

区域经济研究实用方法： 基于*ArcGIS*, *GeoDa*和*R*的运用

QuYu JingJi YanJiu ShiYong FangFa:
JiYu ArcGIS, GeoDa He R De YunYong

王庆喜 蒋烨 陈卓咏 编著



经济科学出版社
Economic Science Press

国家自然科学基金青年项目 (编号71103160)
浙江工业大学“区域经济学”系列教材
浙江工业大学全球浙商发展研究院浙商研究文库
浙江省哲学社会科学重点研究基地资助

区域经济研究实用方法： 基于ArcGIS, GeoDa和R的运用

王庆喜 蒋烨 陈卓咏 编著



经济科学出版社
Economic Science Press

图书在版编目 (CIP) 数据

区域经济研究实用方法: 基于 ArcGIS、GeoDa 和 R 的运用 / 王庆喜, 蒋烨, 陈卓咏编著. —北京: 经济科学出版社, 2014. 5

ISBN 978 - 7 - 5141 - 4606 - 6

I. ①区… II. ①王…②蒋…③陈… III. ①区域
经济 - 研究方法 - 应用软件 IV. ①F061.5 - 39

中国版本图书馆 CIP 数据核字 (2014) 第 083792 号

责任编辑: 张 频

责任校对: 郑淑艳

版式设计: 齐 杰

责任印制: 李 鹏

区域经济研究实用方法: 基于 ArcGIS、GeoDa 和 R 的运用

王庆喜 蒋 烨 陈卓咏 编著

经济科学出版社出版、发行 新华书店经销

社址: 北京市海淀区阜成路甲 28 号 邮编: 100142

总编部电话: 010 - 88191217 发行部电话: 010 - 88191522

网址: [www. esp. com. cn](http://www.esp.com.cn)

电子邮件: [esp@ esp. com. cn](mailto:esp@esp.com.cn)

天猫网店: 经济科学出版社旗舰店

网址: [http: //jjkxcbs. tmall. com](http://jjkxcbs.tmall.com)

北京季蜂印刷有限公司印装

710 × 1000 16 开 12.25 印张 210000 字

2014 年 5 月第 1 版 2014 年 5 月第 1 次印刷

ISBN 978 - 7 - 5141 - 4606 - 6 定价: 42.00 元

(图书出现印装问题, 本社负责调换。电话: 010 - 88191502)

(版权所有 翻印必究)

前 言

在给硕士生上《区域经济学》这门课时,发现学生经常困扰于一些操作性和技术性的问题,本书是为解决这些问题而编写的。区域经济学的实证分析需要用到一些诸如设定空间权重矩阵、计算空间自相关指数等特定方法,而这些方法利用 SPSS、Stata 等常用软件的固定模块无法实现,需要借助 ArcGIS、GeoDa 和 R 软件。在这些软件的帮助下,进行空间数据分析是很方便的事。本书对这些软件操作进行了基本介绍,使得读者能够在看完本书后,进行通常的空间数据分析。

本书内容主要分为三大模块:

第一模块为原理部分。第 1 章介绍区域数据分析的原理,包括空间数据的特征、空间权重矩阵的设定规则、空间自相关指数的计算原理和空间回归模型的形式和估计等。

第二模块为软件操作。分别在第 2、3、4 章介绍 ArcGIS、GeoDa 和 R 软件操作,其内容主要是空间数据的读入与描述、空间图形的表达、空间权重矩阵的生成、空间自相关的测算与判定、空间横截面数据和空间面板数据分析模型等。

第三模块为专题应用部分。具体章节为第 5 章到第 8 章,共 4 章,分别对应区域经济分析中的四个常见问题,即产业地理集中与地区专业化、区域经济发展差距及其收敛性分析、区域知识生产及其溢出分析,以及空间互动模型分析。在这些专题分析中,综合运用了多种指标、方法和计算程序。按照书中给出的操作步骤进行演练,读者能够很快掌握从数据到结果的整个过程。本书的随书数据及代码可到邮箱 qyjyjf@163.com 去提取,密码为 [qyj123456](#)。数据下载下来后,注意在 D 盘建一文件夹 `regionResearchMethods`,将数据解压到其中,以保证程序代码的顺利运行。

本书的写作得到了国家自然科学基金青年项目(编号 71103160)和浙江省哲学社会科学重点研究基地的资助,同时也有幸列入浙江工业大学“区域经济学”系列教材和浙江工业大学全球浙商发展研究院浙商研究文

库，谨致谢意。在本书写作过程中，得到了浙江工业大学经贸管理学院、浙江工业大学全球浙商发展研究院等众多领导和同仁的关心和支持，深表感谢。要特别感谢产业经济学科徐维祥教授和唐根年教授的勉励和教诲，使得本书得以尽快完稿。同时，也要感谢经济科学出版社张频副编审专业而高效的工作，改进了本书质量。当然，书中所有可能的瑕疵，均由作者负责。

如有意见和建议，请即与本人联系：wqx1976@zjut.edu.cn，期待您的指正。



目 录

第 1 章 区域经济实证分析原理	1
1.1 空间数据	1
1.1.1 空间数据简介	1
1.1.2 可变面元问题	3
1.1.3 空间依赖问题	4
1.2 空间权重矩阵	5
1.2.1 相邻规则	5
1.2.2 空间权重矩阵的构建	7
1.3 空间自相关分析	8
1.3.1 全域空间自相关统计量	9
1.3.2 局域空间自相关统计量	12
1.4 空间计量经济模型	14
第 2 章 ArcGIS 在空间数据分析中的应用	19
2.1 编辑数据	19
2.1.1 Shapefile 文件准备	19
2.1.2 读入 Shapefile 地图文件	21
2.1.3 数据表联接	23
2.1.4 图层切割	25
2.2 图形展示	26
2.3 空间探索性数据分析	27
2.3.1 生成空间权重矩阵	27
2.3.2 ArcGIS 空间自相关分析	33

第 3 章 GeoDa 在空间数据分析中的应用	36
3.1 读入数据	36
3.2 生成空间权重文件	37
3.2.1 边界相邻	37
3.2.2 阈值相邻	41
3.2.3 k 个最近邻居	43
3.3 数据处理与图形展示	45
3.3.1 数据合并	45
3.3.2 地图展示	47
3.4 GeoDa 空间自相关分析	49
3.4.1 全域空间自相关	49
3.4.2 局域空间自相关	54
3.5 GeoDa 空间回归分析	55
第 4 章 R 语言在空间数据分析中的应用	60
4.1 R 的基本操作	60
4.1.1 R 介绍	60
4.1.2 R 中的数据对象	62
4.1.3 运算与函数	75
4.1.4 R 图形	79
4.2 读入数据与生成空间权重	81
4.2.1 生成 gal 和 gwt 空间权重文件	81
4.2.2 生成空间权重对象	83
4.3 spdep 空间自相关分析	87
4.3.1 全域空间自相关	88
4.3.2 局域空间自相关	92
4.4 spdep 空间回归分析	94
4.5 splm 空间面板数据分析	100
第 5 章 产业地理集中与地区专业化	111
5.1 结构性指标说明	111
5.2 产业地理集中度测量	113

5.3 地区专业化水平测量	129
第6章 区域经济发展差距及其收敛性分析	135
6.1 区域经济发展水平差距测度指标	135
6.2 中国省份经济发展空间关联分析	136
6.3 中国各省份经济发展水平差距测度	140
6.4 中国各省份经济发展收敛性分析	142
第7章 区域知识生产及其溢出分析	148
7.1 模型与指标说明	148
7.2 区域创新活动的空间自相关分析	151
7.3 区域知识溢出的空间面板数据模型分析	153
第8章 中国省际贸易的空间互动模型分析	159
8.1 空间互动模型介绍	159
8.1.1 空间互动模型的形式	159
8.1.2 空间互动模型数据的矩阵表达	161
8.1.3 流量数据的空间关联性	162
8.2 中国省际贸易的描述性分析	164
8.2.1 基于各省份的流入与流出分析	164
8.2.2 给定省份下货物流量的空间分布分析	167
8.3 中国省际贸易的空间相关性分析	171
8.4 中国省际贸易的空间互动模型分析	173
参考文献	188

第 1 章

区域经济实证分析原理

在区域经济实证分析的开始阶段，首先要明确几个基本问题：一是对什么做研究？即明确研究对象——区域，它们构成了我们的观测样本单元。鉴于区域并不是自然形成的，因此要根据自己的研究需要来予以确定。选择的区域层面及其划分方法不同，得到的结论也会有所差别。二是研究这些区域间有什么关系？揭示什么特征或机理？即明确研究内容，构建刻画这些关系的模型。三是如何测量这些区域的特征？明确了研究对象和内容后，接下来就要围绕这些区域搜集它们的空间位置、各种定性和定量指标的相关信息，以形成观测数据（observations）。四是用什么工具和手段来得到分析结果？即要了解研究方法。本书的主要目的是要使读者熟悉以上过程，在学习完本书后，能够马上从事相关研究。本书内容侧重操作，同时只针对几种常用的方法、软件和主题展开讨论，不求全面，读者可举一反三。如要深入了解操作背后的原理，可以进一步阅读本书所列的文献资料。

1.1 空间数据

1.1.1 空间数据简介

对于空间现象、活动、事物或单元（即空间观测），一般有两个属性去刻画它：一是空间属性，即它们的空间位置和几何形态；二是特征属性，即描述它们特征的各种定性和定量指标，如它们是否地处沿海（分别赋值为 1 或 0）、GDP、总人口数等。区域的空间属性和特征属性一起，构成了所谓的空间数据。

空间数据是“位置 + 属性”的结合，既描述了“怎么样”（what），

也描述了“在哪儿”(where)。我们通常分析的观测数据,只描述了观测单元的特征,至于其位置如何,是无需关心的。而空间数据为了分析空间位置对观测单元行为的影响,需要明确观测单元的位置。认为不同的位置,观测单元的行为和特征表现也有所不同。比如说,离垃圾站远近的距离,对房子的价格有一定的影响。如果光考虑房子自身的因素,如户型、建筑质量等,不考虑它们与垃圾站的距离远近,就无法全面解释房子在价格上的差异性。而房子离垃圾站的距离,就需要利用它们的空间坐标才能计算得出。

空间单元的位置和形状一般通过几何方法来表达。描述地理现象或事件的一个方法就是用三大几何元素——点、线、面来代表真实世界中的客体。点用来代表在给定的地理尺度上不存在或几乎不存在任何空间延伸的要素或事件;线用来代表具有线状延伸或定向移动特征的要素或现象;面用来代表被研究现象的空间延伸所覆盖的区域。撇开线要素不谈,点和面数据通常代表单个观测单元的集聚。也就是说,空间数据会受空间归集(spatial aggregation)的影响。在不同尺度上,可能需要以不同的方式来表示它们。

在空间数据分析中,各对象或要素都在空间中占据着一定的位置。这一位置既可以描述为绝对位置(absolute location),比如用经纬坐标来描述;也可以描述为相对位置(relative location),比如杭州市中心以东 50 公里。在定义绝对位置时,我们只需要坐标;而在定义相对位置时,我们则需要用另一个位置或者空间要素作为参照点。

需要注意的是,地球是圆球体,而我们要研究的地理要素一般用平面表达。因此,需要将球面上的点、线、面等投影到平面。球面上的空间元素其位置一般用经纬度表示,因此,需要将经纬度表示的位置与平面上的位置坐标对应,这一过程就是投影(projection)。

区域经济研究主要是研究区域经济行为与区域经济联系,其基本分析单元是区域,也就是面要素。因此,如何划分是关键问题。只有在清楚界定好区域边界后,才能明确地考察这些区域在经济社会方面的行为和特征,以及区域间的空间关系。

划分区域时,最方便的做法是根据行政区域来界定,这也是最常见的方式。如美国的州、县,中国的省、市、县等。有时,出于研究需要,需要将行政区域进行合并或分割,以形成新的区域,如将中国分成东部、中部、西部三大区域,将国家组合成为欧盟、东盟等。当然,也有根据地理

特性来划分的,如长江以北、长江以南等,但这种做法在区域经济研究中很少见。总之,需要根据自己的研究内容和性质,来对区域做合理的界定。

空间数据搜集好后,一般来说,有三种范式可以用来分析空间数据:

(1) 只分析特征属性数据。如果只分析特征属性数据,那么由于不涉及空间属性信息,因此借助传统的统计方法就够了。

(2) 只分析空间属性数据。如果只分析空间属性数据,那么所用到的就只有地理对象的空间信息。除了位置外,这种分析无法对地理对象做出其他区分。也就是说,除了地理位置,两个地理对象将没有任何不同。

(3) 同时对空间属性数据和特征属性数据进行分析。如果同时对这两种数据进行分析,那么所使用的统计方法就必须既能处理观测单位的位置信息,又能处理描述它们的特征数据。

大多数统计技术和模型都不是为了分析带有地理参照信息的观测单元而发展起来的,而主要是为了分析从总体中随机抽取的样本观测数据。当用传统统计方法分析从这些观测单元中获得的数据时,一般假设它们是相互独立的。空间数据是从相互邻近的观测点或同一研究区域收集的,彼此之间往往存在联系,所以不能假设各观测单元之间是相互独立的。因此,利用传统统计方法分析空间数据就会带来一些问题。在空间数据分析中,可变面元和空间依赖是两个比较大的问题。

1.1.2 可变面元问题

有关空间数据的一个关键问题是可变区域单元问题,简称为可变面元问题(modifiable areal unit problem, MAUP)。任何一块区域都可以以多种方式进行划分,可以按任何给定的标准将其划分为更小的区域单元。例如,中国地理版图既可以按省来划分,也可以按地级市来划分,还可以进一步用县(区)来划分。将一个区域不断细分成更小的子区域,便可得到嵌套层级空间划分系统。在不同层级上进行分析时,所得到的结果可能会不一致,这种不一致称为尺度效应(scale effect)。在对从区域单元数相近的不同区划系统中获得的数据进行分析时,我们可以预期分析结果也可能不一致,这就是所谓的区划效应(zoning effect),如中国版图既可划分为长江以北和长江以南,也可划分为黄河以北和黄河以南。尺度效应和区划效应统称为可变区域单元问题,这是因为这两种效应都与区域单元划分标准的改变有关。

可变面元问题源于一个事实,即存在许多不同的方式将地理平面划分

为互不重叠的面元以进行空间分析。一般情况下，定义这些面元的标准是人为和可变的，其结果是这些划分的空间面状单元常常缺少本质的地理学意义，因此以这些面元为单元的分析结果是依赖于面元划分方式和面元大小的。经济地理学中许多统计分析，如投入产出分析、空间相互作用模型等方面的研究大多揭示了可变面元问题的存在。

在大部分空间统计技术中，MAUP 效应都很普遍。它最大的影响是对变量之间相关性的影响。一般来说，基于较大地理单元（比如中国的省）的变量之间的相关性通常强于基于较小地理单元（比如地级市）的变量之间的相关性，因此在对来自不同尺度或空间分辨率水平的数据进行统计分析时，得到的结果会有差异。如果所分析的数据来自尺度或分辨率水平相近的不同区划系统，那么变量之间的关系也会存在差异。在区域经济研究中，不能期望在某一尺度上的研究结论能够适用于其他尺度上。

1.1.3 空间依赖问题

空间数据分析中的另一个重要问题是单元间的空间依赖问题，常表现为数据间的空间自相关。根据地理学第一定律——所有的事物都是相互联系的，离得越近的事物，彼此之间的联系也越强——空间单元的属性数据或多或少都存在一定关联性。因此，在分析空间数据时，我们不能假设空间中的各观测单位是互无关联、彼此独立的，尽管这违背了大部分传统或经典统计分析的基本假设。

图 1.1 中展示了空间数据分布的三种形态。在随机分布下，数据在各空间单元上的分布是随机性的，好比往一张标靶上自由掷飞镖，其落点是随机的。然而，现实空间数据的地理分布往往是非随机的。要么分散，比如说某个超市在某地选址后，为了避开不必要的竞争，其他超市可能在离其更远的地方选址；要么聚集，如企业扎堆在某地，现实中更多的是后者。

在空间数据分析中，空间自相关的研究历史相对较长。其中影响最大的事件是测度空间自相关水平的指标——Moran's I 的引入。总的来说，空间自相关性不会影响推断统计中参数估计的准确性，但会影响假设检验的结论。具体来说，对于变量之间的关系真实与否，也就是说在判断关系系统计量的统计显著性水平时，它可能会导致错误的结论。这是因为，观测单元彼此相关实际上等于观测单元重复。如果大量观测单元都彼此相关，那就可能抬高变量之间的关系（使得实际为不显著的统计量最终显示为显著）。也就是说，如果观测单元重复或自相关，那么变量之间的实际相关

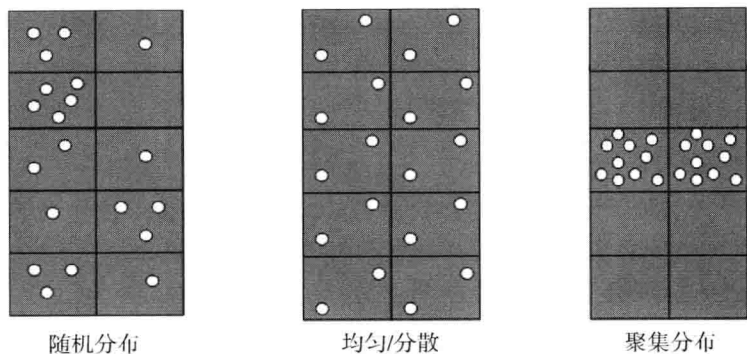


图 1.1 三种空间分布形态

性就应该比数据表现出来的更弱一些。或者说，由于观测单元自相关，有效样本容量将小于实际样本容量。

可以用一些简单的指标来评估空间自相关水平。测度空间自相关主要有两类指标：一类是全域指标；另一类是局域指标。如果空间自相关水平为显著，那么在推断统计中就应该对空间自相关问题的存在加以考虑，比如说运用以吕克·安瑟伦（Luc Anselin）等人为代表发展起来的空间计量经济模型。

1.2 空间权重矩阵

地理单元间的空间结构关系一般通过空间权重矩阵来表达，后者描述了各单元间的邻近关系。可以基于不同规则来确定地理单元间的相邻关系，并因此形成不同的空间权重矩阵。

1.2.1 相邻规则

地理单元间的相邻关系一般根据三种方式来予以确定：一是边界相邻，即两个地理单元有共同边界，则认为它们相邻，称为 Rook 相邻。Rook 是国际象棋里的“车”，按规则可以直走。二是顶点相邻，即两个地理单元有共同顶点，则认为它们相邻，称为 Bishop 相邻。Bishop 为国际象棋里的“相”，按规则可以斜走。三是边界或顶点相邻，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。Queen

是国际象棋里的“后”，按规则既可直走，也可斜走。在图形严重不规则的情况下，可以考虑 Queen 相邻。

此外，相邻还有阶数。在一定规则下，假设某空间单元 A 是另一个单元 B 的邻居，此为一阶相邻；如果 C 又是 B 的邻居，则 A 与 C 是二阶相邻。当然，A 也是 B 的邻居，因此 A 是其自身的二阶邻居。但这是没意义的，在二阶邻居列表中应将其自身剔除。同理，进一步递推下去，还可有 3 阶相邻、4 阶相邻，…，如图 1.2 所示。高阶相邻一般用来分析空间影响效应的衰减过程。

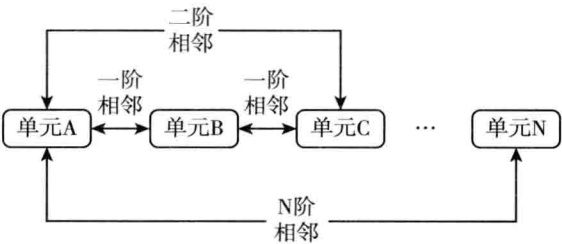


图 1.2 相邻阶数示意

综上，一阶相邻和二阶相邻在三种不同相邻规则下的几何形式如图 1.3 所示。

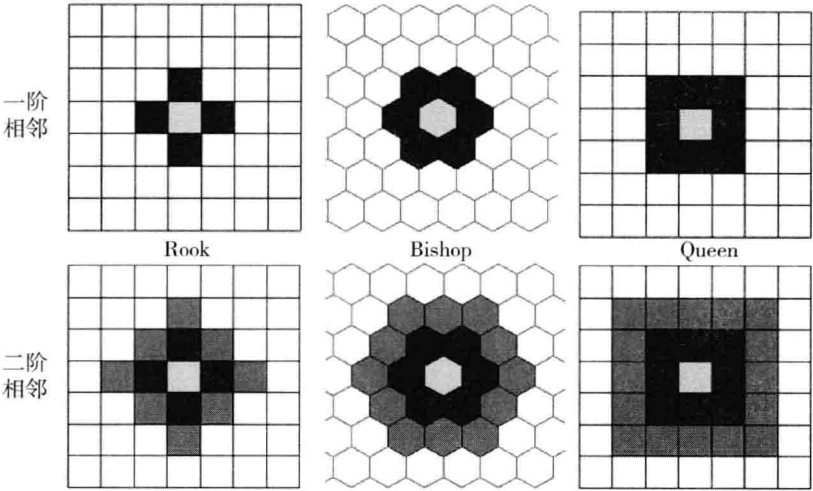


图 1.3 相邻规则示意

1.2.2 空间权重矩阵的构建

空间权重矩阵表达了各空间单元间的邻近关系。它是一个二维矩阵。如研究空间范围内有 n 个单元, 则空间权重矩阵 W 可以表示如下:

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

根据相邻标准, W 中的元素 W_{ij} 为:

$$W_{ij} = \begin{cases} 1 & \text{当区域 } i \text{ 和区域 } j \text{ 相邻} \\ 0 & \text{当区域 } i \text{ 和区域 } j \text{ 不相邻} \end{cases}$$

为了计算空间滞后值, 常常需要将空间权重矩阵进行行标准化, 使得每一行的权重和为 1。图 1.4 展示了表达中国七个省级单元 (安徽、浙江、江西、江苏、河南、湖北和上海) 的邻接关系的空间权重矩阵, 并将其进行了行标准化。

编号	1 安徽	2 浙江	3 江西	4 江苏	5 河南	6 湖北	7 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

(a) 邻接矩阵

编号	安徽	浙江	江西	江苏	河南	湖北	上海	总计
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

(b) 行标准化后的空间权重矩阵

图 1.4 空间权重矩阵示例

除了邻接关系外，空间权重矩阵的元素还可以根据各单元间的距离来确定。一种方式为根据门槛距离 d 来确定两单元间是否相邻。区域之间的距离往往根据各地区间的质心距离或者区域行政中心所在地之间的距离来确定。

$$W_{ij}(d) = \begin{cases} 1 & \text{当区域 } i \text{ 和区域 } j \text{ 在距离 } d \text{ 之内(即区域 } i \text{ 和区域 } j \text{ 相邻)} \\ 0 & \text{当区域 } i \text{ 和区域 } j \text{ 在距离 } d \text{ 之外(即区域 } i \text{ 和区域 } j \text{ 不相邻)} \end{cases}$$

另外一种方式为 K 个最近邻居空间权重矩阵 (K-nearest neighbor spatial weights matrix)，即将地理距离最近的 K 个单元设为自己的邻居，每个单元都有 K 个邻居。使用这一空间权重矩阵的考虑是因为一般使用的基于门槛距离的空间权重矩阵常常会导致不平衡的邻近矩阵结构。譬如，在空间单元的面积相差较大的情况下，就会出现一些小的地理单元有很多邻居，而大的地理单元则很少有邻居的情况。甚至有的单元因为没有邻近单元而成为“飞地”（在门槛距离设置过小的情况下就容易发生这种情况）。

此外，也可直接根据距离来确定空间权重，如距离倒数、距离平方的倒数等，如下所示：

$$W = \begin{bmatrix} 0 & \frac{1}{(d_{1,2})^2} & \frac{1}{(d_{1,3})^2} & \frac{1}{(d_{1,4})^2} \\ \frac{1}{(d_{2,1})^2} & 0 & \frac{1}{(d_{2,3})^2} & \frac{1}{(d_{2,4})^2} \\ \frac{1}{(d_{3,1})^2} & \frac{1}{(d_{3,2})^2} & 0 & \frac{1}{(d_{3,4})^2} \\ \frac{1}{(d_{4,1})^2} & \frac{1}{(d_{4,2})^2} & \frac{1}{(d_{4,3})^2} & 0 \end{bmatrix}$$

1.3 空间自相关分析

一般来说，地理距离近的现象或单元之间，将存在某种联系，从而使得它们的观测值呈现相关性。空间自相关即为测度邻近事物关联强度的方法与指标。具体来说，空间自相关方法处理变量在邻近观测单元之间的协变关系，即比较观测值的相似性和它们空间位置之间的相似性。空间自相关有全域测度和局域测度两种。全域空间自相关描述的是研究空间范围内

所有单元的整体空间关系。在这种情况下,研究范围内空间自相关的空间变异不是我们主要关心的问题。但是,空间自相关水平一般会存在空间差异,即在研究范围内的不同局域地带,其空间相关性程度是不同的。全域指标不能刻画空间自相关的空间变异状况。为了更好地描述空间关系在空间中如何分布和变化,需要借助可以探测局域空间自相关性的局域指标,来提供以某空间单元为中心的空间自相关值。在全域指标不显著的情况下,有可能存在显著的局域空间自相关;而在全域指标显著的情况下,又有可能存在不显著的局域空间自相关。

1.3.1 全域空间自相关统计量

全域空间自相关统计量有 Moran's I、Geary's C 和 Getis-Ord General G 等,用来描述整个研究区域上所有空间单元之间的平均关联程度及其显著性。

(1) Moran's I。

Moran's I 统计量是应用非常广泛的全域空间自相关统计量,其计算公式为:

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

式(1-1)中, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, x_i 表示地区 i 的观测值, n 为地区数, W 为空间权重矩阵。

Moran's I 的取值范围为 $-1 \leq I \leq 1$ 。 I 越接近 1, 表示地区间空间正相关的程度越强; I 越接近 -1, 表示地区间空间负相关的程度越强; 接近 0 表示地区间不存在空间自相关性。在 Moran's I 指数的显著性检验中, 常假设变量服从正态分布, 因此通过服从标准正态分布的 z 统计量来判断地区间的空间相关性。在实际计算中, 空间权重矩阵 W 大多采用基于邻接关系的矩阵, 当第 i 地区与第 j 地区相邻时, W 中的元素 w_{ij} 取值为 1, 否则为 0。

对观测值在空间上不存在空间自相关, 即它们在空间上是随机分布的零假设进行检验, 其统计量一般为标准化之后的 Moran's I 值 (或称 z 值)。

$$Z_I = \frac{I - E(I)}{\sqrt{\text{Var}(I)}} \quad (1-2)$$