 | 1005 | 0 | 595 | 816 | 760 | 935 |
| 溫泉里 | 2763 | 2764 | 1224 | 1519 | 1444 | 1310 | 856 | 966 | 1480 | 595 | 0 | 1208 | 581 | 1214 |
| 大同里 | 3169 | 3618 | 1633 | 1073 | 1265 | 2461 | 1640 | 1049 | 324 | 816 | 1208 | 0 | 882 | 247 |
| 清江里 | 3238 | 3337 | 1645 | 1607 | 1644 | 1869 | 1381 | 1209 | 1203 | 760 | 581 | 882 | 0 | 782 |
| 中央里 | 3375 | 3758 | 1812 | 1316 | 1496 | 2508 | 1762 | 1238 | 522 | 935 | 1214 | 247 | 782 | 0 |

# Nearest Distances (較不常用)

- 兩個Polygon之間最短的點的距離
  - Will be zero for adjacent polygons

| ID | 中和里 | 泉源里 | 開明里 | 文化里 | 智仁里 | 林泉里 | 中心里 | 中庸里 | 豐年里 | 長安里 | 溫泉里 | 大同里 | 清江里 | 中央里 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 中和里 | 0 | 0 | 0 | 51 | 0 | 837 | 403 | 0 | 606 | 464 | 723 | 615 | 934 | 1075 |
| 泉源里 | 0 | 0 | 0 | 920 | 849 | 0 | 0 | 686 | 1236 | 731 | 653 | 1122 | 1095 | 1301 |
| 開明里 | 0 | 0 | 0 | 148 | 0 | 319 | 0 | 0 | 529 | 166 | 304 | 477 | 572 | 773 |
| 文化里 | 51 | 920 | 148 | 0 | 0 | 878 | 299 | 0 | 0 | 0 | 567 | 0 | 549 | 481 |
| 智仁里 | 0 | 849 | 0 | 0 | 0 | 879 | 322 | 0 | 405 | 260 | 600 | 415 | 753 | 870 |
| 林泉里 | 837 | 0 | 319 | 878 | 879 | 0 | 0 | 535 | 986 | 234 | 0 | 609 | 463 | 691 |
| 中心里 | 403 | 0 | 0 | 299 | 322 | 0 | 0 | 0 | 471 | 0 | 0 | 269 | 269 | 477 |
| 中庸里 | 0 | 686 | 0 | 0 | 0 | 535 | 0 | 0 | 89 | 0 | 209 | 89 | 407 | 558 |
| 豐年里 | 606 | 1236 | 529 | 0 | 405 | 986 | 471 | 89 | 0 | 0 | 583 | 0 | 547 | 67 |
| 長安里 | 464 | 731 | 166 | 0 | 260 | 234 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 169 |
| 溫泉里 | 723 | 653 | 304 | 567 | 600 | 0 | 0 | 209 | 583 | 0 | 0 | 70 | 0 | 179 |
| 大同里 | 615 | 1122 | 477 | 0 | 415 | 609 | 269 | 89 | 0 | 0 | 70 | 0 | 0 | 0 |
| 清江里 | 934 | 1095 | 572 | 549 | 753 | 463 | 269 | 407 | 547 | 0 | 0 | 0 | 0 | 0 |
| 中央里 | 1075 | 1301 | 773 | 481 | 870 | 691 | 477 | 558 | 67 | 169 | 179 | 0 | 0 | 0 |

*Spatially Weighted Matrix*

*Using Nearest Distances*

# Spatial Weights Matrix Approaches

**Neighborhood Definition**

- Rook's Definition
- Queen's Definition
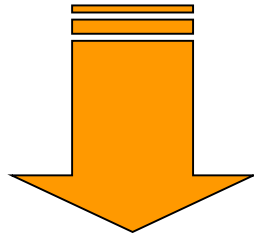
**Spatial Weights Matrix**

- Binary Connective Matrix
- Stochastic or Row Standardized Weights Matrix
- Centroid Distances
- Nearest Distances

# 教科書的研讀教材（不用繳交心得作業，但列入考試範圍）

TEXT_Spatial_Weights.pdf

# Defining spatial neighborhoods and weights

Several of the methods discussed in chapters 3 and 4 show you how to analyze patterns and clusters of feature values. These methods look at both the difference between the values of features and the spatial relationship between the features (distance or other measure).

Specifically, the GIS compares the value of a feature (the "target") to the values of neighboring features. It then moves to the next feature and does the same thing, and so on, for all the features in the study area. In order to do this, the GIS requires that you define the area surrounding each target feature within which feature values are compared—termed the "neighborhood"—and the nature of the spatial relationship between features. The GIS then assigns weights to each feature pair to specify whether the two features are in each other's neighborhoods, and to represent the spatial relationship between the features.

You define the neighborhood based on the interaction between features. Features might influence each other—for example, the value

*Features color coded by value*

*Target feature*

*Neighborhood*
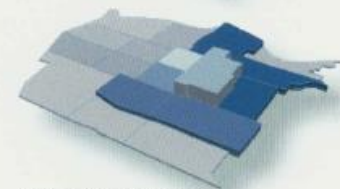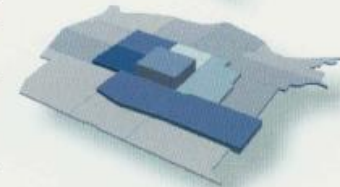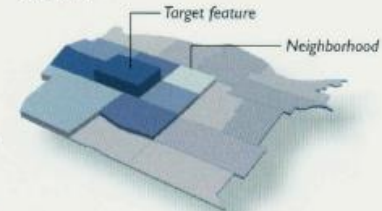
*Each feature in turn is*

# Measuring Spatial Autocorrelation

- Spatial weighting $W_{ij}$

  - ❑ Contiguity [binary or row-standardized]

    - **Common Border**

  - ❑ Distance [centroids or nearest]

    - **Distance band**

    - **K$^{th}$-nearest neighbors**

# 1. Index of Spatial Autocorrelation: Moran's I

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

where

    N    is the number of cases

    $\bar{x}$    is the mean of the variable

    $x_i$  is the variable value at a particular location I

    $w_{ij}$  is a **spatial weight indexing** location of i relative to j

$$W = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \quad \text{(sum of all } w_{i,j}\text{ )}$$

- Applied to a **continuous variable** for polygons or points

# 這個公式是怎麼想出來的？

皮爾森相關係數 Pearson's correlation coefficient (r):

共變異數 cov (X,Y) 的觀念

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})/n}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/n}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2/n}}$$

# 這個公式是怎麼想出來的？ [續]

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})/n}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/n}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2/n}}$$
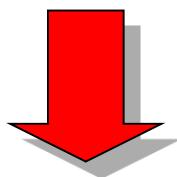
Neighboring values of variable y replace those of x

把 xi 替換成 yi 的鄰居 ($c_{ij}$.yi)

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}(y_i - \bar{y})(y_j - \bar{y})/\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2/n}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2/n}}$$

# 這個公式是怎麼想出來的？ [續]

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}(y_i-\bar{y})(y_j-\bar{y})/\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2/n}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2/n}}$$

$$I = \frac{N}{W}\frac{\sum_i\sum_j w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_i(x_i-\bar{x})^2}$$

# Spatial Autocorrelation: Moran's I Statistic

Product of the deviation from the mean for all pairs of adjacent regions ($w_{ij}=1$)

$$I = \frac{n \sum\limits_{i=1}^{n} \sum\limits_{i=j}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\left( \sum\limits_{i=1}^{n} (y_i - \bar{y})^2 \right) \left( \sum\sum\limits_{i \neq j} w_{ij} \right)}$$

Essentially a measure of variance across the regions

Sum of the weights (count of all adjacent pairs)

- $n$ = number of regions
- $w_{ij}$ = measure of spatial proximity between region i and j

# Moran's I Interpretations

- Similar to correlation coefficient, range between $\pm\,1.0$

  - 0 indicates no spatial autocorrelation, approximate technically it is $-1/(n-1)$

  - Highly auto-correlated, if I is closed to 1 or -1

  - Sign of values indicate negative/positive autocorrelation

- Can be used as index for dispersion/random/cluster patterns

  - 0: random

  - Positive : more toward clustering

  - Negative: more toward dispersion/uniform

# Significance Tests for Moran's I

- Z-score:  $Z = (I - E(I)) / S_{Err}(I)$

  - I: Moran's I of sample

  - E(I): Expected value of I ; $E(I) = -1/(n-1)$

  - $S_{Err}$: Standard error

    - Depend on if free or non-free sampling is used

- α = 0.05 , Critical Z value = $\pm$ 1.96

  - will be $\pm$1.645 for α = 0.1

- 檢定是否為達到顯著差異

  - Ho:無差異（隨機分佈）

  - At p< 0.05, Reject Ho  if  |Z| > 1.96

$$E_N(I) = E_R(I) = \frac{-1}{n-1}$$

$$VAR_N(I) = \frac{(n^2 S_1 - n S_2 + 3W^2)}{W^2(n^2-1)} - [E_N(I)]^2$$

free sampling
(with replacement.)
(normality)

$$VAR_R(I) = \frac{n[(n^2 - 3n + 3)S_1 - n S_2 + 3W^2]}{(n-1)(n-2)(n-3)W^2}$$

$$- \frac{k[(n^2 - n)S_1 - n S_2 + 3W^2]}{(n-1)(n-2)(n-3)W^2} - [E_R(I)]^2,$$

non-free sampling
(without replacement)
(randomization)

where

$$W = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$$

$$S_1 = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (w_{ij} + w_{ji})^2}{2}$$

$$S_2 = \sum_{i=1}^{n} (w_{i.} + w_{.i})^2$$

$$k = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^4}{\left( \sum_{i=1}^{n} (x_i - \overline{x})^2 \right)^2}.$$

# Randomization vs. normality sampling

Randomization sampling assumes that the observed spatial pattern of your data represents one of many possible spatial arrangements—the number of features having a particular value is always going to be the same (based on the observed number of each), but the arrangement can change. Suppose you have the number of cases per census tract of a disease. Some tracts have several cases of the disease; many don't have any. Your null hypothesis would be that the disease strikes randomly. If you could take the case count values for your study area and scatter them on a map of the census tracts, making sure every census tract got a value, you'd create a random pattern. The randomization null hypothesis postulates that if you could perform this operation infinite times, most of the time you would produce a pattern that was not markedly different from the observed pattern. If your significance test indicates that you should reject the null hypothesis at the specified confidence level, then you know that the observed arrangement of values would significantly differ from this randomly produced pattern.

TEXT_Statistical.Significance.pdf （期末考試範圍）

# Randomization vs. **normality** sampling (cont'd)

Normalization sampling, in contrast, assumes that the number of cases associated with any particular census tract could be derived from an infinitely large, normally distributed population of values (through some random sampling process). Rather than scattering the observed values on the map of census tracts, you'd pick values from this hypothetical normal distribution and scatter those values on the map to create the random pattern.

# Randomization vs. normality sampling (比較)

- The **normalizatio**n null hypothesis not only assumes that your data is a sample, but also that the sample was obtained randomly and that the population h m which the sample was obtained has a normal distribution of values. Every time you make an assumption about the data or the sample, you're potentially inducing error into the test**.**

- **Randomization makes fewer assumptions than normalization, so it's safer to use**, unless you know for sure your data matches the assumptions of normalization.

# Test Statistic for Normal Frequency Distribution

2.5%    2.5% | 1%

Reject null  -1.96          $-1/(n-1)$          1.96   2.54
                                 0
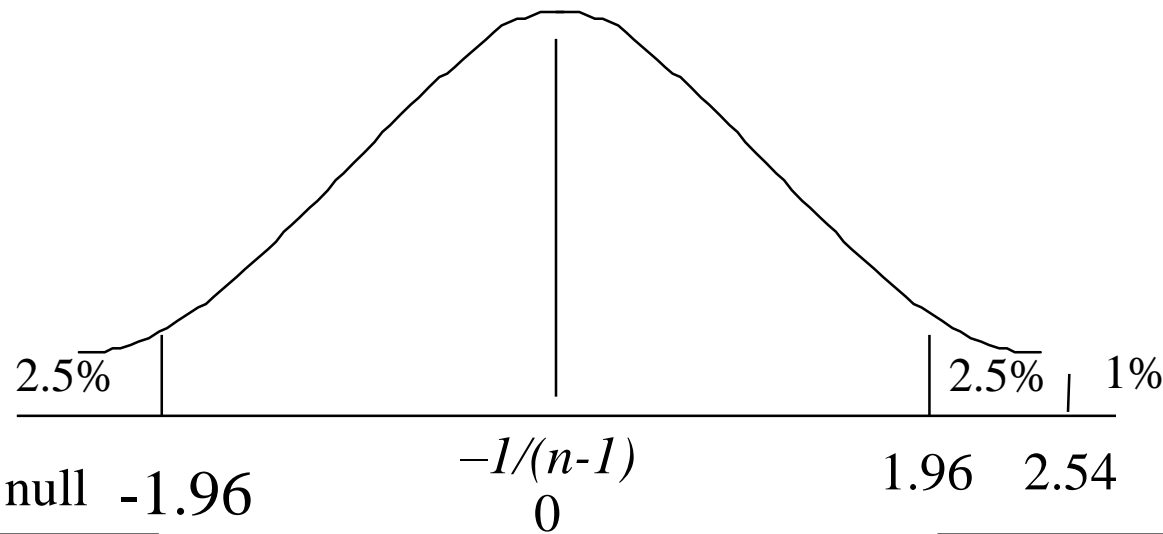
Reject null at 2.5%

Reject null at 1%

*Null Hypothesis:* no spatial autocorrelation

*Moran's I = 0   * technically  $-1/(n-1)$

*Alternative Hypothesis:* spatial clustering exists

*Moran's I > 0（單尾檢定）

Reject *Null Hypothesis* if Z test statistic > 1.96  (or < -1.96)

---less than a 2.5% chance that, in the population, there is no
        spatial clustering

---97.5% confident that spatial clustering exits

# Monte-Carlo Significance Test

- **Permutation test (排列檢定)**
  - ❑ The null hypothesis is that the data were determined and then assigned to their spatial locations at random.
  - ❑ The alternative is that the assignment to each location depended on the assignment at that location's neighbors.
  - ❑ The permutation test does not randomize over the possible sets of data values--it considers them given-- but **conditional on the data observed**, considers all possible ways of reassigning them to the locations.

# Monte-Carlo Significance Test (cont'd)

- **Permutation test (排列檢定)**

- Such a reassignment is a *permutation.* For $n$ data points, there **are n! = n x (n-1) x (n-2) x ... x (2) x (1)** permutations. For $n$ much larger than 10 or so, that's too many to generate.

- There usually is no simple analytical expression for the full permutation distribution.

- Accordingly, we typically resort to sampling from the set of all permutations at random, giving them all equal weight. The distribution of the autocorrelation statistic in a sufficiently large sample (usually involving at least 500 permutations) approximates the true distribution.
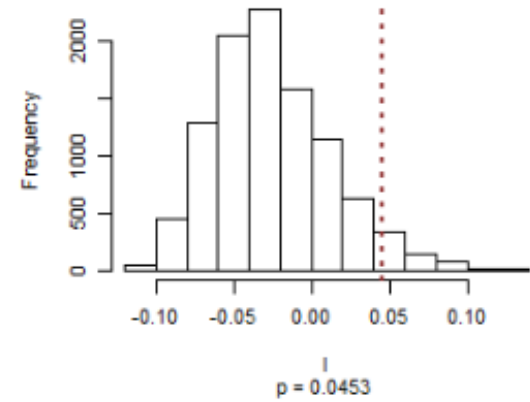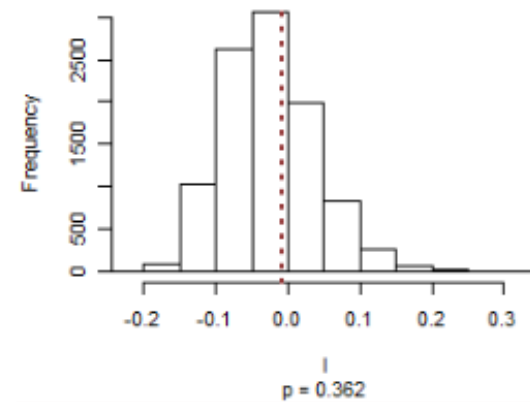
# Examples



Autocorrelated Data

Null Distribution of Moran's I

$p = 0.0453$
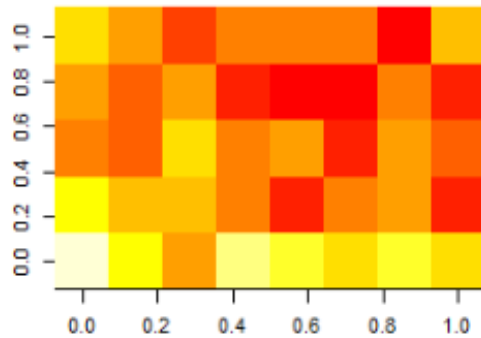
Uncorrelated Data

Null Distribution of Moran's I

$p = 0.362$
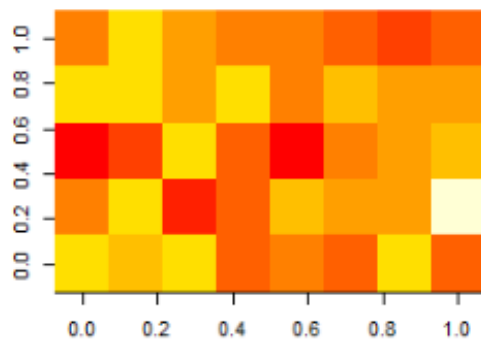
# Output in R: Moran's I statistic

```
> M<-moran.test(Popn, listw=TWN_nb_w, zero.policy=T); M

        Moran I test under randomisation

data:  Popn
weights: TWN_nb_w

Moran I statistic standard deviate = 2.1678, p-value = 0.01509
alternative hypothesis: greater
sample estimates:
Moran I statistic       Expectation          Variance
      0.181359104      -0.025000000       0.009062094
```
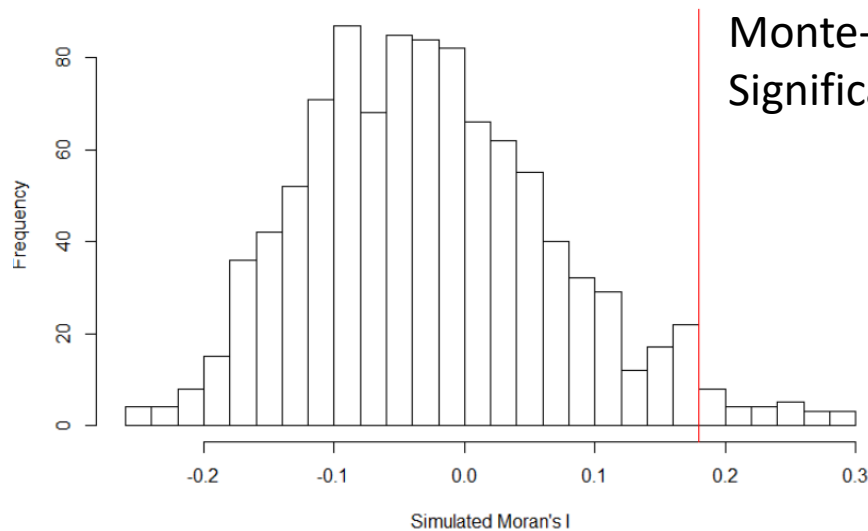
Monte-Carlo simulation Significance Test

# Moran.test()

## Moran's I test for spatial autocorrelation

### Description

Moran's test for spatial autocorrelation using a spatial weights matrix in weights list form. The assumptions underlying the test are sensitive to the form of the graph of neighbour relationships and other factors, and results may be checked against those of `moran.mc` permutations.
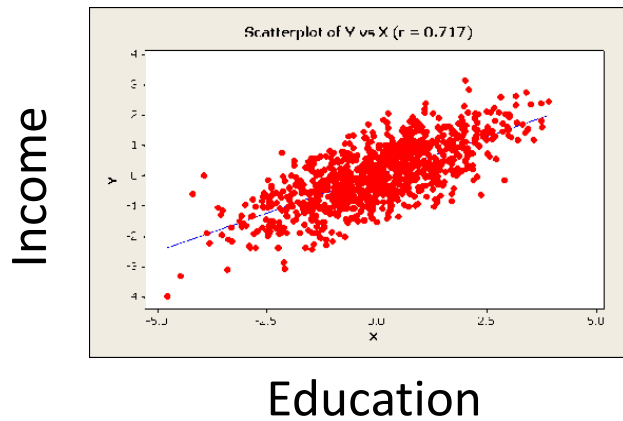
### Usage

```
moran.test(x, listw, randomisation=TRUE, zero.policy=NULL,
  alternative="greater", rank = FALSE, na.action=na.fail, spChk=NULL, a
```

| | |
|---|---|
| `randomisation` | variance of I calculated under the assumption of randomisation, if FALSE normality |

# Concept of Moran Scatter Plots

## Scatter Plot
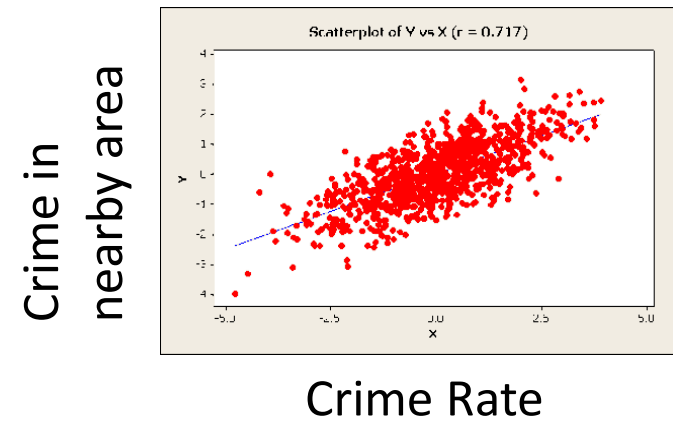
Two variables



Income (vertical axis) vs Education (horizontal axis)

## Moran Scatter Plot

Only one variable



Crime in nearby area (vertical axis) vs Crime Rate (horizontal axis)

# Moran Scatter Plots

Moran's I can be interpreted as the correlation between variable, X, and the "spatial lag" of X formed by averaging all the values of X for the neighboring polygons.
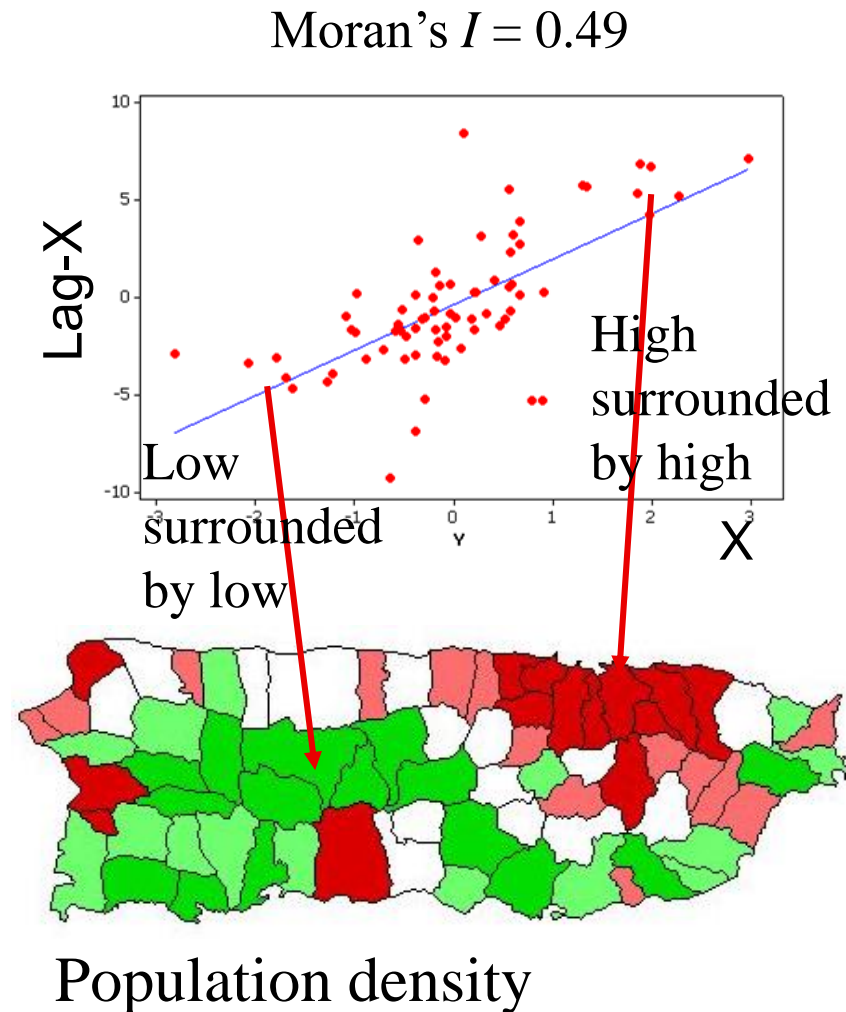


Moran's I= 0.5237

Lag $X_i$ is average of these
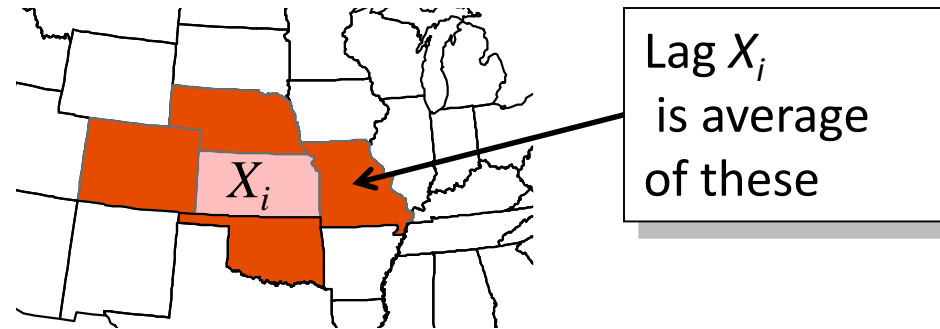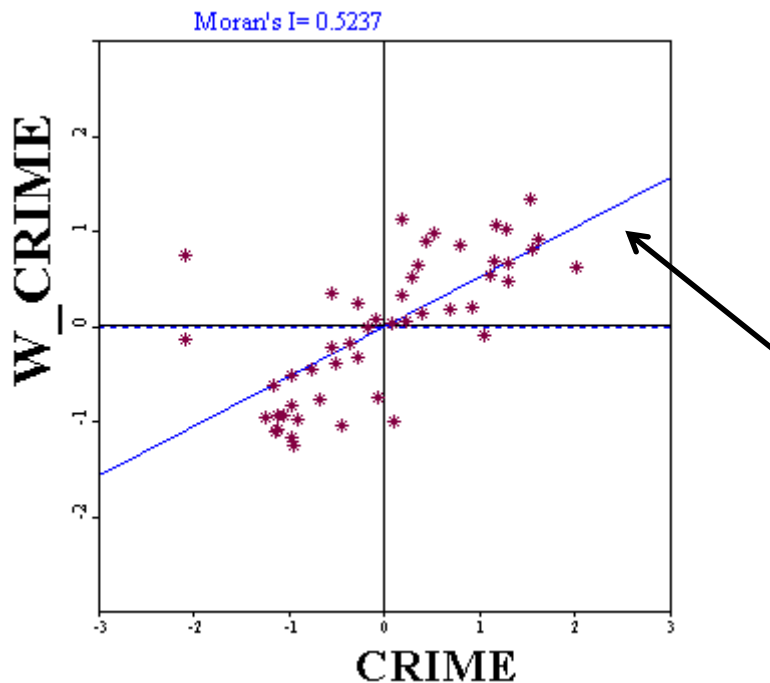
Least squares "best fit" line to the points.
The <u>slope</u> of this *regression line* is Moran's I

# Moran Scatter Plot: example

- Scatter plot of **X** vs. **Lag-X**

- The slope of the regression is Moran's *I*

Moran's $I = 0.49$



High surrounded by high

Low surrounded by low

Population density

# Moran Correlograms

*Correlogram*: plot distance on X-axis against correlation coefficient on Y-axis
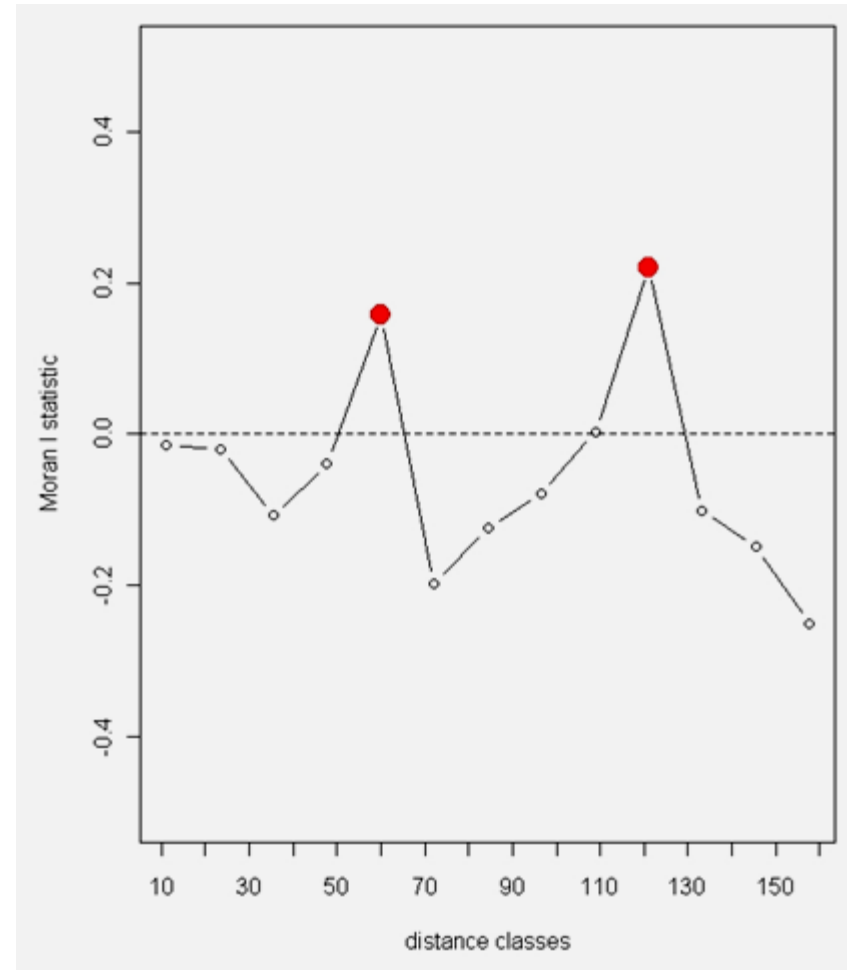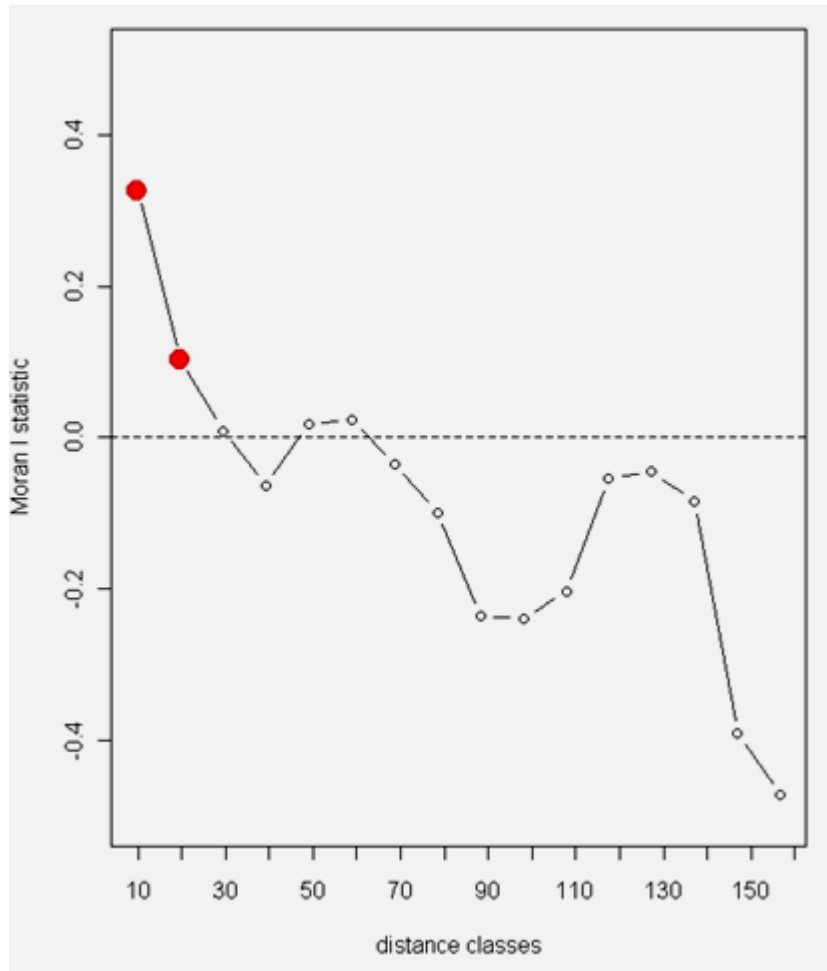
# Getis-Ord General G-statistic

- Moran's I 無法區別

  " **hot spots"** or **"cold spots"**

- *Spatial Concentration* method

- Definition

$$G(d) = \frac{\sum \sum w_{ij}(d) x_i x_j}{\sum \sum x_i x_j}$$

  *d : neighborhood distance*

  *$W_{ij}$ : 1  if it is within d, 0 otherwise*

- Calculation of G must begin by identifying a neighborhood distance within which cluster is expected to occur

# Getis-Ord General G-Statistic

- General G-statistic **can distinguish between hot/cold spots**. It identifies **spatial concentrations**.

  - G is relatively large if high values cluster together

  - G is relatively low if low values cluster together

- G statistic is interpreted relative to its expected value

  - > E(G) ➔ potential "***hot spot***"

  - < E(G) ➔ potential "***cold spot***"

  - = E(G) ➔ no spatial association

- 所謂之larger/smaller不能單從值的大小判斷，
  需以 **Z test statistic** 來檢定差異的統計顯著性。

# Significance Test for Getis-Ord General G

- Statistical Significance Test

$$Z = \frac{G - E(G)}{S_{Err(G)}}$$

- Expected G : $\quad E(G) = \dfrac{W}{n(n-1)}$ ; where $\quad W = \displaystyle\sum_{i}\sum_{j} w_{ij}(d),$

- Standard Error will depend on the sampling method (free / non-free)

# Getis-Ord General G-statistic in R

```
> G<-globalG.test(Popn, listw=TWN_ran1_wb); G

        Getis-Ord global G statistic

data:  Popn
weights: TWN_ran1_wb

standard deviate = 3.4804, p-value = 0.0002504
alternative hypothesis: greater
sample estimates:
Global G statistic          Expectation                Variance
      0.6216802233          0.5402439024              0.0005474983
```

# 教科書的研讀教材（列入考試範圍）

TEXT_Pattern.of.Feature.Value.pdf



## MEASURING THE SPATIAL PATTERN OF FEATURE VALUES

In addition to measuring the pattern formed by the locations of features, you can also measure patterns of attribute values associated with features, such as the pattern formed by median house values. These methods reveal whether similar values tend to occur near each other, or whether high and low values are interspersed.

Legend:
- 0 - 44,360
- 44,361 - 88,720
- 88,721 - 133,080
- 133,081 - 177,440
- 177,441 - 221,800

*Median house value by census tract.*

### The idea behind measuring patterns of feature values

Measuring the spatial pattern of feature values is based on the notion that things near each other are more alike than things far apart, an idea often attributed to geographer Waldo Tobler. The idea is consistent with our

# 實習：介紹 R package: spdep
## Spatial Dependence: Weighting Schemes, Statistics and Models

spdep: Spatial Dependence: Weighting Schemes, Statistics and Models

A collection of functions to create spatial weights matrix objects from polygon 'contiguities', from point patterns by distance and tessellations, for summarizing these objects, and for permitting their use in spatial data analysis, including regional aggregation by minimum spanning tree; a collection of tests for spatial 'autocorrelation', including global 'Morans I', 'APLE', 'Gearys C', 'Hubert/Mantel' general cross product statistic, Empirical Bayes estimates and 'Assunção/Reis' Index, 'Getis/Ord' G and multicoloured join count statistics, local 'Moran's I' and 'Getis/Ord' G, 'saddlepoint' approximations and exact tests for global and local 'Moran's I'; and functions for estimating spatial simultaneous 'autoregressive' ('SAR') lag and error models, impact measures for lag models, weighted and 'unweighted' 'SAR' and 'CAR' spatial regression models, semi-parametric and Moran 'eigenvector' spatial filtering, 'GM SAR' error models, and generalized spatial two stage least squares models.

# spdep 重要函數

- **Spatial Neighbors**
  - Contiguity: QUEEN vs. ROOK  poly2nb(); nb2mat()
  - K-nearest Neighbors (KNN)  knn2nb(); knearneigh(coords, k=2)
  - Distance-based  dnearneigh()
- **From Spatial Neighbors to ListW (Weighting matrix)**
  - nb2listw()
- **Spatial Autocorrelation**
  - Mapping the attribute  **tmap::**tm_shape()
  - Moran's I Statistic  moran.test()
  - Monte-Carlo simulation  moran.mc()
  - Moran correlogram  sp.correlogram()
  - Moran Scatter Plot  moran.plot()
  - Getis-Ord General G Statistic  globalG.test()

■ 台灣鄉鎮市區**人口密度**的空間型態分析

❑ 計算以下統計量與繪製圖表，說明其參數設定，並解釋其意義。
包括：Moran's I coefficient, Monte-Carlo simulation, Moran scatter plot, Correlogram, and General G statistic.

❑ 利用以下三種不同的空間鄰近定義，計算Moran's I coefficient，
比較其數值的差異，並討論可能的原因。

Spatial Neighbors: Contiguity; K-nearest Neighbors (KNN); Distance-based

# 作業-1：實作

資料： **Popn_TWN2.shp**

- 共三題 (以Contiguity定義鄰近)

  - 繪製各鄉鎮的鄰居數的直方圖。

  - 找出台灣本島<u>最多鄰居</u>的鄉鎮是哪一個? (TOWN_ID)
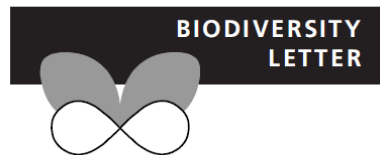
  - 繪製台灣各鄉鎮的<span style="color:red">1st-order鄰居人口密度</span>的面量圖。
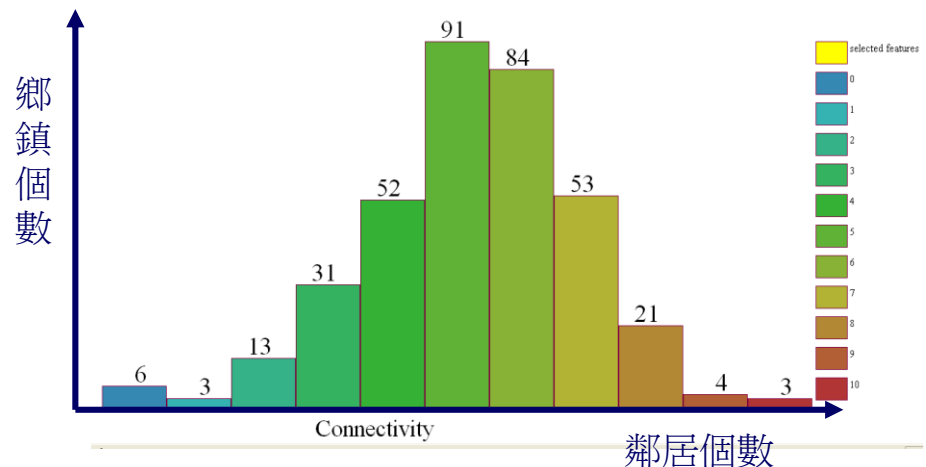
# 作業-2：論文研讀與心得　　Reading_autocorrelation.pdf

內容需包括：

1. 空間自相關在植物物種多樣性分布的意涵

2. 解釋 Figure 1 and Table 1 的意義

**BIODIVERSITY LETTER**

# Incorporating spatial autocorrelation may invert observed patterns

Ingolf Kühn

UFZ — Centre for Environmental Research Leipzig-Halle, Department Community Ecology (BZF), Theodor-Lieser-Str. 4, 06120 Halle, Germany

## ABSTRACT

Though still often neglected, spatial autocorrelation can be a serious issue in ecology because the presence of spatial autocorrelation may alter the parameter estimates and error probabilities of linear models. Here I re-analysed data from a previous study on the relationship between plant species richness and environmental correlates in Germany. While there was a positive relationship between native plant species richness and an altitudinal gradient when ignoring the presence of spatial autocorrelation, the use of a spatial simultaneous liner error model revealed a negative relationship. This most dramatic effect where the observed pattern was inverted may be explained by the environmental situation in Germany. There the highest altitudes are in the south and the lowlands in the north that result in some locally or regionally inverted patterns of the large-scale environmental gradients from the equator to the north. This study therefore shows the necessity to consider spatial autocorrelation in spatial analyses.

Correspondence: Ingolf Kühn, Virtual Institute for Macroecology, Theodor-Lieser-Str. 4, 06120 Halle, Germany. Tel.: (++49)345/558-5311 or (++49)345/558-5329. E-mail: ingolf.kuehn@ufz.de