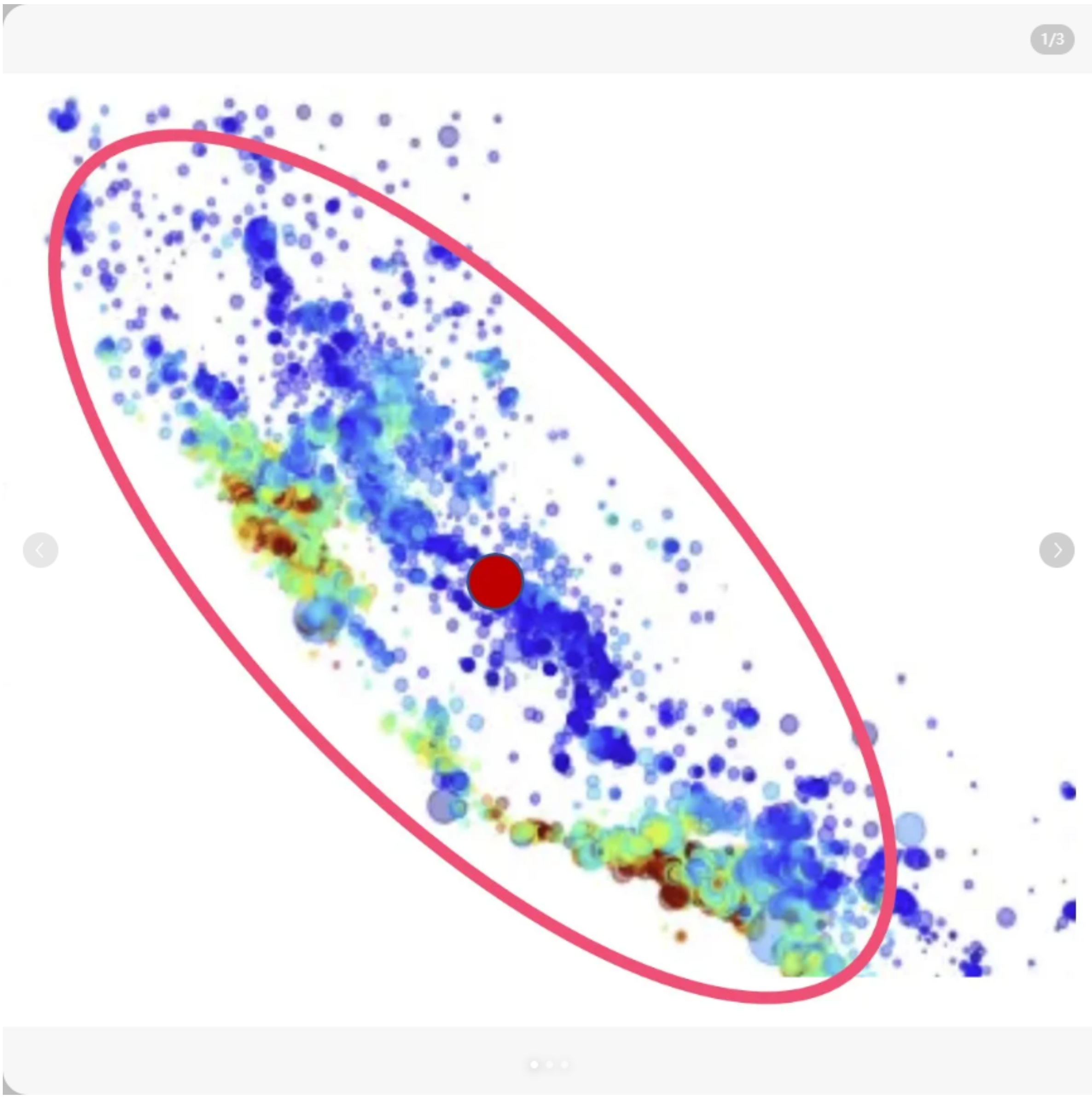
空间数据探索性分析

计算标准误差椭圆

描述一组数据点在空间上的聚集性和方向性，用来理解空间数据的分布特征和不确定性。



$$\bar{X}=\frac{\sum_{i=1}^{n}x_i}{n};\quad \bar{Y}=\frac{\sum_{i=1}^{n}y_i}{n}$$

其中，$(\bar{X},\bar{Y})$表示的平均 x 坐标和平均 y 坐标；n为事件的总数；(x_i,y_i)为第i个事件点的空间坐标。

$$\begin{cases}\theta=\arctan\dfrac{\left(\sum a_i^2-\sum b_i^2\right)+\left(\left(\sum a_i^2-\sum b_i^2\right)^2+4\left(\sum a_i b_i\right)^2\right)^{\frac{1}{2}}}{2\sum a_i b_i}\\ \mathrm{SDE}_x=\sqrt{2}\sqrt{\dfrac{\sum_{i=1}^{n}\left(a_i\cos\theta+b_i\sin\theta\right)^2}{n}}\\ SDE_y=\sqrt{2}\sqrt{\dfrac{\sum_{i=1}^{n}\left(a_i\sin\theta-b_i\cos\theta\right)^2}{n}}\end{cases}$$

其中，θ表示标准差椭圆长轴与竖直方向的夹角；$SDEx$表示标准差椭圆的长半轴，$SDEy$表示椭圆的短半轴；$\mathrm{a}_i=x_i-\bar{\mathrm{X}}$，$b_i=y_i-\bar{\mathrm{Y}}$，$(x_i,y_i)$为第$i$个事件点的空间坐标，$(\bar{X},\bar{Y})$为所有事件点的平均中心$x$坐标和$y$坐标，$n$为事故点的总个数。

然后确定椭圆的方向，以X轴为准，正北方（12点方向）为0度，顺时针旋转，计算公式如下：

$$\begin{aligned}\tan\ \theta &= \frac{A+B}{C}\\ A &= \left(\sum_{i=1}^{n}\tilde{x}_i^2-\sum_{i=1}^{n}\tilde{y}_i^2\right)\\ B &= \sqrt{\left(\sum_{i=1}^{n}\tilde{x}_i^2-\sum_{i=1}^{n}\tilde{y}_i^2\right)^2+4\left(\sum_{i=1}^{n}\tilde{x}_i\tilde{y}_i\right)^2}\\ C &= 2\sum_{i=1}^{n}\tilde{x}_i\tilde{y}_i\end{aligned}$$

其中 $\tilde{x}_i$ 和 $\tilde{y}_i$ 是平均中心和 xy 坐标的差。

最后确定XY轴的标准差，公式如下：

$$\begin{aligned}\sigma_x &= \sqrt{2}\sqrt{\frac{\sum_{i=1}^{n}(\tilde{x}_i\ \cos\ \theta-\tilde{y}_i\ \sin\ \theta)^2}{n}}\\ \sigma_y &= \sqrt{2}\sqrt{\frac{\sum_{i=1}^{n}(\tilde{x}_i\ \sin\ \theta+\tilde{y}_i\ \cos\ \theta)^2}{n}}\end{aligned}$$

空间权重矩阵

地理学第一定律：任何事物都相关，距离越近关系越紧密

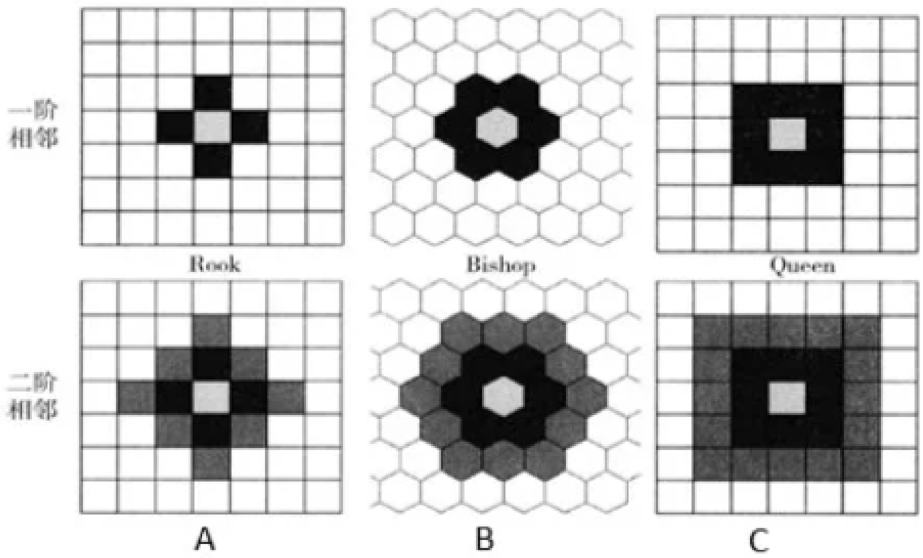
空间权重矩阵：量化地理空间单元之间的联系强度

表 4-1　空间权重矩阵设定方法比较

空间权重矩阵		公式		适用范围
邻接矩阵	Rook 邻接矩阵	$w_{ij}=\begin{cases}1 & \text{区域}i\text{和}j\text{共边}\\0 & i=j\text{或区域}i\text{和}j\text{不共边}\end{cases}$	(4.1.8)	多边形空间单元
	Queen 邻接矩阵	$w_{ij}=\begin{cases}1 & \text{当区域}i\text{和}j\text{共边或共顶点}\\0 & i=j\text{或区域}i\text{和}j\text{无共边且不共顶点}\end{cases}$	(4.1.9)	多边形空间单元
距离函数矩阵	二进制地理距离矩阵	$w_{ij}=\begin{cases}1 & d_{ij}\leqslant d_0\\0 & d_{ij}>d_0\end{cases}$	(4.1.10)	离散点空间单元
	阈值权重矩阵	$w_{ij}=\begin{cases}0 & i=j\\a_1 & d_{ij}<d_0\\a_2 & d_{ij}\geqslant d_0\end{cases}$	(4.1.11)	离散点空间单元
	Cliff-Ord 矩阵	$w_{ij}=(d_{ij})^{-a}(\beta_{ij})^{b}$	(4.1.12)	多边形空间单元
	Decay 权重矩阵	$w_{ij}=d_{ij}\cdot\alpha_i\cdot\beta_{ij}$	(4.1.13)	多边形空间单元
	k 近邻矩阵	$w_{ij}=\begin{cases}1/d_{ij} & d_{ij}\leqslant d_*^{(k)}\\0 & i=j\text{或}d_{ij}>d_*^{(k)}\end{cases}$	(4.1.14)	离散点空间单元
其他距离矩阵	基于引力模型的空间邻接矩阵	$w_{ij}=\begin{cases}\dfrac{m_i m_j}{d_{ij}^2} & i\neq j\\0 & i=j\end{cases}$	(4.1.15)	两两间具有某种联系的空间单元，这种联系的强度可以使用万有引力的形式进行估计

基于邻接关系

- 邻接关系是空间单元间有公共边界的的现象，包括：
  - Rook相邻 $w_{ij}=\begin{cases}1 & \text{区域}i\text{和}j\text{共边} \\ 0 \quad i=j & \text{区域}i\text{和}j\text{不共边}\end{cases}$
  - Bishop相邻 $w_{ij}=\begin{cases}1 & \text{区域}i\text{和}j\text{ 共顶点} \\ 0 & \text{区域}i\text{和}j\text{不共顶点}\end{cases}$
  - Queen相邻 $w_{ij}=\begin{cases}1 & \text{区域}i\text{和}j\text{共边或共顶点} \\ 0 & \text{区域}i\text{和}j\text{无共边且不共顶点}\end{cases}$



基于距离关系

- 反距离权重矩阵 $w_{ij}=\dfrac{1}{\mathrm{d}_{ij}^{p}}$
- 半径距离权重矩阵

$$w_{ij}=\begin{cases}1 & d_{ij}\le d_0\\ 0 & d_{ij}>d_0\end{cases}$$

- K近邻权重矩阵

$$w_{ij}=\begin{cases}1/d_{ij} & d_{ij}\le d_0^{(k)}\\ 0 & i=j\text{或}d_{ij}>d_0^{(k)}\end{cases}$$

- ……

空间权重矩阵的标准化

- 将空间权重矩阵进行行（列）标准化，使得每一行（列）的权重和为1

$$w_{ij}^{*} = w_{ij} / \sum_{j=1}^{n} w_{ij}$$

行标准化　　双重标准化

$$w_{ij}^{*} = w_{ij} / \sum\sum w_{ij}$$

$$W = \begin{bmatrix} 0 & 0.33 & 0.33 & 0.33 \\ 0.5 & 0 & 0.5 & 0 \\ 0.33 & 0.33 & 0 & 0.33 \\ 0.5 & 0 & 0.5 & 0 \end{bmatrix} \quad W = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad W = \begin{bmatrix} 0 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0 & 0.1 & 0 \\ 0.1 & 0.1 & 0 & 0.1 \\ 0.1 & 0 & 0.1 & 0 \end{bmatrix}$$

行标准化保证了矩阵**行元素之和为1**，但其列和不一定为1，因此标准化后的空间权重矩阵不一定是对称阵。

双重标准化保证矩阵**行列元素之和为1**，标准化后的空间权重矩阵是对称阵。

例：邻接关系的空间权重矩阵

	编号	安徽	浙江	江西	江苏	河南	湖北	上海	总计
安徽	1	0	1	1	1	1	1	0	**5**
浙江	2	1	0	1	1	0	0	1	**4**
江西	3	1	1	0	0	0	1	0	**3**
江苏	4	1	1	0	0	0	0	1	**3**
河南	5	1	0	0	0	0	1	0	**2**
湖北	6	1	0	1	0	1	0	0	**3**
上海	7	0	1	0	1	0	0	0	**2**
邻接矩阵(标准化)									
安徽	1	0.00	0.20	0.20	0.20	0.20	0.20	0.00	1
浙江	2	0.25	0.00	0.25	0.25	0.00	0.00	0.25	1
江西	3	0.33	0.33	0.00	0.00	0.00	0.33	0.00	1
江苏	4	0.33	0.33	0.00	0.00	0.00	0.00	0.33	1
河南	5	0.50	0.00	0.00	0.00	0.00	0.50	0.00	1
湖北	6	0.33	0.00	0.33	0.00	0.33	0.00	0.00	1
上海	7	0.00	0.50	0.00	0.50	0.00	0.00	0.00	1

空间自相关指数

莫兰指数：评估一个区域内的观测值与其相邻区域观测值之间的相似程度

正数表示聚集分布，负数离散，0为随机分布

3/4

莫兰指数的计算公式

- 全局指数：

$$I=\frac{n}{S_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}(x_i-\bar{X})(x_j-\bar{X})}{\sum_{i=1}^{n}(x_i-\bar{X})^2}$$

- 局部指数：

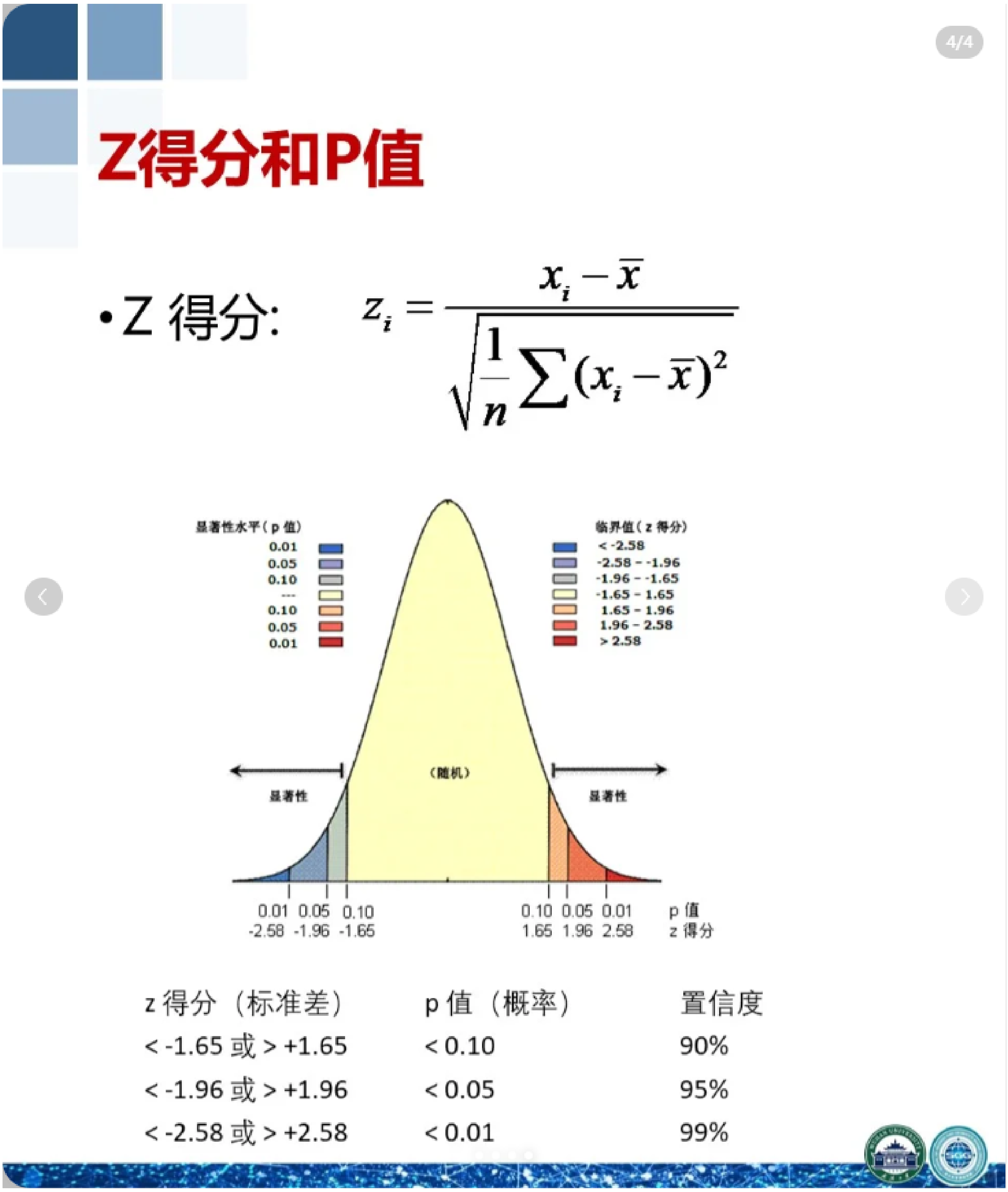
$$\boldsymbol{I_i}=\frac{\boldsymbol{x_i}-\bar{\boldsymbol{X}}}{\boldsymbol{S_i^2}}\sum_{j=1,j\neq i}^{n}\boldsymbol{w_{i,j}}(\boldsymbol{x_j}-\bar{\boldsymbol{X}})$$

x_i 和x_j：分区 i 和j的属性值，$\bar{X}$ 是研究区域属性值的平均值，$w_{i,j}$ 是分区 i 和 j 之间的空间权重，n 为分区总数。

$$S_0=\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}$$

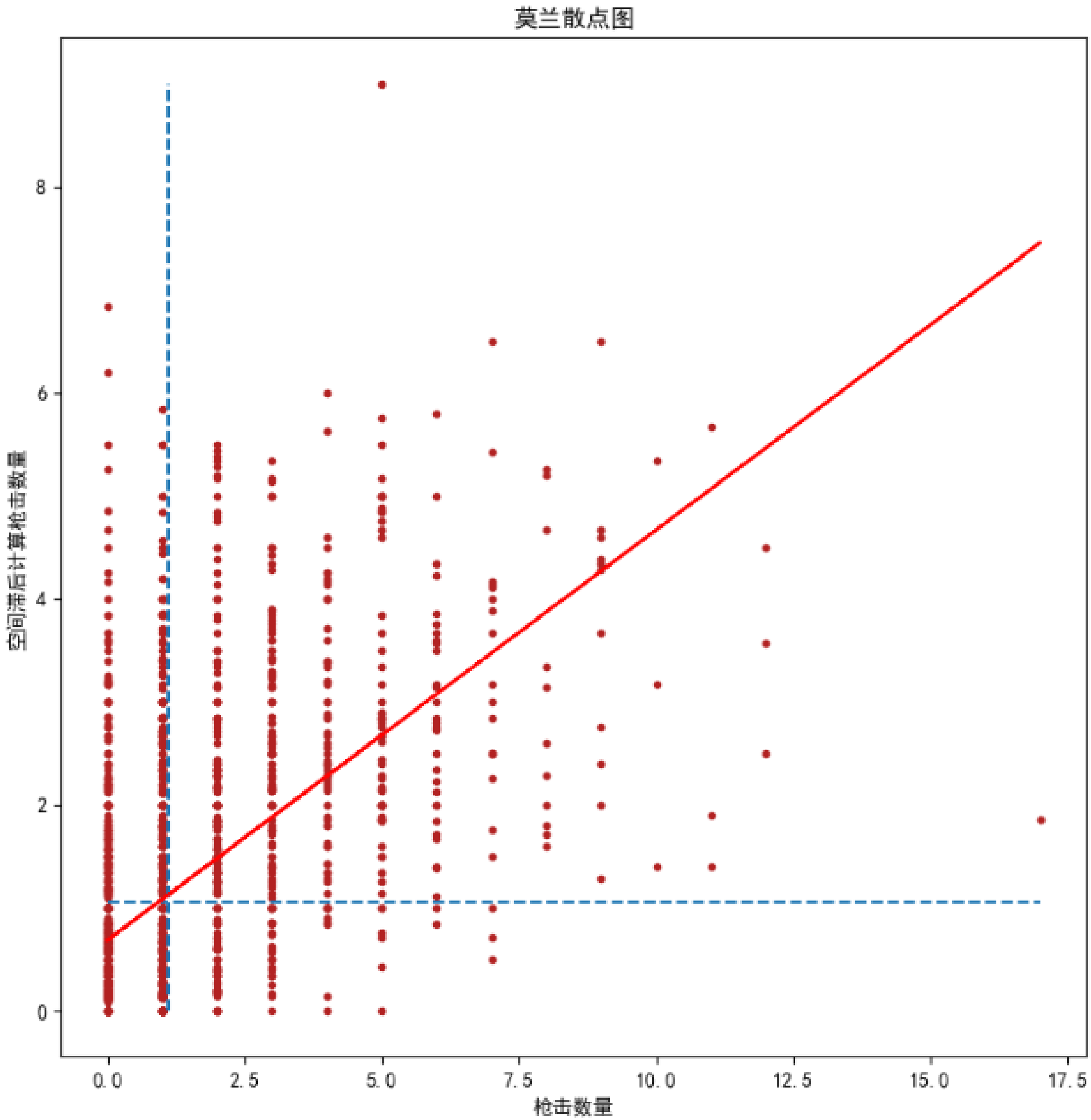
$$S_i^2=\frac{\sum_{j=1,j\neq i}^{n}\left(x_j-\bar{X}\right)^2}{n-1}$$

$$S^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\;;\quad \bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i\text{。}$$

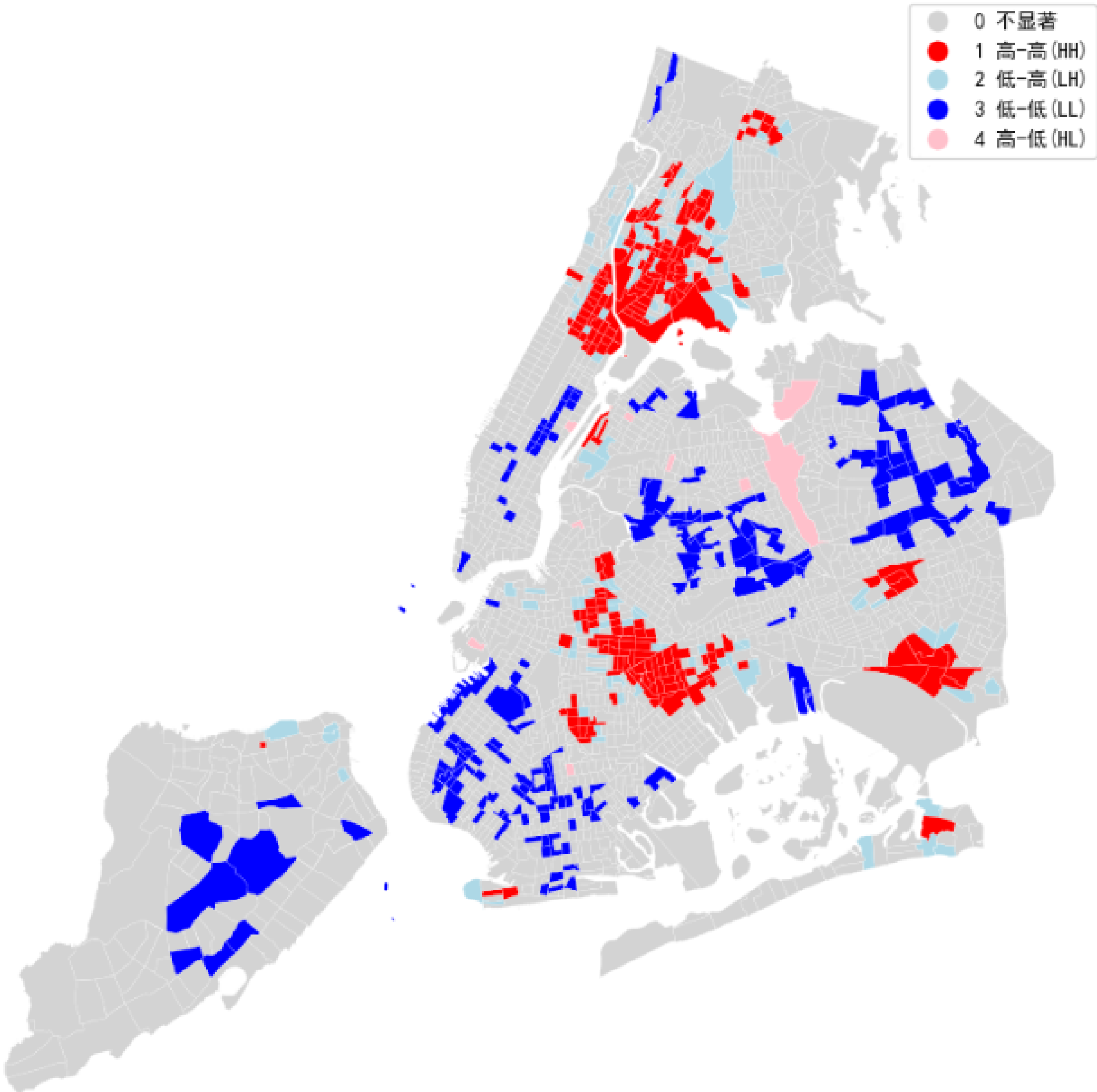

4/4
Z得分和P值
•Z 得分: $z_i = \dfrac{x_i - \bar{x}}{\sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}}$
显著性水平（p 值）
0.01
0.05
0.10

0.10
0.05
0.01
临界值（z 得分）
< -2.58
-2.58 - -1.96
-1.96 - -1.65
-1.65 - 1.65
1.65 - 1.96
1.96 - 2.58
> 2.58
（随机）
显著性
显著性
0.01 0.05 0.10 0.10 0.05 0.01 p 值
-2.58 -1.96 -1.65 1.65 1.96 2.58 z 得分
z 得分（标准差） p 值（概率） 置信度
< -1.65 或 > +1.65 < 0.10 90%
< -1.96 或 > +1.96 < 0.05 95%
< -2.58 或 > +2.58 < 0.01 99%

莫兰散点图：



LISA散点图


0 不显著
1 高-高(HH)
2 低-高(LH)
3 低-低(LL)
4 高-低(HL)