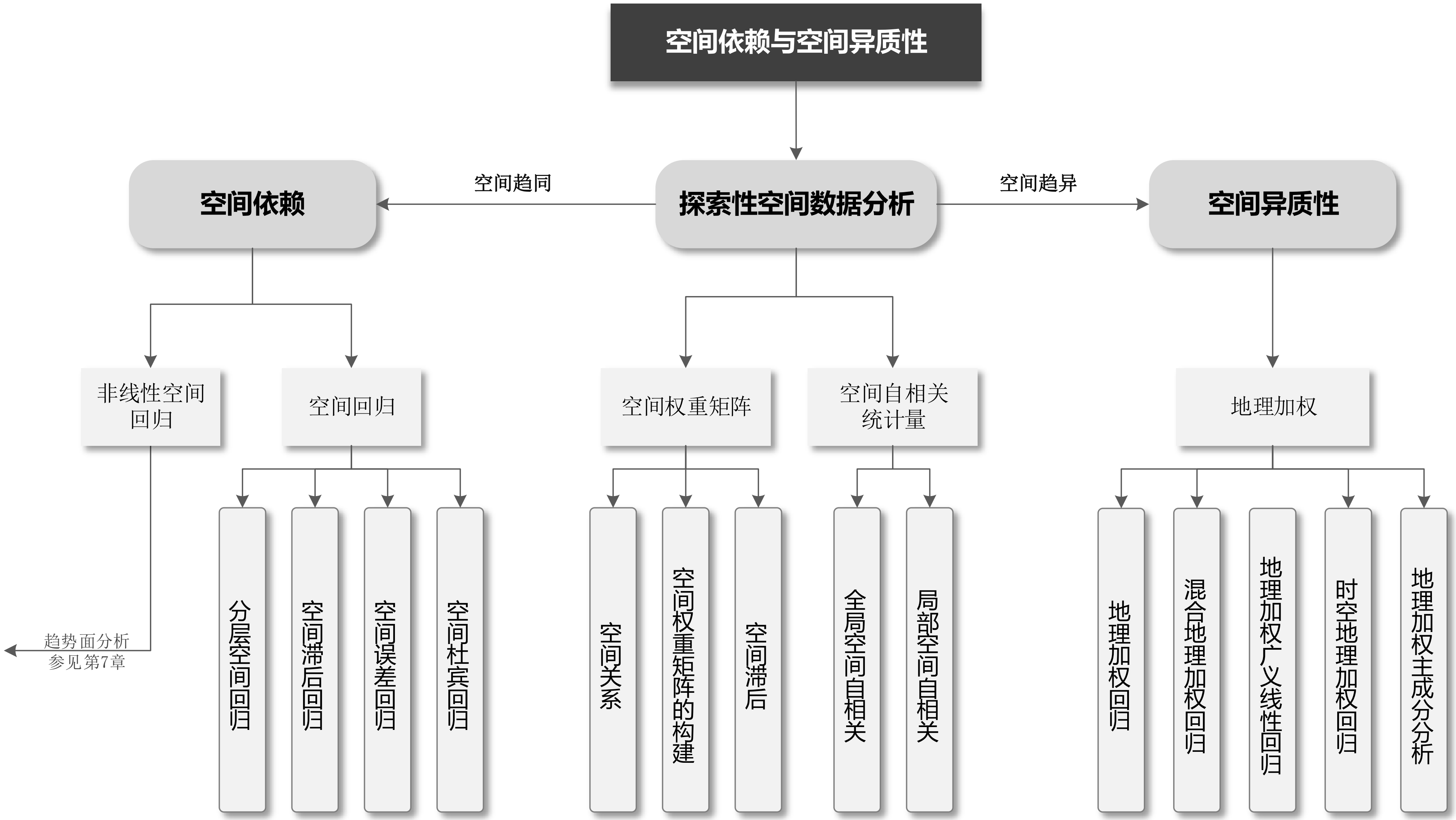
04

空间依赖与空间异质性

本章结构


空间依赖与空间异质性
空间依赖
空间趋同
探索性空间数据分析
空间趋异
空间异质性
非线性空间回归
空间回归
空间权重矩阵
空间自相关统计量
地理加权
分层空间回归
空间滞后回归
空间误差回归
空间杜宾回归
空间关系
空间权重矩阵的构建
空间滞后
全局空间自相关
局部空间自相关
地理加权回归
混合地理加权回归
地理加权广义线性回归
时空地理加权回归
地理加权主成分分析
趋势面分析
参见第7章

空间权重矩阵

空间关系

空间依赖反映的是一个区域某种地理现象或空间实体属性值与邻近区域同一地理现象或空间实体属性值的相关程度。

空间数据蕴含了丰富的位置信息，与空间数据位置有关的重要空间概念是：**邻接（Contiguity）和距离（Distance）**。这些概念都从某些方面描述了空间中对于邻近（Neighborhood）的定义。 Z

01 邻接关系

02 距离关系

空间权重矩阵

邻接关系

邻接关系:空间数据间的邻接关系描述空间单元间有公共边界且公共边界长度非0的的现象，可以认为是名义的、双向的和相等的距离。

Rook相邻

边界相邻，即两个地理单元有共同边界，则认为它们相邻，称为Rook相邻。

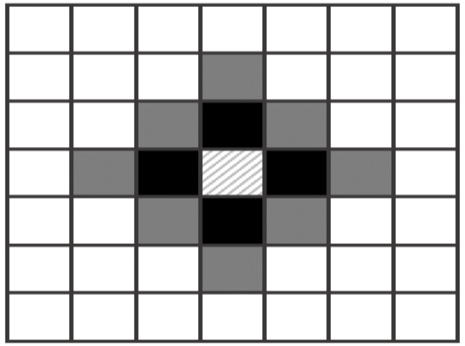
Bishop相邻

顶点相邻，即两个地理单元有公共顶点，则认为他们相邻，称为Bishop相邻。

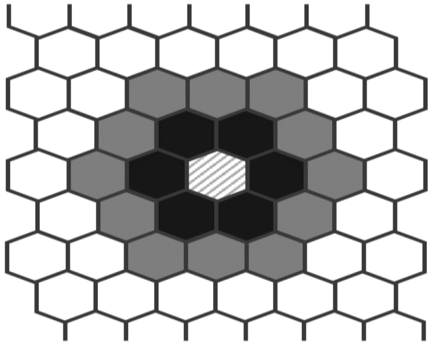
Queen相邻

边界或顶点相邻，即两个地理单元有共同边界或相同的顶点，则认为它们相邻，称为Queen相邻。

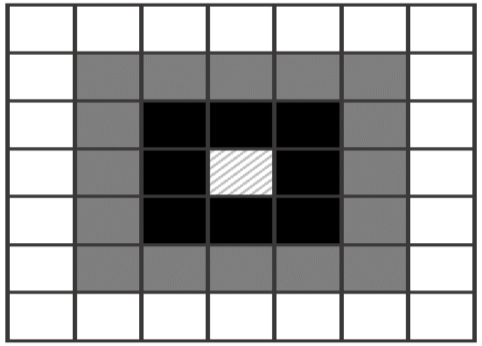
Rook



Bishop



Queen

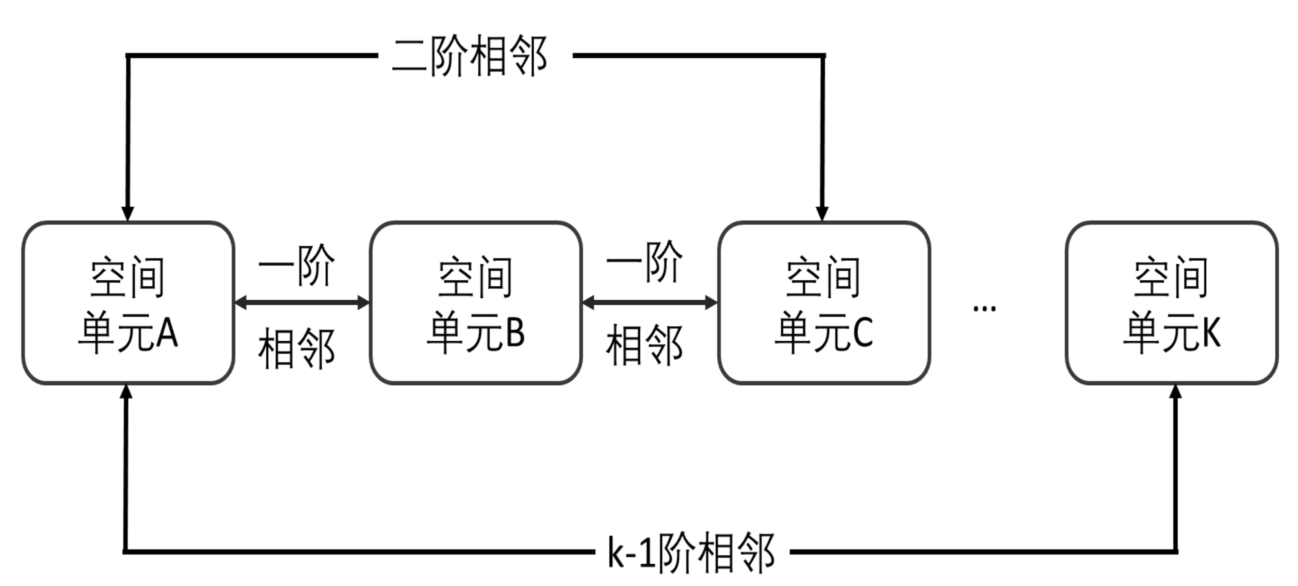


目标空间单元　一阶邻接　二阶邻接

空间权重矩阵

邻接关系

相邻关系示意图



构建空间权重矩阵时还要考虑相邻的阶。根据是否是直接邻接，可分为一阶邻接（First order spatial contiguity，即一阶邻接或直接邻接）、二阶邻接（Second order spacial contiguity，通过一阶邻接区域单元与其它区域单元形成的邻接）、高阶邻接（Higher order special contiguity，二阶邻接的推广。

距离关系

距离关系:空间数据中的距离是指空间对象间的直线距离或者球面距离。

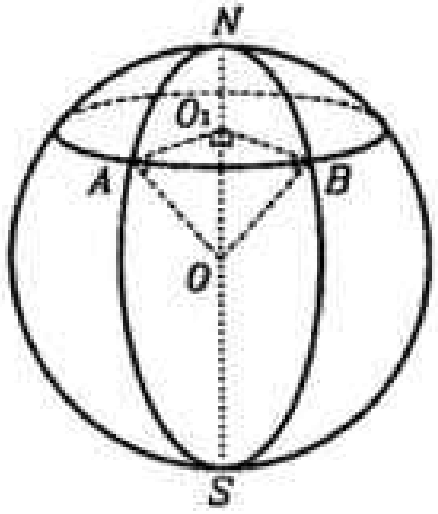
欧氏距离

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

球面距离

$$Lat_r = \frac{(Lat_d - 90) \times \pi}{180}$$

$$Lon_r = \frac{(Lon_d - 90) \times \pi}{180}$$



$$d_{ij} = R \times \arccos[\cos(\Delta Lon) \times \sin Lat_{r(i)} \times \sin Lat_{r(j)} + \cos Lat_{r(i)} \times \cos Lat_{r(j)}]$$

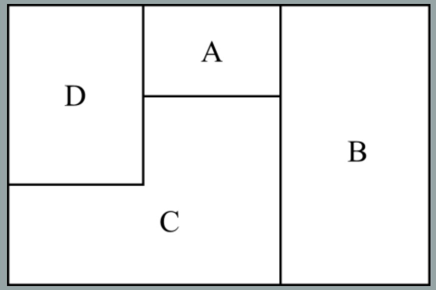
其中，R是地球曲率，$\Delta Lon = Lon_{r(i)} - Lon_{r(j)}$

空间权重矩阵的概念

空间权重矩阵定义了空间单元的相邻关系，决定了任意空间单元的特征对其邻近的空间单元贡献程度。空间权重矩阵通常用一个二元对称阵来表达n个空间单元之间的邻近关系。空间权重矩阵W可以表示如下：

$$\boldsymbol{W}=\begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}$$

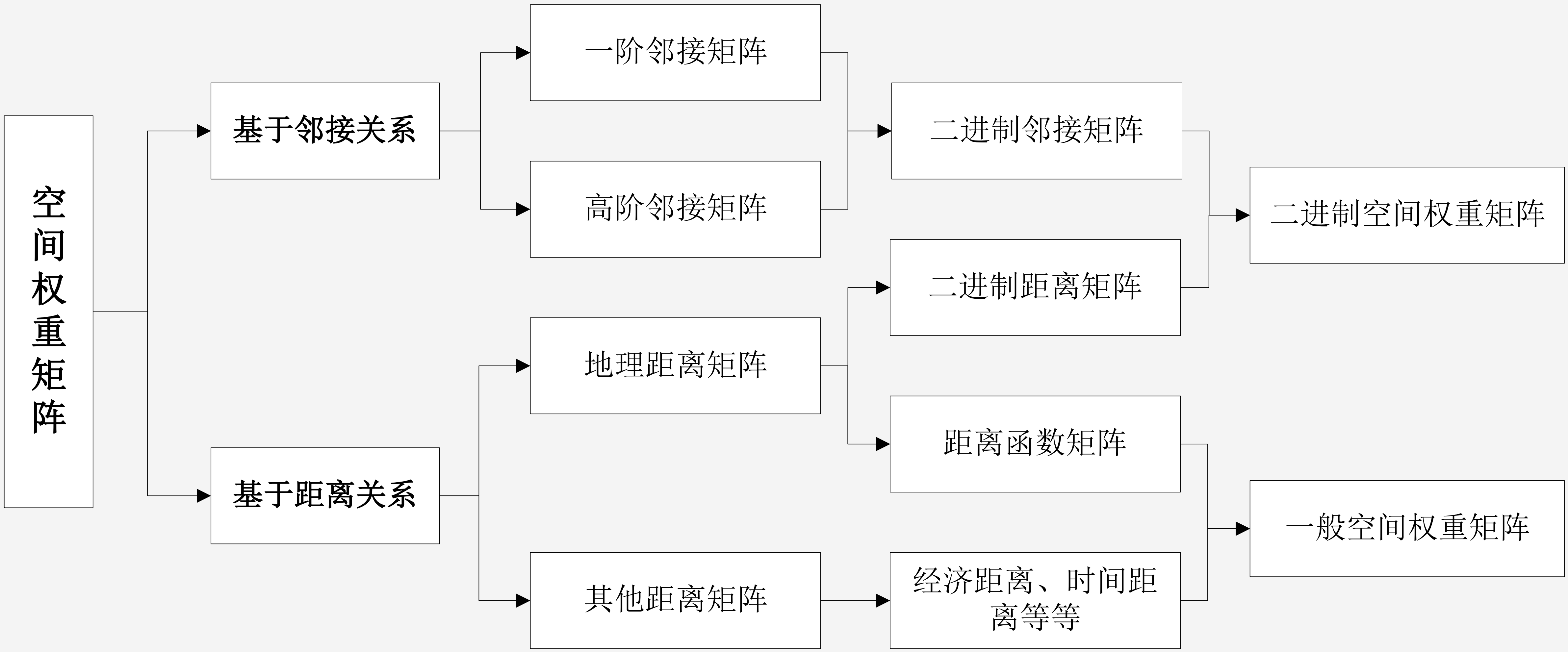
其中，wij表示区域i与j的邻近关系，且wij=wji。可以根据邻接标准或者距离标准来度量，其中对角线上的元素被设为0（即同一区域间的距离为0）。



A、B、C、D四个区域相邻，若采用Rook邻接矩阵描述其空间关系，其空间权重矩阵：

$$\boldsymbol{W}=\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

空间权重矩阵的构建


空间权重矩阵
基于邻接关系
基于距离关系
一阶邻接矩阵
高阶邻接矩阵
地理距离矩阵
其他距离矩阵
二进制邻接矩阵
二进制距离矩阵
距离函数矩阵
经济距离、时间距离等等
二进制空间权重矩阵
一般空间权重矩阵

空间权重矩阵

基于邻接关系构建

空间权重矩阵		公式	适用范围
邻接矩阵	Rook 邻接矩阵	$w_{ij}=\begin{cases}1 & \text{区域}i\text{和}j\text{共边} \\ 0 & i=j\text{或区域}i\text{和}j\text{不共边}\end{cases}$	多边形空间单元
	Queen 邻接矩阵	$w_{ij}=\begin{cases}1 & \text{当区域}i\text{和}j\text{共边或共顶点} \\ 0 & i=j\text{或区域}i\text{和}j\text{无共边且不共顶点}\end{cases}$	多边形空间单元

01 一阶邻接矩阵

与多数情况并不相符。

02 二元邻接矩阵

简单直观、设定方便且计算量小；但灵活性差

空间权重矩阵

基于距离关系构建

空间权重矩阵		公式	适用范围
距离函数矩阵	二进制地理距离矩阵	$w_{ij}=\begin{cases}1 & d_{ij}\le d_0\\ 0 & d_{ij}>d_0\end{cases}$	离散点空间单元
	阈值权重矩阵	$w_{ij}=\begin{cases}0 & i=j\\ a_1 & d_{ij}<d_0\\ a_2 & d_{ij}\ge d_0\end{cases}$	离散点空间单元
	Cliff-Ord矩阵	$w_{ij}=(d_{ij})^{-a}(\beta_{ij})^{b}$	多边形空间单元
	Decay权重矩阵	$w_{ij}=d_{ij}\cdot\alpha_i\cdot\beta_{ij}$	多边形空间单元
	K近邻矩阵	$w_{ij}=\begin{cases}1/d_{ij} & d_{ij}\le d_0^{(k)}\\ 0 & i=j\text{或}d_{ij}>d_0^{(k)}\end{cases}$	离散点空间单元

01 一定程度上克服二元邻接矩阵不能描述离散点间空间关系的缺陷。

02 多边形地理单元之间的距离根据各区域质心之间的欧式距离来确定。

空间权重矩阵

基于其他距离关系构建

空间权重矩阵		公式	适用范围
其他距离矩阵	基于引力模型的空间邻接矩阵	$w_{ij}=\begin{cases}\dfrac{m_i m_j}{d_{ij}^2} & i\neq j\\ 0 & i=j\end{cases}$	经济贸易等往来频繁的空间单元

空间权重矩阵

空间权重矩阵的标准化

行标准化

双重标准化

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

$$W=\begin{bmatrix}0 & 1 & 1 & 1\\ 1 & 0 & 1 & 0\\ 1 & 1 & 0 & 1\\ 1 & 0 & 1 & 0\end{bmatrix}$$

$$W=\begin{bmatrix}0 & 0.1 & 0.1 & 0.1\\ 0.1 & 0 & 0.1 & 0\\ 0.1 & 0.1 & 0 & 0.1\\ 0.1 & 0 & 0.1 & 0\end{bmatrix}$$

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

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

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。

空间权重矩阵

空间滞后

空间单元间存在一定的相互关系，而且距离越近产生这种关系的可能性就越强。空间依赖产生于直接邻接的空间实体间，也随着时间的推移扩散到邻近的区域，继而扩散至更多的空间单元。

空间滞后的数学表达：

$$y^* = Wy$$

其中，空间权重矩阵（这里也称空间滞后算子）*W*类似于时间序列分析的滞后算子值。

01 与时间滞后不同的是，空间滞后算子意味着空间上的推移，通过空间滞后算子可以得出实际上相邻的空间单元观测值依距离加权的平均值。

02 创建空间滞后算子的一个合适方法就是直接使用高阶邻接关系所创建的空间权重矩阵。

空间自相关统计量

空间自相关定义

空间自相关是指同一个变量在不同空间位置上的相关性，是空间依赖的一种度量。空间自相关性使用全局和局部两种指标。

01 **全局指标**

全局指标用于探测整个研究区域的空间模式，使用单一的值来反映该区域的自相关程度。

02 **局部指标**

局部指标计算每一个空间单元与邻近单元就某一属性的相关程度。

空间自相关统计量

全局空间自相关

01. 全局Moran's I

$$Moran's\ I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$$

标准化指数：$Z_{\alpha} = \frac{I - E(I)}{\sqrt{VAR(I)}}$

Moran’s I统计量的取值一般在-1到1之间，小于0表示负相关，等于0表示不相关，大于0表示正相关。

02. Geary’s C系数

$$C = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - x_j)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\sum_{k=1}^{n}(x_k - \bar{x})^2}$$

标准化指数：$Z(C) = \frac{(C - E(C))}{\sqrt{Var(C)}}$

Geary统计量C的取值一般在[0，2] 之间，大于1表示负相关，等于1表示不相关，小于1表示正相关。

空间自相关统计量

局部空间自相关指标

常见的局部空间自相关统计量是空间联系局部指标LISA（Local Indicators of Spatial Association）（Anselin，1995），这是一组统计量的合称，常用的包括局部Moran指数（Local Moran’s I）和局部Geary指数（Local Geary’s C）。

局部Moran指数

局部Moran指数被定义为：

$$I_i = \frac{(x_i - \bar{x})}{S^2}\sum_j \omega_{ij}(x_j - \bar{x})$$

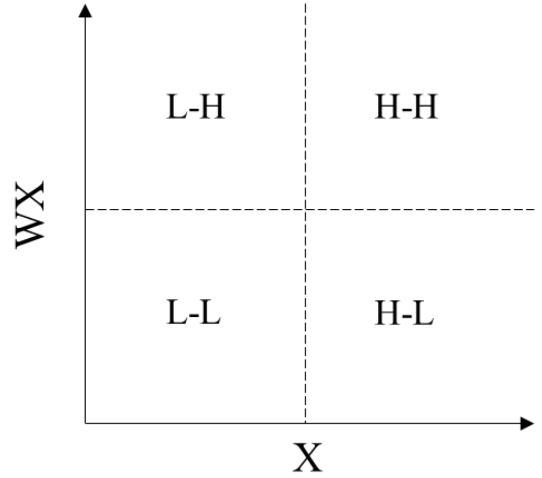
其中，$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$，$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Getis-Ord指数Gi

Gi指数计算公式为：

$$G_i^* = \frac{\sum_j w_{ij} x_i}{\sum_k x_k}$$

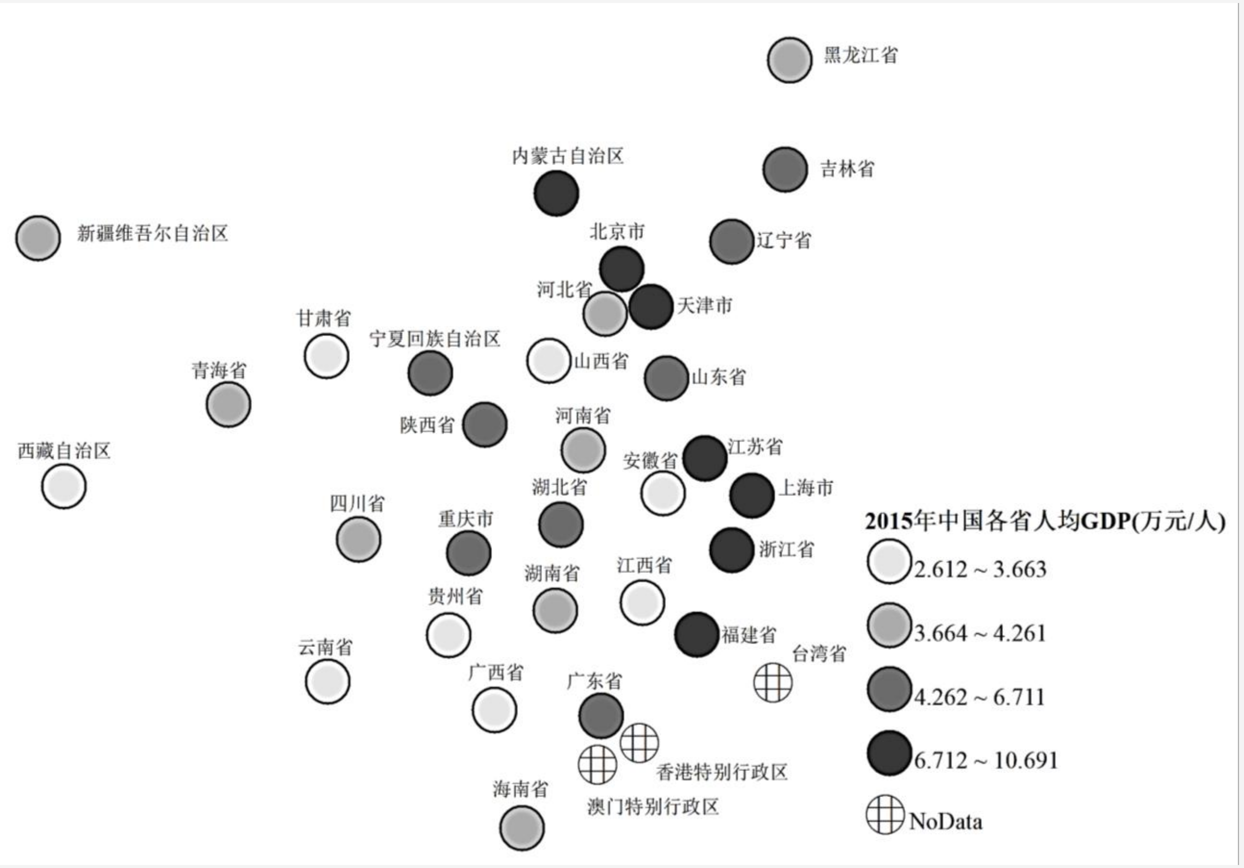
Moran散点图



应用实例

应用实例

本实例以2015年中国各省区人均国内生产总值（人均GDP）为指标探索中国各省区经济发展的空间自相关效应。数据取自国家统计局发布的2016年《中国统计年鉴》。



预处理

首先对2015年中国各省区人均GDP进行地理可视化，并使用四分位分级法进行分级。

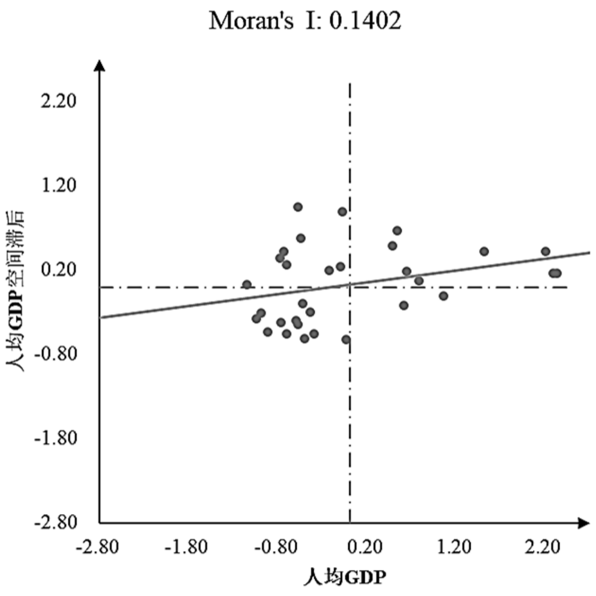
应用实例

应用实例

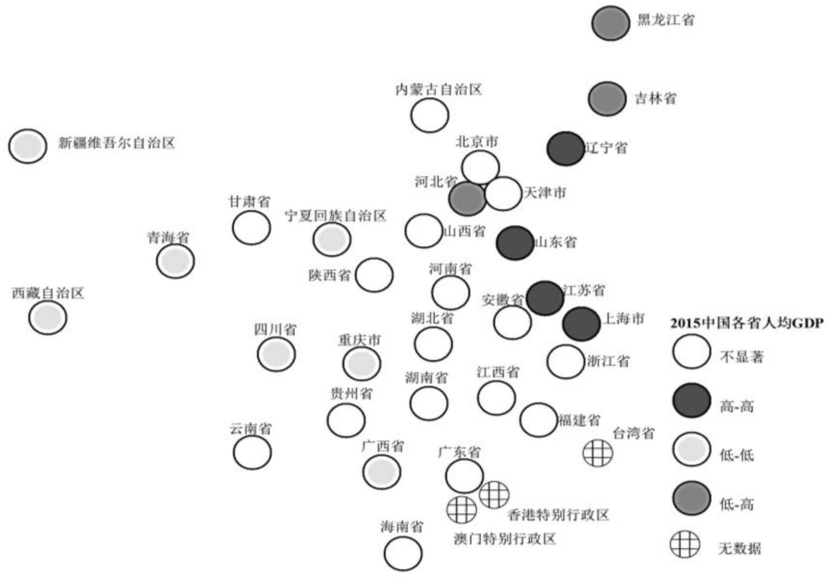
1.构建空间权重矩阵

以中国全境为例，对中国各省市构建一阶空间权重矩阵。

2.绘制Moran 散点图



3.绘制LISA聚类图



空间回归的一般形式

空间回归的一般形式

空间回归:通过在构建模型时显式地引入空间滞后变量，可以估算和检验空间自相关对空间关联的贡献。

Anselin（1988，1990）给出了空间回归模型的一般形式:

$$Y = \rho W_1 Y + X\beta + u$$

$$u = \lambda W_2 \varepsilon + \mu \qquad \mu \sim N[0, \sigma^2 I]$$

其中，Y是因变量；X是解释变量；β表示解释变量的空间回归系数，u是随空间变化的误差项；μ是白噪声；W1是反映因变量自身空间趋势的空间权重矩阵，W2为反映残差空间趋势的空间权重矩阵，通常根据邻接关系或者距离函数关系确定空间权重矩阵。ρ为空间滞后项的系数，其值为0到1，越接近1，说明相邻地区的因变量取值越相似；λ为空间误差系数，其值为0到1，越接近于1，说明相邻地区的解释变量取值越相似。其中，W1可以等于W2。

一般形式的空间自回归模型可以派生出其他几种模型:

01 当ρ=0，λ=0时，模型为普通线性回归模型 。

02 当ρ≠0，β=λ=0时，为一阶空间自回归模型。

03 当ρ≠0，β≠0，λ=0时，为空间滞后模型。

04 当$\rho=0$，$\beta\neq0$，$\lambda\neq0$时，为空间误差模型。

05 当$\rho\neq0$，$\beta\neq0$，$\lambda\neq0$时，为空间杜宾模型。

空间滞后回归

空间滞后回归

模型思想

在模型中引入空间滞后因子*WY*作为解释变量，认为相邻研究区域间的因变量存在空间自相关性

参数估计

Anselin（1988）给出了空间滞后模型的极大似然估计方法:

（1）构造ML统计量

令 $A = I - \rho W$，则空间滞后模型可以写为: $AY = X\beta + \varepsilon$

利用ML估计的一阶极值条件:

$$0 = X'B'\Omega^{-1}BAY - X'B'\Omega^{-1}BX\beta$$

并且令 $B = I$，解一阶条件得到 β的估计量为:

$$b = [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}AY$$

也即

$$b = [X'\Omega^{-1}X]^{-1}X'\Omega^{-1}Y - \rho[X'\Omega^{-1}X]^{-1}X'\Omega^{-1}WY$$

公式:

$$Y = \rho WY + X\beta + \varepsilon$$

$$\varepsilon \sim N[0, \sigma^2 I]$$

空间滞后回归

空间滞后回归

(2)ML估计步骤

1.对模型$Y = X\beta_1 + \varepsilon_1$
进行OLS估计；

2.对模型 $WY = X\beta_2 + \varepsilon_2$
进行OLS估计；

3.计算残差估计量

$$\varepsilon_1 = Y - X\hat{\beta}_1$$

$$\varepsilon_2 = WY - X\hat{\beta}_2$$

4.将残差估计量带入似然函数，找到使得似然函数取极大值的估计量ρ

$$L_c = C - (\frac{n}{2})\ln(\frac{1}{n})(\varepsilon_1 - \rho\varepsilon_2)'(\varepsilon_1 - \rho\varepsilon_2) + \ln(I_n - \rho W)$$

5.给定使L_c最大的$\hat{\rho}$，计算

$$\hat{\beta} = (\hat{\beta}_1 - \rho\hat{\beta}_2)$$

$$\hat{\sigma}_\varepsilon{}^2 = (\frac{1}{n})(\varepsilon_1 - \rho\varepsilon_2)'(\varepsilon_1 - \rho\varepsilon_2)$$

空间误差回归

空间误差回归

模型思想

空间相关性的存在不直接影响回归模型的结构，但此时误差项则存在着类似于空间滞后模型的结构

公式：

$$Y = X\beta + \varepsilon$$

$$\varepsilon = \lambda W\varepsilon + u$$

$$u \sim N[0, \sigma^2 I]$$

参数估计

空间误差模型的一般公式，令 $B = I - \lambda W$，则对数似然函数可以写成：

$$\ln L = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln\{|\Omega| \cdot [|B|]^{-2}\} - \frac{1}{2}[BY - BX\beta]'\Omega^{-1}[BY - BX\beta]$$

利用ML估计的一阶极值条件： $0 = X'B'\Omega^{-1}BAY - X'B'\Omega^{-1}BX\beta$

解一阶条件得到的β估计量为： $b = [X'B'\Omega^{-1}BX]^{-1}X'B'\Omega^{-1}BY$

假设随机项协方差矩阵 $\Omega = \sigma^2 I$，从而得到估计量：

$$b = [X'B'BX]^{-1}X'B'BY, \quad \hat{\Omega} = \frac{1}{n}[Be]'[Be] \cdot I$$

式中， $e = Y - Xb$，将 $\hat{\Omega}$ 和 b 带入似然函数，求解得：

$$\max_{\lambda}\{-\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln\{|\hat{\Omega}| \cdot [|B|]^{-2}\} - \frac{1}{2}[BY - BXb]'\hat{\Omega}[BY - BXb]\}$$

得到估计量 $\hat{\lambda}$，进一步可以利用 $\hat{B} = 1 - \hat{\lambda}W$，重新估计b，并反复迭代直到收敛。

空间杜宾回归

空间杜宾回归

模型公式

$$y = \rho W y + X\beta + W\overline{X}\gamma + \varepsilon$$

模型可以简化为：

$$y = (1-\rho W)^{-1}(X\beta + W\overline{X}\gamma + \varepsilon)$$

使用原因

当对样本区域数据进行空间回归建模时，同时存在：
①普通最小二乘回归模型的误差项中有空间相关性
②当处理区域样本数据的时候会有一些与模型中的解释变量的协方差不为0的解释变量被忽略

空间杜宾模型囊括其他几种模型：

01 当$\gamma = 0$时，它包含了因变量的空间滞后因素，而排除了空间滞后解释变量的因素，称为空间滞后回归模型。

02 当$\rho = 0$时，假设因变量之间的观测值不相关，但是因变量与相邻区域的特性有关，此时模型成为解释变量的空间滞后模型。

03 当$\gamma = 0$且$\rho = 0$，该模型成为标准最小二乘模型。

空间回归模型的检验

空间回归模型的检验

空间回归模型的检验主要是基于拉格朗日乘数（Lagrange Multiplier，LM）检验进行的。

1.不存在空间自回归时空间残差相关的LM检验

原假设: $H_0: Y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I_n)$

构造的检验统计量:

$$LM = \frac{(e'We/s^2)^2}{T} \sim \chi^2(1)$$

2.存在空间自回归时空间残差相关的LM检验

原假设: $H_0: Y = \rho WY + X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I_n)$

构造的检验统计量:

$$LM = \frac{[e'We/s^2 - T(R\tilde{J})^{-1}(e'WY/s^2)]^2}{T - T^2(R\tilde{J})^{-1}}$$

3.不存在空间残差相关时空间自回归效应相关的LM检验

原假设: $H_0: Y = X\beta + \varepsilon \quad \varepsilon \sim N(0, \sigma^2 I_n)$

构造的检验统计量:

$$LM = \frac{(e'We/s^2)^2}{R\hat{J}} \sim \chi^2(1)$$

4.存在空间残差相关时空间自回归效应相关的LM检验

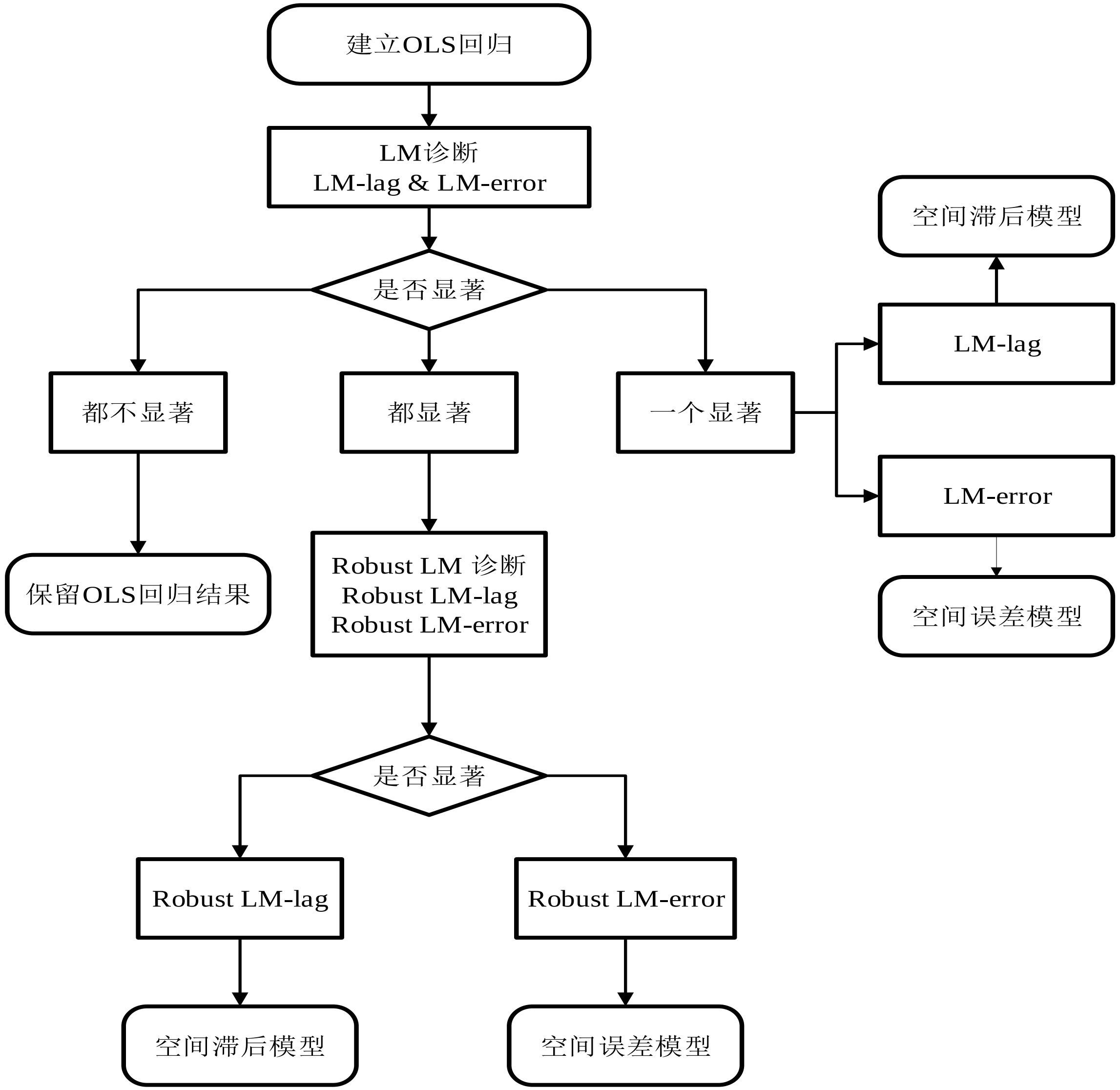
原假设: $H_0: Y = X\beta + \lambda W\varepsilon + \mu \quad \varepsilon \sim N(0, \sigma^2 I_n)$

构造的检验统计量:

$$LM = \frac{(e'WY/s^2 - e'We/s')^2}{R\hat{J} - T} \sim \chi^2(1)$$

空间回归模型的选择

空间回归模型的选择



空间回归模型的选择

空间回归模型的选择

1

首先建立OLS回归模型，借助LM统计量对回归结果作出空间自相关性诊断；

2

比较LM-lag和LM-error检验统计量。若两者都不显著，则保留OLS回归的结果；若LM-lag显著而LM-error不显著，则建立空间滞后模型；相应地，若LM-error显著而LM-lag不显著，则建立空间误差模型。若两者都显著，则转入3;

3

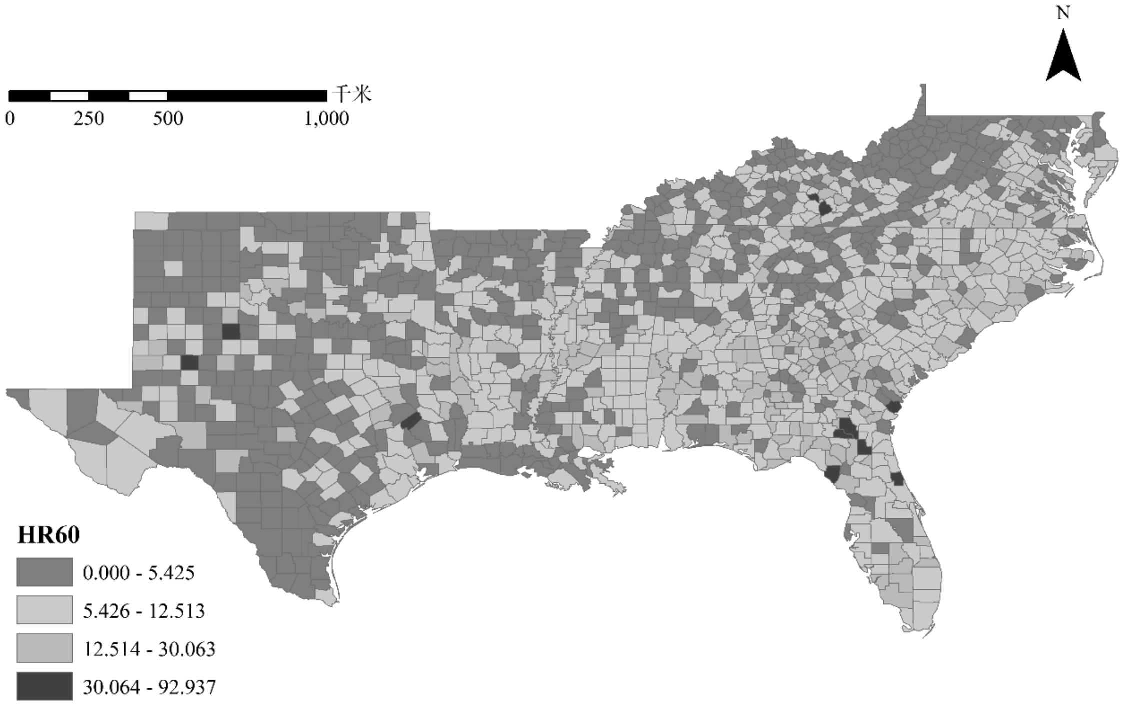
比较Robust LM-lag和Robust LM-error检验统计量。一般地，这两者只会有一个是显著的。如若不然，则比较两者的显著性程度，选择更显著的那个统计量对应的空间模型。

应用实例

应用实例

1.数据收集

数据源自美国南部县的室内极端案件与相关的社会经济数据，以10年为单元分四次进行统计，分别是1960，1970，1980和1990年（Messne et.al.，2000；Baller et.al.，2001）。



1960年美国南部室内案件发生率空间分布图

应用实例

应用实例

2.建立OLS模型

结论：传统的一般线性回归模型并不适用，根据空间回归模型的选择条件，可以发现这里应选用空间滞后模型进行建模。

回归结果总结：最小二乘回归

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

因变量	Dependent Variable : HR60	Number of Observations: 1412	观测样本个数
因变量均值	Mean dependent var : 7.29214	Number of Variables : 6	自变量个数
因变量标准差	S.D. dependent var : 6.41874	Degrees of Freedom : 1406	自由度
判定系数R^2	R-squared : 0.103657	F-statistic : 32.5193	F统计量
调整R^2	Adjusted R-squared : 0.100470	Prob(F-statistic) :1.85631e-031	P值
残差平方和	Sum squared residual: 52144.5	Log likelihood : -4551.5	对数似然值
方差	Sigma-square : 37.0872	Akaike info criterion : 9115.01	赤池信息准则（AIC）
回归标准误	S.E. of regression : 6.08992	Schwarz criterion : 9146.52	施瓦兹准则（SC）
方差（极大似然）	Sigma-square ML : 36.9296		
回归标准误（极大似然）	S.E of regression ML: 6.07697		

	变量名 Variable	回归系数 Coefficient	回归系数标准误 Std.Error	t统计量 t-Statistic	P值 Probability
常数项	CONSTANT	13.2155	1.12457	11.7516	0.00000
自变量	RD60	1.76448	0.198244	8.90057	0.00000
自变量	PS60	0.299302	0.214257	1.39693	0.16266
自变量	MA60	-0.275209	0.0380642	-7.23014	0.00000
自变量	DV60	1.17945	0.243517	4.84341	0.00000
自变量	UE60	-0.291856	0.0711715	-4.10074	0.00004

回归诊断 REGRESSION DIAGNOSTICS

多重共线性条件数 MULTICOLLINEARITY CONDITION NUMBER 18.899123

误差正态性检验 TEST ON NORMALITY OF ERRORS

	TEST	DF	VALUE	PROB
Jarque-Bera检验	Jarque-Bera	2	87427.8750	0.00000

异方差性诊断 DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

	TEST	DF	VALUE	PROB
Breusch-Pagan检验	Breusch-Pagan test	5	599.4759	0.00000
Koenker-Bassett检验	Koenker-Bassett test	5	30.1069	0.00001

空间自相关性诊断 DIAGNOSTICS FOR SPATIAL DEPENDENCE

空间权重矩阵归一化方法 FOR WEIGHT MATRIX : (row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.1356	8.3495	0.00000
Lagrange Multiplier (lag)	1	80.3219	0.00000
Robust LM (lag)	1	19.0122	0.00001
Lagrange Multiplier (error)	1	66.4285	0.00000
Robust LM (error)	1	5.1188	0.02367
Lagrange Multiplier (SARMA)	2	85.4407	0.00000

应用实例

应用实例

3. 构建空间滞后模型

对数似然指数增长，AIC和SC指数都相对于OLS下降，证明空间滞后变量有助于改进模型的拟合程度。

回归结果总结：空间滞后回归——最大似然估计

	SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION	
	Data set : south	
	Spatial Weight : south	
因变量	Dependent Variable : HR60 Number of Observations: 1412	观测样本个数
因变量均值	Mean dependent var : 7.29214 Number of Variables : 7	自变量个数
因变量标准差	S.D. dependent var : 6.41874 Degrees of Freedom : 1405	自由度
滞后系数	Lag coeff. (Rho) : 0.291639	
判定系数R^2	R-squared : 0.158910 Log likelihood : -4518.64	对数似然值
	Sq. Correlation : - Akaike info criterion : 9051.28	赤池信息准则（AIC）
残差平方和	Sigma-square : 34.6531 Schwarz criterion : 9088.05	施瓦兹准则（SC）
回归标准误	S.E of regression : 5.88669	

	变量名 Variable	回归系数 Coefficient	回归系数标准误 Std.Error	t统计量 z-value	P值 Probability
因变量滞后	W_HR60	0.291639	0.0363962	8.01289	0.00000
常数项	CONSTANT	9.26817	1.15997	7.98999	0.00000
自变量	RD60	1.37722	0.197301	6.98029	0.00000
自变量	PS60	0.18666	0.20738	0.900084	0.36808
自变量	MA60	-0.20248	0.0372972	-5.42882	0.00000
自变量	DV60	0.929869	0.235674	3.94557	0.00008
自变量	UE60	-0.199602	0.0689376	-2.8954	0.00379

回归诊断	REGRESSION DIAGNOSTICS
异方差性诊断	DIAGNOSTICS FOR HETEROSKEDASTICITY
	RANDOM COEFFICIENTS
	TEST DF VALUE PROB
Breusch-Pagan检验	Breusch-Pagan test 5 671.3716 0.00000
空间自相关性诊断	DIAGNOSTICS FOR SPATIAL DEPENDENCE
空间权重矩阵归一化方法	SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : south
	TEST DF VALUE PROB
似然比检验	Likelihood Ratio Test 1 65.7278 0.00000

空间异质性

01	地理加权回归
02	混合地理加权回归模型
03	地理加权广义线性回归
04	时空地理加权回归
05	地理加权主成分分析
06	应用实例

空间异质性

- 各种事物和现象在空间上缺乏平稳的结构
- 空间单元本身不是均质的，在面积和形状上具有较大差别

地理加权回归

基本模型

回归模型关系

- 线性回归模型
  - 全局模型
    - 一元线性回归模型
    - 多元线性回归模型
    - …
  - 局部模型
    - 地理加权回归模型
    - …

基本模型

Fortheringham等人（1996）基于局部平滑的思想，提出了地理加权回归模型，将数据的空间位置嵌入到回归参数中，利用局部加权最小二乘方法进行逐点参数估计，其中权是回归点所在的地理空间位置到其他各观测点的地理空间位置之间的距离函数。

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

其中，(ui，vi)是第i个采样点的坐标，βk(ui，vi) 是第i个采样点上第k个回归参数，是地理位置的函数。εi是第i个区域的随机误差，满足零均值、同方差、相互独立等基本假定。

地理加权回归

参数估计

利用加权最小二乘法（Weighted Least Square，WLS）来估计参数，i点的回归参数可通过使如下的式子达到最小来进行估计

$$\sum_{j=1}^{n} w_{ij}(y_j - \beta_{i0} - \sum_{k=1}^{p}\beta_{ik}x_{ik})$$

此时wij为回归点i与其他观测点j之间的地理距离dij的单调递减函数。

令 $\beta_i = [\beta_{1k}\ \ \beta_{2k}\ \ \cdots\ \ \beta_{np}]', W_i = diag(\mathrm{w}_{i1}, \mathrm{w}_{i1}, \cdots, \mathrm{w}_{in})$ ，这里的空间权重矩阵是对角阵。则i点上的回归参数估计为：

$$\beta_i = (X'W_iX)^{-1}X'W_iy$$

$$\hat{y}_i = X_i\beta_i = X_i(X'W_iX)^{-1}X'W_iy$$

令 $\mathrm{S_i} = X_i(X'W_iX)^{-1}X'W_i$ ，称为i点的帽子向量，则该点的残差 $e_i = y_i - \hat{y}_i = y_i - S_iy_i$ 。

按照如上的方法逐点进行回归计算，可以得到各采样点上回归参数的估计矩阵如下：

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_{10} & \hat{\beta}_{20} & \cdots & \hat{\beta}_{\mathrm{n}0} \\ \hat{\beta}_{11} & \hat{\beta}_{21} & \cdots & \hat{\beta}_{n1} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\beta}_{1p} & \hat{\beta}_{2p} & \cdots & \hat{\beta}_{np} \end{bmatrix}$$

其中每一行表示同一个解释变量的回归参数在不同采样点上的估计值。由此，可以求得各个采样点上的回归值为：

$$\hat{y} = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{bmatrix} y = \begin{bmatrix} X_1(X'W_1X)^{-1}X'W_1 \\ X_2(X'W_2X)^{-1}X'W_2 \\ \vdots \\ X_n(X'W_nX)^{-1}X'W_n \end{bmatrix} y = Sy$$

由此可以计算残差：

$$e = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{bmatrix} y = (1-S)y$$

地理加权回归

参数估计

令RSS表示残差平方和，则

$$SSE = (e'e) = y'(I - S)'(I - S)y$$

假设拟合值 $\hat{y}_i$ 为$E(\hat{y}_i)$的无偏估计，即$E(\hat{y}_i)=E(y_i)$，则有:

$$\begin{aligned} SSE = (e'e) &= (e - E(e))'(e - E(e)) \\ &= (y - E(y))'(I - S)'(I - S)(y - E(y)) \\ &= \varepsilon'(I - S)'(I - S)\varepsilon \end{aligned}$$

从而有:

$$\begin{aligned} E(SSE) &= E(tr[\varepsilon'(I - S)'(I - S)\varepsilon]) \\ &= tr[(I - S)'(I - S)E(\varepsilon'\varepsilon)] \\ &= \sigma^2 tr[(I - S)'(I - S)] \\ &= \sigma^2 (n - 2tr(S) + tr(S'S)] \end{aligned}$$

这样随机误差项的方差 σ^2 的无偏估计 $\hat{\sigma}^2$ 为:

$$\hat{\sigma}^2 = \frac{SSE}{n - 2tr(S) + tr(S'S)}$$

在很多情况下， $tr(S)$ 非常接近$tr(S'S)$，因此上述式子可以进一步简化为:

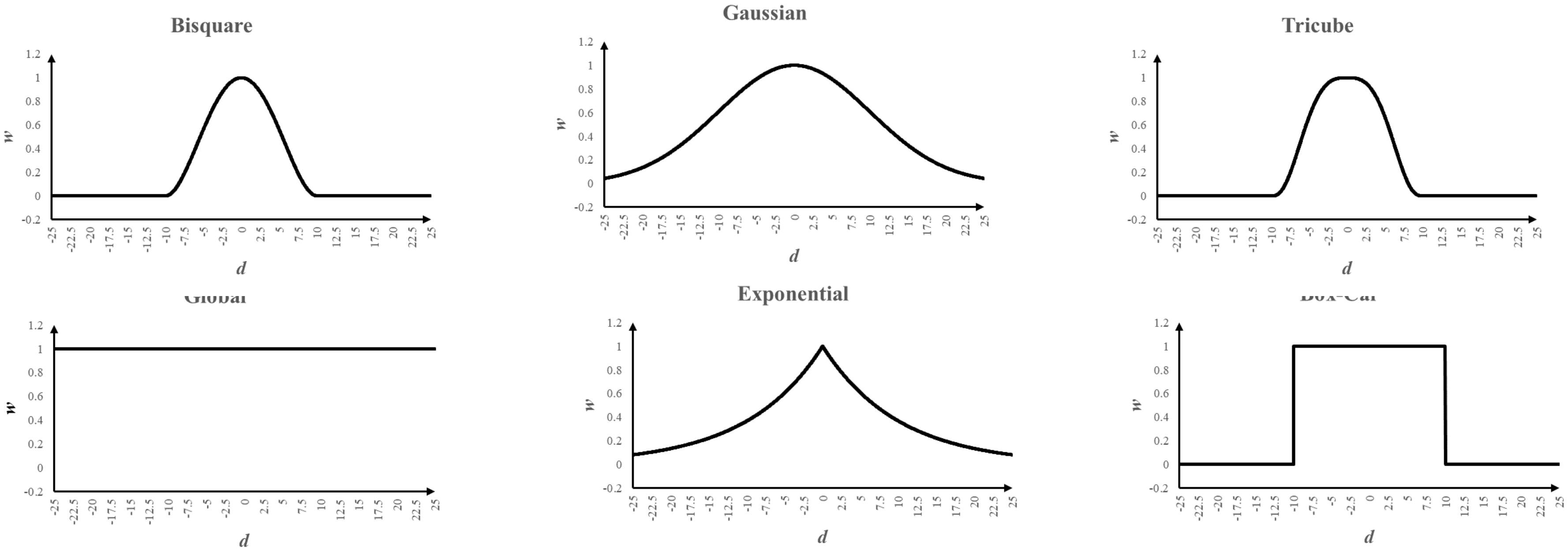
$$\hat{\sigma}^2 = \frac{SSE}{n - tr(S)}$$

空间权函数

常见的空间权函数

核函数名称	空间权函数
全局函数(Global)	$w_{ij}=1$
距离阈值函数(Box-Car)	$w_{ij}=\begin{cases}1 & d_{ij}\le b\\ 0 & d_{ij}>b\end{cases}$
指数型(Exponential)	$w_{ij}=e^{(-\frac{\lvert d_{ij}\rvert}{b})}$
高斯型(Gaussian)	$w_{ij}=e^{-(\frac{d_{ij}}{b})^2}$
双重平方(Bi-square)	$w_{ij}=\begin{cases}(1-(\frac{d_{ij}}{b})^2)^2 & \lvert d_{ij}\rvert<b\\ 0 & \lvert d_{ij}\rvert\ge b\end{cases}$
三次立方(Tri-cube)	$w_{ij}=\begin{cases}(1-(\frac{d_{ij}}{b})^3)^3 & \lvert d_{ij}\rvert<b\\ 0 & \lvert d_{ij}\rvert\ge b\end{cases}$

空间权函数



权函数的几种表现形式（b=10）

地理加权回归

带宽选择优化

1. 交叉验证法（Cross Validation）

$$CV = \sum_{i=1}^{n}[y_i - \hat{y}_{\neq i}(b)]^2$$

把不同的带宽b及其对应的CV值绘制成趋势线，就可以非常直观地找到最小的CV值所对应的最优带宽b。

2. AIC准则

$$AIC = \text{-}2\ln L(\hat{\theta}_L, x) + 2q$$

选择AIC达到最小的模型是“最优”的模型。

3. 贝叶斯信息准则（Bayesian Information Criterion）

$$BIC = -2\ln L(\hat{\theta}_L, \mathrm{x}) + q\,\mathrm{lnn}$$

BIC最小的模型为“最优”模型。

4.平稳指数（Index of Stationarity）

$$SI_i = \frac{IQR(SE_{(GWR)i1}, SE_{(GWR)i2}, \ldots, SE_{(GWR)ij}, \ldots, SE_{(GWR)in})}{2SE_{(OLS)i}}$$

将使得所有变量的曲线都趋于平稳的带宽作为合适的参数对估计GWR模型。

自适应权函数

常见的自适应空间权函数

核函数名称	
高斯型(Gaussian)	$w_{ij}=e^{-(\frac{d_{ij}}{b_{i(k)}})^2}$
双重平方 (Bi-square)	$w_{ij}=\begin{cases}(1-(\frac{d_{ij}}{b_{i(k)}})^2)^2 & \lvert d_{ij}\rvert<b_{i(k)}\\ 0 & \lvert d_{ij}\rvert\geq b_{i(k)}\end{cases}$

其中bi(k)是第i个数据点取周边最邻近的k个空间数据单元。

假设检验

假设检验

(1)回归模型的空间非平稳性检验

H0：$\hat{y}_S = Sy$ 的拟合优度与 $\hat{y}_H = Hy$ 的拟合优度无明显差异；

检验统计量：$F_1 = \frac{SSE_H}{SSE_S}$ 或 $F_2 = \frac{SSE_H - SSE_S}{SSE_S}$

(2)回归参数的空间非平稳性检验

H0：$\beta_{1k} = \beta_{2k} = \cdots = \beta_{nk}$

检验统计量：$V_k^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_{ik} - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{ik})^2$

(3)回归模型的空间非平稳性AIC比较

若 $(AIC_{OLS} - AIC_{GWR}) > 3$，则判定因变量与自变量之间具有明显的空间非平稳性，反之则判定普通线性回归模型比地理加权模型更接近真实模型。

混合地理加权回归模型

混合地理加权回归模型

模型思想

让回归模型中的一部分回归参数随地理位置而变，称为变参数，而其余回归参数为常数，称为常参数

公式

$$y_i = \beta_0 + \sum_{k=1}^{p_a} \beta_k x_{ik} + \sum_{l=1}^{p_b} \beta_k x_{ik} + \varepsilon_i$$

其中，

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \beta_a = \begin{bmatrix} \beta_{1_a} \\ \beta_{2_a} \\ \vdots \\ \beta_{p_a} \end{bmatrix} \quad \beta_b = \begin{bmatrix} \beta_{1_b} \\ \beta_{2_b} \\ \vdots \\ \beta_{p_b} \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$X_a = \begin{bmatrix} 1 & x_{11_a} & \cdots & x_{1p_a} \\ 1 & x_{21_a} & \cdots & x_{2p_a} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1_a} & \cdots & x_{np_a} \end{bmatrix} \quad X_b = \begin{bmatrix} 1 & x_{11_b} & \cdots & x_{1p_b} \\ 1 & x_{21_b} & \cdots & x_{2p_b} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1_b} & \cdots & x_{np_b} \end{bmatrix} \quad \mathrm{m} = \begin{bmatrix} \sum_{l=1}^{p_b} \beta_{1l_b} x_{1l_b} \\ \sum_{l=1}^{p_b} \beta_{2l_b} x_{2l_b} \\ \vdots \\ \sum_{l=1}^{p_b} \beta_{nl_b} x_{nl_b} \end{bmatrix}$$

模型也可以写成矩阵形式： $y = X_a \beta_a + m + \varepsilon$

由此可以看出，若保留$X_a\beta_a$而将m去掉，则混合地理加权回归模型就变为普通线性回归方程，若保留m而将$X_a\beta_a$去掉，则混合地理加权回归模型则变为地理加权回归模型。由此可见，普通线性回归模型和地理加权回归模型都可以看成是混合地理加权回归模型的特殊形式。

地理加权广义线性回归

地理加权广义线性回归

模型思想

地理加权广义线性回归则是对普通广义线性回归模型的扩展，将数据的地理位置嵌入到回归参数中

定义地理加权广义线性回归模型：

若因变量y服从指数分布族，其概率密度函数为：

$$f(y_j \mid \theta_{ij}, \phi_i) = \exp\left(\frac{y_j\theta_{ij} - b(\theta_{ij})}{\phi_i} + c(y_j, \phi_i)\right)$$

$$\eta_{ij} = \beta_{i1}(u_i, v_i)x_{j1} + \beta_{i2}(u_i, v_i)x_{j2} + \ldots + \beta_{ip}(u_i, v_i)x_{jp} + \varepsilon_{ij}$$

i点回归参数的估计为：

$$\sum_{j=1}^{n} W_{ij}(u_i, v_i)(y_i - \beta_{i0} - \sum_{k=1}^{p} \beta_{ik} x_{jk})^2$$

其中，$W_{ij}(u_i, v_i)$为位置(u_i, v_i)的空间权重矩阵，则有：

$$\hat{\beta}(u_i, v_i) = (X' W_{ij}(u_i, v_i) X)^{-1} X' W_{ij}(u_i, v_i) Y$$

对所有样本点进行逐点回归计算，得到所有点回归参数的估计值，由于不同采样点上的估计值不同，它反映了该参数对应的变量间的关系在研究区域内的变化情况，这样就可以探测这种空间关系的非平稳性。

地理加权广义线性回归

地理加权广义线性回归

地理加权逻辑回归

假设地理位置i发生事件的概率为p，则不发生该事件的概率为(1-p)，则:

$$P(y=1)=\frac{\exp(\beta_0(u_i,v_i)+\beta_{i1}(u_i,v_i)x_{j1}+\beta_{i2}(u_i,v_i)x_{j2}+\ldots+\beta_{ip}(u_i,v_i)x_{jp})}{1+\exp(\beta_0(u_i,v_i)+\beta_{i1}(u_i,v_i)x_{j1}+\beta_{i2}(u_i,v_i)x_{j2}+\ldots+\beta_{ip}(u_i,v_i)x_{jp})}=\frac{e^z}{1+e^z}$$

其中, $z=\beta_0(u_i,v_i)+\beta_{i1}(u_i,v_i)x_{j1}+\beta_{i2}(u_i,v_i)x_{j2}+\ldots+\beta_{ip}(u_i,v_i)x_{jp}$

经过Logit变换，有

$$\log it(p)=\ln\frac{p}{1-p}=\beta_0(u_i,v_i)+\beta_{i1}(u_i,v_i)x_{j1}+\beta_{i2}(u_i,v_i)x_{j2}+\ldots+\beta_{ip}(u_i,v_i)x_{jp}$$

其参数估计可以是

$$\hat{\beta}(u_i,v_i)=(X'W_{ij}(u_i,v_i)X)^{-1}X'W_{ij}(u_i,v_i)\log it(P)$$

其中,(u_i,v_i)为位置i的地理坐标，X为解释变量矩阵。

地理加权泊松回归

基本表达式如式:

$$\log(\mu)=\beta_0(u_i,v_i)+\beta_{i1}(u_i,v_i)x_{j1}+\beta_{i2}(u_i,v_i)x_{j2}+\ldots+\beta_{ip}(u_i,v_i)x_{jp}$$

对应的参数估计方程:

$$\hat{\beta}(u_i,v_i)=(X'W_{ij}(u_i,v_i)X)^{-1}X'W_{ij}(u_i,v_i)\log(\mu)$$

其中，(u_i,v_i)为位置i的地理坐标，X为解释变量矩阵。

时空地理加权回归

时空地理加权回归

模型思想

通过在传统的地理加权回归模型中引入时间维度的概念，使得回归系数是地理位置和观测时刻的函数，从而可以将数据的时空特征纳入到回归模型中。

基本表达式：

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_k \beta_k(u_i, v_i, t_i)x_{ik} + \varepsilon_i$$

其中，(ui，vi，ti)是第i个采样点的时空坐标，βk(ui，vi，ti) 是第i个采样点上第k个时空回归参数，是地理位置和时间的函数。εi是第i个时空区域的随机误差。

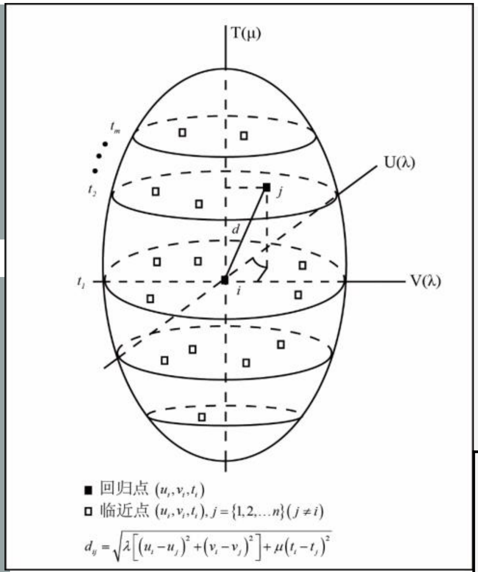
得到的参数估计式：

$$\hat{\beta}(u_i, v_i, \mathrm{t}_i) = (X^T W(u_i, v_i, \mathrm{t}_i) X)^{-1} X^T W(u_i, v_i, \mathrm{t}_i) y$$

时空地理加权回归

时空权函数的选择

椭圆坐标系统　u，v分别代表一个时间截面上的空间位置



时空权重概念模型

时间因素与空间因素对于“时空距离”的影响公式：

$$d^{ST}=d^{S}\otimes d^{T}$$

式中，d^S 和 d^T 分别代表两个样本点之间的空间和时间距离，d^{ST} 是时空距离，代表时间和空间的复合影响，⊗是一个运算符，表征这两种距离复合成为时空距离的运算。

加权方式
- 固定距离核函数：时空距离固定
- 自适应核函数：样本点个数指定

常见的核函数

核函数名称	固定距离核函数	自适应核函数
高斯型 (Gaussian)	$w_{ij}=e^{-(\frac{d_{ij}}{b})^2}$	$w_{ij}=e^{-(\frac{d_{ij}}{b_{i(k)}})^2}$
双重平方 (Bi-square)	$w_{ij}=\begin{cases}(1-(\frac{d_{ij}}{b_i})^2)^2 & \lvert d_{ij}\rvert<b_i\\ 0 & \lvert d_{ij}\rvert\geq b_i\end{cases}$	$w_{ij}=\begin{cases}(1-(\frac{d_{ij}}{b_{i(k)}})^2)^2 & \lvert d_{ij}\rvert<b\\ 0 & \lvert d_{ij}\rvert\geq b\end{cases}$

时空地理加权回归

时空权函数的选择

如果使用欧氏距离和高斯核函数来构造时空权重矩阵，可以得到

$$d_{ij}^{ST}=\sqrt{\lambda[(u_i-u_j)^2+(v_i-v_j)^2]+\mu(t_i-t_j)^2}$$

其中λ和μ为平衡空间距离和时间距离的比例因子，ti 和tj是不同的观测时间。则有

$$\begin{aligned}w(u,v,t)&=\exp(-\frac{d_{ij}^{ST}}{h^2})\\&=\exp(-\frac{\lambda[(u_i-u_j)^2+(v_i-v_j)^2]+\mu(t_i-t_j)^2}{h^2})\\&=\exp(-\frac{(u_i-u_j)^2+(v_i-v_j)^2}{h_1^2})\exp(-\frac{(t_i-t_j)^2}{h_2^2})\\&=w(u,v)w(t)\end{aligned}$$

其中h为时空带宽参数，$h_1=\sqrt{\frac{h^2}{\lambda}}$ 和 $h_2=\sqrt{\frac{h^2}{\mu}}$ 分别为空间和时间带宽参数，可以看出(u，v，t)处的权值为空间上的权值与时间上的权值的乘积。

采用交叉验证的方法确定带宽参数h1和h2，令

$$CV(h_1,h_2)=\sum_{i=1}^{n}[y_i-\hat{y}_i(h_1,h_2)]^2$$

选择h_1和h_2使得:

$$CV(h_{10},h_{20})=\min CV(h_1,h_2)$$

地理加权主成分分析

地理加权主成分分析

模型思想

通过引入地理加权的一系列算式，将变量间地理位置的交互影响纳入到计算中，从而有效地解决了多元数据中的空间异质性问题

地理加权协方差矩阵的计算公式为：

$$\sum(u_i,v_j)=X'W(u_i,v_j)X$$

(ui，vi)点的地理加权主成分可以写作：

$$L(u_i,v_i)V(u_i,v_i)L'(u_i,v_i)=\Sigma(u_i,v_i)$$

局部主成分得分成分得分:

$$T(u_i,v_i)=XL(u_i,v_i)$$

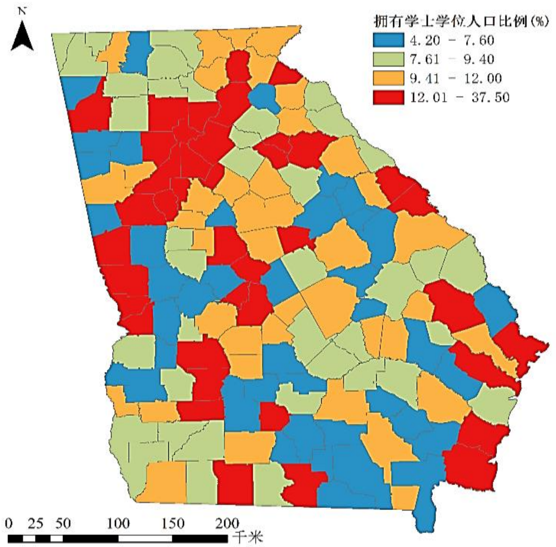
进一步得到局部成分方差与载荷。

应用实例

地理加权回归应用实例

数据

本实例采用佐治亚州人口普查数据（Georgia Census Data Set）。



佐治亚州拥有学士学位的人口比例

佐治亚州人口普查数据字段

AreaKey	**各县的标识码**
Latitude	各县中心点的纬度
Longitude	各县中心点的经度
Totpop90	1990年各县的人口
PctRural	各县的农村人口比例
PctBach	各县拥有学士学位的人口比例
PctEld	各县65岁及以上的人口比例
PctFB	各县在国外出生的人口
PctPov	各县生活在贫困线以下的人口
PctBlack	各县的非裔人口
ID	面ID
X	X坐标
Y	Y坐标

应用实例

地理加权回归应用实例

数据

本例通过对比普通多元线性回归与地理加权回归，探究各县的农村人口、贫困人口、非裔人口比例与该县学士学位比例的关联。

```
*****************************************************************************
 Global regression result
*****************************************************************************
 < Diagnostic information >
Residual sum of squares:                  2639.559476
Number of parameters:                        4
 (Note: this num does not include an error variance term for a Gaussian model)
ML based global sigma estimate:              4.074433
Unbiased global sigma estimate:              4.126671
Log-likelihood:                            897.927089
Classic AIC:                               907.927089
AICc:                                      908.319245
BIC/MDL:                                   923.271610
CV:                                         18.100197
R square:                                    0.485273
Adjusted R square:                           0.471903

Variable                   Estimate    Standard Error      t(Est/SE)
-------------------- --------------- --------------- ---------------
Intercept                 23.854615        1.173043       20.335661
PctRural                  -0.111395        0.012878       -8.649661
PctPov                    -0.345778        0.070863       -4.879540
PctBlack                   0.058331        0.029187        1.998499
```

普通多元线性回归诊断报告

```
*****************************************************************************
 GWR (Geographically weighted regression) result
*****************************************************************************
 Bandwidth and geographic ranges
Bandwidth size:                    49.460274
Coordinate                 Min              Max            Range
--------------- --------------- --------------- ---------------
X-coord           635964.300000  1059706.000000   423741.700000
Y-coord          3401148.000000  3872640.000000   471492.000000

 Diagnostic information
Residual sum of squares:           2312.592458
Effective number of parameters (model: trace(S)):                    8.033359
Effective number of parameters (variance: trace(S'S)):               5.454906
Degree of freedom (model: n - trace(S)):                           150.966641
Degree of freedom (residual: n - 2trace(S) + trace(S'S)):          148.388187
ML based sigma estimate:               3.813739
Unbiased sigma estimate:               3.947752
Log-likelihood:                      876.900473
Classic AIC:                         894.967192
AICc:                                896.184041
BIC/MDL:                             922.689706
CV:                                   17.914091
R square:                              0.549033
Adjusted R square:                     0.516564
```

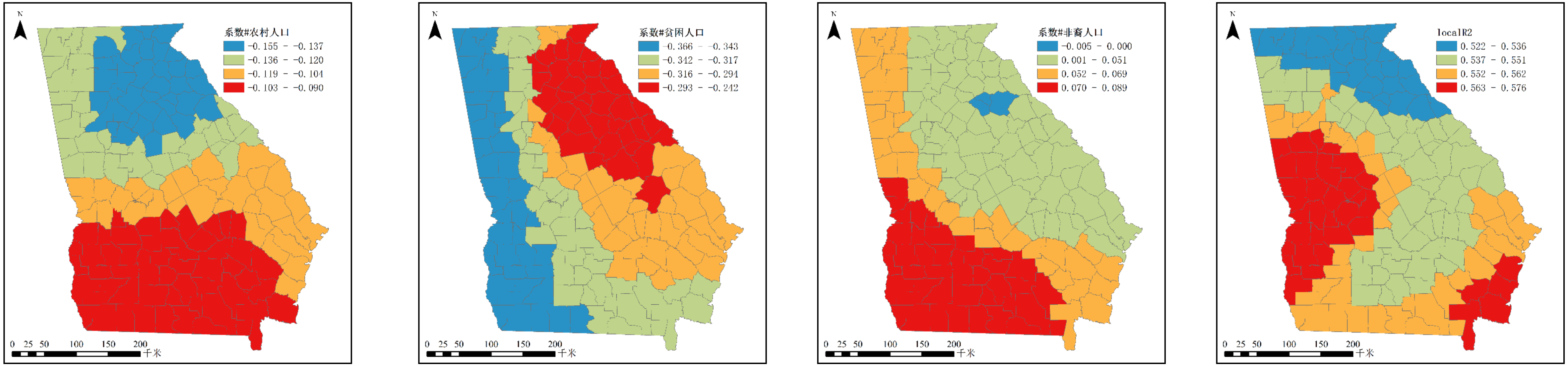
地理加权回归诊断报告

本例中，采用高斯核函数构建空间权重矩阵，使用AICc指标选取最优带宽。通过对比回归诊断报告（可以发现，地理加权回归的AIC和AICc低于普通最小二乘回归，R2要高于普通最小二乘回归，这说明地理加权回归模型较普通最小二乘回归模型显示出更好的拟合优度。其中，地理加权回归模型选取的最优带宽是49.46。

地理加权回归应用实例

应用实例

结论



地理加权回归系数及局部决定系数

通过地理加权回归，可以对局部的数据进行拟合，能够较大限度地挖掘和展现空间异质性。

应用实例

混合地理加权回归应用实例

对于佐治亚州的例子而言，若将各县的在国外出生的人口及65岁以上的老人数量设置为全局变量，重新拟合成为混合地理加权回归。

```
****************************************************************************
 Global regression result
****************************************************************************
 < Diagnostic information >
Residual sum of squares:                2110.569589
Number of parameters:                     6
 (Note: this num does not include an error variance term for a Gaussian model)
ML based global sigma estimate:            3.643353
Unbiased global sigma estimate:            3.714105
Log-likelihood:                          862.366074
Classic AIC:                             876.366074
AICc:                                    877.107796
BIC/MDL:                                 897.848403
CV:                                       15.331159
R square:                                  0.588429
Adjusted R square:                         0.572182

Variable                     Estimate    Standard Error      t(Est/SE)
-------------------- --------------- --------------- ---------------
Intercept                   17.243732        1.753292        9.835062
PctRural                    -0.070323        0.013579       -5.178928
PctPov                      -0.255236        0.072477       -3.521617
PctBlack                     0.049114        0.026485        1.854437
PctEld                       0.011448        0.129535        0.088377
PctFB                        1.852471        0.306830        6.037452
```

普通多元线性回归诊断报告

```
****************************************************************************
 GWR (Geographically weighted regression) result
****************************************************************************
 Bandwidth and geographic ranges
Bandwidth size:                   52.000000
Coordinate                Min              Max             Range
--------------- --------------- --------------- ---------------
X-coord           635964.300000  1059706.000000   423741.700000
Y-coord          3401148.000000  3872640.000000   471492.000000

 Diagnostic information
Residual sum of squares:          1940.327587
Effective number of parameters (model: trace(S)):                        9.658938
Effective number of parameters (variance: trace(S'S)):                   8.412795
Degree of freedom (model: n - trace(S)):                               149.341062
Degree of freedom (residual: n - 2trace(S) + trace(S'S)):              148.094920
ML based sigma estimate:             3.493325
Unbiased sigma estimate:             3.619657
Log-likelihood:                    848.994009
Classic AIC:                       870.311884
AICc:                              871.998744
BIC/MDL:                           903.023142
CV:                                 15.525575
R square:                            0.621627
Adjusted R square:                   0.593575

***********************************************************
 << Fixed (Global) coefficients >>
***********************************************************
 Variable               Estimate        Standard Error  t(Estimate/SE)
-------------------- -------------- -------------- --------------
PctEld                    -0.073003        0.187428       -0.389497
PctFB                      1.685558        0.351772        4.791619
```

混合地理加权回归诊断报告

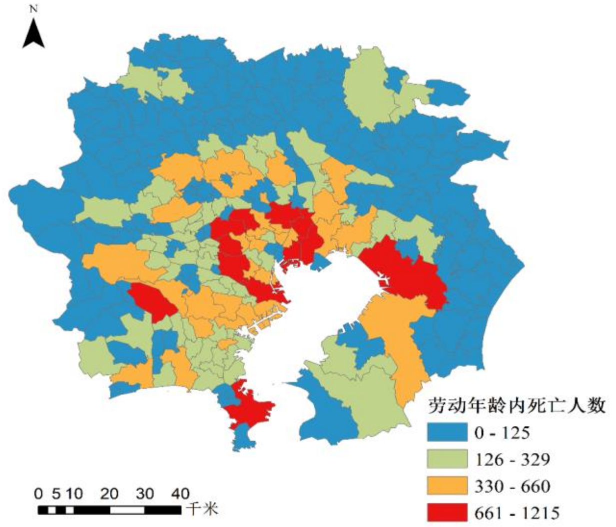
结论：混合地理加权回归的拟合优度依然优于线性回归。

应用实例

地理加权广义线性回归应用实例

数据与分析

将区域劳动年龄内死亡人数的预测值作为偏移量（Offset variable），将该区域专业工人占比和自有住房占比分别作为自变量，建立普通泊松回归模型和地理加权泊松回归模型。



东京各地劳动年龄内死亡人数

东京死亡率数据

字段名	含义
IDnum0	区域标识代码
X_CENTROID	区域中心x坐标
Y_CENTROID	区域中心y坐标
db2564	该区域劳动年龄内（25-64岁）死亡人数观测值
eb2564	该区域劳动年龄内（25-64岁）死亡人数预测值
OCC_TEC	该区域专业工人占比
OWNH	该区域自有住房占比
POP65	该区域老年人占比（大于等于65岁）
UNEMP	该区域无业率

应用实例

地理加权广义线性回归应用实例

结论

```
*****************************************************************************
  Global regression result
*****************************************************************************
  < Diagnostic information >
Number of parameters:                  3
Deviance:                           577.679297
Classic AIC:                        583.679297
AICc:                               583.772320
BIC/MDL:                            594.384330
Percent deviance explained            0.398403

Variable                   Estimate    Standard Error       z(Est/SE)        Exp(Est)
-------------------- --------------- --------------- --------------- ---------------
Intercept                  0.568956        0.034385       16.546642        1.766422
OCC_TEC                   -2.873540        0.147765      -19.446670        0.056499
OWNH                      -0.431167        0.038782      -11.117664        0.649750
```

普通泊松回归诊断报告

```
*****************************************************************************
  GWR (Geographically weighted regression) result
*****************************************************************************
  Bandwidth and geographic ranges
Bandwidth size:                  46.000000
Coordinate                 Min              Max            Range
--------------- --------------- --------------- ---------------
X-coord              276385.400000   408226.180000   131840.780000
Y-coord              -86587.480000    33538.420000   120125.900000

  Diagnostic information
Effective number of parameters (model: trace(S)):                         7.205366
Effective number of parameters (variance: trace(S'WSW^-1)):               4.642877
Degree of freedom (model: n - trace(S)):                                254.794634
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)):          252.232145
Deviance:                          530.986722
Classic AIC:                       545.397453
AICc:                              545.863363
BIC/MDL:                           571.108681
Percent deviance explained           0.447029
```

地理加权泊松回归诊断报告

结果表明，地理加权泊松回归模型的变异值（Deviance）和AIC值相较于普通线性回归模型更低，表明地理加权泊松回归模型的拟合优度较普通泊松回归模型更好。

Thank you