

# 第二次作业：空间自相关实验与分析

王逸

# 第二次作业：空间自相关实验与分析

王逸

## 一、空间权重矩阵在空间自相关分析中的作用研究

### 1. 总体宏观作用

空间权重矩阵定义了空间单位的相邻关系，决定了任意空间单位的特征对其邻近空间单位的贡献程度，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。

空间权重矩阵是空间自相关分析中必不可少的部分，包含 Moran's I 和 Geary's C 在内的各种自相关分析都需要空间权重矩阵做中间的计算以得到最后的自相关系数。空间权重矩阵常用于构建空间滞后项，是解释变量的空间滞后版本，捕捉空间单元之间的相互影响。空间权重矩阵本质是空间数据异质性的体现，即不同空间单位之间在某些特征上的显著差异，这种差异可以代表着地理距离、经济联系和社会网络等等含义。空间权重矩阵本身在空间效应的控制作用也决定了其对空间自相关估计偏误的影响。

### 2. 在公式中的体现

Moran's I 指标

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

其中， $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  为观测值方差， $\bar{x}$  为观测值均值； $w_{ij}$  空间单元  $i$  和  $j$  之间的邻近性度量值（如距离）

Getis-Ord G 指标

$$G(d) = \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij}(d) x_i x_j}{\sum_{i=1}^n \sum_{j=1, j \neq i}^n x_i x_j}$$

其中， $d$  为人为给定的距离阈值； $w_{ij}(d)$  为在给定距离阈值下空间单元  $i$  和  $j$  的空间邻近性度量值。

以 Moran's I 指标和 Getis-Ord G 指标为例，可以发现，空间矩阵元素的值（即  $w_{ij}$ ）直接影响了  $I$  和  $G(d)$  的绝对数值。

### 3. 作用实例验证

为验证空间权重矩阵对于空间自相关分析指标的影响，设计一组实验进行验证和总结。以深圳市交通小区数据为总空间数据，选取其中各交通小区的平均收入属性值进行专项研究，这里列出了图 1：深圳交通小区平均收入分布图。

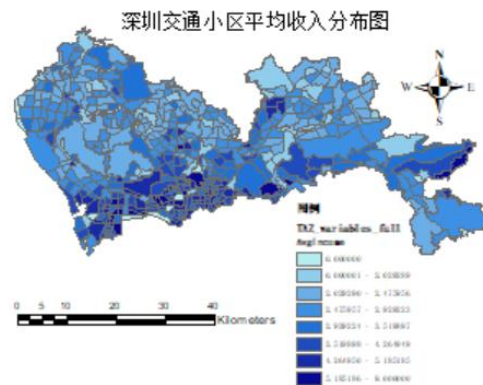


图 1：深圳交通小区平均收入分布图

### (1) 空间权重矩阵的参数对于空间自相关分析的作用

首先，设定空间权重矩阵类型为基于 k 近邻权重的邻接矩阵，对于这个矩阵而言，k 值（邻接多边形元素数量）是这个空间权重矩阵确定的参数，是一个正整数，我们求出一定范围内所有 k 取值对应的 Moran's I 空间自相关指标以及 Geary's C 空间自相关指标散点图。

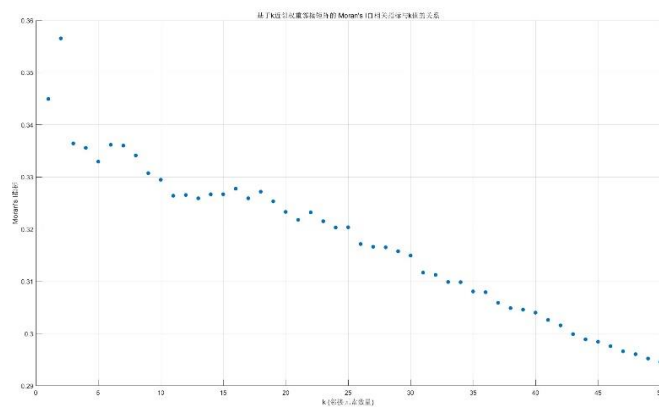


图 2：基于 k 近邻权重邻接矩阵的  
Moran's I 自相关指标与 k 值的关系

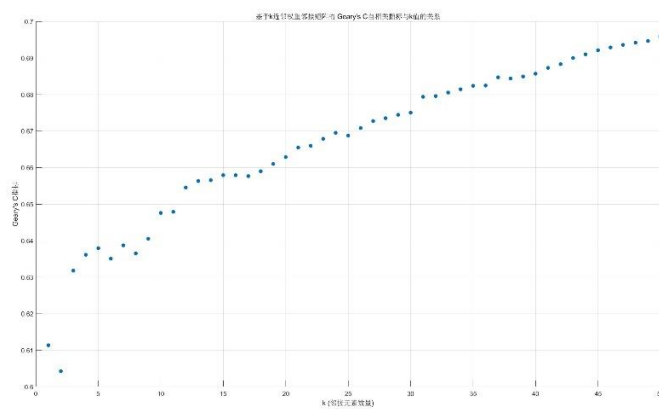


图 3：基于 k 近邻权重邻接矩阵的  
Geary's C 自相关指标与 k 值的关系

可以发现，图 2 和图 3 中  $k$  值的变化都显著影响了 Moran's  $I$  指标以及 Geary's  $C$  指标的变化，而且两者刚开始都是  $k$  值影响波动较大，但是在  $k$  大于 15 之后都趋于稳定。

$K$  近邻权重的邻接矩阵不会出现基于阈值距离的简单矩阵那样的不平衡邻近矩阵结构，是因为  $k$  近邻权值的大小由给定空间单元周围最邻近的  $k$  个单元来计算得到，而不是直接拿绝对欧氏距离决定。如果  $k$  值给的太小，那么对于很多空间单元而言，只将很少的邻近单元纳入计算，这样去求取自相关指标，会使得空间权重矩阵参考量较少，使得总体的结果不是很稳定，对应着图中  $k$  在 15 之内情况。所以说，不稳定情形的根本原因在于空间权重矩阵设定的不合理，这直接说明了空间权重矩阵的参数对于空间自相关分析有着较大的作用。

另外，图 2 直线越来越趋向 0，图 3 直线越来越趋向 0.9，说明  $k$  值较小时空间自相关程度较高，而如果  $k$  值给的较大，空间自相关程度则较小。这是因为随着  $k$  值的增大，所包含周围的邻近单元越来越多，空间随机的一些因素影响越大，导致空间自相关程度逐步下降。

(2) 空间权重矩阵的种类对于空间自相关分析的作用

随后，对于 Moran's  $I$  空间自相关指标以及 Geary's  $C$  空间自相关指标分别求出其在  $k$  近邻权重邻接矩阵和距离邻接矩阵情形下的四种变化图，每两组进行对照如下：

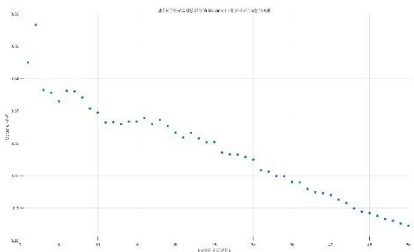


图 2：（与图 2 是一张图，在这里重复放置是为了对照方便查看）  
基于  $k$  近邻权重邻接矩阵的  
Moran's  $I$  自相关指标与  $k$  值的关系

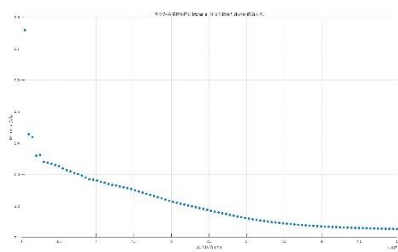


图 4：基于距离邻接矩阵的 Moran's  $I$  自相关指标与距离阈值的关系

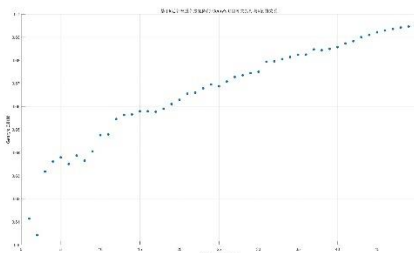


图 3：（与图 3 是一张图，在这里重复放置是为了对照方便查看）  
基于  $k$  近邻权重邻接矩阵的  
Geary's  $C$  自相关指标与  $k$  值的关系

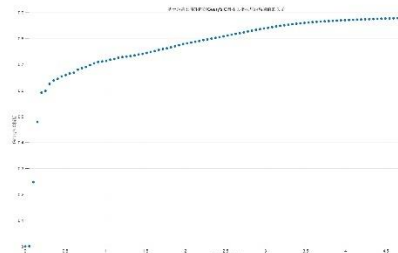


图 5：基于距离邻接矩阵的 Geary's  $C$  自相关指标与距离阈值的关系

将图 2 和图 4 放在一起比较研究，并将图 3 和图 5 放在一起比较研究，可以发现无论是对于哪种类型的空间自相关指标而言，不同类型的空间权重矩阵对于空间自相关指标的影响都是很大的。图 4 中的散点图明显比图 2 中的更加平滑，而图 5 中的散点图也同样比图 3 中的更加平滑，这实质上与空间权重矩阵本身直接挂钩。基于  $k$  近邻权重邻接矩阵由给定空间单元周围最邻近的  $k$  个单元来计算得到， $k$  值是一个离散变量，邻接矩阵在  $k$  值变化时容

易出现跳跃式变化，当增加或减少一个邻接地理元素的数量时，常常会由于某些极端邻接值的出现或者减少造成突变。而距离邻接矩阵直接由附近多边形质点的欧式距离进一步计算权重得到，随着距离阈值的变化，空间权重矩阵的变化也更倾向于平缓进行，稍微扩大或缩小距离阈值，邻接量可能只会有小幅变化。因此，无论从趋势还是绝对量来说，空间权重矩阵的种类的不同都会显著导致空间自相关分析的不同。

(3) 总结

以上两个实验分别以对照的方式，逐步深入，从单纯的参数变化分析到种类以及参数平滑度的比较，发现空间权重矩阵对于空间自相关指标的影响之大。空间权重矩阵从概念来说就是任意空间单元的特征对其邻近空间单元的贡献程度，在不同贡献程度的评定下，一定会导致整体空间自相关指标评价的大不同。

二、空间距离-空间自相关系数关系研究

从上面的探究过程中，实际上我们已经得到了在距离邻接矩阵的情形下，空间距离阈值与空间自相关系数的关系图，参见（1）中所绘制的图 4 和图 5，可以发现原来绘制的图 4 和图 5 都很好反映了 Moran's I 指标和 Geary's C 指标随着距离阈值的增大分别产生了减少和增加的情况，这实质反映了空间自相关指标随着距离阈值的增加而减少，这里的原因其实和之前 k 近邻矩阵的原因很像，都是因为距离阈值的增大导致空间随机性的因素涉及得越来越多，导致空间自相关性越来越低，最终呈现出一种接近于空间随机的结果。

在（1）中指出了空间自相关程度随着空间距离的增加而逐步降低，是一种相对平滑的趋势，为了进一步验证这种趋势关系，可以使用一系列相关性指标来验证，比如 Pearson 系数、Spearman 系数以及 Kendall 系数，我计算的结果如下：（系数结果保留小数点后四位）

	Pearson	Spearman	Kendall
图 4 结果	-0.8559	-1.0000	-0.9996
图 5 结果	0.7429	1.0000	1.0000
显著程度	极显著	极显著	极显著

*（注：该指标是对应之前图 4 和图 5 数据的线性相关关系，不是后面图 6 和图 7 的局部分析。图 4 和图 5 明显直观上就有强烈的线性相关关系，而后面的图 6 和图 7 由于做了极端值的局部放大，会有较大的失真，可能与此线性相关结果不匹配）*

经过计算，可以从一定程度上说明，这种趋势不仅符合线性趋势，也符合等级趋势，呈现一种较好的平滑、线性、分等级关系。

但是，可以发现的一点是，在图 4 和图 5 两个散点图上，虽然大部分都符合这种关系，但是当距离阈值设置得非常小的时候却有着一些极端的异常情况，在图中零星出现，为了更好地研究，将图 4 和图 5 的距离自变量范围取在 0-5km 之内，因为这个范围内就是极端情况的集中出现区，可以得到以下的图 6 和图 7，除了将自变量范围减少，还将自变量的取值变得更加密集，这样极端值就更加明显了。

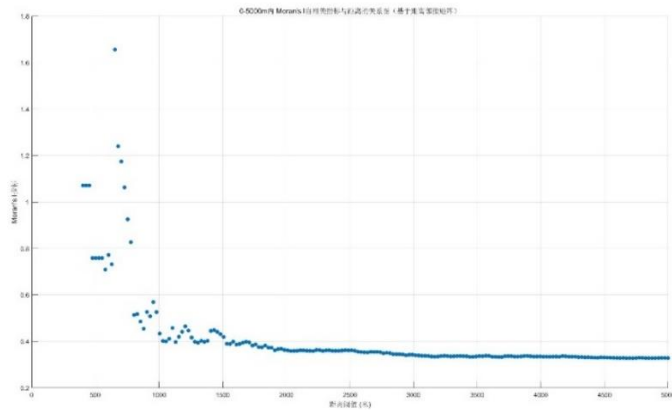


图 6：将图 4 的自变量范围取在 0-5km 之内的散点图

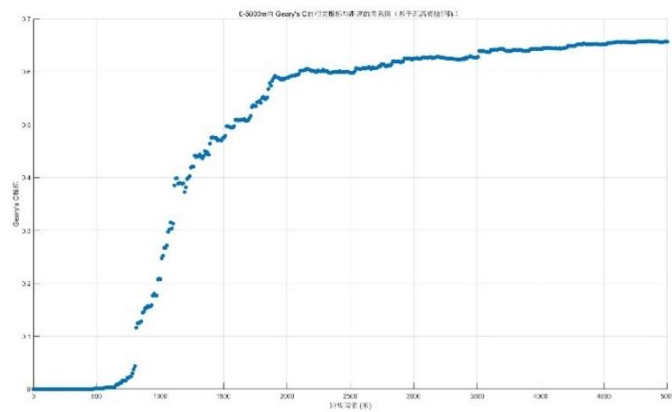


图 7：将图 5 的自变量范围变化取在 0-5km 之内的散点图

研究发现，图 6，也就是 Moran's I 在距离阈值较小的区间出现极端值较为剧烈，在初始段没有明显规律，后续又出现了从突变高值突降低值的情形，随后恢复为正常规律。推测原因是在距离阈值较小时，邻接矩阵的结构高度敏感，少量距离变化会导致某些点突然获得或失去邻居，进而导致邻接矩阵的显著变化。由于 Moran's I 依赖于所有点的全局关系，当邻接关系发生显著变化时，Moran's I 的值会出现剧烈波动。

而图 7，也就是 Geary's C 在距离阈值较小的区间的极端值相对来说有规律一些，从一个极低值，猛烈地增长到一个高值，然后恢复为缓慢增长的正常规律。推测原因是 Geary's C 主要关注相邻点之间的差异度，而不是整体的平均相似性。因此，在距离阈值极小时，邻接关系很少，差异度可能低；随着阈值增加，突然的邻接关系变化会揭示出更高的差异度，使得 Geary's C 的值快速升高，随后随着距离阈值进一步增加，邻接关系趋于稳定，Geary's C 的增长趋于正常。

### 三、普通空间聚集情况分析

在研究了不同空间自相关指标与权重矩阵之间的关系之后，尝试使用 ArcGIS 一般的空间聚集研究指标对于深圳交通小区平均收入的分布进行普通空间聚集情况分析。

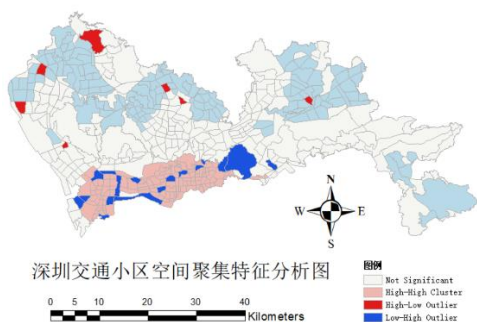


图 8：普通距离矩阵下的  
Anselin Local Moran I 指标分析

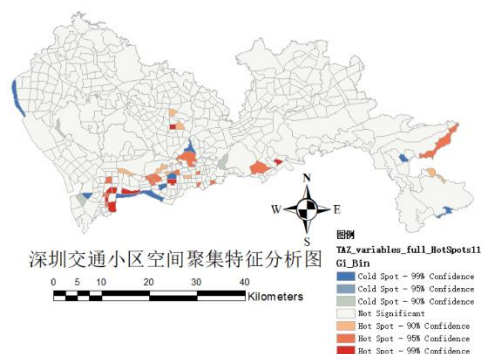


图 10：普通距离矩阵下的  
Getis-Ord Gi\*指标分析

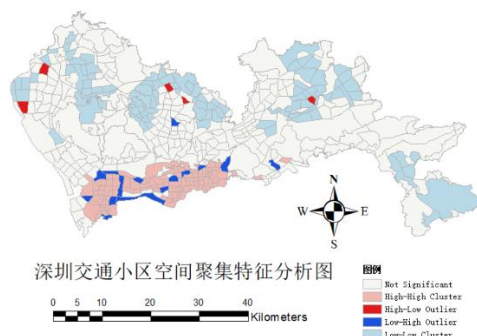


图 9：平方递减距离矩阵下的  
Anselin Local Moran I 指标分析

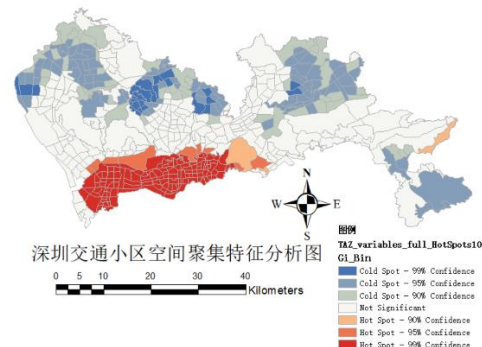


图 11：近邻矩阵下的  
Getis-Ord Gi\*指标分析

想要研究深圳交通小区平均收入分布普通空间聚集情况，需要采用合适的权重矩阵对应合适的空间分析指标。根据 ArcGIS 中的原理文档介绍，对于 Anselin Local Moran I 指标而言，采用距离矩阵是比较合适的，所以我们采用了普通距离矩阵（图 8）和使用了平方来加速递减的距离矩阵（图 9）来采集该项指标，可以发现两者的空间分析在总体上差别不大，但是在局部一些地方还是有区别的。高值区和高值边缘区基本集中在深圳南部地区，而低值和低值边缘区基本集中在深圳北部地区。可以发现使用了平方来加速递减的距离矩阵使得总体的距离衰减加快了，使得高值区和低值区的确定范围相对于普通距离矩阵减少了很多。

而对于 Getis-Ord Gi\*指标而言，ArcGIS 原理文档认为该项指标比较合适近邻矩阵，这种近邻矩阵和 k 近邻矩阵有些区别，这种近邻矩阵是将所有环绕的近邻元素都计入在内并赋以权重，而 k 近邻矩阵选取的环绕的近邻元素则是以 k 为其数量上限。为了验证为什么普通距离矩阵不适合该项指标，直接使用普通距离矩阵（图 10）和近邻矩阵（图 11）对同一指标分别分析一次得到结果，可以发现，使用普通距离矩阵的情况下该指标的有效分析范围大大减小，明显不如近邻矩阵情况下的分析显著。再调大和调小了普通距离矩阵的阈值上限后发现这一现象并没有消失，所以确实近邻矩阵更加合适，我们也能得到相似的结论，即高值区和高值边缘区都在南部，而低值区和低值边缘区都在北部和东南部。

## 四、总结

本次作业利用深圳市交通小区数据，选择一个典型指标，进行空间自相关分析实验，分析该指标的空间聚集特性。

在该作业中，先是通过总体宏观作用（主要是概念上的阐述），再是通过公式的展示简

要说明了空间权重矩阵在空间自相关分析中的作用。随后，为了更好地在实例中验证这种作用，设计了两组实验验证了空间权重矩阵的参数和种类显著作用着空间自相关指标。

随后，在前面两组实验的基础上（实际上前面两组实验的图 4 和图 5 就已完成空间距离-空间自相关系数散点图并分析了原因）进一步分析了空间距离-空间自相关系数的整体关系，采用三种不同的传统统计相关指标来验证了空间自相关程度随着空间距离的增大而稳定、平滑、线性、分级地减弱。额外地，分析了在距离阈值较小的区间，Moran's I 指标和 Geary's C 指标呈现异常值的原因。

最后，我们以 Anselin Local Moran I 指标和 Getis-Ord  $G_i^*$  指标为例，研究了交通小区平均收入的普通空间聚集情况，采用了不同权重矩阵来比对分析结果，发现对于不同的指标而言，有不同合适的权重矩阵，如果采用了不合适的权重矩阵，容易出现分析不显著的情况。